Sampling in quantitative research 2: non-experimental quantitative studies (1)

KEY WORDS
- Case-control
- Causality
- Cohort
- Cross sectional
- Generalisability
- Sampling

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

In the last paper in this miniseries, we identified the approach taken to generating a study sample for experimental studies; more specifically we examined randomised controlled trials (RCTs). We saw how in RCTs it is important to take account of generalisability, homogeneity and the management of bias. We identified the sampling approach as being probability sampling, ensuring all people within the study population have an equal chance of being selected for the study. We also demonstrated how randomisation within the process enables researchers to avoid selection bias.

In this paper, the second on sampling in quantitative research, we will consider the nature and reasons for the sampling methods used in cohort and case control-control studies. We will identify specifically why the given approaches are used and how these contribute to the aims of the studies in question.

COHORT STUDIES

Cohort studies are predominately prospective; that is they collect data during the life of the study and do not use previously collected data. Cohort studies follow a group, or groups, of people over time and measure the incidence (the new occurrences) of predetermined and well-defined outcomes (e.g. new cases of a disease) in that group (Gordis, 2014). Cohort studies tend to go on for years, or even decades, and are often used to demonstrate the cause and effect relationship between lifestyle and health outcomes (Ellis, 2016).

Sample selection in cohort studies depends, as with all sampling, on the exact question, or questions, being asked. For example, if the study is interested in the development of diabetic foot ulcers among people diagnosed with insulin-dependent diabetes in their teenage years, then the sample must only include people who have had such a diagnosis. Moreover, the sample, or cohort, used in the study must not have foot ulcers at the start of the study because, as we saw in the definition, cohort studies seek to uncover the incidence (new occurrences) of an outcome.

On the other hand, if the study is interested in identifying the incidence of several undefined outcomes (e.g. diseases) in a group of individuals over a period of time, then a more general, but well-defined group is chosen to be studied. For example, nurses, as in the justly famous Nurses Health Studies in the US, which in various iterations has followed the wellbeing of nurses over many decades (The Nurses' Health Study Group, ND).

In common with RCTs, cohort studies need a comparator group in order to demonstrate a cause and effect relationship between an exposure (say diabetes) and an outcome (say a diabetic foot ulcer). Again in common with RCTs, the comparator group has to be as broadly similar to the affected group as possible, such that single or small amounts of the difference between the two might be identified as being the cause of a specific disease outcome. In cohort studies, which are studying rare, or undefined, outcomes, this is achieved by studying large groups of people over extended periods of time. Again the people in the groups should be generally homogenous, with detailed information collected about many aspects of their lives, including things like diet, exercise and eating habits in order to identify the one thing that is different and which may cause the outcome of interest (for example, the Nurse Health Studies have been, and are still, collecting data on over 275,000 nurses since 1976).

More specific cohorts with estimable and relatively high outcomes of interest may require much smaller cohorts to study (for example, regular attendance, or not, at an integrated foot care service and the incidence of diabetic foot ulcers [Paisey et al, 2019]).

The big issue with cohort studies is the ability to collect data consistently from members of the cohort over long periods. Losses to follow up may indicate something specific about individuals and may be associated with the outcome of interest – the data thereby being lost.

CASE-CONTROL STUDIES

Case-control studies are not like RCTs and cohort

studies in that they do not seek to prove a cause and effect relationship and they are not prospective nor longitudinal. Case-control studies seek to identify associations between an outcome of interest and potential causes – identifying potential causes lays the foundations for future studies of causality which may include cohort studies (Ellis, 2019).

In direct contrast to cohort studies or RCTs, the sample for a case-control study is drawn from people who have an outcome of interest, again say a diabetic foot ulcer, and comparing them, or more specifically comparing their exposures, with carefully matched individuals who do not, at least at this time, have the outcome of interest (the controls).

In common with all quantitative research methodologies, what constitutes case needs to be well defined. Returning to our example of diabetic foot ulcers, we might define a case as someone experiencing a diabetic foot ulcer for the first time, or perhaps by Wagner grade.

As well as considering the outcome of interest, it is also important to identify where the cases are drawn from. Again this is about considering their representativeness of a given population and the ability to generalise from the findings of the study. For example, there may be a significant and important difference between people who attend diabetic foot clinic in an affluent country and those attending similar clinics in a developing country.

It is also important to understand that in some study situations, and this may include people with diabetic foot ulcers, selecting cases from pre-existing (prevalent) rather than new (incident) cases may have a bearing on the outcomes and generalisability of the study. Perhaps the most important reason for this is that prevalent cases represent people who have survived with an outcome of interest which may have already claimed the lives of others (perhaps through infection and sepsis in the case of diabetic foot ulcers). People who survive with disease outcomes which are associated with increased mortality may be in some way significantly different from those who do not and therefore are not representative of the population of interest. In such instances, it may be better to use incident cases for the study.

CROSS-SECTIONAL STUDIES

Cross-sectional, also known as prevalence, studies. Are used to measure the prevalence of an outcome or exposure of interest in a particular group at a given point in time (Ellis, 2019). Again returning to our example of diabetic foot ulcers an example might be the prevalence of grade 2 ulcers mong people attending a particular foot clinic in a given period of time. The definition of cases of interest might also extend to include the ages, ethnicities or perhaps length of time since a diagnosis of diabetes.

Samples for cross-sectional studies are generally taken from populations among which the exposures or outcomes of interest is already known to highly prevalent. So in the case of diabetic foot ulcers, it would seem wise to draw the population from either a general surveillance clinic for a low grade or incident ulcers or the specific foot clinic for higher-grade or prevalent ulcers.

The whole purpose of prevalence studies is to establish the size of a health problem within a well-defined population. Establishing prevalence is important in designing and resourcing health provisions for example and for considering the utility of a screening programme, especially where the disease being screened for is rare.

CONCLUSION

In this paper, we have examined the approaches to sampling in cohort, case-control and cross-sectional studies. We have seen why the specific approaches have been chosen and what these mean for the outcomes under study. We have seen that issues such as determining causality, looking for associations or merely quantifying an issue of interest drive such study designs and their attendant sampling methods.

In the next paper in this series, we will start to consider the nature of sampling in qualitative research, exploring the different qualitative methodologies and attendant sampling methods. We will also in a future paper in this series explore the meaning and value of incidence and prevalence. WUK

REFERENCES

Brewis J (2014) The Ethics of researching friends. *Brit J Manage* 25: 849–62

Mamotte N, Wassenaar D (2015) Measuring voluntariness of consent to research: an instrument review. *J Empirical Res Human Res Ethics* 10 (2): 121–31

Miller F, Wertheimer A (2010) *The Ethics of Consent: Theory and Practice.* Oxford University Press, Oxford

Usher KJ, Arthur D (1998) Process consent: a model for enhancing informed consent in mental health nursing. *J Adv Nurs* 27: 692–7

Whelehan P, Evans A, Ozakinci G (2015) Informed choice and consent among women attending for breast screening in the UK: data from a qualitative study. *Breast Cancer Research* 17(1): 06