

Sampling rare populations

Sampling rare populations presents many challenges. In caring for those who experience social exclusion, nurses are well placed to conduct research into these hard-to-reach populations. This article by Mary Dawood discusses some of the sampling methods used

key words

- ▶ rare populations
- ▶ sampling techniques
- ▶ social exclusion
- ▶ survey design

Introduction

There has long been an acceptance among academics and policy-makers that rare populations are hard to locate (Sudman and Kalton 1986, Kish 1991). Sometimes a sampling frame covering a rare population is available, an example being people with a rare disease. In most cases, however, such frames are unavailable and special sampling techniques are required.

Heckathorn (1997) defines a 'hidden' population as having two characteristics: first, no sampling frame exists, so the size and boundaries are unknown; and second, there exist strong privacy concerns because membership sometimes involves stigmatised or illegal behaviour, often resulting in individuals refusing to co-operate or giving unreliable answers to protect their privacy. Most hidden populations are rare and so do not fit with traditional methodologies or sampling frames. As a result, innovative non-random sampling methods have gained increasing recognition as social researchers have striven to study these rare groups.

To ensure consistency in this paper, 'rare groups' will be used as an umbrella term to cover other phrases, such as 'hard-to-reach' or 'hidden' groups, which are cited frequently in the literature. The strategies for sampling rare groups, which also appear frequently in the literature, include network sampling or snowball sampling, screening and location sampling.

Strategies

Various strategies have been devised for sampling rare populations. But finding cost-effective methods calls for a degree of inventiveness and this can be complicated further by the topic of research. Lee (1993) suggests sampling becomes more difficult the more sensitive the topic under investigation, as potential informants will have more reason to conceal their activities. Second, and related to this, the less visible an activity, the harder it is to sample (Lee 1993). Researchers may be tempted to use *ad hoc* convenience samples when faced with a rare population, and while such samples may be adequate for the exploratory phase of the research, they may be totally inadequate for making estimates about a rare population (Sudman and Kalton 1986).

Snowball sampling involves constructing lists of rare populations by using initial sets of selected members as informants for names and addresses of unknown members (Kish 1991). It is used for two main reasons, the first being to reach a target population where the numbers may be small, such as illicit drug users. It is a useful method for finding respondents where some degree of trust is a prerequisite to establishing contact. Second, it may be used formally to study populations that have been otherwise difficult to enumerate through other methods (Snijders 1992). It is used most frequently to conduct qualitative research, mainly through interviews (Atkinson and Flint 2001) and is often also known as 'chain-referrals'. This method falls in the category of a wider set of link tracing methods that exploit the social networks of identified respondents to give the researcher an expanding list of potential contacts. It is particularly suited to investigating marginal populations or elites.

Intermediaries who introduce the researcher to respondents are useful in those communities whose members may be vulnerable or highly stigmatised and who could not be easily approached or are unwilling to be interviewed. The approach involves initial contact with an individual connected with the population of interest. This individual introduces other members of the population to the team. These subjects are then interviewed, observed or invited to attend a focus group. In turn, these subjects introduce other members of the population. This continues until either no further sample members can be contacted or the point of saturation is reached.

This strategy has traditionally been sidelined in social research because it

does not adhere to many of the notions underpinning conventional random methods of data collection. Moreover, formal evaluation of snowballing as an effective strategy has been limited (Biernacki and Waldorf 1981). It has, however, been recognised by some as a useful strategy for sampling rare populations (Sudman and Kalton 1986). Indeed it may still be the only way of sampling deviant populations, which are clandestine and therefore concealed from the view of mainstream society (Watters and Biernacki 1989). Faugier and Sargeant (1997), in their study of prostitutes, used non-random and snowball sampling methodologies to gain insight into the lives of drug-using prostitutes as no descending method, such as a census-based sampling frame, nor any other reliable source was available to define and randomly sample this population.

Kish (1991), however, cautions against what he refers to as naïve optimism about the possibilities of snowball sampling because in practice the sampling technique cannot guarantee that the sample will be representative. Indeed, the quality of the data, in particular selection bias, is seen as one of the disadvantages of this method (Kish 1991). Because samples are not randomly drawn but are dependent on subjective choices of the first contact, samples may then tend towards a reflexive bias or what Lee refers to as reciprocity and transitivity (Lee 1993). In this situation some individuals are more likely to be targeted than others. Moreover, this method assumes that some connection exists between the intermediary and the target population; although this may be the case much of the time, relying on the intermediary may result in inevitable bias where some suitable contacts will be overlooked and the researcher may have to sift through large numbers of unsuitable respondents, which may be time-consuming and costly.

Faugier and Sargeant acknowledge the problems of bias inherent in snowball sampling but sees it as the price the researcher has to pay for gaining understanding of these rare groups. She suggests that the confidence that develops between the researcher and the subject is perhaps the best guarantee of sincerity and should enhance the validity of the data (Faugier and Sargeant 1997).

In terms of researching deviant groups such as drug users, snowball sampling may offer a degree of protection and minimise suspicious or guarded

responses that may be prevalent in more traditional methodologies. Lee (1993), for example, highlights the fact that when researching deviant or stigmatised groups, security that would otherwise be a problem for the researcher may be less of a concern because the intermediaries who form the links in the referral chain are known to the potential respondents and will vouch for them. Being accepted in this way or being regarded as an 'insider' can also aid entry to settings where conventional approaches may not succeed (Atkinson and Flint 2001).

However, such a methodology can be a slow, protracted and unreliable process, and may not result in a large sampling frame. The main value of snowball sampling therefore is twofold. It may be used as a method of finding respondents in a rare population where some level of trust is required to initiate contact, or it may be used as a method of sampling in a more formal statistical sense. In both situations, it may have to be accepted that the sample obtained may be unrepresentative (Lee 1993).

Screening

An alternative strategy for identifying a rare population is screening or two-phase samples. This method is also referred to as sift sampling and can be considered when a probability sample is required for a specific group for which no sampling frame is available (Gilbert 2001). It involves collecting data from members of the initial sample to be able to classify them as members or non-members of the rare population (Kalton 2003). The usual approach to selecting the sample in these instances is to start with a large sample of the population – for example, households. Then a short screening survey for all households, which might be done on the doorstep, is carried out, and finally a full interview is completed in households with the relevant people.

As a method, it is time-consuming and costly, and may only produce a limited 'strike rate' (Lee 1993). Realistically it can only be used for sub-groups where the identifying characteristics are not especially sensitive or confidential. One could not, for example, ask about illegal drug use or illegal immigrants, but could ask about the number of children under five with asthma. Indeed, screening was used as a sampling strategy in the 2003 Health Survey for England, which had a focus on child health and included a boost

sample of children. However, this approach is limited and certainly not a suitable strategy for sampling deviant or stigmatised groups. To contain costs, telephone screening may be appropriate for some topics. Gilbert describes how a successful sampling frame was achieved to study off-shore oil workers in the Aberdeen area. From 8,000 households with obtainable telephone numbers, 421 men were identified. Less than 5 per cent of numbers were unanswered after three calls and only 4 per cent refused to answer further questions (Gilbert 2001).

The success of screening largely depends on being able to ask screening questions quickly at the start of interviews and to structure the questionnaire to avoid transparent disclosure of the rare group. Kalton points out that the risk of false negatives is likely to be greater when the screener respondents can deduce the rare population of interest from the contents of the screener questions or from other material supplied to them, and can then avoid the full interview by manipulating their answers accordingly (Kalton 2003). Consideration must be given to the fact that misclassification at the screening stage can give rise to serious levels of non-coverage (Sudman 1972).

Another anxiety for researchers using screening or sift sampling is the potential for non-response at the screening phase. Added to this is the concern that non-response will be higher for the rare population than for the total population. So even a high screener response rate may hide a low response rate for members of the rare population (Kalton 2003).

Another method of sampling a rare population is called location sampling, as some groups can be defined by their activities. This involves visiting places where members of the study population are known to gather. Sudman suggests there are three main aspects of sampling on location that should be considered: where to sample, when to sample, and recognition of the fact that different people visit the location with different frequency (Sudman 1980). For many surveys using location sampling, the visitor is the appropriate unit of analysis – for example, the number of visitors to soup kitchens in one week to estimate the number of homeless (Kalton 2003). The aims are to collect background data on the behaviours taking place and increase rapport with key individuals, which may also help with some network sampling.

Locating oneself at the scene of the 'action' is a well-respected tradition in

social research (Lee 1993) and respondents can be selected as they enter or leave the location. An unbiased sample requires that all entrances have some probability of selection, but there is no guarantee that the sample selected from those in the setting is representative of the wider population, as those who do not use the location cannot be studied.

Different types of people will visit locations at different times of the year so this will have to be taken into account. Once you know how likely people are to visit a location at particular times, the sampling of people can be made inversely proportional to that probability, which means the final sample of visitors is self-weighting to that point (Sudman and Kalton 1986). Location sampling is not dissimilar to those methods that have been proposed for counting mobile animal or human populations at selected locations – called capture-recapture methods – but this method has its own problems. In the case of a complex mobile population, such as homeless people or asylum seekers, identifying those who have already been sampled in the capture period but return in the re-capture period may be difficult.

Conclusion

Sampling rare populations is fraught with difficulty but some nurses, particularly those working in accident and emergency or public health, have greater access to these rare groups and so are better placed than many other professionals to conduct such research. This is an opportunity that should be optimised but ethical guidelines must be strictly adhered to to protect these vulnerable groups.

In random sampling methods, some types of data are so prevalent and occur so frequently that only very small samples are needed. In sampling rare or hidden populations, the term 'hidden' comes into its own and becomes much more of an issue for researchers. In the sampling methods described, two problems emerge continually: representation and practicality. It may well be that nursing researchers in this area have to accept lack of representation to ensure that hidden populations are not even more marginalised but are studied in the most pragmatic and cost-effective way.

Choosing not to undertake such research because of its complexity and potential lack of scientific rigour seems almost a dereliction of duty and ques-

tions nursing's ability to advocate effectively and promote the health of the most vulnerable in society. To be truly professional we should be seeking to help those least able to help themselves ■

Mary Dawood RN, MSc, BSc (Hons), PG Diploma in Teaching and Learning is a nurse consultant at St Mary's Hospital, Imperial College Healthcare NHS Trust, London, UK

This article has been subject to double-blind review

references

Atkinson R, Flint J (2001) Accessing hidden and hard to reach populations: snowball research strategies. *Social Research Update*. 33.

Biernacki P, Waldorf D (1981) Snowball sampling: problems and techniques of chain referral sampling. *Sociological Methods and Research*. 10, 2, 141-163.

Faugier J, Sargeant M (1997) Sampling hard to reach populations. *Journal of Advanced Nursing*. 26, 4, 790-797.

Gilbert N (2001) *Researching Social Life*. Second edition. Sage Publications, London.

Heckathorn D (1997) Respondent-driven sampling: a new approach to the study of hidden populations. *Social Problems*. 44, 2, 174-199.

Kalton G (2003) Practical methods for sampling rare and mobile populations. *Statistics in Transition*. 6, 4, 495-501.

Kish L (1991) Taxonomy of elusive populations. *Journal of Official Statistics*. 7, 3, 339-347.

Lee RM (1993) *Doing Research on Sensitive Topics*. Sage, London.

Snijders TAB (1992) Estimation on the basis of snowball samples: how to weight? *Bulletin Methodologie Sociologique*. 36, 59-70.

Sudman S (1972) On sampling very rare human populations. *Journal of the American Statistical Association*. 67, 335-339.

Sudman S (1980) Improving the quality of shopping centre sampling. *Journal of Marketing Research*. 17, 4, 423-431.

Sudman S, Kalton G (1986) New developments in the sampling of special populations. *Annual Review of Sociology*. 12, 401-429.

Watters JK, Biernacki P (1989) Targeted sampling: options for the study of hidden populations. *Social Problems*. 36, 4, 416-430.

Online archive

For related articles and author guidelines visit our online archive at www.nurseresearcher.co.uk