Sampling in quantitative research (1)

KEY WORDS
- Generalisability
- Non-probability
- Probability
- Random
- Representativeness
- Sample

PETER ELLIS
Registered Manager at The Whitepost Health Care Company; Independent Nursing and Health Care Consultant, Writer and Educator

In this paper we will look at the methods used to generate samples for quantitative research, most specifically experimental studies using the example of a randomised control trial (RCT). It is useful to have some understanding of this process when reading such research as it allows you to make some informed decisions about how well the research might apply to the sort of people that you care for.

The approach to generating samples varies not only between quantitative and qualitative research but, as we will see in this paper and the next in this miniseries, also between quantitative research methodologies. Subsequent papers in the series will look at the various approaches to sample selection within the qualitative research methodologies.

WHAT IS QUANTITATIVE RESEARCH?

In an earlier paper in this series, we identified the key features of quantitative research (Ellis, 2014) as being research which seeks to answers questions that have a numerical element to them; primarily quantitative studies look for associations between variables (so called cause and effect). Quantitative research follows one of two broad paths either being interventional (as in experiments which manipulate variables) or observational (quantifying a phenomenon or the associations between naturally occurring phenomena). In this paper, we are exploring the former, interventional path.

WHAT IS SAMPLING?

On the whole it does not make economic or practical sense to study the whole of population in order to understand the answer to a medical question. For example, it would not make sense to study the effectiveness of a new dressing for everyone with venous leg ulcers in the UK, rather it makes sense to study the effect on a carefully chosen subset, a sample, of such people. Much quantitative research, however, seeks to be generalisable. That is the purpose of quantitative research is to provide answers to questions and then consider "how confidently we can extend the results from a sample to the population from which the sample was drawn" (Murad et al, 2018).

In principle this means sampling for quantitative research needs to be undertaken in a considered way such that we, as clinicians, can be confident that the findings of the research can be used to help us make decisions about the care of the patients in front of us. This generalisability is the essence of evidence-based practice.

There are two broad approaches to selecting a sample for a research study, these are probability and non-probability sampling. Probability sampling creates a sample using methods which are random, while non-probability sampling uses methods to select participants for a study which are non-random (we will discuss non-probability sampling in a later paper in this series).

WHY IS RANDOMNESS IMPORTANT?

Randomness is important in the selection of people for study because randomness ensures that people within the study group(s) are truly representative of the population form which they are drawn. This may seem odd, because we could of course identify the characteristics we want represented within a study, however, this would not account for variables, such as some we cannot see or control, e.g. genetic makeup or psychosocial issues.

Understanding the process of sampling is perhaps best understood by considering how it is applied to some quantitative research methodologies.

SAMPLING IN EXPERIMENTAL STUDIES

In experimental studies the choice of sample is dependent on the hypothesis which is being tested. The most common form of experimental design study in health care are RCTs. At its simplest, a RCT will involve two groups those under test and those they are being compared to, we will assume this is the case in the following worked example.

Let's return to the example of venous leg ulcers and their management. If we want to compare a new compression bandage used in the management of the ulcers, first of all we need to define what we mean by a venous ulcer and where such an ulcer might be situated on the body. Next, we need to define the group of people we are including in the study: do we

include people who are obese, diabetic or very elderly or not? This is important, as the generalisability of the study would be limited to the types of ulcers and the characteristics of the people we include.

When we have decided on the definition of ulcer and the characteristics of the people we will include in the study, we have a study population — in theory this could include everyone who meets the criteria, but in practice it actually means everyone who meets the criteria to whom the researchers have access. While this suggests that such study populations are more convenient (drawn from locally available patients) than random (Murad et al, 2018), researchers do have to work within the constraints of time, money and practicality.

From this study population, we are then in a position to select a study sample. If the approach to selection for this study sample provides everyone within the study population with the same chance of being included, subject to the criteria which we have applied — which may also include ethical criteria such as having the ability to consent — this is called probability sampling.

Probability sampling, therefore, refers to the fact that all people within the study population have an equal chance of being involved in the study. This is important in protecting and promoting representativeness in the study. Quatember (2019) states: *"A sample is called 'representative' with respect to a certain population characteristic (such as a whole distribution of a study variable or a parameter of this distribution) if this characteristic can be (at least approximately) unbiasedly estimated from the available data with a predefined accuracy".* In actual fact, the size of the sample used for the study plays an important role in ensuring the representativeness of the sample; very often the bigger the sample taken from the study population the greater the representativeness and hence the more confident we can be as to the generalisability of the findings.

Next, if our study is setting out to test a new compression bandage against an existing one, we have to choose the groups into which the people selected for the study sample will be divided. What is important at this stage is that the study sample is seen as being homogenous; that is the people share the characteristics we defined as necessary for inclusion, so they are broadly similar in terms of the things we can see and measure, as we approach splitting the sample into two groups the cases and the controls (more on this in a moment).

The division of people into the study groups needs now needs to ensure that the people allocated to each group are broadly similar in all important respects. So, if we call the people who are put into the group having the new compression bandage the cases and those allocated to using the old bandages the controls, then we have created the conditions for an experiment to take place.

This is in fact a little simplistic, since what we really need is to divide the two groups in such a way that they are probably close to identical in all important respects. This is not usually achieved, as one might think, by sharing out the known characteristics between the two groups in a managed way (e.g. matching gender, age, grade of ulcer), rather it is done by dividing the sample population randomly into the study groups. In theory, and with a large enough sample size (the assumptions and process for which is usually clearly delineated in the study write up) the two groups will be broadly similar in all respects; both those which can be seen and measured and those which cannot. Randomisation is achieved through the use of randomisation tables or more commonly is computer generated.

Study staff do not allocate people to separate groups are they most likely, consciously or unconsciously, allocate people to an arm of the study they think most suits the person or the aims of the study. Randomly allocating study subjects therefore voids the introduction of selection bias to the sampling process (Pocock, 1997).

CONCLUSION

We have identified that the approach to selecting a study sample for a RTC needs to take account of various issues such as generalisability, homogeneity and the avoidance of bias. Probability sampling is important and randomisation done properly helps researchers avoid the trap of selection bias. WUK

REFERENCES

Ellis P (2014) Decoding science: the language of research (part 1): research paradigms. *Wounds UK* 10(2):118–9

Murad MH, Katabi A, Benkhadra R, Montori VM (2018) External validity, generalisability, applicability and firectness: a brief primer *BMJ Evid Based Med* 23(1): 17–9

Pocock S (1997) *Clinical Trials: A practical Approach*. John Wiley, Chichester

Quatember A (2019) The representativeness of samples (Chapter 48). In: Darquennes J, Salmons J, Vandenbussche W. eds, *Handbook on Language Contact. Series Handbooks of Linguistics and Communication Science*. de Gruyter: Berlin