Conducting a survey

Wai-Ching Leung explains how to conduct a good survey

Most medical students carry out at least one project involving a survey in either a core module—for example, in public health or primary care—or a special study module. The survey methodology is popular among students for two reasons. Firstly, it seems familiar and easy to do. Most students have taken part in either an interview or questionnaire survey, and many have conducted a survey in their secondary school days. Secondly, people are interesting, and the survey is a useful tool to gather a wide range of information about them. It can be used for a range of disciplines, such as medical or health studies, business studies, education, biology, and sociology.

It is indeed easy to administer questionnaires and carry out interviews. However, there are so many potential pitfalls in conducting a survey that few of them produce valid and meaningful results. The purpose of this article is to outline briefly the basic steps in carrying out a survey and the potential pitfalls.

What is a survey?

A survey is a method of collecting information from a sample of the population or sometimes the organisations we are interested in. This may involve gathering information either at one point in time—that is, cross sectional studies—or following a group of people over a period of time—that is, longitudinal studies. Most non-academic surveys—for example, surveys in market research—are usually of the first type. The type of information that we can gather from people include factual information, their level of knowledge, attitude, personalities, beliefs, and preferences.

The box shows the steps in conducting a survey. Potential pitfalls occur in each of these stages and it is important to consider each of them in detail.

Stages of conducting a survey

- Clarify the purposes
- Define the study population
- Sampling and estimating the sample size
- Decide what information to collect
- Decide how to measure the information
- Collect the data
- Record, analyse, and interpret the data.

Clarifying the purposes

It is important to be absolutely clear and explicit about the purposes at the start. Generally speaking, surveys can be used for two purposes.

Firstly, we may wish to know how common a characteristic is—that is, a descriptive survey. An example would be the proportion and characteristics of students who read the *studentBMJ*, or the proportion and characteristics of people who use a particular product. Such purposes can be achieved by collecting information from a sample of students at one point in time (cross sectional survey). If we repeat the cross sectional survey periodically we can gather information about the time trends—for example, whether the *studentBMJ* is becoming more popular.

Secondly, we may wish to learn something about the causes for these characteristics—that is, analytic survey. An example would be if we wanted to know how students' learning styles at the start of the medical course affect their final course results. A cohort (longitudinal) study following a group of first year students until they graduate is more likely to yield the required information, as this allows the initial learning styles to be accurately assessed without being influenced by the knowledge of the final course results.

Define the study population

The next step is to define exactly whom we are interested in studying. It is vital to ensure that this definition corresponds to the purposes of the survey. This usually includes specific personal criteria, time and place. For example, to determine the extent to which medical students in the United Kingdom read the *studentBMJ*, our study population may include all students who have registered for a medical degree in a United Kingdom university on a specified date. Some surveys require more specific criteria. For example, to study the factors contributing to women's non-attendance for cervical cancer screening we must exclude from our study population women who are not eligible for such screening—for example, those above or below the recommended screening age or those who have had a total hysterectomy. Otherwise, the results would not give a valid answer to the original question.

It would be helpful if we have some baseline information about this study population. For example, if our study population consists of all doctors in the United Kingdom on a particular date, we know the overall age, sex, ethnic, and specialist status of all doctors from various official sources—for example, the General Medical Council's

Sources of pitfalls or biases in conducting a survey

Steps	Possible pitfalls and biases	Possible precautions/ remedies
Clarify the purposes	Intended question not answered	Make explicit the purposes from the start
Define the study population	Relevant people not studied Lack of baseline denominator information	Make explicit the study population Consider sources of baseline denominator information
Sampling and estimating the sample size	Sample not representative of population studied Sample size insufficient to answer the intended question	Use appropriate sampling method Calculate sample size before the survey
Decide what information to collect	Essential information not collected Long questionnaires or interviews (from collecting unnecessary information) causing a low response rate	Consider carefully what information is necessary
Decide how to measure the information	Invalid measurement of information	Use validated measurement tools Carry out pilot studies
Collect the data	Non-consent and non-response bias Social undesirability bias Observer bias	Approach and explain survey carefully and politely Send reminders to non-responders Hold training sessions for interviewers
Record and analyse the data	Inaccurate recording of results Wrong analysis	Use optical scanning, computer assisted questionnaires Use statistical packages

Medical Register or the *Medical Directory*. This information is useful for at least two purposes. Firstly, if only a low proportion of doctors respond—that is, a low response rate—and we would like to know whether those who respond are similar to those who do not. Secondly, if we want to compare two subgroups of doctors—for example, different seniority or different ethnicity—this baseline information is important.

Sampling

If we decide to collect information on the whole study population, the study is called a census. However, our study population is usually so large that we do not have the time and resources to study all individuals. Instead, we collect only information from a proportion—that is, a sample—of the study population. The process of selecting this sample from our study population is known as sampling.

The sample chosen must be representative of the study population. For example, to find out the proportion and characteristics of students who read the *studentBMJ* it is tempting to survey medical students via the medical students' email discussion group, as this is easy, cheap, and convenient to carry out. The results, however, would not be valid. Students who have joined the email group are a self selected group who are probably more likely to read the *studentBMJ*. In other words, they are not representative of all medical students.

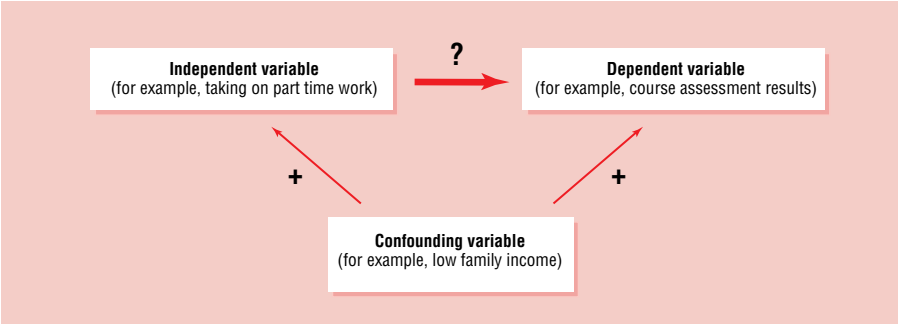
To ensure that the sample is representative of the study population, each student must have an equal chance of being sampled. A common method is by simple random sampling—that is, each student has an equal chance of being selected. This can be carried out using random numbers generated by a computer. This sampling method assumes that we have a list of all subjects in the study population—for example, a list of the names of all medical students in the United Kingdom. However, we often do not, and other sampling methods—for example, cluster method, multistage sampling—must be used.

How many people do we need to survey? Clearly, a larger sample size would yield more precise results. On the other hand, this is often not possible due to limited resources. The sample size needed depends on several factors: the purpose of our survey—for example, descriptive or analytic—how common the main dependent variables are among our sample population; the amount of variation of the factor we are interested in; how precise we need our results to be. Once we are clear of these issues, the formulae for estimating the required sample size are readily available in most statistics textbooks.

What information do we collect?

Let us take for example a survey exploring the effects of students taking on a part time job on their course results. Clearly, we have to collect two types of information: those which we are primarily interested in or dependent variables—for example, performance in the assessments in various components of the course—and those which might explain the dependent variables or independent variables—for example, the number of hours a week the student works.

Suppose we find that students who take on part time jobs do worse in the course



How a third (confounding) variable might create an apparent association between two variables

Epidemiology Notes

In this article, Wai-Ching Leung mentions the various forms of bias which could potentially kill your survey, if you do not consider them before hand. On page 153 David Ogilvie defines bias as 'a feature of the study which makes a particular result more likely – like a football pitch which slopes from one end to the other.'
This glossary on the different forms of bias may help to ensure that you consider them before-hopefully not after- you conduct your survey.

Information Bias This occurs when systematic differences are introduced in the measurement of the response. Two such examples are ***recall bias*** and ***observer bias.***

Recall bias is when a difference occurs because some people are much more likely to remember and event than others. This is typical in a case control study when the cases are more likely than the controls to remember an adverse event.

Observer bias can be as a result of differences between different observers-***inter***- or with the same observer- ***intra***. To eliminate this, it is important that all observers use a standardised method of measuring or collecting data. If you are the only observer, you will still need to have a standard and systematic method of measuring or collecting your data to make sure that your results are not dependant on your mood.

Non-response bias is a researchers nightmare. It arises when those who respond to a questionnaire-responders-differ in some way from those who don't-non-responders. Most researchers try to reduce this as much as possible by trying to a) maximise the response rate- by sending out reminders, having incentives for responding etc-or b) by identifying the characteristics of the non-responders-age, sex, deprivation score etc-so they can see whether they are any different from the responders. They can then make adjustments in their analysis for the non-responders.

Selection bias results when the sample group you have chosen is not representative of the population you want to generalise your results to. Wai-Ching Leung mentions the importance of random sampling to stop this from happening in your survey.

Rhona MacDonald *Editor student BMJ*

assessments. This result may be accounted for by other factors—for example, previous academic achievements, family income, etc—which are related to both dependent variables—for example, how well students do in the course assessments—and independent variables—for example, whether they take on part time work. For example, it might not be the part time work which directly affects the course assessments, but that students from a poor family may be more likely to take on part time work as well as do worse in assessments (see figure). These are known as confounding variables, which we must also collect information on.

How to measure

Some information—for example, course assessment results, number of siblings, income, etc—is easier to measure than others—for example, knowledge about a certain topic, attitudes, experience. Measurement of information is a vast topic and it is not possible to give details here (see further reading below). Generally, it is important to use measurement methods which have been previously validated. Otherwise, pilot studies—that is, testing out the methods with smaller numbers of subjects—are essential.

Methods of collecting data

There are several possible methods of collecting data—for example, postal questionnaires to individuals or via organisations; computer assisted questionnaires; email questionnaires; online questionnaires; face to face interviews, and telephone interviews.

There are advantages and disadvantages of each of these methods, and the ideal method depends on who you are surveying and what the topic is. For example, email questionnaires are ideal for surveying university lecturers, but not for the homeless.

Non-response is an important source of bias. Those who respond to our survey are likely to differ from those who do not. Hence, it is important to maximise the response rates in whatever way we can. This might involve explaining carefully the purpose of the survey, approaching people carefully and courteously, and sending reminders to non-responders.

In an interview survey, it is important to ensure that all the interviewers follow the same interview protocol. All interviewers should adopt the same approach in explaining the survey, phrasing particular questions, and recording the responses. This will minimise any observer bias.

Record and analyse data

For small surveys, results can be easily recorded by hand and analysed using calculators. For larger surveys, more efficient ways of recording data—for example, optical scanning, online questionnaires—may be considered, and statistical packages—for example, SPSS, SAS—may be used for faster, more accurate, and more sophisticated analysis.

Conclusions

The table summarises some of the major potential pitfalls in conducting a survey. To conduct a survey properly, meticulous care is required at all stages.

Further reading:

Abramson JH, Abramson ZH. *Survey methods in community medicine.* Edinburgh: Churchill Livingstone,1999.
Bowling A. *Research methods in health.* Buckingham: Open University Press, 1997.

Wai-Ching Leung *lecturer in public health medicine, University of East Anglia, Norwich*
w-c.leung@uea.ac.uk