

Bringing Scientific Rigor to Survey Design in Health Care Research

Sonja I. Ziniel, PhD, MA,^a Corrie E. McDaniel, DO,^b Jimmy Beck, MD, MEd^b

The use of surveys in health care research has become increasingly popular. Often seen as an inexpensive, quick, and simple way to address a research question, surveys are especially attractive to researchers in training or during early career stages when research funding is limited and the time to complete a research project is limited. Yet, when used as research instruments, surveys should be developed with the same scientific rigor as other measurement instruments used to collect data in health care research.

Traditional research methodologies are known for rigorous protocols, including exact specifications for reproducibility and often the presence of control groups to allow for comparisons between cohorts. Concepts such as reliability and validity are not only relevant for measurement instruments of “hard” data, such as thermometers or mass spectrometers, but also for measurement instruments of “soft” data, such as surveys. Researchers using surveys need to ensure that each question will result in a valid and reliable measure of what the question was intended to measure. Demonstrating the validity and reliability of survey questions requires great effort, and although a thorough psychometric validation is not always warranted, every survey should undergo a methodologically reproducible process through development, selection of potential respondents, testing, and administration so that survey results are reproducible and generalizable.

Despite numerous reference books,¹⁻⁴ guidelines, and best practices on survey design,⁵⁻⁹ surveys are often constructed, administered, and submitted for publication without previous consultation of these references and their recommendations. In this article, we highlight key concepts for survey development, focusing on the critical elements of designing and implementing a survey to yield meaningful and publishable results.

KEY CONCEPTS OF SURVEY DEVELOPMENT To Survey or Not to Survey? That Is the Question

Understanding when to use a survey is the first step in solid survey design. In considering the use of a survey, one ought to ask if the research question can more appropriately be answered by another method.

Surveys are particularly useful when trying to measure concepts that are abstract ideas, such as people's beliefs, attitudes, or opinions. A well-designed survey can capture and quantify measures that otherwise are challenging to observe. For example, this may include concepts such as practitioner beliefs surrounding end-of-life care for medically complex patients or providers' levels of burnout. Appropriately applying this principle, Srivastava et al¹⁰ distributed a survey to pediatric department chairs to clarify the definition of a hospitalist and to assess the practice patterns and views on training needs for pediatric hospitalists in academic centers.

www.hospitalpediatrics.org

DOI: <https://doi.org/10.1542/hpeds.2019-0101>

Copyright © 2019 by the American Academy of Pediatrics

^aSection of Pediatric Hospital Medicine, Department of Pediatrics, Children's Hospital Colorado and School of Medicine, University of Colorado, Aurora, Colorado; and ^bSeattle Children's Hospital, University of Washington, Seattle, Washington

Address correspondence to Sonja I. Ziniel, PhD, MA, Department of Pediatrics, Pediatric Hospital Medicine, University of Colorado School of Medicine, 13123 East 16th Ave, 302, Aurora, CO 80045. E-mail: sonja.ziniel@childrenscolorado.org

HOSPITAL PEDIATRICS (ISSN Numbers: Print, 2154-1663; Online, 2154-1671).

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: No external funding.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

Drs Ziniel, McDaniel, and Beck drafted and reviewed the initial manuscript and approved the final manuscript as submitted.

Often, however, surveys are used to measure constructs such as clinical outcomes, tasks of high cognitive load, and observable behaviors. These data usually are more appropriately obtained through other research methods. For instance, although one could design a question asking people how many times per care encounter that they wash their hands, a stronger approach would be to observe and quantify the actual behavior. Similarly, some inquiries are too challenging to attempt to answer via a survey question. Inquiries such as, "How many hours did you use the Internet last month?" become unanswerable questions, leading to guesses rather than accurate responses.

When considering if a research question is appropriate to answer via a survey, it is important to remember that surveys involve memory recall and to consider how respondents answer survey questions.¹¹ Events are stored in human memory in an autobiographical way, meaning that they are usually, linked temporally to personal experiences, such as family- or job-related events, rather than to objective calendar dates. For example, when asked how often someone has been hospitalized in the past 12 months, respondents typically have to reconstruct their life events over the past year. Furthermore, memories of these events are altered each time they are retrieved, with specific details replaced by more generic information as time passes. The more details a respondent has to retrieve from memory when asked through a survey question, the more cognitively demanding it is to determine an answer. This often makes it more likely that the response given will be based on an estimate or a guess.

Identifying a Sampling Frame and Gauging the Generalizability of the Data

Choosing the correct sampling frame is critical in establishing results with generalizability to a target population.² Although it is usually impossible to survey an entire population, it is also burdensome because of time and money constraints to gather data from every individual of a target population. As such, researchers often

choose a smaller group of individuals intended to represent the population being studied. To adequately represent a larger population, the sampling frame must reflect key characteristics of the target population. A poorly chosen sampling frame will introduce significant bias into the resulting data.

For example, within pediatric hospital medicine, the American Academy of Pediatrics Section on Hospital Medicine Listserv functions as a robust cohort of pediatric hospitalists representing the single largest listing of pediatric hospitalists, currently with over 3800 subscribers (KT Powell, MD, PhD, FAAP, personal communication, 2019). Although the listserv at first glance appears to be a prime sampling frame for capturing pediatric hospitalists, its use is riddled with complications that make it less than the ideal sampling frame. First, the listserv does not collect demographic data on subscribers, making it impossible to determine if survey respondents are truly representative of pediatric hospitalists in the United States. Percentages of pediatric hospital physicians, midlevel providers, subspecialty providers, and trainees among the listserv subscribers are unknown. With variability among practice settings,^{12,13} small samples taken from respondents on the listserv are unlikely to represent a generalizable population. In addition, because the listserv is a dynamic mailing list, subscribers may have more than 1 registered e-mail address, and the listserv may include e-mail addresses that are rarely checked.¹⁴ This makes calculating an accurate response rate difficult.

Another common approach to creating a sampling frame is to collaborate with colleagues, with each serving as a site lead at their given institution. This is preferable to using the listserv because it allows for the identification of unique eligible respondents and the calculation of a final denominator for a response rate. A major limitation, however, is the generalizability of a convenience sample of a few institutions to a larger population in regard to the characteristics measured in the survey.

Ultimately, choosing a sampling frame is a balance of feasibility with generalizability. When designing a survey around provider perspectives on high-flow nasal cannula use, for example, surveying pediatric hospitalists from every hospitalist program across the country would provide a known number and capture differences in region, site practices, local customs, and resource variability, but it would also be expensive, time-consuming, and impractical. Instead, as illustrated by Leyenaar et al¹⁵ in using a stratified sample of hospitals within the United States to characterize the direct admission practices across diverse hospitals, a stratified multisite sample specifically targeting factors that may influence these perspectives (such as resource variability, location within the country, hospital type, and size of hospitalist group) may provide a sampling frame that balances feasibility with a generalizable sample.

Next, understanding response rates and nonresponse bias is critical for the interpretation of survey data.¹⁶ Accurate classification of sample members as respondents, nonrespondents, and ineligible, according to the guidelines of the American Association for Public Opinion Research, allows for correct calculation of response rates and, ultimately, comparability across studies.¹⁷ Although low response rates are often interpreted as indicators of surveys with nonresponse bias, neither low nor high response rates are equivalent to nonresponse bias.²

Nonresponse bias occurs if respondents differ from nonrespondents with regard to a specific survey question with low response rates, increasing the likelihood of nonresponse bias occurring in a statistic of interest, such as a mean, that is based on that survey question. For example, when designing a Web-based survey to determine the prevalence of nocturnists within hospital medicine, if the title of the survey is, "A Survey of Pediatric Hospital Medicine Division Chiefs Regarding Nocturnists," the title itself may predispose division chiefs without nocturnists to disregard the survey. As such, resulting

data collected would disproportionately reflect programs with nocturnists, thus overestimating the prevalence, an example of nonresponse bias.

The assessment of the existence of nonresponse bias is usually difficult because the characteristics of nonrespondents are often unknown when compared with survey respondents. To mitigate this, a well-defined sampling frame with known characteristics allows for statistical comparisons of respondents and nonrespondents after data collection and some approximate assessment of nonresponse bias.¹⁶ An example for the evaluation of nonresponse bias within the statistical analyses can be found in the study described by Joyce et al¹⁸ through the creation of sample weighting based on known demographic factors and subsequently accounted for within regression modeling. A study with a well-defined sampling frame with described characteristics may still have generalizable results even with a low response rate if researchers can demonstrate that (1) the survey respondents adequately reflect the sampling frame, and (2) the sampling frame reflects the target population, with regard to characteristics that are likely to influence the survey outcomes. Additional efforts should still be made to increase response rates as much as possible, because increasing the overall size of the final data set improves the power of statistical analyses.

Lastly, Web-based surveys (using survey software such as REDCap,¹⁹ Survey Monkey,²⁰ Qualtrics)²¹ have gained significant popularity over the past few decades because of their ease of use, availability, and streamlined data collection. Careful planning around design and administration is required to yield adequate Web survey response rates, and details related to survey administration should be determined before the survey is distributed. Potential options to maximize respondent participation when using a Web-based survey include personalizing e-mail contact to the respondent, using multiple reminders, tailoring the invitation and reminders to emphasize the importance of

the survey to each participant, and providing incentives, as appropriate.¹

Survey Design Principles

Once the decision to use survey methodology to answer a research question has been made and the sample frame identified, several straightforward survey design principles can guide development while keeping the end goal in mind: the research question to be answered and the constructs to be measured. Gehlbach et al⁵ summarize a rigorous and solid survey design process developed since the 1940s by survey researchers into a 7-step approach: conducting a literature review, conducting interviews and/or focus groups, synthesizing the results from the literature review and interviews and/or focus groups, developing questions, establishing content validity through expert interviews, conducting cognitive interviews, and conducting pilot testing.^{3,6}

The initial steps help ground the development of a survey. First, starting with a comprehensive literature review ensures that the identified survey topic is relevant within a given field (step 1). By nature of curiosity and through the formation of any research question, researchers often become attached to ideas. Anchoring the research question within the available literature helps to hone a research question. A comprehensive literature review may also uncover already published survey instruments and provide insight into the quality and applicability of developed tools. Publication of a survey does not inherently make the survey a good tool or a tool with valid results. Surveys are often idiosyncratic, in that they produce valid results when used in the same target population as the one the tool was originally developed for. For example, the MCAT is accepted as producing valid results for predicting premedical student success within medical school. However, if one asked graduating fellows in nephrology to take the MCAT, the results would not be predictive of future career success. As such, careful and thoughtful evaluation of a tool before use in a new population is critical. However, if a tool

does exist with established validity evidence for the same sample population, using a previously developed tool or scale is preferable to developing one de novo and will strengthen the survey results.

Next, preliminary discussions or even interviews, if warranted, with the target population will help to ensure that the research question is a topic with relevance and provides a local context in which to appropriately situate the survey questions (step 2). Involving topic experts can provide insight as the research team compiles and synthesizes the literature along with results from interviews or focus groups with the target population to begin formulating a preliminary instrument (step 3).

Step 4 launches into writing survey questions. It is at this stage that early involvement of an expert with survey design expertise and/or psychometric experience can help direct question development, avoiding common pitfalls and planning survey design so that the data obtained reflect the aims of the study. Although a comprehensive review of writing good survey questions is beyond the scope of this article, in Table 1 we present a few common mistakes to avoid when writing survey questions, including double-barreled questions, negatively worded items, using statements instead of questions, response scales that do not match the question, asking socially desirable or anxiety-provoking questions, and choosing answer categories that are not mutually exclusive.

Lastly, when writing survey questions, it is often tempting to include free text responses with the intention of qualitatively analyzing respondents' answers. Given the limited space provided in most surveys, respondents rarely provide rich responses, reducing the quantity and quality of substantial contributions.²² As such, free text responses rarely meet the standard for excellent qualitative research.²³ Although including these responses to perform supplementary analysis to the primary survey research is acceptable, unless rigorous qualitative methods are applied in the design and distribution of the survey, one ought to avoid describing the survey as "mixed methods."²⁴ These supplementary

TABLE 1 Key Mistakes to Avoid When Constructing Survey Questions

Mistakes to Avoid	Definition	Example	Potential Solution
Double-barreled questions	These are questions that ask about 2 areas, issues, or groups but only allow for one general answer.	How often do you talk to your nurses and administrative staff when you have a problem?	Choose one: How often do you talk to your nurses when you have a problem?
Negatively worded items	These are questions that include a negative or even a double negative. Some people confirm negation with a “yes” and some do so with a “no.”	In an average week, have you ever not been able to start rounds on time?	Reframe as a positive question: In an average week, how many times do you start rounds on time?
Using statements instead of questions	Asking a statement requires the respondents to do multiple cognitive tasks instead of just one. The respondent must understand (1) what the statement means, (2) if it applies to them, and (3) which response captures how much the statement applies to them.	I am confident I can do well in this course: not at all, a little bit true, somewhat true, mostly true, completely true.	Reframe as a question rather than a statement: How confident are you in succeeding in this course? Not at all confident, a little confident, somewhat confident, mostly confident, completely confident.
Response scales do not match the question	This may take many forms in regard to discordance, such as number of response options, polarity of the scale, mutual exclusiveness of response options, or missing options.	Would you recommend this hospital to someone? Not likely at all, a little likely, somewhat likely, very likely.	Match the descriptors in your scale with the stem of your item: How likely are you to recommend this hospital to someone? Not likely at all, a little likely, somewhat likely, very likely.
Asking socially desirable or anxiety-provoking questions	Respondents have difficulty responding truthfully if they are afraid of being judged.	Do you think women and children should be given the first available flu shots?	Consider removing or rephrasing so that the stem sets up a nonbiased approach: Who should be given the first available flu shots (pick one)? (1) Infants (<1 y of age), (2) children (1–17 y), (3) adults (18–64 y), or (4) elderly (65 y and older).
Not using mutually exclusive answer categories	These are questions with overlapping response options, so that 2 categories may be simultaneously applicable.	How often have you worked >40 h in the past 30 d? (1) Never, (2) 1–5 times, (3) 5–10 times, (4) 10–15 times, (5) > 15 times.	How often have you worked >40 h in the past 30 d? (1) Never, (2) 1–5 times, (3) 6–10 times, (4) 11–15 times, (5) > 15 times.

Adapted with permission from Wolters Kluwer Health, Inc: Artino A, Gelbach H, and Durning S. AM last page: avoiding 5 common pitfalls of survey design. *Acad Med.* 2011;86(10):1327. <http://www.academicmedicine.org>.

analyses, however, involve coding of all open-ended responses, and it is usually time intensive to yield meaningful analytical results in comparison with using closed-ended responses.

Testing the Survey

Commonly, steps 1 through 4 of the Gehlbach et al⁵ framework are routinely attempted, albeit rarely specifically delineated within articles. Stopping after step 4 with item development and distributing the survey at this point, however, hinders the establishment of a high-quality survey instrument and limits the conclusions that can be drawn from collected data.

After the initial survey instrument is constructed, researchers should reengage content topic experts and survey design experts to establish content validity (step 5). Content validity is the degree to which the items of the survey instruments reflect the construct being measured.²⁵ Items should not only pass the “sniff test,” but the research team should ask the experts to evaluate 4 main areas: individual question clarity, topic relevance, language level, and missing content.²⁶

Once content validity has been established, the next step in survey design is the testing of questions through cognitive interviews, also called cognitive pretesting (step 6).²⁷ A common form of cognitive interviews involves in-person interviews in which the interviewer asks the respondent to “think aloud” when answering the survey question, allowing the survey designer to identify problematic questions before the survey is conducted.²⁸ Although the respondents might be able to verbalize problems, the interviewer can also ask detailed questions after the respondents answer each question, eliciting more information on how they arrived at the answer (eg, how they would ask the question in their own words or how they would define a certain term). In general, cognitive interviews elucidate issues in question comprehension and sensitivity, memory retrieval, and the interest and motivation of the respondent.

Cognitive interview participants should be recruited from the same population that will

TABLE 2 Additional Resources for Different Aspects of Survey Design

If You Want to Learn More About...	Resources
A wide range of topics related to survey design and administration	Ref 2 Ref 1 Ref 3 Ref 29
Writing survey questions	Ref 6 Ref 8 Ref 30 Ref 4
Concepts of validity and reliability	Ref 31 Ref 32 Ref 33
Conducting cognitive interviews	Ref 27 Ref 28
Nonresponse rates and nonresponse bias	Ref 16
Using open-ended questions for qualitative data	Ref 23

receive the final survey, varying the background of the participants with regard to those characteristics that most likely will influence understanding or interpretation of the questions. Obvious problems, like missing response categories, should be corrected immediately before starting another interview, whereas major wording changes should be discussed with the research team and only corrected if more than 1 participant encountered this problem. Any major changes to the survey should be tested through cognitive interviews again.

Similar to qualitative interviews, the number of cognitive interviews that should be conducted is variable. Cognitive interviews can be concluded when no additional major problems are found in the survey. Often, this occurs after a handful of interviews if the survey was carefully constructed. This process should be clearly documented in the article.

The last step before survey administration is piloting the survey (step 7). The goal of piloting is to collect evidence that the data produced by the survey is consistent with the expectations of the research team. Working with a statistician, the researchers should administer the survey to a small

subset of the survey sample. The resulting data can then be analyzed to assess if individual items show the desired variance and if the actual survey completion time matches what was expected.

Taking the Next Steps in Designing a Survey

There are many additional aspects to consider when designing a survey that are not touched on in this article, including survey administration, data management, and use of conceptual models or frameworks. Available resources for additional reading into survey design when planning a research study are listed in Table 2.

CONCLUSIONS

Through choosing appropriate research questions to be answered by a survey, thoughtfully selecting a sampling frame, and using a structured approach to instrument design, including cognitive pretesting and piloting, researchers can design a methodologically rigorous survey with generalizable and reproducible results. Although this process may not be quick or easy, it sets a survey up to collect meaningful data, ultimately increasing potential impact and publishability.

Acknowledgments

We thank Lilliam Ambroggio, Kevin Messacar, and Jeannie Zuk for valuable comments on earlier drafts of this article.

REFERENCES

- Dillman DA, Smyth JD, Christian LM. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4th ed. Hoboken, NJ: John Wiley; 2014
- Groves RM, Fowler FJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R. *Survey Methodology*. 2nd ed. New York, NY: John Wiley & Sons; 2009
- Fowler FJ. *Survey Research Methods*. 5th ed. London, United Kingdom: Sage Publications; 2014
- Fowler FJ. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications; 2006
- Gehlbach H, Artino AR Jr, Durning S. AM last page: survey development guidance

for medical education researchers. *Acad Med*. 2010;85(5):925

- Artino AR Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE Guide No. 87. *Med Teach*. 2014;36(6):463–474
- Artino AR Jr, Durning SJ, Sklar DP. Guidelines for reporting survey-based research submitted to academic medicine. *Acad Med*. 2018;93(3):337–340
- Gehlbach H, Artino AR Jr. The survey checklist (manifesto). *Acad Med*. 2018; 93(3):360–366
- Rickards G, Magee C, Artino AR Jr. You can't fix by analysis what you've spoiled by design: developing survey instruments and collecting validity evidence. *J Grad Med Educ*. 2012;4(4): 407–410
- Srivastava R, Landrigan C, Gidwani P, Harary OH, Muret-Wagstaff S, Homer CJ. Pediatric hospitalists in Canada and the United States: a survey of pediatric academic department chairs. *Ambul Pediatr*. 2001;1(6):338–339
- Tourangeau R, Rips LJ, Rasinski K. *The Psychology of Survey Response*. Cambridge, United Kingdom: Cambridge University Press; 2012
- Snow C. Pediatric Hospitalists in Community Hospitals. In: Proceedings from the Pediatric Hospital Medicine Conference; July 20-23, 2017; Nashville, TN
- Fromme HB, Chen CO, Fine BR, Gosdin C, Shaughnessy EE. Pediatric hospitalist workload and sustainability in university-based programs: results from a national interview-based survey. *J Hosp Med*. 2018;13(10):702–705
- Saleh A, Bista K. Examining factors impacting online survey response rates in educational research: perceptions of graduate students. *J Multidiscip Eval*. 2017;13(29):63–74
- Leyenaar JK, O'Brien ER, Malkani N, Lagu T, Lindenauer PK. Direct admission to hospital: a mixed methods survey of pediatric practices, benefits, and

- challenges. *Acad Pediatr*. 2016;16(2): 175–182
16. Phillips AW, Reddy S, Durning SJ. Improving response rates and evaluating nonresponse bias in surveys: AMEE Guide No. 102. *Med Teach*. 2016; 38(3):217–228
 17. American Association for Public Opinion Research. Response rates - an overview. Available at: <https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx>. Accessed March 25, 2019
 18. Joyce CM, Scott A, Jeon SH, et al. The “medicine in Australia: balancing employment and life (MABEL)” longitudinal survey—protocol and baseline data for a prospective cohort study of Australian doctors’ workforce participation. *BMC Health Serv Res*. 2010;10:50
 19. REDCap. REDCap. Available at: <https://projectredcap.org/>. Accessed April 14, 2019
 20. SurveyMonkey. SurveyMonkey. Available at: <https://www.surveymonkey.com/>. Accessed April 14, 2019
 21. Qualtrics. Qualtrics. Available at: <https://www.qualtrics.com/>. Accessed April 14, 2019
 22. Tracy S. Qualitative quality: eight “big-tent” criteria for excellent qualitative research. *Qual Inq*. 2010;16(10):837–851
 23. LaDonna KA, Taylor T, Lingard L. Why open-ended survey questions are unlikely to support rigorous qualitative insights. *Acad Med*. 2018;93(3):347–349
 24. Creswell JW, Klassen AC, Plano Clark VL, Clegg Smith KC; Office of Behavioral and Social Sciences Research. *Best Practices for Mixed Methods Research in the Health Sciences*. 2nd ed. Bethesda, MD: National Institutes of Health; 2011
 25. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: Amer Educational Research Association; 2014
 26. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales a Practical Guide to Their Development and Use*. 5th ed. Oxford, United Kingdom: Oxford University Press; 2015
 27. Peterson CH, Peterson NA, Powell KG. Cognitive interviewing for item development: validity evidence based on content and response processes. *Meas Eval Couns Dev*. 2017;50(4):217–223
 28. Willis GB. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA: Sage Publications; 2010
 29. Fink A. *How to Conduct Surveys: A Step-by-Step Guide*. 6th ed. Thousand Oaks, CA: Sage Publications; 2017
 30. Boynton P, Greenhalgh T. Selecting, designing, and developing your questionnaire. *BMJ*. 2004;328(7451): 1312–1315
 31. Slattery EL, Voelker CCJ, Nussenbaum B, Rich JT, Paniello RC, Neely JG. A practical guide to surveys and questionnaires. *Otolaryngol Head Neck Surg*. 2011; 144(6):831–837
 32. Beckman TJ, Cook DA. Developing scholarly projects in education: a primer for medical teachers. *Med Teach*. 2007; 29(2–3):210–218
 33. Raykov T, Marcoulides GA. *Introduction to Psychometric Theory*. New York, NY: Routledge/Taylor and Francis Group; 2011