Surveys: An Introduction

Gordon D Rubenfeld MD MSc

Surveys are a valuable research tool for studying the knowledge, attitudes, and behavior of a study population. This article explores quantitative analyses of written questionnaires as instruments for survey research. Obtaining accurate and precise information from a survey requires minimizing the possibility of bias from inappropriate sampling or a flawed survey instrument, and this article describes strategies to minimize sampling bias by increasing response rates, comparing responders to nonresponders, and identifying the appropriate sampling population. It is crucial that the survey instrument be valid, meaning that it actually measures what the investigator intends it to measure. In developing a valid survey instrument, it can be useful to adapt survey instruments that were developed by other researchers and to conduct extensive pilot-testing of your survey instrument. *Key words: research, survey methods, surveys, research methodology, statistics.* [Respir Care 2004;49(10):1181–1185. © 2004 Daedalus Enterprises]

Introduction

In our everyday lives, we are surrounded by survey results. Political surveys (polls) tell us who is going to win elections, consumer surveys (market research) tell companies how to sell more soda pop, and telephone surveys determine the winners on *American Idol*. Surveys of clinicians and patients can yield answers to important questions: How does residency education affect knowledge of mechanical ventilation?[1] What are respiratory care managers' attitudes toward baccalaureate and master's degree education for respiratory therapists?[2] At first glance it might seem that a survey is an extremely productive and easy research method. One need only write down some interesting questions, distribute the survey to the subjects, and collect the responses. To some degree this is true, and survey studies can usually be performed for considerably less than, for example, multicenter randomized clinical trials. The results of valid surveys can inform important debates about many issues in medicine, including health care reform, medical error, and end-of-life decision making.[3–5] However, well designed, pilot-tested, valid surveys of a representative population are usually more difficult to perform than a novice investigator—or untrained reader—might imagine.[6,7] The present article briefly describes the types of research questions that can be answered by surveys, and the pitfalls in the design, implementation, and analysis of surveys.

For the most part this article explores quantitative analysis of written questionnaires as instruments for survey

Gordon D Rubenfeld MD MSc is affiliated with the Division of Pulmonary and Critical Care Medicine, Harborview Medical Center, Seattle, Washington.

Gordon D Rubenfeld MD MSc presented a version of this article at the RESPIRATORY CARE Journal symposium, "The ABCs of Research," at the 49th International Respiratory Congress, held December 8–11, 2003, in Las Vegas, Nevada.

Correspondence: Gordon D Rubenfeld MD MSc, Division of Pulmonary and Critical Care Medicine, Box 359762, Harborview Medical Center, 325 9th Avenue, Seattle WA 98104-2499. E-mail: nodrog@u.washington.edu.

research. Written questionnaire is not the only technique for performing a survey. In-person or telephone surveys can also be used and can be analyzed quantitatively or qualitatively. Qualitative analysis of written responses, particularly "free text" responses, is extremely useful for generating hypotheses and permits a richer, less hypothesis-driven approach to complex research questions.[8] However, those techniques are beyond the scope of this article.

What Questions Do Surveys Answer?

Surveys are primarily used to measure the knowledge, attitudes, and behavior of subjects. Clinicians are quite experienced with a common type of "survey" that is used to measure knowledge: the standardized multiple-choice tests used for clinical accreditation. These are a specific kind of survey used to measure *knowledge* about a selected subject. The responses are compared to "correct" answers determined by the survey designers, and scores are reported as correct answers or to an accepted scale. Surveys of *attitudes* usually do not have correct answers, and they are designed to elicit the subject's feelings about a topic. For example, political polls are designed to measure respondents' attitudes about a given policy or politician. Patient and family satisfaction surveys measure their attitudes about a recent interaction with the health care system.[9] Peer assessment surveys are used to measure one clinician's attitudes about another's clinical practice.[10] Although there is no accepted standard for measuring a respondent's attitudes, there are techniques to assess the validity of surveys designed to measure attitude.[11]

Surveys can also measure *behavior*. Behavior surveys are usually self-reported and therefore suffer that limitation. Respondents may be reluctant to admit to certain behaviors (smoking, for example) and may not remember other behaviors (what did you have for breakfast this morning?).[12]

Investigators hope that their survey results will reflect the true knowledge, attitudes, or behavior of the surveyed population. To the extent that the survey results deviate from the truth, the survey results are *biased*. There are 2 important sources of bias in a survey study: the sample population and the survey instrument. In *sampling bias*, the results of the survey systematically deviate from the truth because the respondents who were sampled are different from the population that the investigator intended to describe, in ways that affect the survey responses. Bias from the survey instrument occurs when the wording, order, and/or coding of the questions or instructions causes the respondent to enter inaccurate responses.

Minimizing Sampling Bias

Ideally, the respondents are a random sample of the population the investigator wishes to study, and the sample is sufficiently large to estimate with precision the variables of interest. For example, say an investigator is interested in determining respiratory therapists' attitudes about using noninvasive mechanical ventilation for exacerbations of acute chronic obstructive pulmonary disease. Consider the following sampling strategies: (1) sample attendees at an American Association for Respiratory Care symposium on noninvasive mechanical ventilation and obtain a 90% response rate, (2) mail a survey to Association members and obtain a 30% response rate, and (3) conduct a telephone survey of a 5% random sample of Association members and obtain a 95% response rate.

Sampling attendees at conferences is a convenient technique because the attendees are usually motivated and to some extent they are a captive audience. Surveys of conference attendees are inexpensive and usually yield a high response rate. If the investigator is primarily interested in the attendees' responses, for example when assessing whether the conference improved attendees' knowledge, then this is an appropriate sample. Unfortunately, if the goal of the survey is to understand the attitudes of a broad population of respiratory therapists, then meeting participants are probably a poor choice. A mailed survey to a broad population based on membership lists of a national organization is a reasonable sampling approach for reaching clinicians; however, the 30% response rate in the example suggests that respondents are only a minority of the population. The risk is that the respondents will differ in important ways from the nonrespondents. There is a large body of literature on techniques to deal with a poor survey response rate. These include strategies to increase response rate, such as reminders, hand-stamped envelopes, small incentives, and short, simple, surveys.[7] Investigators can also try to compare respondents to nonrespondents, to argue that the respondents are not significantly different demographically from the nonrespondents. However, demographic data on the entire sampled population is frequently unavailable, and investigators rarely know which type of respondent will hold which views.

If possible, identifying a small, random sample from a large population and then focusing resources on achieving a 100% response rate from the sample ensure a random representation of the population. However, it is not always possible to enumerate and identify the entire population before sampling, which is required in that approach.

There is no specific response rate that guarantees an unbiased representation of the population. To some extent it depends on how different the responses of the unsampled population are from the sampled population. As a general rule of thumb, most reviewers look for a response rate $\geq$ 70%; however, even with a response rate of over 70%, it is possible for a study to be biased if the minority that is unrepresented (the nonresponders) have strong opinions that significantly differ from the majority. The expected

response rate depends on the population and the nature of the questions. Clinicians and patients are thought to be motivated respondents, and surveys about medical care are nonthreatening, so a high response rate is expected. But surveys on some topics (eg, high-risk sexual behavior in difficult study populations) may never be able to achieve 70% response rates.

There are times when a biased sample is useful. When the direction of bias is known or can be assumed, the survey results can be helpful if they indicate the opposite of the presumed bias. The presumption is that a sample that included the entire population would find results that are even more extreme. For example, consider the research question, "What are the barriers encountered by clinicians when they try to implement lung-protective ventilation in patients with acute lung injury?" This research question can be approached by studying actual behavior, reported behavior, and/or by surveying people about their attitudes or knowledge. There are some aspects of this research question that make a random sample of clinicians problematic. If you are interested in *actual* barriers to implementing lung-protective ventilation (as opposed to barriers people think they might encounter), you must survey clinicians who have actually used or tried to use lung-protective ventilation. A random sample of clinicians, particularly early after the release of the study that showed the benefits of lung-protective ventilation, might have found very few clinicians who had tried to use it. Furthermore, even if a random sample identified clinicians with experience, the survey respondents might not have *expertise* in administering lung-protective ventilation. Alternatively, the investigator might specifically sample clinicians who have substantial experience and expertise in providing lung-protective ventilation. If that study identified no barriers, then it would not be particularly instructive; however, the barriers identified by such experienced respondents would probably also be barriers encountered by less experienced clinicians. A recent study took just that approach in selecting a "purposive sample" that was designed not to be a random sample, with the assumption that barriers identified by experienced clinicians at centers that have substantial experience would under-represent barriers likely to be encountered at less experienced sites.[13]

Minimizing Instrument Bias

Asking the right question the right way is the most difficult part of developing a survey. The easiest, and often overlooked, approach to avoid asking the wrong question or asking the right question in the wrong way is to use or adapt (with permission) someone else's survey. Assuming they knew what they were doing, this approach can save a lot of time. Questions can be either "open-ended" (the respondent can write in his or her answer in his or her own words) or "fixed-response" (the respondent is only allowed to choose from a list of responses on the survey form, and is thus forced to choose a specific type of response) (Table 1). There are advantages and disadvantages to both open-ended and fixed-response formats. Open-ended responses provide a rich source of detailed information, but they are difficult to analyze quantitatively. Fixed-response questions are relatively easy to analyze, but the questions must be fashioned very carefully. Poorly written questions will provide little useful information. Survey developers recommend using a fifth grade reading level in a survey instrument, even when sampling a well-educated population of clinicians.

Table 1. Types and Examples of Survey Questions

Type	Example Question
Open-ended	Can you explain why you did not use lung-protective ventilation in a recent case of ALI?
Fixed-Response	
Yes/No	Is physician recognition of ALI a barrier to using lung-protective ventilation?
Multiple-choice	Which of the following are reasons you don't use lung-protective ventilation for ALI (check all that apply)
Rating scale	Rate each of the following barriers on a scale from 1 (not a barrier) to 5 (very important barrier)
Ranking	Rank the following from most important to least important barrier to implementing lung-protective ventilation

ALI = acute lung injury.

Certain types of survey question should be avoided (Table 2). The layout and format of the written survey are important. The form should be easy to read and follow, and should use a large ($\geq$ 12-point) font. Shading and borders should be used so that it is easy for respondents to link the question stem to the response area. Middle and "other" categories of answers do not provide much information. For example, some surveys use a 5-point or 7-point scale to assess respondents' attitudes about a topic. The problem with that construct is that the mid-point number on the scale may be interpreted differently by different respondents: some will assume that the mid-point number indicates "I do not feel strongly one way or the other," some will assume that it means "I have no opinion on the question or do not feel qualified to answer," and some might be indicating "I prefer not to answer the question." Note that the latter answer-options can be offered next to (ie, in addition to) the point-scale. If the investigator feels obliged to add an "Other" category to allow for additional responses, the item should read "Other (please specify): _____" so that the information can be categorized.

Table 2. Types of Questions to Avoid in Surveys

Question Type	Reason to Avoid	Problematic Example Question
"Double-barreled" question	Combines 2 questions	To what extent do you think this report has changed your ability to understand and perform surveys?
"Halo effect"	Invokes influential source that may influence respondent's answer	Do you agree with David Pierson MD that respiratory therapists are effective clinicians?
"Loaded" question	Inflammatory language can influence respondent's answer	True or False: Increasing the role of intensivists will reduce the epidemic of medical error that is claiming the lives of thousands of critically ill patients?
Double-negative	Can make the question difficult to interpret	Which of the following do you not disagree with: I never use lung-protective ventilation
No comparator	Can make the question difficult to interpret	Do you prefer to use noninvasive ventilation?
Jargon	Can make the question difficult to interpret	With regard to the use of APRV, do you believe the P_{high}, T_{low}, T_{high} settings should be determined by RTs, MDs, or RNs?
Complex vocabulary	Can make the question difficult to interpret	Indicate your predilections regarding the ensuing options

A common research question in clinical medicine is how clinicians manage specific cases. A common approach is to provide a case vignette to the clinician and to ask how he or she would manage the case. Of course, the case vignette never has the depth and richness of an actual case, and investigators never know whether the respondents' survey answers actually represent how they really practice. Standardized tests for accreditation frequently use case vignettes, in which a "correct" answer to the case can be identified. However, investigators are often interested in studying differences in practice and the causes of those differences. In those studies it can be difficult to separate differences in practice style from differences in interpretation of the case vignette. A powerful, if underused, tool that uses vignette survey questions to study the effect of specific factors on survey responses is the *randomized questionnaire*.[4,14] With this survey design, some respondents receive one version of the questionnaire while other respondents receive another version that has slightly different questions. The respondents are randomized as to which version they receive. For example, one respondent might receive a case vignette that reads, "A patient with bilateral pulmonary opacities, no evidence of heart failure, and a P_{aO_2} of 120 mm Hg on F_{IO_2} 0.5. . . " and another respondent would receive a nearly identical vignette, except the patient is described as having a P_{aO_2} of 70 on F_{IO_2} 0.5. Since every other aspect of the vignette is the same, differences in the survey responses are assumed to reflect the difference between the vignettes (in this case, the severity of hypoxemia). This allows the investigator to test hypotheses about which specific aspects of cases drive decision making.

The most important step in preventing instrument bias and developing a questionnaire that asks the right question in the right way is to pilot-test the survey and revise it based on the feedback. This step should never be skipped. Even if you can only pilot-test the survey with a few respondents, it is better than not pilot-testing the survey at all. It is acceptable, and even preferable, to pilot-test the survey with individuals who know something about your research question and survey design.

Sample Size

Statistical power is important in survey research, just as it is in clinical trials. There are 2 issues to consider in deciding on the number of people to sample: the precision of estimates of prevalence and the power to make comparisons and correlations. The more subjects that are sampled, the more precise the investigator can be about the estimates in the population. For example, a recent CNN (Cable News Network) poll published just after Senator John Kerry announced his vice-presidential running mate found, "The survey of 706 likely voters, conducted between Thursday and Sunday, found 50% would support the Kerry-Edwards ticket, and 46% would vote for President Bush and Vice President Dick Cheney. The margin of error was plus or minus 4 percentage points."[15] This means that, assuming the poll respondents were an unbiased sample of the public, the 95% confidence interval for the percentage of people who support the Bush-Cheney ticket is between 42% (46% minus 4%) and 50% (46% plus 4%). The "margin of error" is simple to calculate, with the equation:

$$1.96 \times \sqrt{\frac{pq}{n}}$$

in which p is the percent that responded "yes" (46% in this example), q is $(1 - p)$, and n is the sample. Solving that equation, we see that the actual margin of error in the poll was plus or minus 3.6%, so apparently someone rounded

up to be conservative. You can also use the latter formula to see that if the pollsters had surveyed only 100 subjects, the margin of error would have been 10%. Similarly, a poll of 5,000 people would give very precise results, with a margin of error of only 1.4%. However, this does not take into account the possibility of sampling and instrument bias; it merely reveals the statistical uncertainty in the poll.

Investigators can use this formula to decide how big a sample they need, by solving for *n* in the equation and making some assumptions about *p*. For example, say an investigator wishes to do a survey on the percentage of respiratory therapists who believe that there is a disagreement between nurses and respiratory therapists about suctioning practices. The investigator believes that about 80% of respiratory therapists will say "yes, nurses and respiratory therapists disagree about suctioning practices." Obtaining a 2% margin of error would require 1,500 respondents, whereas obtaining a 5% margin of error would require only 250 respondents.

There are other factors besides the precision of the estimates of prevalence that drive the sample size. For example, to continue the example about suctioning practices, let's say that the investigator wants to compare the responses of nurses and respiratory therapists, or wants to compare the genders of those who believe there is disagreement and those who do not, or wants to correlate the perception of disagreement with respondent age. Each of those *comparisons* requires a statistical test that must be appropriately powered. Even a survey with relatively few items can generate hundreds of such statistical comparisons. Investigators should be wary about the potential for false-positive (Type I) errors when they perform hundreds of statistical tests at the $p = 0.05$ level of statistical significance.

Summary

Surveys are invaluable research tools for studying knowledge, attitudes, and behavior. They can often be performed for relatively little money and therefore are an accessible research method for new investigators. Survey responses can be analyzed and presented quantitatively or qualitatively. However, as with all research methods, even when performed well, surveys have limitations and tradeoffs. They can only reveal, and in fact are only designed to reveal, the respondent's perceptions. In many cases, that is the research question; however, respondents' answers may not reflect their true behavior. Despite these limitations to surveys, for many important research questions there are no alternatives.

REFERENCES

1. Cox CE, Carson SS, Ely EW, Govert JA, Garrett JM, Brower RG, et al. Effectiveness of medical resident education in mechanical ventilation. Am J Respir Crit Care Med 2003;167(1):32–38.
2. Becker EA. Respiratory care managers' preferences regarding baccalaureate and master's degree education for respiratory therapists. Respir Care 2003;48(9):840–858.
3. Blendon RJ, DesRoches CM, Brodie M, Benson JM, Rosen AB, Schneider E, et al. Views of practicing physicians and the public on medical errors. N Engl J Med 2002;347(24):1933–1940.
4. Cook DJ, Guyatt GH, Jaeschke R, Reeve J, Spanier A, King D, et al. Determinants in Canadian health care workers of the decision to withdraw life support from the critically ill. Canadian Critical Care Trials Group. JAMA 1995;273(9):703–708.
5. Asch DA. The role of critical care nurses in euthanasia and assisted suicide. N Engl J Med 1996;334(21):1374–1379.
6. Boynton PM. Administering, analysing, and reporting your questionnaire. BMJ 2004;328(7452):1372–1375.
7. Boynton PM, Greenhalgh T. Selecting, designing, and developing your questionnaire. BMJ 2004;328(7451):1312–1315.
8. Patton MQ. Qualitative evaluation and research methods, 2nd ed. Newbury Park, California: Sage Publications; 1990.
9. Heyland DK, Rocker GM, Dodek PM, Kutsogiannis DJ, Konopad E, Cook DJ, et al. Family satisfaction with care in the intensive care unit: results of a multiple center study. Crit Care Med 2002;30(7): 1413–1418.
10. Wenrich MD, Carline JD, Giles LM, Ramsey PG. Ratings of the performances of practicing internists by hospital-based registered nurses. Acad Med 1993;68(9):680–687.
11. Groves RM. Survey methodology. Hoboken, New Jersey: John Wiley; 2004.
12. Montano DE, Phillips WR. Cancer screening by primary care physicians: a comparison of rates obtained from physician self-report, patient survey, and chart audit. Am J Public Health 1995;85(6):795–800.
13. Rubenfeld GD, Cooper C, Carter G, Thompson BT, Hudson LD. Barriers to providing lung-protective ventilation to patients with acute lung injury. Crit Care Med 2004;32(6):1289–1293.
14. Redelmeier DA, Tversky A. Discrepancy between medical decisions for individual patients and for groups. N Engl J Med 1990;322(16): 1162–1164.
15. Poll finds apparent Kerry lead in close contest. Available at: http://www.cnn.com/2004/allpolitics/07/12/presidential. Accessed August 11, 2004.