Survey methodology for the uninitiated

Cristen R. Wall, MD; Mark J. DeHaven, PhD; Kevin C. Oeffinger, MD; Dallas, Texas

From The University of Texas Southwestern Medical Center at Dallas, the Department of Family Practice and Community Medicine, Dallas, TX. Presented at the 33rd Annual Conference of the Society of Teachers of Family Medicine, Orlando, FL, May 2000, and at the North American Primary Care Research Group, Amelia Island, FL, November 2000.

Mailed, self-administered questionnaires are a common data collection method used in family practice research. However, little guidance exists in the medical literature about survey methodology specifically designed for family practitioners. As a result, primary care physicians sometimes struggle with questionnaire design and interpretation. Our goal was to synthesize general survey methodology guidelines (from other disciplines as well as our own) in a manner that would be meaningful to novice family practice researchers.

key words Data collection; family practice; questionnaires; research methodology; surveys. (J Fam Pract 2002; 51:573)

Research using self-developed questionnaires is a popular study design in family practice and is frequently used for gathering data on knowledge, beliefs, attitudes, and behaviors. A Medline literature search from 1966 to 2000 identified 53,101 articles related to questionnaires, of which 2088 were directly related to family practice. Despite the large number of questionnaire-related articles, however, only 2 in the general medical literature and 1 in the family practice literature were directly related to research methodology.

To obtain guidance on survey research methodology, novice family practice researchers often must go through volumes of information by specialists in other disciplines. For example, a search of a psychology database (PsychInfo) from 1966 to 2000 produced 45 articles about questionnaire methodology. The goal of this article is to synthesize pertinent survey research methodology tenets from other disciplines as well as from family practice-in a manner that is meaningful to novice family practice researchers as well as to research consumers. This article is not aimed at answering all questions, but rather is meant to serve as a general guideline for those with little formal research training who seek guidance in developing and administering questionnaires.

Avoiding common pitfalls in survey research

Although constructing a questionnaire is not exceedingly complex, simple mistakes can be avoided by following some basic rules and guidelines. The Figure is a checklist for conducting a survey research project that combines guidelines and suggestions from published survey research literature, and the cumulative experience of the authors. Two of the authors (M.J.D. and K.C.O.) are experienced survey researchers who have published, in peer-reviewed journals, numerous studies that used questionnaires. One of the authors (MJD) has been teaching research to residents and junior faculty for over a decade, and has been an advisor on scores of resident, student, and faculty research projects. The perspective of the novice researcher is represented by 1 author (C.R.W.).

Getting started

The “quick and dirty” approach is perhaps the most common pitfall in survey research. Because of the ease of
administration and the relatively low cost of survey research, questionnaires can be developed and administered quickly. The researcher, however, should be sure to consider whether or not a survey is the most appropriate method to answer a research question. Adequate time must be given to thoroughly searching the relevant literature, developing and focusing on an appropriate research question, and defining the target population for the study (see Figure A. Getting Started). Large, multisite surveys are more likely to be generalizable and to be published in peer-reviewed journals.

One way to avoid undertaking a project too rapidly and giving inadequate attention to the survey research process is for novice researchers to avoid independent research. Those with little or no experience must realize that researchers in both family practice and other fields perform research in teams, with the various participants bringing specific skills to the process. Oversights, mistakes, and biases in the design of questionnaires can always occur, whether a researcher is working independently or as a member of a team. It seems reasonable to assume, however, that significant problems are much less likely to occur when a multidisciplinary team approach is involved rather than an individual researcher undertaking a study independently.

Ideally, a research team should include a statistician, a professional with experience in the content areas of the study, and a senior investigator. The desirable area of expertise, however, is often not readily available to family physicians, especially those in community-based settings. Individuals with some training in research who are interested in being involved can usually be found in colleges and universities, hospitals, and at the local public health department. Psychologists, sociologists, health services researchers, public health epidemiologists, and nursing educators are all potential resources and possible collaborators. Establishing the necessary relationships to form an ad hoc research team is certainly more time and labor intensive than undertaking research independently, but generally results in the collection of more useful information.

Novices should consult survey methodology books before and during the study. Excellent resources are available that provide a comprehensive overview of survey methods, means for improving response rates, and methods for constructing relatively brief but thorough survey questions. Academic family practice fellowships often provide training in survey methodology. In addition, many family practice researchers respond favorably to requests for information or advice requested by telephone or email contact. The novice author of this article reports excellent success in contacting experts in this manner. With the advent of the Internet, a “cyberspace” team comprised of experts in the topic and the methodology is a reasonable and helpful option for the novice.

Survey content and structure

Novice researchers often assume that developing a questionnaire is an intuitive process attainable by virtually anyone, regardless of their level of research training. While it is true that questionnaires are relatively simple to construct, developing an instrument that is valid and reliable is not intuitive. An instrument is valid if it actually measures what we think it is measuring, and it is reliable if it measures the phenomenon consistently in repeated applications. By following a few basic guidelines, those with limited research training can develop survey instruments capable of producing valid and reliable information. The 3 primary concerns for developing appropriate questions (items) are: (1) response format; (2) content; and (3) wording and placement (see Figure B, Survey Questions; and Figure C, Designing and Formatting the Survey).

Format

Questionnaires generally use a closed-ended format rather than an open-ended format. Closed formats spell out response options instead of asking study subjects to respond in their own words. Although there are many reasons for using closed formats, their primary advantages over open formats is that they are more specific and provide the same frame of reference to all respondents, and they allow quantitative analysis. A disadvantage is that they limit the possible range of responses envisioned by the investigators. Questionnaires with closed formats are therefore not as helpful as qualitative methods in the early, exploratory phases of a research project.

Closed-ended items can be formatted into several different categories (classes) of measurement, based on the
relationship of the response categories to one another. Nominal measurements are responses that are sorted into unordered categories, such as demographic variables (ie, sex, ethnicity). Ordinal measurements are similar to nominal, except that there is a definite order to the categories. For example, ordinal items may ask respondents to rank their preferences among a list of options from the least desirable to the most desirable.

Survey items that ask for respondents to rank order preferences are often a more useful than items that state, “check all that apply.” While checking all relevant responses may be necessary for certain items, such questions often lose valuable information as they can only supply raw percentages without supplying any comparison between responses. If a survey uses a rank order response, it enables determining the relative importance of the different categories during data analysis Table 1.

Two additional tools used on questionnaires are continuous variables and scales. Continuous variables can be simple counts (eg, the number of times something occurred) or physical attributes (eg, age or weight). A general rule when collecting information on continuous variables is to avoid obtaining the information in ranges of categories unless absolutely necessary. Response categories that reflect ranges of responses can always be constructed after the information is gathered, but if the information is gathered in ranges from the start, it cannot later be expanded to reflect specific values.

Scales are used by survey researchers to assess the intensity of respondents’ attitudes about a specific issue or issues. Likert scales are probably the best known and most widely used for measuring attitudes. These scales typically present respondents with a statement and ask them to indicate whether they “strongly agree,” “agree,” “neither agree nor disagree,” “disagree,” or “strongly disagree.” The wording of the response categories can be changed to reflect other concepts (eg, approval or disapproval), and the standard 5-response format can be expanded or abbreviated if necessary.

There are no hard and fast rules for determining the number of response categories to use for scaled items, or whether to use a neutral category or one that reflects uncertainty. Research indicates that the reliability of respondents’ ratings declines when using more than 9 rating scale points. However, the reliability of a scale increases when the number of rating scale points is increased, with maximum benefit achieved with 5 or 7 scale points. Since the objective of using scales is to gauge respondent’s preferences, it is sometimes argued that a middle point or category of uncertainty should not be used. Odd-numbered rating scales, however, conform better with the underlying tenets of many statistical tests, suggesting the need for including this category. As the number of rating scale points increases, respondents’ use of the midpoint category decreases substantially. Thus, based on the available literature, it is generally advisable to use between 5 and 7 response categories and an uncertainty category, unless there is a compelling reason to force respondents to choose between 2 competing perspectives or alternatives.

Content

Items should not be included on questionnaires when the only justification for inclusion is that the investigator feels the information “would be really interesting to know.” Rather, for each item, you should ask yourself how it addresses the study’s research question and how it will be used in the data analysis stage of the study. Researchers should develop a data analysis plan in advance of administering a questionnaire to determine exactly how each question will be used in the analysis. When the relationship between a particular item and the study’s research question is unclear, or it is not known how an item will be used in the analysis, the item should be removed from the questionnaire.

Wording and placement

The wording of questions should be kept simple, regardless of the education level of the respondents. Questions should be kept as short and direct as possible since shorter surveys tend to have higher response rates. Each question should be scrutinized to ensure it is appropriate for the respondents and does not require or assume an inappropriate level of knowledge about a topic. Since first impressions are important for setting the tone of a questionnaire, never begin with sensitive or threatening questions. Questionnaires should begin with simple, introductory (“warm-
questions to help establish trust and an appropriate frame of mind for respondents. Other successful strategies are: (1) when addressing multiple topics, insert an introductory statement immediately preceding each topic (e.g., "In the next section we would like to ask you about "); (2) request demographic information at the end of the questionnaire; and (3) always provide explicit instructions to avoid any confusion on the part of respondents.

Additional, clear information on survey content and structure is available in 2 books from Sage Publications. By following simple guidelines and common sense, most family practice researchers can construct valid and reliable questionnaires. As a final safeguard, once a final draft of the questionnaire is completed, the researcher should always be the first respondent. By placing yourself in the respondent's role and taking the time to think about and respond to each question, problems with the instrument that were overlooked are sometimes identified.

Analyzing surveys

It is not within the scope of this project to address statistical analysis of survey data. Before attempting data analysis, investigators should receive appropriate training or consult with a qualified professional. There are 3 topics that can and should be understood by novice researchers related to data analysis (Figure D, Developing a Framework for Analysis).

Coding

Before analyzing survey data it is necessary to assign numbers (codes) to the responses obtained. Since the computer program that is used for analyzing data does not know what the numbers mean, the researcher assigns meaning to the codes so that the results can be interpreted correctly. Coding refers to the process of developing the codes, assigning them to responses, and documenting the decision rules used for assigning specific codes to specific response categories. For example, almost all questionnaires contain missing values when respondents elect to not answer an item. Unique codes need to be assigned to distinguish between an item's missing values, items that may not be applicable to a particular respondent, and responses that have a "none" or "no opinion" category.

Data can be entered into appropriate data files once codes have been assigned to responses and a codebook compiled that defines the codes and their corresponding response categories. It is important to ensure that the data are free of errors (are clean) prior to performing data analysis. Although many methods can be used for data cleaning (i.e., data can be entered twice and results compared for consistency), at a minimum all of the codes should be checked to ensure only legitimate codes appear.

Frequency distributions are tables produced by statistical software that display the number of respondents in each response category for each item (variable) used in the analysis. By carefully examining frequency tables, the researcher can check for illegitimate codes. Frequency tables also display the relative distribution of responses and allow identification of items that do not conform to expectations given what is known about the study population.

Sample size

Since it is usually not possible to study all of the members of the group (population) of interest in a study, a subset (sample) of the population is generally selected for study from the sampling frame. Sampling is the process by which study subjects are selected from the target population, while the sample frame is the members of a population who have a chance of being included in the survey. In probability samples, each member of the sampling frame has a known probability of being selected for the study, whereas in nonprobability samples, the probability of selection is unknown. When a high degree of precision in sampling is needed to unambiguously identify the magnitude of a problem in a population or the factors that cause the problem, then probability sampling techniques must be used.

When conducting an analytical study that examines precisely whether statistically significant differences exist between groups in a population, power analysis is used to determine what size sample is needed to detect the differences. Estimates of sample size based on power are inversely related to the expected size of the differences (effect size)—that is, detecting smaller differences requires a larger sample. If an analytical study is undertaken to determine the magnitude
of the differences between 2 groups, it is necessary to work with a statistician or other methodology expert to perform the appropriate power analysis. For a basic but valuable description of sample size estimation, see chapter 13 of Hulley and Cummings.  

In contrast to analytical studies, exploratory and descriptive studies can frequently be conducted without the need for a power analysis. While some descriptive studies may require the use of probability techniques and precise sample estimates, this often is not the case for studies that establish the existence of a problem or estimating its dimensions. When conducting an exploratory or descriptive study using a survey design and a nonprobability sampling technique, considerations other than effect size or precision are used to determine sample size. For example, the availability of eligible respondents, limitations of time and resources, and the need for pilot study data can all contribute to selecting a nonprobability sample. When these types of sampling techniques are used, however, it is important to remember that the validity and reliability of the findings are not assured, and the findings cannot be used to demonstrate the existence of differences between groups. The findings of these types of studies are only suggestive and have limited application beyond the specific study setting.

Response rate

The response rate is a measure indicating the percentage of the identified sample that completed and returned the questionnaire. It is calculated by dividing the number of completed questionnaires by the total sample size identified for the study. For example, if a study is mailed to 500 physicians questionnaires and 100 returned a completed questionnaire, the response rate would be 20% (100/500).

The response rate for mailed questionnaires is extremely variable. Charities are generally content with a 1% to 3% response rate, the US Census Bureau expects to achieve a 99% rate, and among the general population, a 10% response rate is not uncommon. Although an 80% response rate is possible from an extremely motivated population, a rate of 70% is generally considered excellent.

The effect of nonresponse on the results of a survey depend on the degree to which those not responding are systematically different from the population from which they are drawn. When the response rate is high (ie, 95%), the results obtained from the sample will likely provide accurate information about the target population (sampling frame) even if the nonrespondents are distinctly different. However, if nonrespondents differ in a systematic way from the target population and the response rate is low, bias in how much the survey results accurately reflect the true characteristics of the target population is likely.

When calculating the response rate, participants who have died or retired can be removed from the denominator as appropriate. Nonrespondents, however, who refuse to participate, do not return the survey, or have moved should be included. This bias tends to be more problematic in “sensitive” areas of research than in studies of common, nonthreatening topics. Imputing values for missing data from nonrespondents is complex and generally should not be undertaken.

Given the importance of response rate, every effort must be made to obtain as many completed questionnaires as possible and strategies to maximize the response rate should be integrated into the study design (see Dillman for a useful discussion of successful strategies). Some simple means for improving response rates include constructing a short questionnaire, sending a well-written and personalized cover letter containing your signature, and emphasizing the importance of the study and the confidentiality of responses. It is also advisable to include a self-addressed, stamped envelope for return responses, and sometimes a small incentive is worthwhile. The National Center for Education Statistics notes that all surveys require some follow-up to achieve desirable response rates. Survey researchers, therefore, should develop procedures for monitoring responses and implement follow-up plans shortly after the survey begins.

Generally, 2 or 3 mailings are used to maximize response rates. Use of post card reminders is an inexpensive, though untested, method to increase response. Several randomized studies have reported an increase in response rate from physicians in private practice with the use of monetary incentives, although the optimum amount is debated. Everett et
al\(^{40}\) compared the use of a $1 incentive vs no monetary incentive and found a significant increase with the incentive group (response rates: 63% in the $1 group; 45% in the no incentive group; \(P < .0001\)).\(^{41}\) Other studies have compared $2, $5, $10, $20, and $25 incentives and found that $2 or $5 incentives are most cost effective.\(^{42,45}\) Similar findings have been reported for physician surveys in other countries.\(^{31,46}\) In an assessment of incentive for enrollees in a health plan, a $2 incentive was more cost effective than a $5 incentive.\(^{47}\) A $1 incentive was as effective as $2 in significantly increasing response rate in a low-income population.\(^{48}\) Quality of responses have not varied by use of incentives and there does not appear to be an incentive-bias.

Use of lottery appears to also increase response rate in both physicians and the lay public, although there are no studies comparing lottery to a monetary incentive enclosed for all participants.\(^{31,49}\) Use of either certified or priority return mail appears to increase response rates, and may be more cost effective when used for the second mailing.\(^{45,48}\)

Pilot testing

Though pilot testing is generally included in the development of a survey, it is often inadequately conducted Figure F (Final Preparation). Frequently, investigators are eager to answer their research question and pilot testing is synonymous with letting a few colleagues take a quick look and make a few comments. Table 2 illustrates a problem that could have been avoided with proper pilot testing.\(^{10}\) One of the questions in the survey asked about how time is allotted for faculty to pursue scholarly activities and research (Format A). Unfortunately, the question mixes 2 types of time in 1 question: extended time away from the institution (sabbatical and mini-sabbatical) and time in the routine schedule. This was confusing to respondents and could have been avoided by separating the content into 2 separate questions (Format B).

Investigators should consider carefully whom to include in the pilot testing. Not only should this include the project team and survey “experts”, but it should also include a sample of the target audience. Pilot testing among multiple groups provides feedback about the wording and clarity of questions, appropriateness of the questions for the target population, and the presence of redundant or unnecessary items.

### CONCLUSIONS

One of the authors (C.R.W.) recently worked on her first questionnaire project. Among the many lessons she learned was the value of a team in providing assistance, the importance of considering if the time spent on a particular activity makes it cost effective, and the need to be flexible depending on circumstances. She found that establishing good communication with the team cuts down on errors and wasted effort. Rewarding the team for all of their hard work improves morale and provides a positive model for future projects.

The mailed self-administered questionnaire is an important tool in primary care research. For family practice to continue its maturation as a research discipline, family practitioners need to be conversant in survey methodology and familiar with its pitfalls. We hope this primer-designed specifically for use in the family practice setting will provide not only basic guidelines for novices but will also inspire further investigation.

### ACKNOWLEDGMENTS

The authors thank Laura Snell, MPH, for her thoughtful review of the manuscript. We also thank Olive Chen, PhD, for research assistance and Janice Rookstool for manuscript preparation.

### REFERENCES

31. Kalantar JS, Talley NJ. The effects of lottery incentive and length of questionnaire on health

Address correspondence to Cristen R. Wall, MD, The University of Texas Southwestern Medical Center, Department of Family Practice and Community Medicine, 6263 Harry Hines Boulevard, Dallas, TX 75390-9067. E-mail: Cristen.Wall@UTSouthwestern.edu.

To submit a letter to the editor on this topic, click here:jfp@fammed.uc.edu.