Pediatric Anesthesia

Pediatric Anesthesia ISSN 1155-5645

SPECIAL INTEREST ARTICLE

Survey research: it's just a few questions, right?

Alan R. Tait[1,2] & Terri Voepel-Lewis[1]

1 Department of Anesthesiology, University of Michigan Health System, Ann Arbor, MI, USA
2 Center for Bioethics and Social Sciences in Medicine, University of Michigan Health System, Ann Arbor, MI, USA

Keywords
survey research; psychometrics; validity; reliability

Correspondence
Prof. Alan R. Tait, Department of Anesthesiology, University of Michigan Health System, 1500 E. Medical Center Drive, Ann Arbor, MI 48109, USA
Email: atait@umich.edu

Section Editor: Mark Thomas

Accepted 8 April 2015

doi:10.1111/pan.12680

Summary

While most anesthesiologists and other physician- or nurse-scientists are familiar with traditional descriptive, observational, and interventional study design, survey research has typically remained the preserve of the social scientists. To that end, this article provides a basic overview of the elements of good survey design and offers some rules of thumb to help guide investigators through the survey process.

Introduction

Although survey research represents a relatively small subset of the anesthesia literature, it has been our experience as a former Chair of the Research Committee for the Society for Pediatric Anesthesia (SPA) and as journal reviewers that the number of anesthesia-related surveys submitted for review has increased over the last few years. While many of these surveys have undoubtedly resulted in important work, some are poorly constructed, suffer from low response rates, and provide limited generalizable data. Furthermore, surveys that are conducted without thought to progress beyond an 'abstract' presentation can be wasteful and impose an unnecessary burden on the survey respondents. The reasons for poor survey design likely include a lack of training in survey methodology and, perhaps, the belief that surveys are 'easy to do'. In truth, good survey research is difficult and requires, at its heart, an important overall objective or question together with a thorough understanding of the *psychometrics* of survey design including item construction, choice of response options, pilot-testing, and, if appropriate, measures of reliability and validity. This article is designed to provide anesthesiologists and other health professionals with a basic understanding and overview of the elements of good survey design and to offer some *rules of thumb* to help guide the investigator through the survey process.

Survey development

Like all good research, surveys should begin with an important question and set of objectives. Consideration should be given to the uniqueness of the question and the degree to which answering that question might contribute to generalizable knowledge. This question should begin with a thorough literature search. The purpose of this step is to separate what is known from what is not known and to determine how the survey will serve to close the gap in knowledge. Even with a good research question, however, many surveys fall short of their promise in that they either do not fully explore the question(s) at hand or ignore the power of surveys to 'measure' things. For example, while simple yes/no questions are easy to design and can be informative from a descriptive perspective, they do not always allow the investigator to explore underlying constructs or the qualitative nature of the responses.

Table 1 Checklist for developing and conducting a self-administered survey

Develop an answerable question: What are the objectives?
Perform a literature review: Is the question novel?
Item construction
• Discuss potential items with colleagues or focus groups
• Can previously validated items/surveys be used?
• Closed-ended vs open-ended?
Questions should
• Contain no typographical errors
• Be purposeful (related to the subject at hand)
• Be concrete
• Utilize complete sentences
• Avoid jargon and abbreviations
• Ordered in a logical manner with appropriate transition and skip patterns
• Avoid double negatives
• Avoid double-barreled questions
• Be culturally sensitive
• Unbiased
• Clearly formatted
Response options
• Consider dichotomous scales, numbers scales, visual analog scales, and Likert scales
• Should be exhaustive and mutually exclusive
• Ordered from negative to positive
• Ordered vertically
Conduct pre- and pilot-testing
Obtain IRB approval
Cover letter should
• Be well written with no typographical or punctuation errors
• Generate interest and enthusiasm
• Contain instructions and anticipated time for completion
• Acknowledge IRB approval
• Include a 'Thank you'
Sampling considerations
• Convenience sampling
• Simple random sampling
• Stratified sampling
• Systematic sampling with a random start
Are there potential sources of bias, and, if so, how can they be minimized?
• Nonresponse bias
• Recall bias
• Report bias
Are tests of reliability and validity appropriate? If so, what tests?
• Test–retest
• Internal consistency
• Face validity
• Content validity
• Criterion validity
• Construct validity

Constructing the items

The generation of questions or items is perhaps the most difficult, yet important aspect of good survey research and, as such, deserves thoughtful consideration.[1] In general, the choice of questions will be driven by the objectives of the study. Depending on the purpose of the survey, questions can be designed to elicit different types of information including: facts, e.g., demographics (age, practice, etc.); values, e.g., attitudes and behaviors; and knowledge, e.g., quizzes or tests. While discussing potential items with knowledgeable colleagues is helpful, it is also worthwhile to review the extant literature to see if there are existing validated items and constructs that can be incorporated or adapted for use in the survey. Informal focus groups can also be invaluable in developing items and in pretesting drafts of the survey. A checklist that summarizes the process of developing and conducting a survey is described in Table 1.

Open- vs closed-ended questions

Most self-administered questionnaires utilize structured or closed-ended questions wherein the respondent selects from a list of known potential responses.[2,3] The advantage of this approach is that the responses will tend to be more standardized, take less time to complete, and are easier to interpret and code. Lists of closed-ended responses should, where possible, be *exhaustive* and *mutually exclusive*, yet not so extensive that respondents will become fatigued. Adding a response option of 'other' with an opportunity to describe or clarify ('please specify' or 'please describe') is helpful in allowing the respondent to provide additional information not covered in the list of responses.

Open-ended or nonstructured questions have the advantage of allowing the respondent to answer in his/her own way and are helpful in exploratory research when the response options are unknown.[2,3] While these questions can add a nice dimension to a self-administered survey, they are generally more suited to interviews and focus groups. The disadvantages of open-ended questions are that they can take more time and effort to complete and, thus, are sometimes left unanswered and, further, may require specific expertise on the part of the investigator to identify themes and interpret the data. Typically, self-administered questionnaires should contain a preponderance of closed-ended questions as over-use of open-ended questions can lead to response fatigue and potentially, a higher likelihood of noncompletion.

Regardless of whether open or closed formats are used, the questions must be well written. As a starting point, using existing items from previously validated questionnaires can help to ensure that the items are well constructed and measure what they are supposed to measure. Furthermore, using standard item batteries saves time in development and permits comparison with other studies. Although many anesthesiology surveys

are directed at practitioners, others may be directed at patients and, as such, it is important that the items be written at a level consistent with the reading ability of the lay individual (approximately 8th grade). According to the National Adult Literacy Survey, over 90 million Americans have limited literacy skills.[4] Reading formulae such as SMOG (Simple Measure of Gobbledygook)[5] and the Flesch-Kincaid, and Flesch Reading Ease[6] are useful in reducing grade reading levels and are accessible both on the Internet and within Microsoft Word.®

Regardless of the anticipated reading level of the target audience, attention to the formatting of the survey can be critical. Surveys should be ordered in a logical manner, employ consistent spacing, and utilize bolding, underlining, or capitalization to clarify instructions or highlight important elements.[2,7] A font size of 11–12 point is recommended although larger fonts may be better for those with low literacy or limited vision. It is also important that transitions and skip sequences are used consistently and clearly indicate a logical path through the survey.[8]

When constructing items, consideration of the BRUSO acronym, i.e., **B**rief, **R**elevant, **U**nambiguous, **S**pecific, and **O**bjective, is helpful.[9] Items should be brief but should also utilize complete sentences. Questions should also be related to the subject at hand. Surveys that include items not seemingly relevant to the primary objective are often returned incomplete. In addition, the use of concrete or specific questions are preferable, e.g., 'In the past 6 months, how would you describe your health?' rather than 'How would you describe your health?' Surveys should also avoid the use of abbreviations, jargon, and acronyms. Negative questions, e.g., 'Parents should not be allowed to be present during anesthetic induction' should be avoided as they introduce the specter of the double negative. Double-barreled questions should also be avoided, e.g., 'How satisfied were you with your anesthetic and surgical care?' Clearly, this addresses two separate questions. In addition, questions should never be leading, evocative, or biased toward a particular point of view. Finally, if the questionnaire is determining knowledge, it is better to start with easier questions and end with the more difficult.

Choosing the response categories

Although some questions may be answered by a dichotomous 'yes/no' response, the use of nominal (named) or ordinal (ordered) response categories will likely provide significantly richer information.[2,10] These typically take the form of (i) Likert-type scales, e.g.,

How satisfied were you with your anesthetic care?

- Extremely dissatisfied
- Neither satisfied nor dissatisfied
- Extremely satisfied

(ii) Number scales wherein the respondent selects a number (usually 0–10, where 10 = maximum response), e.g., *Please rate your overall health by circling a number from 0 to 10*

0	1	2	3	4	5	6	7	8	9	10
Extremely poor health										Extremely healthy

(iii) Visual analog scales that require the respondent to place a mark on a continuous 10 cm (or 100 mm) line to indicate their level of response. Often too, respondents are required to use numerical values to rank order responses to indicate preferences. Combinations of these types of scales are often used. When using Likert scales, a minimum of three options should be provided although five to seven response options will garner significantly greater detail. In any case, when ordinal categories are used, they should be balanced (i.e., endpoints are mirror opposites) and ordered from negative to positive, e.g., 'Extremely dissatisfied' to 'Extremely satisfied'. When several response categories are offered, it is preferable to present them vertically rather than horizontally. Vertical formats reduce errors due to confusion over selecting a response category and are generally easier to code.

Over the years, there has been much debate regarding the advantages and disadvantages of offering an odd vs even number of response options.[1,11,12] Whereas, many survey researchers believe that using an odd number of response options provides a natural middle ground or 'fence' on which the respondent can sit, e.g., 'neither dissatisfied nor satisfied,' others believe that a neutral response simply provides an excuse for not answering. Advocates of even numbered responses believe that it forces the respondent to decide one way or the other. Similar rationales exist for including a 'don't know' option. Although this may be a valid recall response, some respondents choose this as an easy option or a way to satisfice an attitude question to which they are ambivalent. Ultimately, however, these decisions come down to whether a valid conceptual midpoint or natural ambivalence exists.

A question often asked is where to place the demographic items? Conventional wisdom suggests that the demographics be placed at the end of the survey as many individuals are sensitive about revealing socio-demographic information and may be wary if they see this first. Interestingly, however, one study found that placing socio-demographics at the beginning may actually

increase both demographic and nondemographic item response rates.[13] In practice, however, it likely depends on the focus of the survey and how important demographics (or other potentially sensitive information) are to the overall survey results. If socio-demographic data are central to the survey, it may be prudent to place them early in the survey. This ensures that the most important data will be captured in the event that the respondent subsequently tires of the survey and drops out without full completion. If demographics are less important, these items should be placed at the end, so that if respondents choose not to respond, the loss will be less critical.

Pre- and pilot-testing

Once all the items have been developed and the response options selected, it will be important to pre- and pilot-test the survey prior to distribution. Pretesting is important to ensure that the questions in the survey are clear and unambiguous, make sense, have appropriate response options, and are measuring what they are designed to measure. Once the items have been pretested, the resultant survey can undergo pilot-testing among a small group of individuals who resemble the target population. Pre- and pilot-testing are important as a means to establish the *face* and *content* validity of the questionnaire and to establish the time needed to complete the survey.[10,14] Face validity refers to how good items or groups of items in the questionnaire appear to lay individuals with no specific training, whereas content validity relies on input from individuals with expertise in the subject matter at hand. Pilot-testing also provides an opportunity to evaluate how subjects navigate through the survey.

Regardless of whether the survey is administered through online programs such as Qualtrics® or SurveyMonkey® or through the regular mail, attention to the structure and quality of the items at the early stages of development is critical for success. Poorly constructed surveys are less likely to engender interest and may thus, suffer from poor response rates. In addition, a well-constructed survey is easier to analyze and interpret and is more likely to provide meaningful results. As a note, inclusion of the survey as an appendix when submitting for publication is always good practice as it allows journal reviewers and readers to evaluate the items and serves as a potential resource for future surveys.

The cover letter

The importance of a cover letter to introduce the survey cannot be over-stated.[7,8] The cover letter serves as the 'carrot' to pique interest and encourage participation. The cover letter, whenever possible, should be written on letterhead and include: an introduction to the investigator(s) with contact information; a description of the purpose and rationale for the survey; instructions for completion of the survey (e.g., link to an *e*-survey) including expectations for return; and a statement regarding the potential significance of the results, i.e., why the study is important. It should be well written with no typographical errors. Given that survey respondents (particularly physicians) are typically busy people, the cover letter should also include information with respect to how long the survey might take to complete and a statement of assurance regarding the confidentiality of the data. Assurance of Institutional Board (IRB) approval should also be included and a 'thank you' in anticipation of completion is always good practice, both in the cover letter and at the end of the questionnaire.

Sampling and sample size

Sampling

For surveys that use the membership lists of the smaller component anesthesia societies (e.g., SPA), surveys can be sent out to all members of the target population (convenience sampling). However, for larger parent societies (e.g., ASA) or large patient populations, this may become too unwieldy and, as such, it may be necessary to employ some type of probability sampling.[3,15] This approach should provide a representative sample without the need to survey every member of the target group. Examples of probability sampling include simple random sampling of the target population using computer-generated tables of random numbers or stratified random sampling in which random selection occurs among different subgroups or strata of the target (e.g., by race/ethnicity). This approach ensures that different subgroups are appropriately represented. Systematic sampling is another approach wherein every *nth* subject is included in the sample. Determination of the *n*th interval between selected subjects is made based on the anticipated size of the sample in relation to the size of the target population. For example, if a sample of 500 is required from a target population of 4000, then every eighth subject would need to be surveyed. In systematic sampling, the first subject is identified by random assignment and every *n*th subject selected thereafter.

Sample size

As with all research, the robustness of the findings is generally a function of the sample size. In survey

research, it is important that the sample be representative of the target population. Small samples resulting from poor planning or large nonresponse rates may bias the results and interpretation of the findings. Although some surveys, through necessity, utilize convenience samples, it is always good practice to perform an *a priori* sample size estimation and formulate a plan to optimize subject participation.

One simple rule of thumb is to base the sample size on the ratio of subjects to variables or items in the survey. Ratios of 10–20 subjects per item have been suggested by some authors as sufficient to allow for higher level statistics such as multivariate or factor analyses using this approach.[16,17] Alternatively, some researchers believe that a set minimum sample size of 100–500 is sufficient for most surveys.[18–20] A more robust method, however, is to estimate the size of the target population from which the sample will be drawn and then establish a confidence level and confidence interval for the data.[3,21,22] The confidence interval establishes the investigator's level of confidence in the data (usually 95%) and the confidence interval, the margin of error. Typically, the margin of error is set at + or −4 or 5%. This type of sampling is similar to that seen in political or opinion polls. Thus, if the survey shows that 75% of patients in your sample were 'very satisfied' with their anesthetic care and you set a confidence level of 95% and a confidence interval of + or −5%, then one would be 95% confident that if you had sampled the entire target population, between 70% (75−5) and 80% (75 + 5) would have responded in the same way. Thus, if you can estimate the size of your target population and establish a confidence level and confidence interval, it will be possible to generate a sample size that should be representative of that target population. A number of free sample size calculators or tables are available online that can help with this. If this is not possible, it will be important, at minimum, to assess representativeness based on the degree to which the demographics of the respondents reflect those of the target population.

Survey bias

There are three main types of biases that are important to consider when performing a survey; self-report bias, recall bias, and nonresponse bias.[3,8] *Self-report or social-desirability* bias can manifest when respondents deliberately downplay or exaggerate characteristics or behaviors that place them in a potentially negative or positive light, respectively. Examples of this might include questions related to smoking or alcohol use, or individual clinical practices. To obviate this potential bias, the survey should be anonymized so that respondents are more likely to respond honestly. *Recall* bias can occur when relying on information that occurred in the distant past. Asking questions that respondents are unlikely to remember will generally engender a guess rather than facts. Unless a distant event was memorable, e.g., heart attack, consider shorter time references as appropriate, e.g., 'in the last month'.

The third and perhaps most important potential bias is *nonresponse*. There is no set standard for what represents a good response rate but obviously the greater the response, the more likely the data will be representative. In general, online surveys tend to have poorer response rates compared with paper-based surveys.[23,24] If a survey has a poor response rate, there is concern that the nonrespondents are in some way different from the respondents and, as such, may bias the results. For example, if subjects do not respond because they had a bad outcome, then the outcome data will be underreported and thus may be misleading.

To optimize response rates, surveys should be interesting, relevant, visually pleasing, and well constructed. Some investigators will also include small monetary incentives to help increase response rates but these necessitate additional costs and must be IRB approved. Because of the importance of this potential bias, there should be a plan for mitigating nonresponse by employing strategies such as e-mail reminders and providing several opportunities for subjects to access a survey link or receive additional mail surveys.[24] In general, there should be no more than three follow-up reminders. Dillman recommends that these occur at 2, 4, and 8 weeks following the initial survey distribution.[8] If the response rates remain low after three attempts, one useful technique is to send out a short survey (four to five questions) to the nonrespondents to determine if they differ substantively from the respondents. This short survey could contain some basic demographics with an opportunity for the subject to provide reasons for their nonresponse. This approach helps to determine if the nonresponse was simply due to a lack of interest or time or, more importantly, outcome.

Reliability and validity

Some variables in a survey cannot be measured by a single item. For example, many psychological or behavioral traits are unobservable and can only be measured in a survey by developing constructs (latent variables) that underlie that behavior or trait. Constructs can be identified by asking a series of questions that address similar behaviors that are thought to define the construct. For example, if we are interested in knowing whether subjects

participating in research studies are altruistic, a single yes/no response might tell us only part of the story. Instead, it would be more valuable to identify a number of questions that measure the presence or absence of different but associated behaviors or traits that describe 'altruism.' For example, questions such as: 'I think of myself as a generous person' and 'I go out of my way to help others if I can' describe two possible altruistic behaviors.

As with all types of research, it is important that any instrument used to measure something is both reliable and valid and this is no less true when the survey is 'the instrument'.[14] In survey research and instrument development, reliability refers to the reproducibility of the data. There are many different types of reliability measures and not all are appropriate for all surveys, but for this review, we will mention some of the more commonly used measures. *Test–retest* reliability measures the stability of responses over time. This requires that the survey be administered at two different times and the responses compared. When this is done in the same individual, it is termed *intraobserver* reliability. *Interobserver* reliability measures how different respondents rate the same measure or construct. These reliability measures are typically expressed as correlation coefficients such as Spearman's rho or Pearson's coefficients. Correlation coefficients of >0.7 generally indicate good reliability. Another important measure of reliability is *internal consistency*. This is used to measure how several items in a scale or construct vary together and are typically expressed using Cronbach's alpha (α). Values of >0.7 again indicate good internal consistency.

The validity of a survey determines how well it measures what it is supposed to measure. For simple descriptive surveys, validity testing may not be appropriate, but if the survey is designed to measure something, e.g., 'patient satisfaction', it will be important to determine if the items accurately capture the intended constructs. We have discussed earlier the importance of pilot-testing a survey to establish face and content validity, but other types of validity such as criterion and construct validity are also important. *Criterion* validity is a measure of how well the items or scales in a survey correlate with a 'gold standard,' if one is available. *Construct* validity is a harder concept to understand but refers to how meaningful the items and constructs are in practice, i.e., how do the items or scales in a survey actually reflect the true theoretical meaning of the concept? Items in a survey with good construct validity should correlate well with different methods of obtaining the same information (*convergent* validity) and not correlate with related but distinct traits or concepts (*divergent* validity). For example, items in a survey that measure satisfaction should correlate well with other methods of measuring satisfaction but not with other related concepts such as unhappiness.

Analyzing the data

Although a detailed review of statistical methods is beyond the scope of this article, most survey data can be described in terms of simple frequency distributions and measures of central tendency such as means, medians, and modes. Surveys that employ validity and reliability testing will generally utilize correlation coefficients as described above. When defining constructs (e.g., satisfaction with care) or validating an instrument or tool (e.g., obstructive sleep apnea) that are measured using multiple items, a factor analysis may be a useful technique as a means of reducing the number of items or factors to only those that explain the largest proportion of the variance.[18]

Summary

Although survey research represents only a small portion of the pediatric anesthesia literature, its value in examining such things as anesthesia management practice patterns, individual approaches to different anesthetic dilemmas, and patients'/parents' perceptions of their anesthesia experience can be significant. While many physician-investigators are trained to conduct the traditional observational and analytical studies, survey research has typically remained the preserve of the social scientists. With this in mind, this review describes the 'nuts and bolts' of survey research and offers some rules of thumb as a means to help the budding survey researcher navigate the survey process. An understanding of these basic elements should help ensure that anesthesia-related surveys are well conceived and conducted and increase the likelihood of generating meaningful and publishable data that progress beyond the annual conference poster presentation.

Disclosures

There are no conflicts of interest to report. IRB approval was not required.

Funding

This review was supported by the Department of Anesthesiology, University of Michigan Health System.

References

1 Passmore C, Dobbie A, Parchman M *et al*. Guidelines for constructing a survey. *Fam Med* 2002; **34**: 281–286.
2 Fink A. How to Ask Survey Questions. Thousand Oaks, CA: Sage Publications, 2002.
3 Krosnick J. Survey research. *Annu Rev Psychol* 1999; **50**: 537–567.
4 Kirsch I, Jungrlblut A, Jenkins L *et al*. Adult Literacy in America: A First Look at the Findings of the National Adult Literacy Survey. Washington, DC: US Department of Education, 2002.
5 McLaughlin G. SMOG grading – a new readability formula. *J Reading* 1969; **12**: 639–646.
6 Flesch R. A new readability yardstick. *J Appl Psychol* 1948; **32**: 2211–2223.
7 Bourque L, Fielder E. How to Conduct Self-Administered and Mail Surveys. Thousand Oaks, CA: Sage Publications, 2002.
8 Dillman D. Mail and Internet Surveys: The Tailored Design Method. Hoboken, NJ: John Wiley & Sons, 2007.
9 Peterson R. Constructing Effective Questionnaires. Thousand Oaks, CA: Sage Publications, 2000.
10 Fink A. How to Design Surveys. Thousand Oaks, CA: Sage Publications, 2002.
11 Nowlis S, Kahn B, Dhar R. Coping with ambivalence: the effect of removing a neutral option on consumer attitude and preference judgments. *J Consum Res* 2002; **29**: 319–334.
12 Si S, Cullen J. Response categories and potential cultural bias: effects of an explicit middle-point in cross-cultural surveys. *Int J Org Anal* 1998; **6**: 218–230.
13 Teclaw R, Price M, Osatuke K. Demographic question placement: effect on item response rates and means of a Veterans health administration survey. *J Bus Psychol* 2012; **27**: 281–290.
14 Litwin M. How to Measure Reliability and Validity. Thousand Oaks, CA: Sage Publications, 2002.
15 Fink A. How to Sample in Surveys. Thousand Oaks, CA: Sage Publications, 2002.
16 Arrindell W, van der Ende J. An empirical test of the utility of the observer-to-variables ratio in factor and component analysis. *Appl Psychol Meas* 1985; **9**: 165–178.
17 Hair J, Anderson R, Tatham R *et al*. Multivariate Data Analysis. Saddle River, NJ: Prentice-Hall, 1995.
18 Comfrey A, Lee H. A First Course in Factor Analysis. Hillsdale, NJ: Erlbaum, 1992.
19 Gorusch R. Factor Analysis. New York: McGraw-Hill, 1983.
20 MacCallum R, Widaman K, Zhang S *et al*. Sample size in factor analysis. *Psychol Methods* 1999; **4**: 84–99.
21 Krejcie R, Morgan D. Determining sample size for research activities. *Educ Psychol Meas* 1970; **30**: 607–610.
22 Bartlett JI, Kotrlik J, Higgins C. Organizational research: determining appropriate sample size in survey research. *Inform Technol Learn Perform* 2001; **19**: 43–50.
23 Dommeyer C, Baum P, Hanna R *et al*. Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations. *Assess Eval Higher Educ* 2004; **29**: 611–623.
24 Nulty D. The adequacy of response rates to online and paper surveys: what can be done? *Assess Eval Higher Educ* 2008; **33**: 301–314.