The language of research (part 9)—research methodologies: cohort studies

KEY WORDS
- Cause and effect
- Cohort
- Incidence
- Prospective study

PETER ELLIS
Nursing Director, Hospice in the Weald, Pembury, Tunbridge Wells

Previous papers in this series have explored research paradigms, quantitative and qualitative methodologies, and some of the terminology associated with research. In this paper, we will explore the third of the quantitative methodologies that this series will decode: cohort studies.

We should remember that a methodology is general research approach (the blueprint for a way in which the research is done) and reflects the academic and philosophical underpinnings of the research.

WHAT IS A COHORT STUDY?

A cohort study is one of only two accepted methodologies for demonstrating a cause and effect relationship in epidemiological research, the other being the experimental study trials (such as randomised controlled studies). Usually, cohort studies are prospective rather than retrospective (they go forward in time, as opposed to using previously collected data) and follow a group of individuals, measuring the incidence of predetermined specified outcomes (Gordis, 2014).

For example, a cohort study might follow a group of patients after a specific surgery to measure the incidence of wound infections. The purpose of a cohort study is to link an exposure or cause (in this case, surgery) with an outcome or effect (in this case, a wound infection). Cohort studies are able to determine the causality of an outcome of interest, because they ascertain its absence before the study starts and then quantify exposure and outcomes.

WHAT IS A COHORT?

A cohort is simply a group of people who share one or more characteristics — in the case of cohort studies, an exposure of interest. Unlike experimental studies, cohort studies do not require the researchers to undertake an intervention – they are purely observational. In our example, people undergoing a specific surgery might constitute a 'cohort'.

A FAMOUS COHORT STUDY

One famous cohort study is the Nurses' Health Study (NHS I), which started in 1976 and was further extended in 1989. In 1976, NHS I enrolled married registered nurses aged between 30 and 55 who lived in US, where the nursing boards had agreed to supply the nurses' names and addresses. Approximately 122,000 of the 170,000 nurses approached enrolled in the study. From 1976, every two years, members of the cohort have completed questionnaires about illness, disease and health-related topics, including potential causes of ill health, such as smoking, and the use of medical interventions like hormone replacement therapy.

Since 1980, the cohort have also received questions about their diet every four years, and since 1990, they have answered questions about their quality of life. Response rates to the questionnaires have been high, at about 90% throughout (www.channing.harvard.edu/nhs).

In 1989, a second study (NHS II) commenced using younger participants, aged 25–42, to collect new data, such as the use of contraception, diet and lifestyle.

Since the initial cohort was large in size and has been followed for many years, the NHS studies can provide over 4.5 million years of data about nurses' lives, exposures, diets, health-related behaviours, illness and disease. Notably, these studies have produced hundreds of research papers exploring disease risk factors for women.

DOING A COHORT STUDY

The question(s) that a cohort study sets out to answer will determine the sample for the study. For a study interested in a specific exposure and outcome in a precise group of people — e.g. people who have undergone surgery and might develop a wound infection — only people who are exposed to the criteria of interest — surgery — can be selected to be in the sample.

Conversely, if a study is set up to measure a number of outcomes, then a more general group of individuals are chosen and followed over time, and a range of exposures are measured and observed in order to see what outcomes occur (the NHS studies are a good example of this approach).

In order to demonstrate cause and effect, a comparison group is needed so that the researcher can compare exposure in one group to non-exposure in the other, and thus determine the extent to which the outcome of interest is caused

by the exposure (demonstrating causality). One might determine the incidence of wound infection in patients undergoing conventional surgery and compare the incidence of infection to that seen in patients having the same or similar surgery using a keyhole technique. This way, the incidence of infection in the conventional group minus the incidence of infection in the keyhole group might be claimed to demonstrate that conventional surgery causes x-amount of wound infections.

The two groups need to be as similar as possible in order to be certain that the exposure of interest (e.g. conventional surgery) might be said to be the cause of the outcome of interest. In studies like the NHS I and II, the comparison is achieved by studying a large group of people of similar ages over a period of time and measuring their lifestyle (exposures). Then, when an outcome of interest arises, such as a disease, the incidence of the disease in the group not exposed to the potential cause and the incidence of disease in the group exposed to the potential cause are compared. A good example in the NHS studies was the higher incidence of deep vein thrombosis in women taking the mini pill as opposed to those not taking the mini pill.

Of course, as well as there being a statistical association between an exposure and an outcome, there has to be some biological plausibility too. That is to say, the incidence of wound infection following conventional surgery as opposed to keyhole surgery is likely to be a result of the surgical approach and not the fact that one surgeon plays classical music in theatre and the other plays jazz.

Cohort studies usually collect their data via self-completion questionnaires. The use of these questionnaires is a practical one as cohort studies are so large and take such long periods of time that individual visits, or data collection by study staff, would be prohibitively time-consuming and expensive (Ellis, 2016).

Sometimes cohort studies collect extra data from a selection of participants, e.g. NHS I collected toenail clippings (to examine mineral content) and blood samples from many participants over a period of time to test some hypotheses.

Sometimes cohort studies are undertaken retrospectively (they look back in time) and collect existing data about an exposure and an outcome. Researchers place much less faith in retrospective studies because the data are often incomplete and tend to rely on the recall of participants about things such as their smoking, exercise or dietary history.

Relying on memory can lead to a form of bias (anything in the design or undertaking of a study that causes an untruth to occur) called recall bias. One well-known example is given by Last (1995) who shows that mothers of children with leukaemia are better at remembering their history of having had X-rays while pregnant than mothers of children without leukaemia.

ANALYSING A COHORT STUDY

Data from cohort studies tell us in crude terms how many people develop an outcome of interest during the period of the study (the incidence of the outcome of interest). The data also tell us how many people in the cohort who were similarly subjected to the exposure of interest did not develop the outcome. On the other hand, cohort studies also supply data about people who were not exposed and who did or did not develop the disease of interest.

Take the example of post-surgical infection. Some people who have conventional surgery will develop an infection and some will not; the same is true for the keyhole surgery. What is of interest, however, is how much bigger the odds are of getting an infection in one group or the other. The measure of the difference of the odds of developing an infection in either group is called the odds ratio. The bigger the odds ratio, the stronger the association between the exposure and outcome of interest — although, like all statistics, this is subject to the study being of sufficient size.

CONCLUSION

Cohort studies are an important way of measuring cause and effect in the development of disease, illness and other healthcare phenomenon. They are best undertaken prospectively and for rare outcomes (like cancers) undertaken over a long period of time. As well as there being a statistical association, the outcome has to make biological sense. WUK

REFERENCES

Ellis P (2016) *Understanding Research for Nursing Students*. 3rd edn. Sage, London

Gordis L (2014) *Epidemiology*. 5th edn. Saunders, Philadelphia

Last JM (1995) *A Dictionary of Epidemiology*. 3rd ed. Oxford University Press, Oxford