Understanding sampling: representativeness matters

Julian PL Davis, Iain K Crombie, Huw TO Davies

Sampling, or selecting a group of people to represent a whole population, lies at the heart of almost all research designs. There are many ways of going about this, each of which presents its own problems. The trick is to obtain a good-sized sample that is truly representative of the population as a whole.

INTRODUCTION

What is sampling?

The use of sampling is seen in many aspects of our daily lives: 'the world's favourite airline', or '8 out of 10 owners said their cats preferred it'. These are statements based upon data collected from a sample of people chosen from the population at large. Sampling will determine what is put on supermarket shelves, or what programme is shown at peak time on TV. These statements do not, of course, actually represent everyone's opinion. They are the result of asking a small number of people and then assuming that what they say will also (broadly) hold true for everyone else.

In medicine, the situation is no different. Research often intends to discover new knowledge about large groups. However, it is usually impractical to conduct studies on all the individuals in these groups. What is commonly done is to select a smaller group of people, the sample, upon whom to conduct the research, and to select the sample in such a way that the results will apply to the larger group of interest. It is in this assumption of generalizability that many of the pitfalls of sampling lie.

Some terminology

Let us pause here briefly to clarify a few terms. First of all, the *population* is the overall group of people to whom the results of a given piece of research may apply. This is often assumed to mean everyone, but can be much more restricted — think of it as shorthand for 'the population of interest'. For example, a study looking at the usefulness of a new type of foam rubber pad in artificial limbs would not intend to be generalizable to everyone, only to those who actually have artificial limbs. And so the 'population' in this case would be everyone who has an artificial limb. Equally, in examining the effectiveness of a new kidney dialysis fluid, the results would only be intended to be applicable to those patients on dialysis.

Then there is the *sample*. The sample is a smaller group of people chosen from within the population of interest. They are the ones who will actually take part in the research, and they should have been chosen to be representative of the population as a whole. Sampling is the process of choosing the sample from within the population, and a representative sample is one that is similar in all respects to the target population.

Sampling can allow powerful conclusions to be drawn about the population of interest by collecting data on only a relatively small number of people. Indeed, the data collected from a well-chosen and carefully selected sample can be almost as informative and reliable as that taken from a complete census. All this begs the questions: how should researchers choose who to include and who to reject, and what are the consequences if they get it wrong?

TYPES OF SAMPLE, AND HOW TO OBTAIN THEM

Random (or probability) samples

Random samples (also called probability samples) are the most commonly encountered in medical research, particularly in surveys (Streiner and Norman, 1996). This is because they are generally thought to offer the best approximation to wider populations of interest, and therefore to allow generalization of the research findings.

A randomly selected sample is one in which the decision about who to include is left to chance. In the simplest case, you could roll a die and use the result to decide the fate of each individual: whether to include (say, roll a one) or exclude (roll two-to-six). This would generate a sample on average only one sixth the size of the target population. In practice, probability sampling is usually done using computer-generated random numbers. All the possible subjects (the population) are listed, and random numbers are used to select the sample. *Figure 1* shows this in action.

Random numbers can be quirky, however. What this means is that randomness does not always guarantee representativeness, and the sample, once chosen, should be checked to see if it is indeed representative. If not, the researchers should allow for any differences in their analysis.

Even if we accept that probability sampling is the best way to select a sample, there are other stages to sampling that should be reported by researchers. In order to select randomly a group of study subjects, the starting point is a clear definition of

Dr Julian PL Davis is Research Fellow and **Professor Iain K Crombie** is Professor of Epidemiology and Public Health, Department of Epidemiology and Public Health, University of Dundee, Ninewells Hospital and Medical School, Dundee, and **Dr Huw TO Davies** is Reader in Health Care Policy and Management, Department of Management, University of St Andrews, St Andrews, Fife KY16 9AL

Correspondence to: Dr HTO Davies

the population from which the sample is chosen. The process of deciding who should be in this population is fraught with difficulty, and is the point at which many sampling procedures go awry. To give a simplistic example, if we are attempting to test a new treatment for angina, and we choose to select a sample from those people attending a fitness centre, we are unlikely to include many people who actually suffer from the condition. It matters not how carefully random our sampling is, our sample will not represent the population who may benefit from the new treatment.

Although random sampling can be used for most applications, there are a number of alternative methods. However, these vary both in their complexity and in the degree of bias that may arise. Here are some of the more common variations on the theme, and some of the attractions and pitfalls associated with using them.

Samples of convenience

It may sometimes be expedient to use an easily available population from which to choose samples. This could be, for instance, the patients attending a particular clinic, people registered at a conference, or inmates at a prison. This can simplify the process of obtaining the sample, and if the researcher is actually interested only in the characteristics of individuals in a single setting, then it may be an appropriate method. However, it is more often the case that the resulting sample will not be representative of any particular larger population, and thus the conclusions that are drawn from the work will not readily be generalizable.

Lack of generalizability is always a problem, but sometimes convenience sampling can provide useful results. An Australian study looking at iron status in city children has shown that sampling children from childhood centre registers provided as good a result as random sampling. However, they also found that for a companion survey of lead levels, this method did not work so well (Ranmuthugala et al, 1998).

Quota samples

This is the method favoured by market research companies, and can be one way of ensuring that different categories of the population are represented in the sample. The population is categorized, for example by age, sex, social class or ethnicity. Samples are then taken from each of these groupings in proportion to the mix of these factors in the population. That is, if 10% of the population are white males aged 20–40 years in social class 1, then individuals with these characteristics are selected until a quota of 10% is filled. The serious problem with this approach is that the persons selected within each category may not actually be representative of that category at all. The result is an unpredictable sample composition — and consequently unreliable research findings.

Close agreement is possible between random and quota sampling. In Sydney, Australia, two surveys were carried out as a part of a policy development exercise for health promotion programmes. One used a quota sample of 1700 people, and the other selected a random sample of 480 individuals. There were 15 questions common to the two, and in no case was the result significantly different (Cumming, 1990). Thus quota sampling may indeed provide acceptable results, but it is often hard to tell from published reports whether it will do so or not.

Imagine a researcher is selecting a sample of 10 individuals from an electoral register containing 1234 people. The register is listed and numbered from 1 to 1234.

Ten random numbers are generated lying between 1 and 1234, and these are used to select the corresponding subjects. Thus each person has the same chance of being chosen as the next. This approach is likely to lead to representative samples — more so the larger the sample selected

For example, we might end up selecting individuals numbered: 4 .... 27 .... 34 .... 154 .... 221 .... 456 ..... 565 .... 768 .... 987 .... and 1232

Figure 1. Random selection.

Cluster sampling

In cluster sampling, the sample is chosen from groups that already exist in some convenient form. An example might be choosing a sample of patients from general practice (Sayer, 1999). A clustering approach would first of all involve selection of a number of specific general practices, followed by selection of patients from within those practices.

An important point to note here is that the clusters (in this case, specific general practices) may have very different characteristics, and it is essential to have enough clusters for any important differences to be allowed for. However, so long as a reasonable number of clusters are sampled, this approach can lead to samples that are representative of the target population. The method can also have hidden advantages for the researchers, such as being able to restrict the geographical distribution of a sample to facilitate visits, without compromising the generalizability of the findings.

Stratified sampling

Suppose that we are examining research on possible links between radiation from mobile telephones and skin lesions on the ear. A random sample of those with phones would probably produce more younger and less older people. A stratified sampling technique would first divide the population into age categories, and then sample randomly from within these groups.

In general, if the population being used for a study can easily be divided into groups that differ in terms of the characteristic in which the researchers are interested, then this can be used to improve the precision of any estimates. The population is divided up into groups according to the characteristic in question. Next, a similar fraction is taken from each grouping. This ensures that each group is equally represented in the final sample. However, the selection of individuals from within each group must be random for this system to work.

This method is also suited to populations that are complex in make up. For example, in a population such as Saudi Arabia, which has great social complexity and many layers of structure, a study of the prevalence of brucellosis required the use of stratified cluster sampling to compensate for the complex nature of societal structure (al-Sekait et al, 1992). Thus, in reading published research, one should try to assess whether the structure of the population of interest is well suited to a stratified approach.

Sequential or systematic sampling

Population listings may already be organized according to some system that may be helpful to those who keep the records, but not to those who want to select samples. In such circumstances it is not unusual for researchers to select every nth individual, where 'n' is chosen so as to generate a sample of the requisite size. In cases where the original lists are ordered only by name, this systematic approach usually generates a result as good as random sampling (Hagino and Lo, 1998). However, when systematic sampling is used, the reader should be aware that the approach is more prone to bias than simple random sampling.

Qualitative research and purposive sampling

The increased interest in qualitative research methods has also resulted in the wider use of different sampling systems. At a basic level, qualitative research is often concerned with gathering opinions and experiences, and as such it is often appropriate to choose people to take part in a way which makes sure that all groups are included.

Although by no means all qualitative research is done on samples chosen in this way, purposive sampling is an important alternative to more probability based methods (Curtis et al, 2000). However, qualitative research does not seek to generalize in the same way as quantitative approaches. Specifically, qualitative research is not concerned with making probabilistic assertions that the patterns seen in the sample will be similar to those seen in the population. Hence assessment of the adequacy of samples used in qualitative research is not based on ideas of representativeness. Instead, different types of generalizability are sought, such as the likely transferability of ideas, theories, understanding and insight from one setting to another (Marshall and Rossman, 1995).

WHAT CAN GO WRONG?

We have already mentioned some ways in which a selected sample can end up being unrepresentative of the population from which it comes. There are also ways in which even the best sample may provide unreliable results.

Small sample size

The size of a sample is very important, as all events are subject to the play of chance. A small sample may not show a distribution of characteristics that is all that similar to the target population. To illustrate this, say we randomly select four groups of 20 young people. The first might contain 15 women and 5 men, the second 8 women and 12 men, the third 11 women and 9 men and the fourth 6 women and 14 men.

None of the samples is evenly divided between men and women, yet we know that there are roughly equal numbers of men and women in our target population. Because the numbers in the samples were small, none of these samples turned out to be particularly representative. The key point here is that research findings from small samples are uncertain: the larger the sample, the more this uncertainty is reduced.

Unrepresentative populations

If the population from which the sample is selected is itself unrepresentative, the sample will clearly not be any better. For example, suppose a sample is selected at random from a list. If that list is inaccurate, because some people are not on it who should be, or are on it and shouldn't be, the sample will not represent the population. The list may be inaccurate if people have moved away, died, or simply not bothered to register. Importantly, it may be that the very people who do not register are the ones who are of prime interest in the study.

Poor response

Even if the population from which the sample is chosen is accurately described, and the sampling is well carried out, there is still the problem of non-response. Consider a study that involved sending out a questionnaire to a sample of people who had a particular condition. If everyone included in the sample responded to the questionnaire, then the results will reflect accurately the characteristics of the population. However, if the only people who respond are those with, for example, more severe symptoms, or those who live alone, then the results will not reflect the characteristics of the population as a whole.

Misleading conclusions

The inappropriateness of a sampling technique can lead to misleading interpretations of data. For example, random sampling of individuals can lead to underestimates of risk of infection in surveys of sexually transmitted infection (Ghani et al, 1998). Here the authors also suggested that allowing for the peculiar nature of sexual partner networks can be achieved by using variants on the theme of cluster sampling.

SUMMARY

In summary, a few simple questions can reveal much about the quality of the sampling in medical research.

How representative is the sample?

It is easy to obtain an unrepresentative sample. What is important when reading and assessing studies is to look to see how hard the researchers tried to make sure their sample was representative. Did they approach the problem systematically? Did they use an appropriate sampling method? Was this rigorously applied, and did they make multiple attempts to contact non-responders?

How good are the data?

It is equally hard to obtain good quality data. How did the researchers go about gathering the data? Did they use previously validated methods (for example questionnaires that had been used in another survey previously)? Did they test their data collection by doing a pilot study (using a smaller sample from an appropriate group)? Were the methods they used sensible (were they trying to use a questionnaire when an interview would have been more appropriate)?

As we have seen, it is the nature of the sample that is the key to the success of much research. Randomized trials, cohort studies, cross-sectional surveys and other study designs all depend upon the appropriateness of their samples for their success. With poor sampling strategies, research findings may prove to be true for the group in which they were determined, but not that generalizable to any wider population of interest. **HM**

al-Sekait MA, Bamgboye EA, al-Nasser AN (1992) Sampling in epidemiological research: a case study of the prevalence of brucellosis in Saudi Arabia. *J Roy Soc Health* **112(4):** 172–6

Cumming RG (1990) Is probability sampling always better? A comparison of results from a quota and a probability sample survey. *Community Health Studies* **14(2):** 132–7

Curtis S, Gesler W, Smith G (2000) Approaches to sampling and case selection in qualitative research: examples in the geography of health. *Soc Sci Med* **50(7-8):** 1001–14

Ghani AC, Donnelly CA, Garnett GP (1998) Sampling biases and missing data in explorations of sexual partner networks for the spread of sexually transmitted diseases. *Statistics in Med* **17(18):** 2079–97

Hagino C, Lo RJ (1998) Random *vs* systematic sampling from administrative databases involving human subjects. *J Manipulative Physiolog Therapeut* **21(7):** 454–9

Marshall C, Rossman GB (1995) *Designing Qualitative Research.* Sage, London

Ranmuthugala G, Karr M, Mira M (1998) Opportunistic sampling from early childhood centres: a substitute for random sampling to determine lead and iron status of pre-school children? *Aust NZ J Public Health* **22(4):** 512–4

Sayer GP (1999) Estimating and generalizing with clustered sampling in general practice. *Aust Family Physician* **28** Suppl 1: S32–4

Streiner DL, Norman GR (1996) *PDQ Epidemiology.* Mosby, St Louis

KEY POINTS

- Taking a sample from a population and then drawing conclusions from it is a common feature of most clinical research.
- Samples can be taken using many methods, and it is important to use the right one.
- Each method has its own features, and its own shortcomings.
- Using the wrong method can produce biased and unreliable results.
- The sampling method used should be a factor in the interpretation of research findings.