Clinical research 3: Sample selection

Ruth Endacott RN, DipN(Lond), MA, PhD (Professor of Clinical Nursing) a,b,*, Mari Botti RN, RM, BN, PhD, MRCNA (Professor, School of Nursing) a,c

a University of Plymouth, Faculty of Health and Social Work, Earl Richards Road North, Exeter EX2 6AS, United Kingdom
b La Trobe University, PO Box 199, Bendigo, Victoria 3552, Australia
c Deakin University, Burwood, Victoria 3125, Australia

Abstract This research series is aimed at clinicians who wish to develop research skills, or who have a particular clinical problem that they think could be addressed through research. The series aims to provide insight into the decisions that researchers make in the course of their work and to also provide a foundation for decisions that nurses may make in applying the findings of a study to practice in their own Unit or Department. The series emphasises the practical issues encountered when undertaking research in critical care settings: readers are encouraged to source research methodology textbooks for more detailed guidance on specific aspects of the research process.

© 2007 Elsevier Ltd. All rights reserved.

KEYWORDS Methodology; Sample size; Sample selection; Qualitative design; Quantitative design

Introduction

This paper addresses two key principles of sampling: adequacy (the number of participants) and appropriateness (the characteristics of the participants). The research aims and design will determine whether you need a statistically based or saturation based sample size (adequacy). The appropriateness of the sample is addressed through inclusion and exclusion criteria. These two aspects are discussed below in relation to quantitative and qualitative designs.

Quantitative designs

By definition, quantitative designs rely on the control or exclusion of all factors that might bias the outcome of a study. In the reality of clinical research, this is rarely achievable. However, a crucial early stage is the identification of confounding variables and, for each one, to identify:
whether it can be controlled;
(b) if not, can it be measured or;
(c) should it be used as an exclusion criterion?

Confounding variables that are measurable but not controllable provide useful data for sub-set analysis. These are usually referred to as secondary outcomes and phrased as secondary hypotheses. For example, when studying the intensity of post-operative pain in a group of patients, a useful secondary analysis would be to explore whether patients with pre-existing chronic pain conditions (a confounding variable that can be measured) experience more post-operative pain. A confounding variable that could be used as an exclusion criterion for such a study might be different cultural background, unless the focus of the study was to compare groups of patients from different cultural backgrounds. If decisions made by the researcher are not made explicit, the rigour of the study may be called into question.

Sampling is necessary because it is not usually possible to collect data from every individual, event or unit in the population we are interested in. Data are collected from a smaller sub-set of the population — a sample — which ideally corresponds to the larger population of interest in terms of particular key characteristics. The intent and the requirements of sampling differ depending on the design of a study.

The intent of sampling in quantitative designs is to estimate or predict outcomes about the larger population based on data from a sample of that population (Schofield, 2004). In other words, generalisations are made about individuals or units from whom data has not been collected. In order to be able to make these generalisations it is necessary to ensure that the sample is in fact representative of the population. To do this, attention is paid to the method used to sample the population and the size of the sample.

**Sampling methods**

Ensuring that the sample accurately represents the larger population is more important than the size of the sample in quantitative research. The representativeness of the sample in terms of the larger population has a direct impact on the external validity of the conclusions of the research.

The first step in sampling then is to identify the specific characteristics of the target population, i.e. the population to which you wish to generalise the results. For example, a target population may include all adult, post-operative day surgery patients in Australia. Because it may not be possible to study every post-operative day surgery patient it is then necessary to identify an *accessible population*. This is a sub-set of the target population that reflects specific characteristics such as age, sex and diagnosis, and is accessible (LoBiondo-Wood and Haber, 2004). For example, an accessible population may include all post-operative day surgery patients in a specified geographical area.

The next major consideration in the sampling process is to define the inclusion and exclusion criteria of the accessible population. Inclusion

---

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Sampling methods in quantitative research</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random sampling</strong></td>
<td></td>
</tr>
<tr>
<td>Simple random</td>
<td>All elements in a sampling frame have an equal chance of selection. Selection is through the use of random numbers or similar method</td>
</tr>
<tr>
<td>Systematic</td>
<td>Samples are drawn by beginning at a random point in the sampling frame and then selecting each nth element, e.g. Every 20th event or individual. Its advantage is its simplicity</td>
</tr>
<tr>
<td>Stratified</td>
<td>Participants are grouped according to strata that are important in a study such as age, sex or diagnosis. Equal numbers of participants are randomly selected from each group</td>
</tr>
<tr>
<td>Cluster</td>
<td>The population is divided into smaller sub-sets or clusters. These clusters are randomly selected and all or a random selection of individuals within these clusters are selected. Commonly used in large national surveys to enable random sampling in diverse populations</td>
</tr>
<tr>
<td><strong>Non-random sampling</strong></td>
<td></td>
</tr>
<tr>
<td>Convenience</td>
<td>The use of the most readily accessible individuals or units in a study. Used in exploratory research to obtain an estimate of a particular element of interest</td>
</tr>
<tr>
<td>Consecutive</td>
<td>A version of convenience sampling where every available individual or event within an accessible population is chosen. The best choice of non-random sampling</td>
</tr>
<tr>
<td>Snowball</td>
<td>Individuals recruited for a study refer other potential participants. Useful when participants maybe known within networks rather than generally or when a desired study characteristic is rare</td>
</tr>
<tr>
<td>Quota</td>
<td>This method ensures that specific individuals or units are included equally in a convenience sample, for example according to age group or sex</td>
</tr>
</tbody>
</table>
criteria are generally based on the research question and the research plan; they are applied to enable selection of homogenous samples and improve the feasibility of conducting a study. For example, it may be considered that pain management outcomes for day surgery patients may be better studied in patients undergoing abdominal and pelvic procedures. Exclusion criteria are generally applied to exclude unique characteristics that may confound the results or to deal with ethical considerations relating to the research. For example, it may be necessary to exclude patients who have additional medical problems that may affect their outcomes, or to exclude patients who have cognitive impairment that impairs their ability to provide informed consent to participate in the study.

Once the target and accessible population characteristics have been defined, a sample is selected from the accessible population. There are two main types of samples: random and non-random samples, also referred to as probability and non-probability samples (see Table 1 for definitions of common techniques used to sample within these categories).

Random or probability sampling refers to sampling processes that guarantee that each of the potential participants, events or units under investigation have an equal chance of being selected. Random selection is important for two reasons. First it reduces the risk of selection bias, and second, a randomly selected sample is a requirement for inferential statistical analyses.

In non-random or non-probability sampling, the sampling technique is not random; therefore, all members of the population do not have an equal chance of being selected for recruitment into the study. Non-random samples cannot be assumed to fully represent the target population and consequently, conclusions about the generalisability of results to the target population should be qualified.

Although random sampling is the preferred sampling method for quantitative research, it can be very difficult to achieve due to time, cost and ethical considerations. It is often necessary to use other sampling techniques. The use of appropriate sampling methods is one part of the process of ensuring a representative sample, the size of the sample is also important.

Calculating sample size

Sample size calculation in quantitative research depends on a number of factors. These include:

1. research design;
2. sampling method;
3. the degree of precision required;
4. the variability of the factors being investigated;
5. the incidence of a particular variable in the population.

As a general statement, the larger the sample the higher the likelihood that the findings will accurately reflect the population because larger samples have lower sampling error. Sampling error refers to the notion of estimating population characteristics from data collected from a sample. If 10 samples were drawn from an accessible population, there would be 10 different patterns of responses to the same question. This is known as sampling error, as the sample size increases the lower the sampling error.

In descriptive studies, where the intent is to identify the proportion of a particular phenomenon in a sample, sample size calculation is based on the level of precision and confidence required of the results (Altman, 1980). This is determined by confidence intervals. Confidence intervals are a range of values that surround a value obtained from a sample that have a known probability (confidence) that the true population value is contained within that range. Normally the specified confidence is set at 95% and the interval width is determined by the clinical significance of the precision of knowing the occurrence of a particular phenomenon. For example if we were interested in determining the percentage of patients who experience moderate to severe pain in the first day after surgery, we may specify a level of confidence of 0.95 and an interval width of 10%. These criteria form the basis for calculating the sample size for the study. If the findings show that 50% of the sample experience moderate to severe pain we can have 95% confidence that 50% ± 5% (between 45% and 55%) of the population of patients experience moderate to severe pain.

In explanatory studies where hypothesised cause–effect relationships are tested, researchers base sample size calculations on three desired factors:

1. The significance level — normally set at 0.05 or 0.01 (p value). Statistical significance refers to reducing the probability of finding an effect when the effect does not exist, referred to as Type I error. When the level of significance is 0.05, the probability of a Type I error is 5%.
2. Power — normally set at greater than 0.8 (80%). Power refers to the probability of finding an effect when the effect does exist. In situations
of low power (inadequate sample size) important effects may not be detected. This is referred to as Type II error.

3. Effect size refers to the minimum size of the difference to be detected and is determined by the clinical importance of the difference. The effect size may be derived from a similar study, pilot work or estimates of the likely effect.

It can be seen that sample size determination in explanatory studies is a crucial part of the overall validity of the study (for an example see Botti et al., 1998). In studies where the effectiveness of a particular intervention is tested, the size of the sample will determine the likelihood of detecting an effect of the intervention if that effect exists. Inadequate sample sizes may result in failure to detect clinically important small to medium effects of interventions.

Controlling for bias

The best way to evaluate the suitability of a sample in terms of its representativeness of the target population is to scrutinise the sampling method and evaluate the size of the sample. Potential biases in sampling can be reduced by rigorous recruitment strategies that increase response rate and reduce the likelihood of systematically excluding some people from the sample or over representing others.

Qualitative designs

Qualitative designs rely on saturation based sample selection to satisfy the sampling principle of adequacy. Appropriateness of sampling is usually achieved through purposive or theoretical sampling.

The concept of saturation

Sampling to the point of saturation requires the researcher to continue to recruit participants until no new data emerge. However, ethics committees require an indication of likely recruitment; five to eight participants are usually sufficient for a homogenous sample and 12–20 for a heterogenous sample, where it is important to maximise variation across the sample (Kuzel, 1999). When exploring the information needs of relatives in ICU, you may wish to include relatives of both long and short stay patients, requiring a sample of five to eight in each category. If you were to explore the same topic with nurses of varying experience and ICU qualification, a single focus group of eight is unlikely to be adequate. A key strategy to ensure saturation is the seeking out of confirming and disconfirming cases.

Purposive sampling

Purposive sampling is commonly used to allow the researcher to include participants with key experience of the issue to be studied. Decisions about whom to sample are typically made before data collection commences. Again, a key decision is whether you need a homogenous or heterogeneous sample. Also bear in mind that the synergy you would wish to create in a Focus Group interview may require you to sample a multi-professional team.

Specific qualitative methodologies also assist in purposive sampling decisions:

1. Grounded theory studies require a homogenous sample and are referred to as ‘theory-based’;
2. phenomenology studies require sampling those who have experienced the phenomenon – a ‘criterion based’ sample;
3. ethnography studies require sampling of the cultural group – a ‘representative based’ sample.

Random sampling is inappropriate for qualitative studies as key informants may be missed through the randomisation process; randomisation also assumes that the issue to be explored is normally distributed across the population sampled (Kuzel, 1999).

Theoretical sampling

With theoretical sampling, the process of data collection is guided by the emerging theory and sampling decisions are taken during data collection.

Controlling for bias

The very nature of qualitative research, with its focus on deliberate targeting of the sample, lends itself to criticisms of bias. Unfortunately, discussion related to the process of identifying saturation is frequently omitted from published research. One approach frequently used to reduce bias is triangulation. The traditional definition of triangulation uses the analogy of navigation, where knowledge
of two points is enables calculation of a third. This is often referred to as triangulation for convergence. Within qualitative studies the term is also used to indicate the use of multiple strategies/sources to provide a fuller picture of the situation being studied, referred to as triangulation for completeness (Redfern and Norman, 1994). For a fuller description of triangulation, see the first paper in this series (Endacott, 2004).

**Summary**

Regardless of the research approach taken, adequacy and appropriateness of the sample are essential components of research design. It is equally essential that these processes are described for consumers of research, both to de-mystify the steps involved and, more importantly, to allow the reader to judge the credibility of the findings (Selby et al., 1990). All decisions made by the researcher about sample size should be made explicit and the steps taken to control bias should be described.

**References**