

Guidelines for Constructing a Survey

Cindy Passmore, MA; Alison E. Dobbie, MD;
Michael Parchman, MD; James Tysinger, PhD

Many researchers in family medicine use surveys to gather data from colleagues, learners, and patients on their demographics, personal histories, knowledge, behaviors, and attitudes. Well-written surveys are easy for respondents to complete, gather information accurately and consistently, and obtain data that can be analyzed to answer research questions. All levels of family medicine researchers can follow eight steps to develop surveys that produce useful and publishable results: (1) state the problem or need, (2) plan the project, (3) state the research question, (4) review the literature, (5) develop or adapt existing survey items, (6) construct the survey, (7) conduct pilot tests, and (8) administer the survey. After completing this article, readers should be able to (1) state the appropriate uses of survey instruments as research tools and (2) construct and administer a well-designed survey instrument.

(Fam Med 2002;34(4):281-6.)

Surveys are the most common method used by family medicine researchers to obtain data from peers, learners, and patients. Seasoned researchers employ surveys that are easy to complete, gather data consistently and accurately, and produce results that answer specific research questions. Less-experienced researchers can also develop such quality instruments by following a systematic approach. This article describes such a systematic approach and provides strategies for overcoming common problems in survey development. This article will equip readers to (1) identify the appropriate uses of surveys as research tools and (2) construct and administer a well-designed survey

Survey Instruments as Research Tools

Use of Surveys

Survey instruments, or questionnaires, are used to collect data about subjects' demographics, personal histories, knowledge, behaviors, and attitudes. Many researchers administer a survey when they need information or wish to answer a question about their patients, colleagues, or learners. Surveys are suitable for use by all levels of researchers, but quality instruments are more challenging to develop than many novice researchers realize. Poorly designed surveys usually

produce unsatisfactory data that cannot answer the desired question.

Limitations of Surveys

Surveys have limitations as research tools. First, they rely on subjects' honesty and memory. Thus, respondents' desire to give socially and culturally acceptable responses to items that address sensitive topics (eg, sexual practices) may bias survey results. Second, the quality of data obtained depends on how well respondents understand the survey items or questions. Such understanding is affected by respondents' reading level, cultural perspective, and language skills. Researchers must address these limitations when designing a survey. For example, if subjects have a primary language other than English, a version of the survey in a second language may be needed.

The response rate can also limit a survey's usefulness. The lower the return rate, the more likely it is that the characteristics of respondents differ from those of nonrespondents. Therefore, the results may not be trustworthy. For example, if a survey of family practice residency directors about teaching ethics to residents has a response rate of 40%, and 80% of respondents report formal ethics teaching in their programs, one cannot conclude that 80% of all family practice residency programs have a formal ethics curriculum. The low response rate casts doubt that the survey results accurately reflect the amount of ethics teaching in the 60% of residency programs that did not reply. Many investigators consider a response rate of 70% adequate for

From the Faculty Development Center, Waco, Tex (Ms Passmore); and the University of Texas Health Science Center at San Antonio (Drs Dobbie, Parchman, and Tysinger).

generalization to the population studied, though this may vary according to the purpose and nature of the study.

Types of Surveys

Descriptive Surveys. Descriptive surveys report factual data (eg, number of articles read per week by residents) or opinions (eg, which journals contain the most interesting articles). Descriptive surveys can provide results that lead to more-sophisticated studies.

Explanatory Surveys. Explanatory surveys attempt to link cause and effect (for example, whether a resident's current rotation is associated with or influences the number and content of journal articles that residents read).

Unidimensional and Multidimensional Scales. Surveys can address one or more underlying construct(s) (ie, idea, attribute, or measure). One-construct surveys, such as an instrument to measure residents' knowledge, are known as unidimensional scales. Surveys measuring more than one construct (eg, residents' knowledge and attitudes) are multidimensional scales.

Ways to Administer Surveys

Surveys can be administered by interviewers (by telephone or in person) or can be self-administered by the respondent. Telephone surveys and personal interviews are expensive and time-consuming to administer, but they permit clarification or explanation of items, reduce the number of blank or incorrectly completed items, and may increase the response rate. Interviewers must be trained to standardize questioning techniques.

Self-administered surveys, distributed by mail or e-mail, are less expensive to administer than in person or telephone surveys and can provide privacy and anonymity to respondents (Table 1). However, these surveys typically yield numerous unusable or incomplete responses and may require multiple mailings to obtain a response rate high enough to generalize the data gathered to the whole target population.

Steps in Developing a Survey

Surveys should be developed using a systematic process. Careful attention to all development stages will strengthen the instrument and enhance the quality of the data. The steps in developing a survey are summarized in Table 2.

Step 1: State the Problem

Describe an identified problem or need. For example, a faculty member may state this problem: "Several colleagues have approached me about presenting a computer skills workshop at the STFM Annual Spring Conference. I can design a workshop, but I have a problem.

Table 1

Pros and Cons of Interview Versus Self-administered Surveys

<i>Interview Surveys</i>	<i>Self-administered Surveys</i>
More expensive (due to personnel, training costs)	Less expensive
Require interviewer recruitment, training, and standardization	No interviewers needed
Responses not private	Privacy conserved if survey is anonymous
Single administration	Multiple mailings/contacts
Clarification and explanation of items possible	No clarification possible
Few incorrectly completed surveys	Many incorrectly completed surveys
Higher response rate	Lower response rate
Interviewer bias	No interviewer bias

I need to know what computer skills my STFM colleagues possess and need."

Step 2: Plan the Project

Plan all stages of the project, including assembling the research team, establishing timelines, and selecting the target journal for publication before preparing the survey. Determine the project's costs before proceeding, and ensure that sufficient time and funding are available.

Working as a team member rather than as a solo investigator improves the likelihood of a successful project. Include individuals with particular types of expertise on the survey development team. For example, a statistician will help state the research question and identify the data needed to answer that question.

Table 2

Steps in Developing a Survey

1. State the problem.
2. Plan the project.
3. State the research question.
4. Review the literature.
5. Develop/adapt survey items.
6. Construct the survey.
7. Pilot test the draft survey.
8. Administer the survey.

If there is no statistician available on the study team, a statistical consult is often helpful at this planning stage to help refine the research question, identify required data, and plan data analysis.

Data entry should also be considered at this step. Respondents' data will need to be entered into an electronic file for analysis. Data may be entered by hand, but this is laborious, time-consuming, and prone to data entry errors. An alternative is to enlist the help of individuals who can aid in the development of computerized scannable answer sheets and who know how to operate the hardware/software needed to scan these answer sheets into an electronic file.

Step 3: State the Research Question

The investigator must state the pertinent research question that will be answered by the survey data. Good research questions are clear, simple, important, interesting, and answerable. A clear, simple question helps identify the required data. Important and interesting questions increase the likelihood that the investigator will complete the project. A specific and concrete question also allows the investigator to restrict the data gathered to manageable amounts.

An example of an unrefined research question posed by our faculty developer is "What computer skills do family medicine faculty members need?" The improved research questions might read: "What computer applications do family medicine faculty routinely use? What is their self-reported competence with those applications? What are their perceived learning needs?"

Even these questions need further refinement. First, we must describe the individuals who compose family medicine faculty. Then, we can define "routinely" as a specific time interval (eg, every day or once or twice per week). We must also define "competence" for the study. This process is called stating the operational definitions of the terms in the research question. Detailed operational definitions minimize vagueness and permit the investigator to state refined research questions.

Step 4: Review the Literature

Having clearly stated a research question, the investigator must review the literature to become familiar with existing published work. Such efforts allow researchers to ensure that their question has not been previously answered, identify gaps in the literature, and note possible research methods for their own studies. The literature review may also reveal an existing survey instrument that the investigators can use or adapt to their situation or population. Using or adapting a validated survey saves time and expense and improves the likelihood that the study results will be valid and reproducible.

Step 5: Develop the Survey Instrument

After clearly stating the research question and identifying the required data, select or construct the survey. Investigators may either use or modify an existing survey or develop a new instrument.

To adapt an existing survey, investigators must generally obtain permission from the developer and should add new items at the end, where they will not interfere with the existing validity of the instrument in its original form. To construct a new survey, clearly define the data or constructs to be gathered or measured. Two strategies to identify item content are to conduct focus groups of potential subjects or to convene an expert panel. A third approach, the Delphi technique, is an iterative process in which items are nominated and rated by a group of experts until consensus is reached. Regardless of how items are developed, several colleagues should complete and critique the items to lend some face validity to the instrument before it is pilot tested with subjects from the target group.

Components of an Item. Items are the questions or statements in a survey that elicit specific responses from individual subjects. The quality of the items will determine the quality of the data.

Items should discriminate among respondents. That is, respondents who feel similarly about a question should choose the same answer, and those whose views differ should choose contrasting options. Surveys measuring the degree of underlying constructs (eg, resident knowledge, skills, and attitudes) require at least three items for each construct to promote the validity and reliability of the responses. Using several items per construct also permits item analysis.

Each item is composed of a "stem" and a "response format." The stem is the question or statement to which the subject provides a response. Response formats provide a framework for subjects' answers. These answers compose the data the investigator will analyze to answer the research question.

Stems. Acceptable stems are short (ie, shorter than 20 words) and clearly stated in a simple sentence or statement. They should not contain modifiers such as "almost everyone" or "usually," which can confuse respondents as in the following example:

"In the opinion of most people, the consumption of alcoholic beverages may be in some cases detrimental to the overall emotional and physical well-being of the consumer." Strongly disagree-1, disagree-2, neutral-3, agree-4, strongly agree-5

A shorter, clearer stem statement is: "Drinking alcohol is harmful to my health." Strongly disagree-1, disagree-2, neutral-3, agree-4, strongly agree-5.

Having definitive statements as stems allows the scale to better discriminate among respondents. Although stem

statements should be unambiguously stated, they should neither prompt a particular answer from the respondent nor encourage dishonest responses. To avoid offending respondents and to encourage accurate and honest answers, items should be nonjudgmental and socially and culturally inoffensive. For example, replace "How often did you use marijuana as a college student?" with "Have you ever used marijuana?"

State item stems positively whenever possible. Negative item stems can confuse respondents and may fail to obtain their true opinion. For example, replace "Driving under the influence of alcohol is not socially responsible behavior" with "It is irresponsible to drive while drunk."

Item stems should focus on only one variable. Asking subjects to select a response containing two answers is confusing and reduces the accuracy of the survey data. For example, replace "How many articles and book chapters did you read last week?" with "How many articles did you read last week?" and "How many book chapters did you read last week?"

Response Formats. Response formats may be open or closed. An open response format allows subjects to answer a question in free text. For example, a survey assessing faculty computer skills might ask, "Please list the computer applications you use at least twice per week" and leave space for respondents to write their answers. An advantage of the open format is that subjects are not limited in their responses. However, answering open questions requires more time and effort from subjects than checking responses on a list. In this example, faculty may forget or fail to record every computer program they use or choose not to complete the survey at all. Thus, the response rate may be lower and the data less complete than that obtained using a closed format. In addition, free-text data can be difficult and time consuming to analyze.

Closed-response formats give subjects a structured way to answer items by requiring the subject to choose from a list of options. These structured responses greatly facilitate data analysis. A researcher might ask:

Please check the computer applications you use at least twice per week: Microsoft Word __, PowerPoint __, Excel __, Access __, Word Perfect __, Other __.

Including the "other" option allows subjects to provide an answer that the investigator may not have anticipated. For example, some respondents may use SPSS more than twice per week. If on the pilot testing of our survey, the investigator received a large number of SPSS answers in the "other" category, he/she would add SPSS to the list of options in the final survey.

The investigator may choose from a number of formats for closed-response scales. Each scale has its strengths and weaknesses.

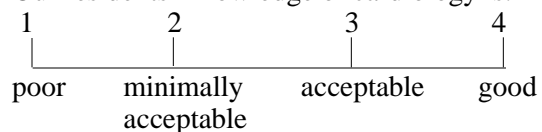
Likert-type Scales. The traditional Likert scale, with a statement as a stem followed by responses ranging from strongly disagree to strongly agree, is commonly used in surveys. This format is familiar to respondents, and it lends itself well to measuring constructs like attitudes. The number of scale points for responses can vary, but five is usually satisfactory. Investigators should decide in advance whether to use an odd or even number of response points. An odd number of points permits respondents to adopt a neutral position, but an even number of points forces a commitment. Novice investigators may assume that increasing the number of points on a scale widens the range of responses and improves item discrimination, but this is usually not so.

In a phenomenon called the "floor" or "ceiling" effect, subjects tend to choose responses that cluster at either the top or bottom of any scale. As a result of this clustering, the instrument may not capture a significant amount of the true variability in opinion among respondents. Increasing the number of points on a scale may worsen this phenomenon. For example, if 10 participants rate an STFM presentation on a 5-point scale (1=poor, 5=excellent), the ratings may range from 3 to 5. Given a 7-point scale, participants might rate the session from 5 to 7. The top 3 points in both scales capture all the variability. Investigators anticipating such a problem might experiment with different response formats such as:

How satisfied were you with the residents' performance on the inpatient service? Very dissatisfied, dissatisfied, satisfied, very satisfied

Rating Scales. Rating scales, similar to Likert scales, allow degrees of expression of an underlying opinion. Item stems need not be statements, but they should be as neutral as possible to allow the scale to discriminate among respondents' answers.

Our residents' knowledge of cardiology is:



Pictorial Scales. Simple pictorial scales do not require respondents to be able to read, write, or count. This advantage makes them especially suitable for use with children.

How do you feel about visiting the dentist?

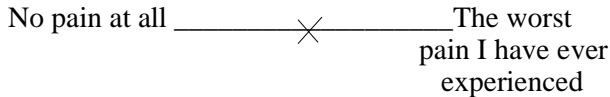


Visual Analog Scales. A visual analogue scale consists of a line, usually 10cm long for easy measurement, between two opposite anchors. Subjects check or cross the line to indicate their response, and the investigator measures the mark to obtain a score. For example:

Our residents' knowledge of cardiology is:



Visual analogue scales are useful for measuring degree. In pretests and posttests, these scales reduce the likelihood of respondents remembering which numbers they circled on the pretest, but their meaning must be clear. In the example below, the meaning of the mark is difficult to interpret because it is not clear if the patient has pain of moderate intensity or has pain about half the time.



Rank Lists

A rank list of items can be useful for market research, but it should be short because respondents quickly become fatigued after too many choices. For example, a pizza restaurant might ask their customers:

Please indicate your preferred pizza toppings by ranking the following items 1 (most liked) to 5 (least liked)

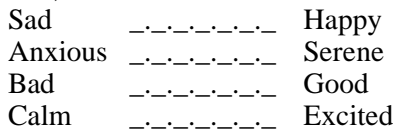
- 1 pepperoni
- 3 3 sausage
- 2 2 cheese
- 5 5 tomatoes
- 4 4 mushrooms

A problem with ranking lists is that they do not indicate relative degrees of liking or disliking. For example, does the above respondent like pepperoni twice as much as cheese or three times more than sausage? Also, what if the respondent's favorite pizza topping is anchovy, which is not listed as a choice? Adding an "other" category to the list addresses this problem.

Semantic Differential Scales

Semantic differential scales consist of a pair of opposite responses, with several check points between them. These scales are excellent for measuring attitudes. For example:

When I think about my residents' in-training exam scores, I feel:



A challenge with semantic differential scales is finding pairs of true opposites. For example, is the opposite of angry "pleased," "serene," "calm," or none of these?

Step 6: Construct the Survey Instrument

Having chosen a response scale and written the items, the investigator can now compile the individual items into a survey instrument. The instrument should be visually inviting to increase the return rate of correctly completed responses. Give subjects clear instructions on how to complete the items, and consider including a correctly completed sample item.

Items may be grouped into major subject areas to assist respondents' thought processes and memory. If using more than one response format, group and label items with the same format.

Ask demographic questions first, since these are simple, nonthreatening questions that "warm up" the respondent. Place items concerning sensitive topics such as sexual behaviors or beliefs toward the end of the survey. By that time, the subject is feeling comfortable and familiar with the survey format and is more likely to respond honestly.

Use the minimum number of items needed to obtain the data desired, and resist the temptation to gather data not pertinent to the research question. Long instruments tire respondents and can lower response rates. Data can be gathered to answer most research questions in 25 or fewer items.

Step 7: Pilot Test the Draft Survey

Identify problems with items or responses by pilot testing the survey with colleagues and subjects from the target audience. Pilot tests help identify redundant or poor questions and provide an early indication of the reproducibility of the responses. For example, the investigator may rewrite or drop an item if it confuses several respondents. At least two pretests are advised, and with each revision, the instrument should become shorter, not longer.

Statistical methods to assess survey validity and reliability are beyond the scope of this article. If no member of the study team is an expert on statistics, a second statistical consultation is helpful at this point.

Step 8—Administer the Survey

Introduce mailed or e-mailed surveys with a brief, simple cover letter, thanking the subject and explaining (1) the purpose of the survey, (2) why that person was chosen to complete it, and (3) why that person's participation is important to the study.

For mailed questionnaires, include a stamped, self-addressed envelope, and conduct at least one follow-up mailing. Follow up with subjects by telephone if surveys are returned incompletely or incorrectly filled out. In general, a response rate of 70% is considered adequate for generalizability to the population of interest.

Summary

Survey or questionnaire research is practical and achievable for all levels of family medicine researchers. Following the guidelines in this paper can help investigators enhance the quality of their surveys, thus improving the chances of successfully completing a study that answers an important research question and adds to the knowledge base of family medicine.

Acknowledgments: This paper was presented as a workshop at the Society of Teachers of Family Medicine 2000 Annual Spring Conference in Orlando, Fla, and at the 1999 Annual National Primary Care Research and Methods Conference in San Antonio, Tex.

Corresponding Author: Address correspondence to Dr Dobbie, University of Texas Health Science Center at San Antonio, Department of Family and Community Medicine, San Antonio, TX 78299-3900. 210-567-4555. Fax: 210-567-2443. dobbie@uthscsa.edu.

REFERENCES

1. Berg AO, Gordon MJ, Cherkin DC. Practice-based research in family medicine. Kansas City, Mo: American Academy of Family Physicians, 1986.
2. Fowler FJ. Survey research methods. Newbury Park, Calif: Sage Publications, 1990.
3. Frary RB. A brief guide to questionnaire development. www.ericae.net/ft/tamu/vpiques3.htm. ERICAE.net, 1998.
4. Calder J. Survey research methods. *Med Educ* 1998;32:636-52.
5. Carmines EG, Zeller RA. Reliability and validity assessment. Beverly Hills, Calif: Sage Publications, 1979.
6. Crocker L, Algina J. Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston, 1986.
7. DeVellis RF. Scale development: theory and applications. Newbury Park, Calif: Sage Publications, 1991.
8. McIver JP, Carmines EG. Unidimensional scaling. Beverly Hills, Calif: Sage Publications, 1981.
9. Spector PE. Summated rating scale construction: an introduction. Newbury Park, Calif: Sage Publications, 1992.