

REVIEW ARTICLE



Fundamentals of designing high-quality surveys: revisiting neonatal perinatal medicine survey applications

Kadriye O. Lewis¹✉, Howard W. Kilbride¹, Carl Bose² and David J. Burchfield³

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

Surveys in neonatal perinatal medicine are practical instruments for gathering information about medical practices, and outcomes related to the care of newborns and infants. This includes research for identifying needs, assessing requirements, analyzing the effects of change, creating policies, and developing curriculum initiatives. Surveys also provide useful data for enhancing the provision of healthcare services, assessing medical specialties, and evaluating training programs. However, creating a high-quality survey can be difficult for many practitioners, particularly when deciding how to formulate the right questions, whether to utilize various types of questions and how best to arrange or format the survey tool for effective responses. Problems with design principles have been evident in many surveys submitted for dissemination to the members of the Section of Neonatal Perinatal Medicine (SoNPM). To prevent potential measurement errors and increase the quality of surveys, it is crucial to follow a systematic approach in developing surveys by adhering to the principles of effective survey design. This review article provides a brief summary of survey use within the SoNPM, and offers guidance for creating high-quality surveys, including identifying important factors to consider in survey development and characteristics of well-written and effective questions. We briefly note techniques that optimize survey design for distribution through digital media.

Journal of Perinatology (2024) 44:777–784; <https://doi.org/10.1038/s41372-023-01801-6>

INTRODUCTION

Surveys have been used for centuries to gather information and have evolved significantly over time, from early paper-based questionnaires to modern online surveys [1, 2]. They are now a crucial instrument in healthcare and medical education for doing scientific research that includes assessing needs, evaluating demands, developing curricula, analyzing the effects of change, obtaining stakeholder input on various topics, and determining policy content [3, 4]. Surveys also provide useful data for improving health services delivery, serving as a medium to mitigate potential conflicts in patient care, assessing the quality of medical practices, and evaluating training programs. With these multiple applications of surveys in our professional practices, we should be cognizant of the elements and crucial steps taken by a well-designed survey tool in order to collect reliable, replicable, and accurate data. Such effort is the only way to ensure measuring progress and guiding solutions.

Depending on the goal, target audience, and information sought, the formats and contents of surveys can range from non-systematic questionnaires to rigorous, validated surveys. Creating a high-quality survey that produces rich, nuanced, and usable data requires significant effort [5]. However, many practitioners who use surveys to inform practice are challenged by designing a high-quality survey, specifically how to write questions, when to use various question types, and how to organize or format a survey tool. Applying effective design principles in the survey development process is often either poorly understood or overlooked.

Developing questions that accurately assess the opinions, experiences, and behaviors of respondents is a critical aspect of survey methodology [2, 3]. Poorly designed surveys confuse respondents and yield inaccurate information, unreliable feedback, and low response rates. Careful planning in survey development can lead to accurate measures and improves data quality.

As a service to its membership, the Section on Neonatal Perinatal Medicine (SoNPM) of the American Academy of Pediatrics distributes surveys through its email directory, collecting data and obtaining feedback on a wide range of topics related to neonatal care, healthcare practices, and outcomes. Because of the increasing demand for this service, in 2017, the SoNPM created a template for survey submission requests and initiated a review by a designated committee. The committee has identified poor survey quality as a limiting factor for accepting surveys for distribution to the membership. The purposes of this review article are: 1) to describe the current state of survey use within the SoNPM, 2) to examine question quality and structure to determine the best design principles for the development of surveys within neonatal perinatal medicine practices, 3) to bring attention to critical design issues in survey development tailored to specific objectives for medical education programs, including question types, response formats, layout, sampling, pretesting and pilot tests, and 4) to briefly describe the benefits of online surveys and techniques that optimize survey design in the digital medium.

¹Children's Mercy Kansas City, Department of Pediatrics, University of Missouri-Kansas City School of Medicine, Kansas City, MO, USA. ²University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ³University of Florida, Gainesville, FL, USA. ✉email: kolewis@cmh.edu

Received: 21 June 2023 Revised: 22 September 2023 Accepted: 6 October 2023
Published online: 20 October 2023

Table 1. Examples of survey topics submitted to SoNPM.

Topics	Content
Variation in Clinical Practices	<ul style="list-style-type: none"> • Use of sedation for intubation • Enteral feeding on therapeutic hypothermia • Use of Continuous Positive Airway Pressure (CPAP) • Delayed cord clamping in preterm infants • Steroids for Bronchopulmonary Dysplasia (BPD) • Definition and treatment for “culture-negative sepsis” • COVID-19 case management
Variation in NICU Policies	<ul style="list-style-type: none"> • Use of Quality Improvement bundles for Intraventricular hemorrhage (IVH) • Target O2 levels for various gestational ages • Hypoglycemia policies in well nursery and Neonatal Intensive Care Unit (NICU) • Infection control policies
Attitudes and Practices Regarding Controversial Issues	<ul style="list-style-type: none"> • Breastfeeding with maternal drug use, particularly marijuana • Palliative Care initiatives in the NICU • Resuscitation of extremely preterm infants in the “grey zone” • Moral distress in neonatology with regard to T18 infants or others with serious genetic differences
Educational Programs	<ul style="list-style-type: none"> • Trauma-informed care • Global health

THE CURRENT STATE OF SURVEYS WITHIN THE SONPM

We retrospectively reviewed survey applications sent to SoNPM for distribution to the members from 2016 through 2021. Approximately, 8–10 surveys were submitted annually, primarily from academic institutions. The applicant was often a trainee (16/40, 40%), usually a neonatal fellow, and occasionally a pediatric resident or medical student, supported by a neonatal faculty member. Surveys most often assessed variation in clinical practice, but some were designed to assess NICU policies, neonatologists' attitudes on controversial topics, or information collected to assist with the development of an educational program (Table 1).

The SoNPM survey review committee did not have strict survey acceptance criteria. Each committee member reviewed the submission, and through discussion, they collaborated to determine “acceptability of the submission”. In determining acceptability, the committee assessed whether the topic was of importance, whether survey questions appropriately addressed that topic, and whether responses would likely provide useful data. Surveys were rejected if there were currently published data on the topic or if the content overlapped significantly with recently published data on the subject or a recent survey. Among concerns identified with the surveys, the most frequent was that the topic and questions referred to NICU policy, which would be better answered by the NICU director or Nurse Manager, rather than surveying all neonatologists. Problem areas or common concerns about the SoNPM surveys are noted in Table 2.

Although approximately 60% of surveys were approved, nearly all required revision before acceptance for distribution, including addressing issues as noted in Table 2. Some of the submissions were initially rejected but subsequently resubmitted, often months or a year or more later, for another review after substantial rewriting. There was no formalized follow-up assessment to determine how successful surveys had been. Unsolicited reports from some surveyors indicated that response rates were about 10 to 15%. One author (HWK) searched PubMed for “neonatal survey research”, published since 2016 in *Journal of Perinatology*, and found 12 references [6–17]. The median reported response rate for these was 18%.

Being able to submit and provide responses to surveys and having an opportunity to provide input to other neonatologists are important benefits of membership in the SoNPM. However, there has been concern about “survey fatigue,” particularly when members have been asked to complete surveys that are poorly written or deemed by some to be of little value. Most clinicians do not have education in survey planning or design, which are necessary skills to develop an effective research tool. Improved

survey quality among submissions and perhaps a more focused review process by the SoNPM, should result in better response rates and more meaningful outcomes, with improved publication rates for those using this mode of research (e.g., survey methodology).

PLANNING A SURVEY FOR CLINICAL RESEARCH USE

Survey vs. Questionnaire

Many individuals use the terms “survey” and “questionnaire” interchangeably. These two are closely related because they involve collecting data or gathering specific information, but there are nuanced differences or distinctions between these two terms [18–20]. A survey refers to scientific investigation with the goal of collecting unbiased data for rigorous analysis for research purposes. Surveys often use one or more questionnaires, that focus on one aspect of the objective of the survey, as a data collection tool. A good example would be patient satisfaction surveys since they involve administering various questionnaires (e.g., about hospital experience, and doctor-patient communication) to a group of patients to gather insights into their overall satisfaction with healthcare services. Another good example would be the National Health and Nutrition Examination Survey conducted by the Centers for Disease Control and Prevention in the United States [21]. It aims to assess the health and nutritional status of the U.S. population. The survey provides data on a wide range of health-related topics, including chronic diseases, nutrition, physical activity, environmental exposures, and more. Thus, a survey is a method of measuring and analyzing relevant opinions and behaviors from a larger population to arrive at relevant research results. On the other hand, a questionnaire is a list of questions to explore multiple constructs simultaneously and can be either descriptive (providing factual facts) or explanatory (drawing conclusions about constructions or concepts). For example, a health-related quality of life questionnaire can ask a series of questions designed to assess an individual's physical, mental, and social well-being. It may ask about aspects such as pain, mobility, emotional well-being, and social functioning. Another example would be a medical student burnout questionnaire, which aims to measure the levels of burnout experienced by medical students. It might include questions about emotional exhaustion, depersonalization, and personal accomplishment. Thus, a questionnaire is a set of written questions that usually inquire about a single subject in detail [18]. Questionnaires can be informal, for example, to gather data for upcoming research, or formal, with specific objectives and

Table 2. Concerns identified with submitted surveys to SoNPM.

Area of Concerns	Examples
Question Content	<ul style="list-style-type: none"> • Questions that requested facts were not available without patient/chart review—answers may have reflected “guesses” or perceptions, not reality. • Questions did not address the stated problem that was the basis for the survey—data were not usable.
Vague Questions or Question-Wording	<ul style="list-style-type: none"> • Survey items/questions were ambiguous or too general; many additional variables were needed to provide an appropriate answer. • Questions were too general to differentiate the specific population (e.g., asks about trainees, not specifying whether medical students, residents, fellows, and each group may require different responses.). • In some instances, language structure and word choices were problematic or complicated. • Language parallelism was not consistent; some sentence structures used passive voice and some of them were written with an active voice in the matrix questions.
Question Types and Response Options or Scales	<ul style="list-style-type: none"> • The types of questions were not well aligned with the purpose of the questions. • The purpose of some questions was not clear and/or it was out of scope. • Some questions had forced choices without “Not Applicable” or “Other: Please specify” options or providing a comment box. • Some suggestions were not applicable to all respondents, and there were double-barreled questions that contained two questions in one question or assumed two things were linked together although they were separate entities. • The scale format was not consistent with the type of the questions.
Question Sequencing	<ul style="list-style-type: none"> • Demographic questions appeared as the first section of the survey, which may potentially distort the main purpose of the survey. At the same time, this may have impacted participant responses to survey questions regarding the topic being studied. • Questions were not listed under the correct themes which may have swayed participants’ attention from one theme to other themes.
Survey Format	<ul style="list-style-type: none"> • There were a variety of survey formats presented to participants, including different arrangements of questions, response options, and question types, without best practice considerations. • Some of the surveys were too long (e.g., If the surveys take > 5–10 min, they are likely not going to get a reasonable response rate.). • The format-related flaws included: <ul style="list-style-type: none"> • Asking for detailed information in multiple clinical scenarios that may cause survey fatigue. • Requesting mandatory answers or forced choices, which may produce unreliable “made up” data and/or participants may not complete the survey. • Some questions were related to personal or institutional information such as asking demographics or specific institution questions that may have been perceived to be sensitive to participants.
Survey Topics	<ul style="list-style-type: none"> • Asking management questions were more appropriate for those in leadership positions rather than individual neonatologists—such a survey should be directed to NICU director (a separate listserv). • Questions related to NICU management (e.g., naming convention for newborns, visiting policies for COVID, unit approach to cord clamping for preterm) were not appropriate for many neonatologists since they may not know or be able to answer questions about unit policy, resulting in inaccurate or uninterpretable data. • Even if accurate information, the data may have overrepresented large NICUs since more of those neonatologists would likely be responding to the survey compared to smaller units. • Questions address areas in which neonatologists were not supposed to make the final decision (management of NEC totalis or timing of shunt placement with hydrocephalus—questions better directed to surgery.) • There were infrequent clinical scenarios that could not produce generalizable data. Only a few neonatologists may have experience with the issue and those that do, may only have knowledge of a single case from years before and responses may not be current. Some of the topics were not clinically important. • Our topic evaluation questions were: 1) What will the investigators do with data regarding clinicians’ perceptions of a problem? 2) Can findings help develop a research study of the topic? 3) Will it provide data to support an educational or advocacy program; if not, why do this? These questions may not have been sufficient to evaluate a survey quality.

outcomes [22]. In essence, both surveys and questionnaires gather information from a target audience. A detailed Table that identifies and contrasts the characteristics of surveys compared to questionnaires is presented in the appendix (see Appendix 1).

In the context of clinical research, most applications submitted for distribution to the SoNPM email directory are considered in the survey category. A questionnaire may or may not be administered as part of a survey. Understanding the differences between these two data collection instruments will help us better determine which is more beneficial, depending on research goals and objectives. If we are looking to collect and interpret data to help us understand opinions, attitudes, and behaviors, then a survey is a valid option for us. However, if we want a flexible instrument to

collect responses from a target audience, it is better to use a questionnaire.

Question quality as an element of survey design

Surveys require a set of predetermined questions. The quality of surveys depends on how well the questions are constructed with the use of appropriate scales. From a survey instrument design point of view, survey questions can be categorized both by their content and format. Robinson and Leonard [5] described four categories of survey content: 1) attributes, 2) behaviors, 3) abilities, and 4) thoughts. Attributes refer to the demographics and traits of a respondent; behavioral questions are self-reports of past or current behaviors with frequency and duration quantification;

abilities refer to training, knowledge, and skills that can be reported or assessed. Thought questions have to do with attitudes and level of satisfaction. More details regarding these four categories of survey content, including sample questions for each, are available in the appendix (See Appendix 2).

When planning and designing surveys, along with the content types we need to consider various design features that may have a substantial and differential impact on both data quality and the cost of a survey. There are several steps to follow keeping in mind the survey content, structure, and question types. The following seven steps provide a solid foundation in item design and best practices for the creation of robust and high-impact surveys:

Step 1: Determine the survey goals and objectives. The goal of a survey, which is related to the study purpose, should be a broad statement that explains what the investigator wants to achieve. The process begins by writing an overall goal. Then the concept is broken down into multiple measurable objectives. Objectives are essential to provide a framework for asking the right questions.

Step 2: Define the population and sample. To make a survey reliable and representative for a study purpose, it is crucial to determine the population of interest and an ideal sample size among members of the population, assuming that the entire population will not be surveyed. A population refers to an entire group of people to be studied, whereas a sample is only a subset of the population being studied [5]. When a sample is not reflective of the population, this may threaten external validity, specifically population validity, or if it is too small, we cannot gain reliable insights. Using a correct survey sample size reduces sampling bias, which is important to provide generalizability of findings. The key questions in this step require answering questions such as “What is the population that I would like to survey?” “Who will participate in the survey? And how many people should my sample consist of?”

Two types of sampling may be considered, probability and non-probability sampling. With probability sampling, every single individual within the target group (which can be either random or representative) has an equal chance of being selected for the survey. Some common types of probability sampling techniques include simple random sampling, systematic sampling, stratified sampling, cluster sampling, multistage sampling, and systematic random sampling [23–26]. Probability sampling techniques have the advantage of ensuring that the sample is representative of the entire population. As a result, validity is ensured for the statistical conclusions. Non-probability sampling (also known as non-random sampling) involves the process of selecting a sample group on the basis of judgment or the convenience of accessing data. Not everyone has an equal chance to participate in a study. Therefore, this type of sampling may not accurately represent the population. The most common non-probability types include convenience sampling, quota sampling, purposive sampling, snowball sampling, and voluntary sampling [27, 28]. Non-probability sampling’s biggest advantage is cost-effectiveness and time-effectiveness compared to probability sampling but risks a lower level of generalization of study findings, including difficulties in estimating sampling variability and identifying possible bias [29]. Surveys that are submitted to the SONPM use convenience sampling by accessing the large database of practicing neonatologists and accepting information from those who chose to take the survey.

Step 3: Determine “exactly” how you plan to use the data. Survey data provide valuable insight to make decisions from multiple perspectives of stakeholders. During the survey planning stage, it is important to know how data will be used and/or analyzed (qualitative or quantitative). Knowing exactly which data will

answer your bigger questions will be a roadmap for how to analyze the data and distribute the results through appropriate channels and reports. If parts of the survey data will not be used or will not benefit the research purpose, it is better to eliminate those segments of the survey.

Step 4: Conduct a literature review before creating a draft of the survey questions. An extensive literature search should be performed before writing a new survey to determine if a similar survey has already been validated or if any measures of the same constructs or related constructs already exist [4, 30]. Using a previously validated survey can be more cost-effective than designing a new survey. However, any modification to the survey may reduce its validity. Despite this disadvantage, it is often preferable to use an existing validated survey compared to designing a new one [4]. Even if no existing surveys are available, a literature review ensures the definition of the correct constructs that align with related research in the field and provides guidance for the development of better content of self-designed survey items and scales that may be adapted to the current study [4].

Step 5: Start developing survey items. A survey item is a question or statement included in a survey that elicits specific responses from the participants. The quality of the items determines the quality of the data [31]. Developing high-quality surveys should include the following vital elements:

Survey introduction: An introduction to the survey should include answers to the following questions:

- What is the purpose of the survey?
- Who will use the information?
- How will the data be used?
- Will the responses be anonymous?
- How will the confidentiality and anonymity of respondents be assured?
- Approximately how long will it take to complete the survey?

Structuring items by choosing the right question types: The most important part of survey development is the creation of items that accurately measure respondents’ opinions, attitudes, and/or behaviors. Survey statements or questions should be organized and aligned with study objectives. If there are varying types of questions, they should be grouped into sections with instructions for how the respondent should answer them at the beginning of each section.

Survey items should be written in plain language and mean the same thing to every respondent. The wording should also express what is being measured and guide respondents in how to respond [32]. Careful wording will increase readability and clarity while decreasing bias. Ideally, each item should focus on one construct, and should not exceed 20 words in length, be easy to understand, nonjudgmental, and unbiased [17]. Certain language or terminologies should be avoided, such as ambiguous questions (various people may interpret some words or phrases in different ways), conditional sentences (“... if...”), negative or double negative phrasing (“Don’t you think.....the hospital should not.....”), absolute terms (e.g., “always,” “none” or “never”), and overly general expressions (“All children.....”). It is crucial to include the definitions of abbreviations, technical terminology, or complex terminology. Survey items that are poorly worded, confusing, or contradictory are likely to generate unreliable data and suffer from low response rates, thus undermining the effort associated with designing and administering the survey. A table is available in the appendix listing common problems with survey questions, and possible solutions (See Appendix 3).

Table 3. Types of Scales in Likert-type questions.

Types of Scales	Description	Advantages	Disadvantages
Unipolar Example: • Not at all agree • Slightly agree • Somewhat agree • Extremely agree	This rating scale prompts a respondent to evaluate the degree for a single quality or attribute. • Starts from zero • No natural midpoint • Goes to an extreme	• Measures the intensity in one direction of an attribute (on a more granular level) • Gives the respondent a fair chance to choose a rating between 0-100%	Difference levels are so minute that makes it hard for participants to select an answer, making the data useless.
Bipolar Example: • Strongly disagree • Somewhat disagree • Neither agree nor disagree • Somewhat agree • Strongly agree	This scale is characterized by a continuum between two opposite endpoints. • Starts at extreme negative • Has a natural midpoint • Goes to the opposite extreme positive	• Measures intensity from both directions of an attribute with a neutral option in the middle • Gives the respondent a fair chance to rate their positive and negative sentiment	Neutral option allows users to lazily answer, making the data useless.

Types of Questions: Varying statement or question types helps respondents actively engage with the survey and may improve response rate, while also tapping into a range of thinking processes with opportunity to obtain richer information. The most common question types can be categorized into the following three themes:

1. *Close-ended questions:* These types of questions ask respondents to choose from a distinct set of pre-defined responses such as "Yes/No," one-word answers, and multiple-choice questions. They are great for collecting quantitative data. Closed-ended questions come in a multitude of forms but are defined by their need to have explicit options from which a respondent may select. A table in the appendix provides details of types of close-ended questions with the pros and cons of each (see Appendix 4).
2. *Open-ended questions:* These questions are also called "unstructured questions" that offer no response options to provide an answer. Respondents write their own answers from a single word or number to one or more paragraphs [22]. Open-ended questions facilitate collection of qualitative data that is rich, nuanced, and detailed, reflecting differences in opinions, feelings, attitudes, and experiences, which may not be captured by the quantitative data. However, responses to open-ended questions require more time and effort on the part of respondents and statistical analysis is more complex and time-consuming than for close-ended questions. These questions are rarely used in surveys submitted to the SoNPM. A table in the appendix details pros and cons in using open-ended questions (see Appendix 5).
3. *Visually delineated questions:* These types of questions use visual images, pictures, graphics, charts, illustrations, and formulas represented as answer options. Question responses require reasoning over visual elements of the image and general knowledge to infer the correct answer [33]. Therefore, visual question answering is significantly more complex, and useful responses are dependent on well-designed images. [2, 34, 35]. A table in the appendix outlines the types of visually delineated questions with pros and cons for each (see Appendix 6).

While having a variety of items as part of a survey may be beneficial for data collection, it also will require greater attention from respondents, which may increase the cognitive load for those being surveyed. Consideration of this potential additional stress should be taken when designing the survey, such as reducing the number of items when there are multiple types involved.

Response Scales: Regardless of the question type, choosing a survey response scale that is suitable for a survey's question prompt is an important component of survey design and if done poorly may lead to poor data quality [2]. To perform statistical data analysis, there are four fundamental measurement scales [36].

1. Nominal Scale (determine quality or categorize data into mutually exclusive categories or groups)
2. Ordinal Scale (determine greater or less or measure variables in a natural order, such as rating or ranking)
3. Interval Scale (measure variables with equal intervals between values to see differences)
4. Ratio Scale (compare and compute ratios, percentages, and averages)

In medicine, many surveys use Likert Scales or Likert Type of scales [37, 38], which is the derivation of these four fundamental levels of variable measurement. Those surveys are commonly used to measure attitudes, attributes, opinions, beliefs, perceptions, and actions of the respondents. As seen in Table 3, Likert Scales can be presented in two ways: using a bipolar or unipolar scale [38, 39].

Likert scales usually range from a 3-10 point scale, 4, 5, and 7 being the most common points. The 3-5 point scales provide fewer response options that make it easier to analyze the data, but harder to differentiate the stronger sentiments versus the neutral ones. As for the 6-10 point scales, they produce more segmented responses to analyze the stronger sentiments versus neutral ones, but it is harder for respondents to distinguish between minute levels of differentiation for the answer options. Scale selection should be related to a study's purpose, the type of analysis to be performed, and the planned interpretation of the study results.

Sequencing of items: Survey items should be organized based on content within domains to facilitate the thought process of respondents since their responses can be impacted by previous items [17]. Therefore, statement or question flow should be arranged in a logical/psychological order so that one leads easily and naturally to the next one or next section smoothly. Each set of items can be grouped together under a heading with descriptive language that tells responders what to expect from the following set of items. Surveys often start with non-threatening items but placing the most important ones near the beginning may better capture respondents' attention.

In sequencing item order, one of the most controversial issues is where to place demographic questions. Placing demographic questions at the beginning of the survey may cause potential respondents to stop their participation because they may feel the questions are too personal. This outcome may affect the response rate. Therefore, we prefer to include demographic questions at the

end of a survey unless there is a specific reason to place them early. Informing respondents that their responses will remain anonymous and/or strictly confidential is one way to encourage them to answer demographic questions truthfully/honestly. Moreover, when in doubt, it is better to provide a “Prefer Not To Answer” option.

Summary of common errors related to item development: As items are being developed, common sources of errors [40, 41] are listed below:

- *Comprehension problem*: The respondent does not understand the question or statement.
- *Multiple interpretations*: The respondent identifies alternate, but equally reasonable, interpretations of the item.
- *Difficulty with recall*: The respondent finds it difficult to recall with the degree of precision necessary.
- *Wrong type of response*: Response to the items will not provide the type of data that the survey designer intended.
- *Social pressure*: The respondent feels pressure to respond in a certain way (not honestly).
- *Format of the question*: The format of the statement or question is tedious to the respondent—questions with multiple response options for example.
- *No response*: For some items, not everyone who is being surveyed may be able to provide a response. If such an item is included, there should be an option for “no response” such as “Not Applicable” or “I do not know” or “I prefer not to answer”.

Step 6. Consult expert review. During the design of a survey, input should be sought from those who will be using the survey data and those who will be conducting the survey [42]. An expert review has two primary objectives [43]. The first is to identify problems with the survey instrument so that these problems can be rectified prior to use. The second is to identify item groups (questions) that are more or less likely to exhibit measurement errors. Several expert reviewers or colleagues should be used to review survey questions individually to capture different perspectives on the same survey. Without content review and question evaluations, a survey may omit key areas or not reflect recent studies or advancements in the area. Thus, expert reviews are valuable in identifying question problems and question alignment related to the purpose of the survey that may affect the quality of the survey data.

Less than 50% of surveys submitted to the SoNPM indicated expert review in the methodology description. Although the survey review committee to some extent provided this step, extensive feedback and revisions were often needed, delaying the evaluation and decision regarding acceptance.

Step 7: Pretesting and Piloting the Survey. While pretesting and pilot testing are frequently used interchangeably, there is a significant distinction between the two. Pretesting focuses on validating the instrument and its measurement to ensure that questions meet their intended purpose. On the other hand, pilot testing is considered to be a “dress rehearsal” by administering the entire survey process to detect overt problems that are not captured during the pretesting [17, 44]. Both pretesting and pilot testing are needed to modify the survey instrument.

Pretesting a survey: The quality of survey data depends on how well respondents understand the question. By pre-testing, we initiate the process of reviewing and revising questions. The main purpose is to assess whether respondents interpret questions in a consistent manner, as intended by the researcher, and to judge the appropriateness of each question included [24]. It is important to note that if there is a small number of respondents in a pretest,

this may reduce the possibility of making a valid conclusion and may increase the margin for error. There are multiple methods for pretesting. Sample questions that may be used for pretesting are available in a table in the appendix (see Appendix 7).

Piloting a survey: A pilot survey is administered to a small sample of the participant population (30–100 participants suggested). This is the process of assessing everything under survey conditions not only to meet the goals and objectives of the study but to analyze various aspects of the whole survey, such as survey structure, adequacy of length and completion time, and question confusion.

Finally, it is crucial to include some questions in a pilot survey designed to get specific and relevant feedback from respondents. Sample questions that can be used in a pilot survey are available in a table in the appendix (see Appendix 8).

RELIABILITY AND VALIDITY

Two important factors in developing and testing a survey instrument are reliability and validity. Consideration of these factors during survey development can help to ensure the quality of a survey instrument and the data to be collected for analysis and interpretation.

Reliability. When an instrument is reliable, it will provide the same results if the measurements are repeated under identical circumstances —i.e., consistent measures in comparable situations [45]. There are four general classes of reliability assessments: 1) test–retest reliability, 2) parallel forms reliability, 3) interrater reliability, 4) internal consistency, and each one estimates reliability in a different way [46–51].

Validity. When evaluating a survey tool, validity refers to how accurately it measures what it is supposed to measure — i.e., answers correspond to what they are intended to measure [45]. There are four main types of validity: 1) face validity, 2) content validity, 3) construct validity, and 4) criterion validity, which have been formulated as a part of accepted research methodology [51]. In addition, the following questions are crucial when testing for validity [52]:

- Are the questions measuring what they are supposed to measure?
- Do the survey questions fully represent the content?
- Are the questions appropriate for the sample/population?
- Are the survey questions sufficient to gather all the information needed to meet the purpose and goals of the study?

In brief, survey questions that are not reliable do not produce consistent responses at different points in time and across respondents. This inconsistency introduces measurement error. Survey questions that are not valid do not capture the true meaning of the construct and therefore do not measure what was intended.

SURVEYS IN THE DIGITAL AGE

Application of online surveys and questionnaires has become a popular data collection method in all spectrums of medical disciplines and research. Online surveys are sent to a sample of respondents and those audiences complete them over a Web platform generally by filling out a form. The main benefit of online surveys is to increase productivity by saving time. Data are instantly available and can easily be transferred into specialized statistical software or spreadsheets when more detailed analysis is needed. Online surveys also may allow for use of visual sliders, in which a respondent moves a slider within a preselected range to indicate a value in response to a statement or question [53]. This is a quick, interactive way for the respondent to provide their

perception or opinions following a statement or question [54, 55]. While visual sliders primarily collect quantitative data due to their nature as numerical scales, some survey platforms might incorporate a combination of quantitative slider questions along with open-ended text fields to gather both types of data in a single survey. This allows for a more comprehensive understanding of respondents' opinions and perspectives.

In recent years, we have witnessed the rapid spread of surveys delivered through social media, smartphones, and numerous other digital devices, enabling us to collect/process data about human behaviors on visual scales or multi-channel formats. Conducting surveys through mobile devices/smartphones and social media is convenient and inexpensive, but there are concerns about possible incomplete or inattentive responses and security risks [56–58]. The visual design of the survey is important when considering the various screen sizes and devices that may be used for access. Changes in visual design properties may have an unintentional effect on responses, skewing survey results. Understanding visual design principles, and graphic interfaces for online and mobile devices is an additional prerequisite for the development of digital surveys [59].

DISCUSSION AND CONCLUSION

In this review paper, we described the current state of survey use within the SoNPM, and highlighted challenges that have been experienced and opportunities for improvement. We noted critical steps that must be considered in the development of a quality survey and the importance of including details of these considerations within the methodology description of the resulting manuscript. A literature review is an important planning step to understand gaps in knowledge and to help define the goals and objectives of the study. It is important to clarify the population to be assessed and sampling methods to avoid selection bias and potential loss of generalization of results. Question design is the most critical step in survey development which will determine survey quality. We have emphasized considerations regarding question types, format and sequencing of questions, and the use of appropriate response scales to best quantify results. The reliability and validity of the responses depend on the consideration of these important factors in survey development. Expert review, as well as pre-test and piloting of the survey, are discussed as important factors to improve survey quality and assure that results are meaningful and meet study objectives. These details of the survey design process are important to clarify within the methods section of the research manuscript.

In conclusion, surveys in medical practice can help identify areas where improvements can be made in care practices and outcomes. They can also provide valuable information for researchers studying the effects of different interventions on patient outcomes. When using the survey as a tool for research, investigators should use the seven steps of survey design to create high-quality surveys which will yield reliable and valid data. These steps help ensure that the survey questions are relevant, unbiased, and understandable and that the survey method is appropriate for the research purpose and objectives. By following the practical guidance for survey development provided in this review, investigators can both increase survey response rates and improve the quality and accuracy of the outcome data for SoNPM studies, or for other medical and public health research. Finally, the appendices provided with this review offer detailed information which may be used as a resource for those interested in high value survey design.

REFERENCES

- Rosenbaum S. Precursors of the Journal of the Royal Statistical Society. *J R Stat Soc: Ser D (Statistician)*. 2001;50:457–66.
- Dillman DA, Christian LM, Smyth JD. Internet, phone, mail, and mixed-mode surveys: The Tailored Design Method. Hoboken, NJ: Wiley; 2014.
- Halbesleben JR, Whitman MV. Evaluating survey quality in health services research: A decision framework for assessing nonresponse bias. *Health Serv Res*. 2013;48:913–30.
- Nikiforova T, Carter A, Yecies E, Spagnoletti CL. Best Practices for Survey Use in Medical Education: How to Design, Refine, and Administer High-Quality Surveys. *South Med J*. 2021;114:567–71. <https://doi.org/10.14423/smj.0000000000001292>.
- Robinson SB, Leonard KF. Designing quality survey questions. Los Angeles, Calif: Sage; 2019.
- Baker SF, Smith BJ, Donohue PK, Gleason CA. Skin Care Management Practices for Premature Infants. *J Perinatol*. 1999;19:426–31.
- Feltman DM, Du H, Leuthner SR. Survey of neonatologists' attitudes toward limiting life-sustaining treatments in the neonatal intensive care unit. *J Perinatol*. 2012;32:886–92.
- Wei D, Osman C, Dukhovny D, Romley J, Hall M, Chin S, et al. Cost consciousness among physicians in the neonatal intensive care unit. *J Perinatol*. 2016;36:1014–20.
- Placencia FX, Ahmadi Y, McCullough LB. Three decades after Baby Doe: how neonatologists and bioethicists conceptualize the Best Interests Standard. *J Perinatol*. 2016;36:906–11.
- Kane SK, Diane EL. The Amount of Supervision Trainees Receive during Neonatal Resuscitation Is Variable and Often Dependent on Subjective Criteria. *J Perinatol*. 2018;38:1081–1086.
- Kurepa D, Perveen S, Lipener Y, Kakkilaya V. The use of less invasive surfactant administration (LISA) in the United States with review of the literature. *J Perinatol*. 2019;39:426–32.
- Hwang JS, Friedlander S, Rehan VK, Zangwill KM. Diagnosis of congenital/perinatal infections by Neonatologists: A national survey. *J Perinatol*. 2019;39:690–6.
- Manja V, Guyatt G, Lakshminrusimha S, Jack S, Kirpalani H, Zupancic JAF, et al. Factors influencing decision making in neonatology: inhaled nitric oxide in pre-term infants. *J Perinatol*. 2019;39:86–94.
- Horowitz E, Feldman HA, Savich R. Neonatologist salary: Factors, equity and gender. *J Perinatol*. 2019;39:359–65.
- Horowitz E, Randis TM, Samnaliev M, Savich R. Equity for women in medicine—neonatologists identify issues. *J Perinatol*. 2020;41:435–44.
- Horowitz E, Samnaliev M, Savich R. Seeking racial and ethnic equity among neonatologists. *J Perinatol*. 2021;41:422–34.
- Callahan KP, Flibotte J, Skraban C, Wild KT, Joffe S, Munson D, et al. How neonatologists use genetic testing: findings from a national survey. *J Perinatol*. 2022;42:260–261.
- Sheppard V. 8.2 Understanding the Difference between a Survey and a Questionnaire. *Research Methods for the Social Sciences An Introduction*. 6 Apr 2020. <https://pressbooks.bccampus.ca/jibcresearchmethods/chapter/8-2-understanding-the-difference-between-a-survey-and-a-questionnaire/>.
- Survey vs. Questionnaire: What Are the Differences?. Indeed, Indeed Editorial Team. 24 June 2022. <https://www.indeed.com/career-advice/career-development/survey-vs-questionnaire>.
- Surbhi S. Difference Between Survey and Questionnaire. *Key Differences*. 2016. Accessed 19 Sep 2023. <https://keydifferences.com/difference-between-survey-and-questionnaire.html#ComparisonChart>.
- NHANES - National Health and Nutrition Examination Survey Homepage. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention; 2023 Accessed 19 Sep 2023. <https://www.cdc.gov/nchs/nhanes/index.htm>.
- Burns KEA, Duffett M, Kho ME, Meade MO, Adhikari NKJ, Sinuff T, et al. A guide for the design and conduct of self-administered surveys of clinicians. *Can Med Assoc J*. 2008;179:245–52. <https://doi.org/10.1503/cmaj.080372>.
- Daniel J. Sampling essentials: Practical guidelines for making sampling choices. Los Angeles: Sage; 2012.
- Ruel EE, Wagner WE, Gillespie BJ. The practice of survey research: Theory and applications. Thousand Oaks, CA: Sage Publications, Inc.; 2016.
- Arnab R. Survey sampling theory and applications. London: Academic Press; 2017.
- Latpate R, Kshirsagar J, Gupta VK, Chandra G. Advanced sampling methods. Singapore: Springer; 2021.
- Setia MS. Methodology series module 5: Sampling strategies. *Indian J Dermatol*. 2016;61:505.
- Dhivyadeepa E. Sampling techniques in educational research. Lulu.com; 2015.
- Schillewaert N, Langerak F, Duharnel T. Non-probability sampling for WWW surveys: A comparison of methods. *Mark Res Soc J*. 1998;40:1–13.
- Artino AR, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for Educational Research: AMEE guide no. 87. *Med Teach*. 2014;36:463–74.
- Passmore C, Dobbie AE, Parchman M, Tysinger J. Guidelines for constructing a survey. *Fam Med*. 2002 34:281–6.

32. Smyth JD, Olson K. The effects of mismatches between survey question stems and response options on data quality and responses. *J Surv Stat Methodol*. 2018;7:34–65.
33. Banchhor M, Singh P. A survey on visual question answering. 2021 2nd Global Conference for Advancement in Technology (GCAT), 2021. <https://doi.org/10.1109/gcat52182.2021.9587797>.
34. Chien Y-T, Chang C-Y. Exploring the impact of animation-based questionnaire on conducting a web-based educational survey and its association with vividness of respondents' visual images. *Br J Educ Technol*. 2012;43. <https://doi.org/10.1111/j.1467-8535.2012.01287.x>.
35. Wu Q, Teney D, Wang P, Shen C, Dick A, van den Hengel A. Visual question answering: A survey of methods and datasets. *Comput Vis Image Underst*. 2017;163:21–40.
36. Stevens SS. On the theory of scales of measurement. *Science* 1946;103:677–80.
37. Sullivan GM, Artino AR Jr. Analyzing and interpreting data from Likert-type scales. *J Graduate Med Educ*. 2013;5:541–542.
38. Jebb AT, Ng V, Tay L. A review of key Likert Scale Development Advances: 1995–2019. *Front Psychol*. 2021;12. <https://doi.org/10.3389/fpsyg.2021.637547>.
39. Höhne JK, Krebs D, Kühnel S-M. Measurement properties of completely and end labeled unipolar and bipolar scales in Likert-type questions on income (in) equality. *Soc Sci Res*. 2021;97:102544.
40. Biemer PP. Measurement error in sample surveys. In: Pfefferman D, Rao CR, editors. *Handbook of Statistics 29A*, North-Holland, Amsterdam, 2009: 281–315. [https://doi.org/10.1016/S0169-7161\(08\)00012-6](https://doi.org/10.1016/S0169-7161(08)00012-6).
41. Groves RM. *Survey Errors and Survey Costs*. New York: John Wiley & Sons; 2004.
42. Glasgow PA. *Fundamentals of Survey Research Methodology*: MITRE. Virginia, USA: Washington C3 Center; 2005.
43. Olson K. An Examination of Questionnaire Evaluation by Expert Reviewers. *Field Methods*. 2010;22(4):295–318. <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1138&context=sociologyfacpub> Available from
44. Rothgeb JM. Pilot test. In: Lavrakas PJ, editor. *Encyclopedia of survey research methods* (pp. 583–585). Thousand Oaks, CA: Sage; 2008.
45. Fowler FJ. *Survey Research Methods* (4th ed.). SAGE Publications, Inc., 2009. <https://doi.org/10.4135/9781452230184>.
46. Carmines EG, Zeller RA. *Reliability and validity assessment*. Newbury Park: SAGE Publications, Inc.; 2013.
47. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A practical guide to their development and use*. Oxford, United Kingdom: Oxford University Press; 2015.
48. Jhangiani RS, Chiang IA, Cuttler C, Leighton DC. *Research Methods in Psychology*, 4th edition. B.C.; Kwantlen Polytechnic University Surrey; 2019.
49. Krippendorff K. *Content analysis an introduction to its methodology*. 4th Edition. Los Angeles: SAGE; 2019.
50. Portney LG, Watkins MP. *Foundations of clinical research: Applications to practice*. 2nd Edition. Upper Saddle River: Prentice Hall Health; 2000.
51. DeVellis RF, Thorpe CT. *Scale development: Theory and applications*. SAGE Publications, Inc: Thousand Oaks, CA, 2022.
52. Bourke J, Kirby A, Doran J. *Survey & Questionnaire Design: Collecting Primary data to answer research questions*. Ireland: NuBooks, an imprint of Oak Tree Press; 2016.
53. Couper MP, Tourangeau R, Conrad FG, Singer E. Evaluating the effectiveness of visual analog scales. *Soc Sci Computer Rev*. 2006;24(2):227–45. <https://doi.org/10.1177/0894439305281503>.
54. Roster C, Lucianetti L, Albaum G. Exploring Slider vs. Categorical Response Formats in Web-Based Surveys. *Canada J Res Pract J Res Pract*. 11;2015. Accessed 19 Sep 2023. <http://jrp.icaap.org/index.php/jrp/article/download/509/433>.
55. Tourangeau R, Maitland A, Rivero G, Sun H, Williams D, Yan T. Web surveys by smartphone and tablets. *Public Opin Q*. 2017;81:896–929. <https://doi.org/10.1093/poq/nfx035>.
56. La Polla M, Martinelli F, Sgandurra D. A Survey on Security for Mobile Devices. *IEEE Commun Surv Tutor*. 2013;15:446–71. <https://doi.org/10.1109/surv.2012.013012.00028>.
57. Khan J, Abbas H, Al-Muhtadi J. Survey on Mobile User's Data Privacy Threats and Defense Mechanisms. *Procedia Comput Sci*. 2015;56:376–83. <https://doi.org/10.1016/j.procs.2015.07.223>.
58. Mishra S, Thakur A. A Survey on Mobile Security Issues. *SSRN Electronic Journal*, Elsevier BV, 2019. <https://doi.org/10.2139/ssrn.3372207>.
59. Sandesara M, Bodkhe U, Tanwar S, Alshehri MD, Sharma R, Neagu BC, et al. Design and Experience of Mobile Applications: A Pilot Survey. *Mathematics*. 2022;10(Jul):2380 <https://doi.org/10.3390/math10142380>.

AUTHOR CONTRIBUTIONS

KOL conceptualized and designed the perspective manuscript, wrote the initial draft, and subsequently revised the manuscript. HWK conceptualized the manuscript, added input regarding the use of surveys within the SoNPM, reviewed and revised the manuscript. CB and DJB reviewed and provided critical revision of the manuscript. All authors reviewed and approved the final version of the manuscript for submission and agreed to be accountable for all aspects of the work.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41372-023-01801-6>.

Correspondence and requests for materials should be addressed to Kadriye O. Lewis.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.