

READERS' TOOLBOX

Understanding Research Methods

Survey Research

David A. Story, M.D., F.A.N.Z.C.A., Alan R. Tait, Ph.D.

ANESTHESIOLOGY 2019; 130:192–202

“It’s always further than it looks. It’s always taller than it looks. And it’s always harder than it looks.”

—The Three Rules of Mountaineering

Elements of Good Survey Research A Readers' Guide



When assessing survey research studies,
readers should ask...

- ✓ Is the survey's aim important?
- ✓ Is the population surveyed relevant?
- ✓ Are there any ethical concerns?
- ✓ Was the survey validated?
- ✓ Was the survey pilot tested?
- ✓ Is the actual survey available for review?
- ✓ Is there a clear analysis plan?
- ✓ Are the results systematically reported?
- ✓ Was a statistician involved?
- ✓ Was there an adequate response rate (>40%)?
- ✓ Are the results precise (95% CI excludes group differences)?
- ✓ Are there themes from free-text responses presented?
- ✓ Do the results support the conclusions?

The ideal survey has...



Important
Question



High
Response
Rate



Precise
Results



Systematic
Answer

SUMMARY

Surveys provide evidence on practice, attitudes, and knowledge. However, conducting good survey research is harder than it looks. The authors aim to provide guidance to both researchers and readers in conducting and interpreting survey research. Like all research, surveys should have clear research question(s) using the smallest possible number of high-quality, essential, survey questions (items) that will interest the target population. Both researchers and readers should put themselves in the position of the respondents. The survey questions should provide reproducible results (reliable), measure what they are supposed to measure (valid), and take less than 10 min to answer. Good survey research reports provide results with valid and reliable answers to the research question with an adequate response rate (at least 40%) and adequate precision (margin of error ideally 5% or less). Possible biases among those who did not respond (nonresponders) must be carefully analyzed and discussed. Quantitative results can be combined with qualitative results in mixed-methods research to provide greater insight.

(*ANESTHESIOLOGY* 2019; 130:192–202)

Surveys remain the foundation of social science research but can be employed in almost any discipline, including medical research. However, good survey research is harder than it looks. Anesthesiology researchers use surveys to research behaviors, attitudes, and knowledge of both physicians and patients or determine population characteristics, such as disease states, practices, or outcomes. Examples of surveys include transfusion practices among American Society of Anesthesiologists (ASA) members,¹ use of ultrasonography for regional anesthesia,² and parental understanding of informed consent for research.³ However, many journals are reticent about publishing survey research because of poor quality.^{4–6} Some organizations, such as the Australian and New Zealand College of Anaesthetists and the Society for Pediatric Anesthesia, have introduced formal

Corresponding article on page 181.

Submitted for publication January 8, 2018. Accepted for publication July 27, 2018. From the Centre for Integrated Critical Care and Melbourne Clinical and Translational Sciences, The University of Melbourne, Melbourne, Victoria, Australia (D.A.S.); and the Department of Anesthesiology, Center for Bioethics and Social Sciences in Medicine, and Michigan Institute for Clinical and Health Research, Michigan Medicine, Ann Arbor, Michigan (A.R.T.).

Copyright © 2019, the American Society of Anesthesiologists, Inc. Wolters Kluwer Health, Inc. All Rights Reserved. *Anesthesiology* 2019; 130:192–202

vetting processes to improve the quality of survey research and decrease respondent fatigue and burden.^{7,8}

Several survey research errors (biases that divert from the truth) were seen in what is widely regarded as the greatest survey disaster: the *Literary Digest* survey of 10 million Americans that incorrectly predicted that Roosevelt would lose the 1936 Presidential election in a landslide when, in fact, the absolute opposite occurred.⁹ Problems (errors) with this survey included an unrepresentative sample (affluent Americans with phones), a low response rate (20%), and nonresponder bias (Roosevelt voters tended not to respond). As we will discuss, all these errors can be avoided or at least minimized. As Dillman *et al.*¹⁰ note, the entire survey process (from design to reporting) needs to be tailored to the question asked, which in turn is the first step: ask a clear question.

In the absence of international consensus guidelines on conducting and reporting survey research, we aim to discuss the elements of good survey research, outline some of the pitfalls, and introduce some newer approaches to collecting and analyzing survey data. We provide pragmatic toolboxes for survey researchers (table 1) and survey report readers (table 2) and suggest minimum standards for submitting a survey (table 3).

Design Considerations

The primary aim of any survey is to answer a good research question that is interesting for the broader target population.^{4-6,10,11} A good, clear survey has further interrelated advantages: shorter, simpler items that decrease the time to complete and enhance the response rate. Further, effective surveys focus exclusively on “need to know” questions, not those that might be simply “nice to know.”⁵ The aims of

Box 1. What to Look for in Research Using This Method

- Important topic
- Appropriate participants
- Valid questions
- Good response rate
- Precise results
- Conclusions consistent with results

any survey should also be clearly stated in concrete terms, *e.g.*, “To describe the current practice patterns of Nordic anesthesia departments in anesthetic management of endovascular therapy in acute ischemic stroke.”¹²

The choice of survey design will depend on the questions being asked, the population of interest, and available resources.^{5,10} Each type of survey has advantages and disadvantages (table 4). The questions (items) in a survey should reflect the objectives of the study.^{4,6,13} Whereas some surveys are designed to simply measure knowledge, others measure constructs, practices, or behaviors. Thus, researchers should consider the research goals when writing and formatting the questionnaire (instrument). In general, surveys should be short, relevant, focused, interesting, easy to read, and complete. Surveys that lack these attributes often suffer from poor response rates and decreased reliability.^{10,14}

When designing a survey, it is important to know your audience. Researchers and readers should put themselves in the position of the intended respondents. How might they react to being approached and how might they respond to the questions asked? Motivated participants, for example, may be more willing to answer more detailed or probing questions. Questions should be written by using simple

Table 1. Toolbox for Survey Researchers

1. Have a clear research question and/or hypothesis.
2. Clearly define the study population.
3. Discuss potential questions with colleagues or conduct focus groups with representatives of the target audience. Use or modify existing validated survey questions or test the reliability and validity of new questions (if appropriate).
4. Ensure appropriate human research ethics approval.
5. Consider the BRUSO approach (brief, relevant, unambiguous, specific, and objective) when constructing questions. Minimize respondent burden by asking as few questions as possible with simple language, complete and unambiguous sentences, and without abbreviations, jargon, or acronyms. Avoid negative questions.
6. Consider a preponderance of closed-ended responses for postal or online surveys. Lists of response options should be exhaustive and mutually exclusive.
7. Have at least one open-ended question (“Please clarify/describe,” or “Any further comments?”) to provide or clarify additional information regarding specific questions or general final comments on the topic or survey. Be sure to have an analysis plan for these questions.
8. Always pilot-test the cover letter, instructions, and survey before full implementation.
9. An online or postal survey should take less than 10 min to complete.
10. Consider a statistician collaborator for design and analysis. Alternatively, use website/statistical package resources. Beware of overestimating your statistical skills.
11. Calculate a sample size with a maximum 95% CI margin of error of $\pm 5\%$ on the primary question.
12. Modify the sample size for the anticipated response rate. Conservative: 30 to 40% response rate and a 2 to 3 \times multiplier.
13. Select a sampling approach that is both practical and statistically robust. If needed, employ stratified or oversampling techniques for groups that may be typically underrepresented.
14. Use as many strategies as possible to maximize response rates and mitigate bias caused by nonresponse, recall, and social desirability.
15. Statistical analyses including reliability and validity testing should be based on the type of data collected (qualitative and/or quantitative) and the goals of the survey. A mixed-methods approach will often provide richer and more robust information but may require consultation with a statistician or social scientist.
16. The survey should be included with any manuscript submission.

Table 2. Toolbox for Survey Readers

1. Is the aim important? The introduction and discussion should outline the importance. Well answered, important questions are likely to influence practice, attitudes, or policy. The greater the possible effect, the greater the importance.
2. Does the intended study population represent all those relevant to the survey research aim? Who has been left out? For example, medical surveys may exclude those outside academic centers, and patient surveys may exclude the disadvantaged. This is an example of possible bias: something that deviates results away from the truth.
3. Are there any ethical concerns? Institutional Review Board approval? Beyond unacceptable professional behavior, bad ethics equals bad science. Unethical approaches, such as coercion or planned exclusion, can bias results.
4. Has the survey been used before or validated? Beware the novel survey; it may not measure what is wanted. Assess for reliability and validity if appropriate.
5. Was the survey piloted? Without piloting, persisting unrecognized biases are more likely. Ask yourself: how you think you would have navigated/answered the survey? If you think you would have struggled to finish and then written "This survey is a dud" in the further comments, you are probably not alone.
6. Are the survey questions, as asked, readily available? Assessing the report without the survey is like navigating without a compass.
7. Is there a clear analysis plan? A chaotic approach can lead to unreliable results.
8. Are the results systematically or selectively reported (cherry picked)? Cherry picking increases bias.
9. Is one of the investigators a statistician or is a statistician acknowledged? If not, the analysis may be less reliable.
10. What is the response rate? The higher the response rate, the less the likelihood of nonresponder bias. Articles should always discuss response rate as a strength if it is high (e.g., > 70%); if it is low (< 50%), the authors should make some effort to rule out obvious nonresponder bias, e.g., gender gap. How was bias mitigated?
11. How precise are the results? The total number of respondents will be an important factor for precision. Consider the range of possibilities the 95% CIs cover. Beware of imprecise results that cover more than one conclusion, e.g., if the authors conclude there is no difference but the 95% CI includes the possibility that groups differ, then the results are less reliable than if the 95% CI included only a result of no difference.
12. How were free-text answers assessed? Good articles systematically look for themes.
13. Do the results clearly support the conclusions? The ideal survey has precise results and a high response rate for an important question that is systematically answered.

language⁴ at a reading level commensurate with the literacy of the intended audience. In the United States and most developed countries, surveys of the general population should be written at no more than an eighth-grade reading level and avoid abbreviations, jargon, colloquialisms, acronyms, or unfamiliar technical terms. Further, language and cultural differences may also be important considerations, including tendencies to want to please, or conversely avoid, perceived authority figures such as doctors. Surveys for professionals, such as physicians, can have more complex technical words, but simple and clear structure and wording help everyone.

Questions (items) validated in previous research should be used whenever possible.^{4,15} For new or revised questions, Peterson¹⁶ developed a guide with the acronym BRUSO: brief, relevant, unambiguous, specific, and objective.⁴ First, questions should be *brief* to reduce the length of the survey. Questions should include complete sentences but not be long-winded. Questions should also be *relevant* to the

survey's purpose and focus on "need to know" information. Questions that may not appear intuitively relevant but are deemed necessary require a brief explanation about why the questions are important. Questions must be *unambiguous*. For example, asking respondents how often they check social media on a "typical work day" may mean different things to different people, i.e., what is "typical?" Questions that evoke a double-negative require logical thinking and are often answered incorrectly. Questions should also be *specific* so that the respondent is clear as to their intent; questions should be unidimensional. For example, "Do you consider yourself an empathetic and sympathetic person?" could evoke different responses because one can be sympathetic without being empathetic. This example would be better split into two questions addressing sympathy and empathy separately. Unless a primary focus of the study, demographic questions should be placed at the end of the survey and kept to a minimum. *Objective* questions should not contain words that "nudge" the answer or reveal the researchers' beliefs or opinions.

The choice of questions and response options (scales) depends on the type and goals of the survey.^{5,10,11} Interviews and certain types of written surveys are better served by open-ended questions with responses that can be electronically recorded or manually transcribed. Online and postal surveys typically employ closed-ended questions in which the respondent chooses a response from a structured list of options. Both open and closed responses have advantages and disadvantages. Open-ended responses allow the respondents to answer in their own words in a manner that reflects their personal experiences or beliefs and are less likely to be influenced by the expectations of the investigator.^{10,17–19} Open-ended questions are particularly

Table 3. Suggested Minimum Standards for Manuscript Submission

- Clear research question/hypothesis
- Well defined population
- Ethics board approval
- Clear *a priori* sampling and analysis plan
- Piloted survey as an appendix
- ≥ 40% response rate
- Precision: 95% CI with no more than ±5% margin of error on the primary question
- Analysis and discussion of potential sources of bias

Table 4. Advantages and Disadvantages of Different Survey Methods¹⁰

Survey Type	Advantages	Disadvantages
Online surveys	Minimal Cost Fast transmission/response times Access to large populations More candid responses No interviewer biases	Confidentiality concerns Technical problems Low response rates Require internet access
Postal (mail) surveys	Intermediate cost Convenience Access to large populations No interviewer bias	Lower response rate Difficulties completing the survey, e.g., literacy, language, and vision
Interviews	Personal contact with the ability to engender more personal and detailed responses Increased response rate	Cost Small samples Interviewer bias Questions limited to fewer response-option types

helpful when the researchers are unclear how respondents might respond and for developing new response options for closed-ended questions. One example is “Under what circumstances would you cancel anesthesia for the child with an upper respiratory tract infection?” The major disadvantages of open-ended questions are that responses can be long, difficult to transcribe, and difficult to classify and may need experts to identify underlying themes. Further, surveys with a lot of open-ended questions may have incomplete or missing answers because of response fatigue.

Closed-ended (structured) questions differ from open-ended by providing a list of options to choose from.^{5,10,20} Closed-ended questions are optimal for postal and online surveys because they provide standardized responses, take less time to complete, and are easier to analyze. The major disadvantage of closed-ended questions is that they can be more difficult to write^{5,10} because the response options must be both exhaustive (include all important options) and mutually exclusive (each option should be distinct). Including every possible option can result in excessively long lists of responses that increase survey fatigue and nonresponse. One strategy to limit the number of responses while avoiding missing important data is to include an “other” response with a clarifying “please describe/specify.” Further, for all surveys, a final open question of “Any further comments?” allows respondents to freely comment on both the topic and the survey itself.²¹

Including too many questions can result in satisficing, where respondents increasingly fail to carefully consider the questions and subsequently provide answers that are not well thought out.^{10,19} Survey Monkey (<http://surveymonkey.com>) report from their data²² that respondents will spend an average of 5 min to answer 10 questions in an online survey but only 10 min to answer 25 questions. This suggests that as the number of questions increases, the time spent on each question decreases, *i.e.*, satisficing. Further, if a survey takes 10 min to complete, data show that up to 20% of respondents will abandon the survey before completing it.²² Respondents may also be more likely to abandon surveys with compulsory

questions, particularly if they do not include a “Don’t know” type option. Compulsory questions should be minimized and, if used, should always include a “Don’t know,” “Not applicable,” or “Don’t wish to answer” option.

Closed Format Responses

1. Multiple choice

How many minutes do you usually wait to see a doctor in our clinic?

☐ less than 5 ☐ 5-14 ☐ 15-30 ☐ more than 30

2. Rating scales

How would you rate the waiting time to see a doctor in our clinic?

1 2 3 4

 poor average good very good

3. Pictorial scales

How happy are you about the waiting time to see a doctor in our clinic?



4. Agreement scales (Likert-type) scales

The waiting times in our medical clinic are acceptable, (please circle one)

strongly disagree disagree uncertain agree strongly agree

5. Visual analogue scale (10 cm line)

How would you rate the waiting time to see a doctor in our clinic?

very poor -----X----- very good

Fig. 1. Examples of open- and closed-format response options for questions (reproduced from *Anaesthesia and Intensive Care*⁵ with the kind permission of the Australian Society of Anaesthetists).

Table 5. Selecting Categorical and Ordinal Questions and Responses

- Questions and scales should be meaningful, i.e., address the aims of the study
- Categorical scales should be exhaustive and mutually exclusive
- Choose 3 to 7 response options (5 is typical)
- Options should be balanced (negative to positive)
- Present lists of options vertically
- Choose a neutral response if a conceptual midpoint is valid

Response scales (fig. 1)⁵ are typically categorical/nominal (e.g., male/female, true/false); ordinal, in which the responses are ordered (e.g., very anxious to very calm); or numerical (e.g., age, height). Categorical and ordinal response options (table 5) typically take the form of Likert scales²³ with different levels of response, for example, “I preoxygenate patients before general anesthesia” could be answered by using the following list formatted vertically:

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

When formatting these scales, the endpoints should be mirror opposites, be balanced, be presented from negative to positive, include equal intervals, and be presented as a vertical rather than horizontal list. Vertical formatting is less subject to mistakes when responding and easier to code.

Depending on the degree of precision required, questions should offer three to seven responses, with five probably optimal. Some survey researchers omit a “neutral” response option to force respondents one way or another or because the researchers argue that a neutral option discourages respondents from answering. Others argue that a neutral response provides a natural choice. Decisions regarding the number of response options and inclusion/exclusion of a neutral response should be made during pilot pretesting. Pilot testing with and without a neutral response option can provide a sense of whether responses tend to cluster around a middle point.

Other survey formats include visual analog scales (e.g., visual analog pain scales²⁴) that ask respondents to either circle or electronically mark a number (typically 0 to 10) or a 100-mm scale to indicate their level of response. Again, like pain scales, there should be descriptive anchors at each end of the scale to provide context. Other types of scales include ranking scales, where the respondent ranks a set of ideas or preferences; matrix scales, where the respondent evaluates one or more row items using the same set of column choices; magnitude estimation scales; and factorial questions, in which a vignette is presented that requires a judgement or decision-making response.

In addition to consideration of the types of questions and response scales, it is also important to consider how

questions transition from one to another. Skip or branch logic is a feature that routes the participant to subsequent questions or page/sections based on their response to a particular question. This is an important process that allows participants to avoid questions that do not apply to them, e.g., “If you replied ‘No’ to question 3, please skip to question 8.” The routes used in skip logic should be thoroughly pre-tested before implementation. For readers, the easiest way to test the flow of questions is to imagine answering the survey.

Reliability and Validity

Not all surveys require formal reliability and validity testing, e.g., simple descriptive surveys (table 6). However, for surveys that are designed to describe or measure constructs, e.g., pain, sleep quality, altruism, empathy, it is critical to ensure that the items in the survey or instrument actually measure what they are designed to measure. All survey measures, whether quantitative or qualitative, are subject to error.²⁵ These errors can either be due to random chance and/or errors in the survey itself: measurement error.¹⁰ Measurement errors reflect the accuracy of the survey, i.e., do the questions measure what they are supposed to measure (validity), and are they reproducible across individuals over time (reliability)? The validity and reliability of questions can be quantified statistically, often by strength of association with other metrics.²⁶

As with all research, the first step in survey research is to review the literature for existing surveys or survey questions that have already been formally tested. It makes no sense to generate a new set of questions as substitutes for ones that have already been validated. Therefore, it is preferable to use or adapt existing questions or surveys that have demonstrated validity, with appropriate acknowledgment or citation. Although some reliability/validity testing may still be required, the burden of formal testing of new questions is greatly reduced.

For developing *de novo* survey questionnaires (instruments), Sullivan²⁶ provides sage advice: “Researchers who create novel assessment instruments need to state the development process, reliability measures, pilot results, and any other information that may lend credibility to the use of homegrown instruments. Transparency enhances credibility.” Readers’ should look for these points in novel surveys and whether the validity of previously reported items/surveys has been demonstrated.

Reliability

Reliability is the degree a measurement yields the same results over repeated trials or under different circumstances.^{25–27} *Test–retest reliability* reflects the stability of the survey instrument and can be measured by having the same group of respondents complete the identical survey at two points in time. Surveys with good test–retest reliability typically have little variance between the two sets of data. *Interrater reliability*

Table 6. Reliability and Validity Estimates

	Measures
Reliability	
Test–retest	Stability of responses over time in the same group of respondents
Interobserver	How well two or more respondents answer the same questions
Intraobserver	Stability of responses over time in the same individual respondent
Alternate form	Ability of differently worded questions to evoke similar responses
Internal consistency	How well questions in a survey vary together in a sample
Validity	
Face	How questions appear to lay individuals (face value)
Content	How questions appear to individuals with relevant expertise
Concurrent	How well the questions perform against a “gold standard”
Convergent	How well questions perform against other questions that are theoretically associated with the same construct of interest
Divergent/Discriminant	The degree to which questions perform independently of questions describing similar but distinct concepts
Construct	The degree to which the questions in a survey measure what they are supposed to measure

refers to how two or more respondents respond to the same questions and *intraobserver reliability* refers to the stability of responses over time in the same individual.

Similar question wording or order is subject to “practice” effects that can be overcome by rewording a question or reordering the responses. Questions with similar responses regardless of wording or order are said to have good alternate-form reliability.

Because not all traits or behaviors are observable or can be measured by a single question, researchers often use several questions to describe the same behavior or trait of interest (constructs). *Internal consistency reliability* is the degree to which these questions vary together as a group, *i.e.*, the degree to which these different questions consistently measure the same construct. For example, because depression is hard to measure by using a single question (Are you depressed?), researchers employ several different questions that address different but related aspects of depression, *e.g.*, fatigue, trouble concentrating.

Validity

Validity measures the degree to which questions in a survey measure what they are intended to measure.^{25–27} For example, questions designed to measure pain should measure pain and not something else, such as anxiety. Although some validity metrics are relatively easy to measure, some are more complex. Two types of validity that are easy to measure are face and content validity. Face validity refers to how the questions appear (on “face value”) to individuals with little expertise in the survey topic. Although face validity is a somewhat casual assessment, it nonetheless reassures the investigator that the questions will make sense at a layperson’s level. Content validity, on the other hand, requires input from content experts. Neither face nor content validity is statistically quantifiable, yet both can provide important information to ensure that

questions are relevant. For example, a survey of pain techniques by anesthesiologists might benefit from pretesting with a small group of surgeons (face validity) and pain medicine specialists (content validity). The value of expert consultation before implementing any survey cannot be understated.

Construct validity is harder to conceptualize but is a measure of the degree to which survey questions, when applied in practice, reflect the true theoretical meaning of the concept. Construct validity is typically established over years of use in different settings and populations. Although there is no simple metric for construct validity, social scientists typically use other quantifiable measures, such as comparing against an existing “gold standard.” This type of validity is termed *concurrent criterion validity*.

Where there is no gold standard, construct validity can be established by measuring the degree to which the questions in a survey correlate with other measures that should theoretically be associated with the same construct (convergent validity). For example, to validate a new survey instrument to measure sleep quality, it might be important to compare it with other measures of sleep quality (*e.g.*, direct observation). If convergent validity is established, a natural follow-up test would be to see whether the same questions are able to discriminate between sleep quality and other related, but different, measures such as sleep quantity. If these two measures do not correlate, we assume (if other validity measures confirm) that they are measuring two separate constructs and that the sleep-quality questions demonstrate good divergent or discriminant validity.

Ethics Review

Ethics committee or institutional review board approval is typically required before testing and implementing any survey. The primary ethical concerns of surveys relate to content (*e.g.*, could items be psychologically damaging?) and how confidentiality will be maintained. Although surveys may not be

identifiable by the participant's name, there are other sources of information, *e.g.*, IP addresses and email addresses, that could potentially link the survey with the participant. This is particularly important when using third-party software services, *e.g.*, Survey Monkey and Qualtrics. Investigators should thus be aware of the security agreements of each company and assure participants that their information will be maintained in a confidential manner, *e.g.*, stored and maintained on password-protected computers or cloud storage and/or how any identifying information will be delinked.

Pretesting (Piloting) the Survey

Although there is no such thing as a perfect survey, pretesting or pilot testing can significantly enhance the effectiveness of any survey. Unfortunately, this step is often missing.⁶ Pretesting is typically conducted in two phases. First, the research team reviews all aspects of the survey, *i.e.*, the instructions, the order and flow of questions, whether it contains skip or branch logic, how long the survey should take to complete, and whether specific questions are ambiguous and/or are being consistently missed. Second, the survey should be distributed among a small subset of the intended audience before it is administered to the larger target group. This can be done somewhat informally but can also involve structured focus groups followed by thorough debriefing. Even if previously validated surveys are used, questions should be pretested because meaning can often be affected by the context of the survey. No matter what the design, the piloted survey should be submitted as part of any manuscript, possibly as an appendix.

Sampling

Precise estimates of large populations, up to millions of people, can be derived from survey samples of fewer than 2,000 people.^{10,28,29} Thus, because it is not always practical to survey an entire population, sampling provides an efficient way to collect data that, if done correctly, can be representative of the population of interest. A representative sample should mirror the characteristics of the broader population, ensuring generalizability and reducing the effect of sample bias. However, although representativeness is a primary goal, the sampling approach will also depend on the type of survey, the target population, inclusion of subgroups, and resources/cost.

Because a survey of the entire ASA membership (53,000 members in 2016) might be impractical, an option would be to generate a sample that is representative of important characteristics of the ASA membership, such as sex, ethnicity, and training. This is best achieved by employing some type of simple random sampling.^{17,29} However, there may be instances in which investigators may want to focus on a subgroup or oversample groups that are underrepresented, *e.g.*, rural practitioners. In these cases, stratified sampling can be employed in which random samples are drawn from each subgroup or strata, *e.g.*, ASA membership by geography. In cases in which

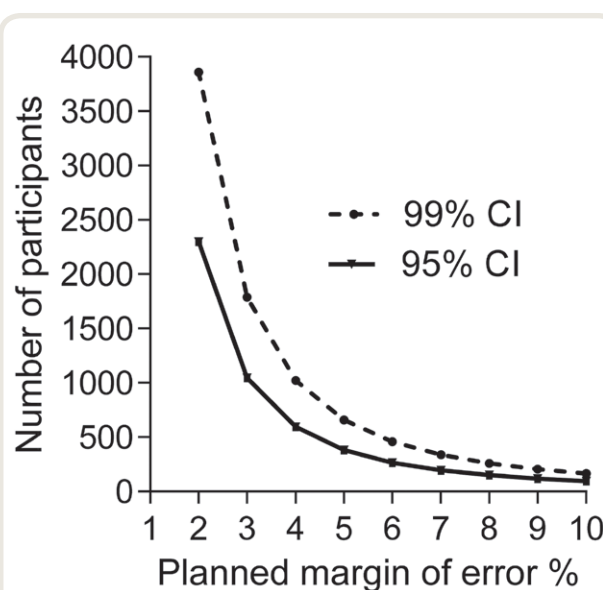


Fig. 2. Effects of different planned margins of error ($\pm\%$) and the 95% and 99% CIs on sample-size estimations for a survey of the entire membership of the American Society of Anesthesiologists ($N = 52,905$). Note: The actual required sample size will also be affected by the response rate.

there may be underrepresentation of certain groups, other methods such as oversampling should be employed.

Sample-size Estimates

Sample-size estimates should be based on the primary question²⁹ and large enough to be confident (usually 95 or 99%) that results from the entire population will lie within the desired margin of error of the sample (fig. 2).²⁹ Typically, the maximum acceptable margin of error for a proportion (percentage) of the population is set around $\pm 5\%$ (most political polls quote 3 to 5%). That is, if the margin of error is $\pm 5\%$, and 25% of respondents from a sample of 325 ASA members reply that they use thiopental for induction, the 95% CI (the results of 95 of 100 repeated samples of the sample) shows that between 20 and 30% of ASA members use thiopental. Small samples typically produce wider CIs. By using the same example, a sample of 30 ASA members that produces a margin of error of $\pm 20\%$ would result in a 95% confidence estimate that 5 to 45% of ASA members use thiopental: an unhelpful estimate ranging from very few to almost half. However, although increasing the sample size to reduce the margin of error increases the precision of the data, there is an effect of diminishing returns (fig. 2). As the margins of error are tightened to less than 4%, the number of participants required increases disproportionately. This is important when balancing precision with the practicality, availability of resources, and costs of surveying large numbers of subjects.

For all investigators, we strongly advise working with a statistician for both planning and analysis. For those with a

background in statistics, there are several online resources^{28,31} and statistical packages such as R (available free from R Foundation for Statistical Computing, Austria), STATA (StataCorp LLC, USA), and SPSS (SPSS Statistics, IBM, USA).

Sample-size calculations can also be based on anticipated proportions, but when comparing groups, the anticipated difference may be important. Notably, calculated sample sizes are for the number of completed surveys. Although a response rate of more than 60% is considered good, less than 50% is common. Recent surveys of anesthesiologists and anesthesia fellows, for example, reported 54 and 33% response rates, respectively.^{32–36} A conservative approach, therefore, is to send the survey to approximately two to three times the calculated sample size. In general, leading journals are unlikely to publish a survey with a response of less than 30 to 40%, except in exceptional circumstances.

Survey Bias

In addition to sampling bias, there are several important ways in which error can creep into survey research.^{10,17}

Researcher Bias

Just as observer bias can adversely affect results in a randomized trial, researcher bias (subtle or overt) can affect the way questions are asked. Care must be taken to ensure that questions are objective and that personal opinions do not bias framing questions, *e.g.*, “Do you feel guilty about accepting a Do Not Resuscitate order?”^{33,36} Interviews can evoke implicit personal bias by both the interviewer and the interviewee. Researchers must avoid words that are potentially charged or could generate an emotional response.¹⁰ An extreme version of biased questions is push polling, where the hidden purpose is to drive opinions rather than ask questions,³⁷ *e.g.*, “For fluid resuscitation do you use *Dodgy-sol*, which is both dangerous and expensive?” Readers’ should look for these biases in questions.

Nonresponse Bias

Along with precision, the response rate is a central metric of survey quality.⁶ Nonresponse is one of the most frustrating aspects of all survey research, and physicians are among the worst offenders.³⁸ Topics that have widespread practice implications may enhance response rate, *e.g.*, video laryngoscopes (67% response)³⁹ or the effect of fatigue in trainees (59% response).⁴⁰ However, even well designed, hot-topic studies suffer from nonresponse. Although some nonresponse is expected and acceptable, surveys that have large nonresponse rates are subject to bias, particularly if the nonresponse is related to the survey topic (outcome) or if the nonresponders differ substantively from responders. For example, individuals who have experienced a bad or sensitive outcome may be less

willing to report it (report bias), and as such, the true outcome may be underreported. Although nonresponse is often simply a function of a lack of respondent time, its impact is often survey-specific. For example, whereas a response rate of 50% may be adequate for postal and online surveys, 85% would be considered minimally adequate for interviews.⁴¹ In any case, it is important to determine whether the nonrespondents differ substantively from respondents. The most pragmatic way to do this is to compare the demographics of the responders with the known demographics of the target population. Another way is to send a brief follow-up survey to the nonrespondents requesting basic demographics and the reason(s) for nonresponse. Using this approach for a survey project, one of us (A.R.T.) found that nonrespondents had similar characteristics to respondents and that most nonresponse was due to a lack of participant time.⁴² This follow-up may be less appropriate with patient surveys. Because nonresponse bias is an important limitation, it should always be discussed in any written publication.⁶

There are several tactics to improve response rates and mitigate the effects of nonresponse.^{43–45} Importantly, prenotification of the survey by email or postcard has increased response rates.⁴³ A professionally written cover letter that explains the importance of the study is also critically important to pique interest. Techniques such as increasing “white space,” emphasizing important points with bolding/underlining, and use of color tend to engender better response rates.^{15,43} For surveys dealing with sensitive topics, response rates will be greater if the data are anonymized or if confidentiality is assured.

For online surveys, there should be email reminders with opportunities to receive additional surveys or access to the online survey link (maximum of three reminders/follow-up attempts).¹⁰ Often researchers will provide small (noncoercive) incentives to encourage respondents to complete their surveys, *e.g.*, gift cards, money, or lottery tickets, but these need to be in the planned budget.⁴⁶ Online surveys typically have poorer response rates than postal surveys^{44,47} and may also be subject to a “speed through” phenomenon, where respondents satisfice by rushing through the survey without due thought. For some online surveys, it is, however, possible to measure the time taken to complete the survey. If this time is deemed too quick based on pilot-testing estimates, the results may be unreliable. Online surveys are also limited to those individuals with online access and thus may evoke a selection bias. Despite these concerns, however, online surveys are supplanting traditional postal surveys. Although response rates can be lower than with postal surveys,^{44,47} online surveys also tend to be quicker and cheaper to administer and reach larger or dispersed audiences.

In addition to the issues posed by total nonresponse, problems can also occur when participants choose not to answer certain questions (item nonresponse). Typically, missing values are automatically excluded from the analysis and do not pose a problem. However, if the percentage of missing responses is high, *e.g.*, more than 20%, the investigator may choose to

correct for this by imputing the missing data. In any case, it is important that missing data are reported to allow the reader to estimate the potential impact of the item nonresponse.

Recall Bias

Recall bias refers to error associated with respondents being unable to adequately recall past events. To minimize recall bias, questions should be framed in time periods calibrated for the events, *e.g.*, “difficult intubations in the last 3 months.”

Self-report Bias

Often called social desirability bias, this type of bias refers to the tendency for individuals to downplay negative attributes. Asking parents whether they smoke in the house, for example, is likely to be underreported because parents often know that second-hand smoke is inherently bad for their children. Assuring that responses are either anonymized or that confidentiality will be honored will typically reduce the potential for self-report bias.

Analysis

Analysis of survey data should be based on a predefined end-point and will depend on the type of data collected and the question(s) asked.¹⁷ Most quantitative survey research involves descriptive frequency data involving proportions and measurements of central tendency, *e.g.*, means and medians, and variability, *e.g.*, SD and range. Comparisons between groups will again depend on the type of data collected, *i.e.*, continuous data *versus* categorical data. For these data, simple statistics, such as Student's *t* tests, ANOVA, and the chi-square test, can be used, as appropriate. Analyzing categorical data, such as Likert scales, can present challenges. For example, imagine a five-point Likert scale of “extremely dissatisfied,” “dissatisfied,” “neither dissatisfied nor satisfied,” “satisfied,” and “extremely satisfied” used to test the attitude of 1,000 Australian anesthetists to a new laryngoscope: the Bonza-Scope. The proportions giving each response could be stated and compared by using the chi-square test. Another option (with greater statistical power) is to combine “extremely dissatisfied” with “dissatisfied” and “satisfied” with “extremely satisfied.” The summed results could thus be that 60% were satisfied, 10% neutral, and 30% dissatisfied with the Bonza-Scope. A simple analysis would be to just compare the proportion who are satisfied with the proportion who are dissatisfied. This provides “headline” statistics, *e.g.*, “In a survey of 1,000 Australian anesthetists, 60% were satisfied with the Bonza-Scope, whereas 30% were dissatisfied (difference 30%, 95% CI: 26 to 34%, $P = 0.002$).”

Another approach is to create dummy variables for categorical data.^{17,19} For example, data using the same five-point Likert scale of “extremely dissatisfied” to “extremely

satisfied” can also be coded from 1 to 5 (*e.g.*, 1 = extremely dissatisfied, 5 = extremely satisfied). These are not continuous data and should not be assumed to be evenly spaced ordinal data. Parametric statistics, including mean and SD descriptive statistics, are not appropriate. These data can be analyzed as numerical data by using comparative statistics, such as the nonparametric Mann–Whitney U test, which examines rank and not magnitude. In another Bonza-Scope research project, the attitudes of Australian anesthetists might be compared with the attitudes of American anesthesiologists. If the Australian group had a median score of 4 and the American group a median score of 3 on a Likert scale question for satisfaction, rather than saying there is a median difference of 1, it is probably more meaningful to say Australians were more satisfied than Americans ($P < 0.005$).

With the advent of powerful desktop statistical programs, more complex statistical analysis can also be applied to survey research, including logistic regression for analysis of predictive factors and factor analyses that identify which individual questions or factors explain most of the variance in the data.^{48,49} This process is important in identifying which factors in a survey are important and which can be safely removed (data reduction). Again, collaborating with statisticians is likely to produce better survey design and analysis.

Open-ended questions from both oral interviews and written surveys are analyzed to identify themes.^{21,50} A theme is a patterned response within the survey data, *e.g.*, repetitions, recurring topics. For example, the question, “Under what circumstances would you cancel anesthesia for the child with an upper respiratory tract infection?” is likely to evoke different responses that can be sorted into themes. These themes might be related to patient, parent, anesthetic, or surgical factors. The importance of a theme is typically determined by its prevalence or how many respondents articulated that theme. Unfortunately, like many aspects of survey research, the importance and difficulty of thematic analysis is often underestimated.^{10,21,50}

Mixed Methods

Mixed-methods research (table 7) represents a relatively new approach to analyzing survey data.¹⁰ Although most clinical survey data are primarily quantitative, mixed methods allow researchers to integrate both qualitative and quantitative data. By integrating both data types, mixed-methods research provides richer information. Typically, mixed methods are used to corroborate results by using other approaches, develop a theory about a phenomenon, complement the strengths and/or overcome the weaknesses of a single design, or develop and test a new instrument.⁵¹

The choice of mixed methods requires a systematic approach, including determining the sequence of data collection, *e.g.*, quantitative precedes or follows qualitative; identifying what method will take priority during data collection and analysis; deciding what the integration of qualitative and quantitative data stage might involve; and deciding whether

Table 7. Mixed-methods Designs

Design	Approach	Purpose
Sequential explanatory	Collection and analysis of quantitative data followed by collection and analysis of qualitative data	Helps explain and contextualize quantitative findings
Sequential exploratory	Collection and analysis of qualitative data followed by collection and analysis of quantitative data	To explore a phenomenon and explain qualitative findings, to test a new instrument
Concurrent triangulation	Concurrent but separate collection and analysis of qualitative and quantitative data	To cross-validate or corroborate findings within the same study, overcoming weaknesses of one approach with strengths of another
Concurrent nested	The predominant collection method (qualitative or quantitative) nests or embeds the method with the lower priority (qualitative or quantitative)	Gains a broader perspective on a topic, seeks information at different levels

a theoretical perspective will be used.⁵¹ The advantages of mixed-methods designs are that they combine the strengths and diminish the weaknesses of a single design, can provide a more comprehensive understanding of the questions asked, and may be more effective in developing survey instruments. The disadvantages are that they can be complex, time-consuming, and difficult to integrate and interpret. Continuing our example: Americans may be less satisfied ($P < 0.005$) (quantitative) with a Bonza-Scope with a qualitative theme of “The handle is too big.” Again, we strongly recommend collaborating with a biostatistician or social scientist or both.

Conclusions

Poor methodologic quality of survey research is often a (negative) factor in decisions regarding publication. Producing good-quality survey research is a complex process that is harder than it looks. We hope that this article will provide investigators with useful tools (table 1) to successfully navigate the survey process and publication and provide readers with useful points to judge survey research (table 2). We also have provided a short list of suggested minimum standards (table 3) that we think can be a threshold for submitting surveys to journals. Survey reports failing these minimums will have far less likelihood of success, and submission to major journals will probably be futile. Therefore, the toolboxes and minimum standards can be used by researchers, editors, and the component anesthesia societies (e.g., Australian and New Zealand College of Anaesthetists, Society for Pediatric Anesthesia) to ensure conduct, submission, and publication of high-quality surveys for informed readers.

Box 2. Where to Find More Information on This Topic

- Survey Research, Colorado State University: www.writing.colostate.edu/guides/research/survey/
- The Tailored Design Method (Don Dillman); Social & Economic Sciences Research Center, Washington State University: www.sesrc.wsu.edu/about/total-design-method/

Research Support

Dr. Story was funded solely from institutional sources. Dr. Tait is supported by grant No. UL1TR000433 from the National Center for Advancing Translational Sciences of the National Institutes of Health (Bethesda, Maryland) and from departmental sources. The funding agency had no role in the concept, design, data collection, interpretation, or writing of this article.

Competing Interests

The authors declare no competing interests.

Correspondence

Address correspondence to Dr. Story: L1 202 Berkeley Street, Building 207, The University of Melbourne, Melbourne, Victoria 3010, Australia. dstory@unimelb.edu.au. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

References

- Nuttall GA, Stehling LC, Beighley CM, Faust RJ; American Society of Anesthesiologists Committee on Transfusion Medicine: Current transfusion practices of members of the American Society of Anesthesiologists: A survey. *ANESTHESIOLOGY* 2003; 99:1433–43
- Matava C, Hayes J: A survey of ultrasound use by academic and community anesthesiologists in Ontario. *Can J Anaesth* 2011; 58:929–35
- Tait AR, Voepel-Lewis T, Malviya S: Do they understand?: Part I. Parental consent for children participating in clinical anesthesia and surgery research. *ANESTHESIOLOGY* 2003; 98: 603–8
- Tait AR, Voepel-Lewis T: Survey research: It's just a few questions, right? *Paediatr Anaesth* 2015; 25:656–62
- Jones D, Story D, Clavisi O, Jones R, Peyton P: An introductory guide to survey research in anaesthesia. *Anaesth Intensive Care* 2006; 34:245–53
- Story DA, Gin V, na Ranong V, Poustie S, Jones D; ANZCA Trials Group: Inconsistent survey reporting in anesthesia journals. *Anesth Analg* 2011; 113:591–5

7. ANZCA Clinical Trials Network: Survey research. Available at: www.anzca.edu.au/research/anzca-clinical-trials-network/survey-research. Accessed November 18, 2017
8. Society for Pediatric Anesthesia: Call for surveys. Available at: www.pedsanesthesia.org/call-for-surveys/. Accessed December 22, 2017
9. Squire P: Why the 1936 Literary Digest poll failed. *Public Opin Q* 1988; 52:125–33
10. Dillman D, Smyth J, Christian L: *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 4th edition. New York, Wiley, 2014
11. Kelley K, Clark B, Brown V, Sitzia J: Good practice in the conduct and reporting of survey research. *Int J Qual Health Care* 2003; 15:261–6
12. Rasmussen M, Simonsen CZ, Sørensen LH, Dyrskog S, Rusy DA, Sharma D, Juul N: Anaesthesia practices for endovascular therapy of acute ischaemic stroke: A Nordic survey. *Acta Anaesthesiol Scand* 2017; 61:885–94
13. Passmore C, Dobbie AE, Parchman M, Tysinger J: Guidelines for constructing a survey. *Fam Med* 2002; 34:281–6
14. Klabunde CN, Willis GB, McLeod CC, Dillman DA, Johnson TP, Greene SM, Brown ML: Improving the quality of surveys of physicians and medical groups: A research agenda. *Eval Health Prof* 2012; 35:477–506
15. Boynton PM, Greenhalgh T: Selecting, designing, and developing your questionnaire. *BMJ* 2004; 328:1312–5
16. Peterson R: *Constructing Effective Questionnaires*. Thousand Oaks, California, Sage Publications, 2000
17. Groves R, Fowler F, Couper M, Lepkowski J, Singer E, Rourangeau R: *Survey Methodology*. New York, Wiley & Sons, 2009
18. Fink A: *How to Ask Survey Questions*. Thousand Oaks, California, Sage Publications, 2002
19. Writing@CSU: Colorado State University: Survey research. Available at: www.writing.colostate.edu/guides/guide.cfm?guideid=68. Accessed January 8, 2018
20. Fink A: *How to Design Survey Studies*. Thousand Oaks, California, Sage Publications, 2002
21. O'Cathain A, Thomas KJ: "Any other comments?" Open questions on questionnaires: A bane or a bonus to research? *BMC Med Res Methodol* 2004; 4:25
22. Chudoba B: How much time are respondents willing to spend on your survey? Available at: www.surveymonkey.com/curiosity/survey-completion-times. Accessed December 18, 2017
23. Likert R: A technique for the measurement of attitudes. *Arch Psychol* 1932; 140:1–55
24. Schug SA, Palmer GM, Scott DA, Halliwell R, Trinca J, APM:SE Working Group of the Australian and New Zealand College of Anaesthetists and Faculty of Pain Medicine: *Acute Pain Management: Scientific Evidence*, 4th edition, ANZCA and FPM, Melbourne, 2015
25. Litwin M: *How to Measure Survey Reliability and Validity*. Thousand Oaks, California, Sage Publications, 2002
26. Sullivan GM: A primer on the validity of assessment instruments. *J Grad Med Educ* 2011; 3:119–20
27. Cook DA, Beckman TJ: Current concepts in validity and reliability for psychometric instruments: Theory and application. *Am J Med* 2006; 119:166.e7–16
28. National Statistical Service: Sample size calculator. Available at: www.nss.gov.au/nss/home.nsf/pages/Sample+size+calculator. Accessed January 8, 2018
29. Bartlett J, Kotrlik J, Higgins C: Organizational research: Determining appropriate sample size in survey research information. *Technol Learn Perform J* 2001; 19:43–50
30. Fink A: *How to Sample in Surveys*. Thousand Oaks, California, Sage Publications, 2002
31. Creative Research Systems: Sample size calculator. Available at: www.surveysystem.com/sscalc.htm. Accessed January 5, 2018
32. Toledo P, Duce L, Adams J, Ross VH, Thompson KM, Wong CA: Diversity in the American Society of Anesthesiologists leadership. *Anesth Analg* 2017; 124:1611–6
33. Keon-Cohen Z, Myles PS, Story DA: A survey of Australian and New Zealand anaesthetists' attitudes towards resuscitation orders in the perioperative setting. *Anaesth Intensive Care* 2017; 45:396–402
34. Braun AR, Leslie K, Merry AF, Story D: What are we telling our patients? A survey of risk disclosure for anaesthesia in Australia and New Zealand. *Anaesth Intensive Care* 2010; 38:935–8
35. Goulding KR, Peyton PJ, Story DA, Parker A, Leslie K: Survey of research capability at Australian and New Zealand College of Anaesthetists accredited training sites. *Anaesth Intensive Care* 2017; 45:196–201
36. Gibbs NM: Anaesthesia, not for resuscitation orders, and shared decision-making. *Anaesth Intensive Care* 2017; 45:289–90
37. Roy A, Chattopadhyay S: Stealth marketing as a strategy. *Bus Horiz* 2010; 53:69–79
38. McLeod CC, Klabunde CN, Willis GB, Stark D: Health care provider surveys in the United States, 2000–2010: A review. *Eval Health Prof* 2013; 36:106–26
39. Cook TM, Kelly FE: A national survey of videolaryngoscopy in the United Kingdom. *Br J Anaesth* 2017; 118:593–600
40. McClelland L, Holland J, Lomas JP, Redfern N, Plunkett E: A national survey of the effects of fatigue on trainees in anaesthesia in the UK. *Anaesthesia* 2017; 72:1069–77
41. Babbie E: *The Practice of Social Research*. Belmont, California, Wadsworth, 2004
42. Tait AR, Reynolds PI, Gutstein HB: Factors that influence an anesthesiologist's decision to cancel elective surgery for the child with an upper respiratory tract infection. *J Clin Anesth* 1995; 7:491–9
43. Edwards P, Roberts I, Clarke M, DiGiuseppi C, Pratap S, Wentz R, Kwan I: Increasing response rates to postal questionnaires: Systematic review. *BMJ* 2002; 324:1183
44. Nulty D: The adequacy of response rates to online and paper surveys: What can be done? *Assess Eval Higher Educ* 2008; 33:301–14
45. Smeeth L, Fletcher AE: Improving the response rates to questionnaires. *BMJ* 2002; 324:1168–9
46. Singer E, Cong Y: The use and effects of incentives in surveys. *Ann Am Acad Pol Soc Sci* 2013; 645:112–41
47. Dommeyer C, Baum P, Hanna R, Chapman K: Gathering faculty teaching evaluations by in-class and online surveys: Their effect on response rates and evaluations. *Assess Eval Higher Educ* 2004; 29:611–23
48. Pett M, Lackey N, Sullivan J: *Making Sense of Factor Analysis*. Thousand Oaks, California, Sage Publications, 2003
49. Comfrey A, Lee H: *A First Course in Factor Analysis*. Hillsdale, New Jersey, Erlbaum, 1992
50. Ryan G, Bernard H: Techniques to identify themes. *Field Methods* 2003; 15:85–109
51. Creswell J: *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 4th edition. Los Angeles, Sage Publications, 2014