

A guide for the design and conduct of self-administered surveys of clinicians

Karen E.A. Burns MD MSc, Mark Duffett BScPharm, Michelle E. Kho PT MSc, Maureen O. Meade MD MSc, Neill K.J. Adhikari MDCM MSc, Tasnim Sinuff MD PhD, Deborah J. Cook MD MSc, for the ACCADEMY Group

Survey research is an important form of scientific inquiry¹ that merits rigorous design and analysis.² The aim of a survey is to gather reliable and unbiased data from a representative sample of respondents.³ Increasingly, investigators administer questionnaires to clinicians about their knowledge, attitudes and practice^{2,4,5} to generate or refine research questions and to evaluate the impact of clinical research on practice. Questionnaires can be descriptive (reporting factual data) or explanatory (drawing inferences between constructs or concepts) and can explore several constructs at a time. Questionnaires can be informal, conducted as preparatory work for future studies, or formal, with specific objectives and outcomes.

Rigorous questionnaires can be challenging and labour-intensive to develop, test and administer without the help of a systematic approach.⁵ In this article, we outline steps to design, develop, test and administer valid questionnaires with minimal bias and optimal response rates. We focus on self-administered postal and electronic surveys of clinicians that are amenable to quantitative analysis. We highlight differences between postal and electronic administration of surveys and review strategies that enhance response rates and reporting transparency. Although intended to assist in the conduct of rigorous self-administered surveys, our article may also help clinicians in the appraisal of published surveys.

Design

Determining the objective

A clear objective is essential for a well-defined survey. Refining initial research objectives requires specification of the topic, respondents, and primary and secondary research questions to be addressed.

Identifying the sampling frame

It is often impractical for investigators to administer their questionnaire to all potential respondents in their target population, because of the size of the target population or the difficulty in identifying possible respondents.⁴ Consequently, a sample of the target population is often surveyed. The “sampling frame” is the target population from which the sample will be drawn.⁶ The “sampling element” refers to the respondents from whom information is collected and analyzed.⁶ The sampling frame should represent the population of interest. To this end, certain sampling techniques (e.g., surveying con-

ference attendees) may limit generalizability compared with others (e.g., surveying licensed members of a profession). Ultimately, the sampling technique will depend on the survey objectives and resources.

Sample selection can be random (probability design) or deliberate (nonprobability design).⁶ Probability designs include simple random sampling, systematic random sampling, stratified sampling and cluster sampling.

- Simple random sampling: Every individual in the population of interest has an equal chance of being included in the sample. Potential respondents are selected at random using various techniques, such as a lottery process (e.g., drawing numbers from a hat) and random-number generator.⁷
- Systematic random sampling: The investigator randomly selects a starting point on a list and then selects individuals systematically at a prespecified sampling interval (e.g., every 25th individual). In systematic random sampling, both the starting point and the sampling interval are determined by the required sample size.
- Stratified random sampling: Potential respondents are organized into strata, or distinct categories, and randomly sampled using simple or systematic sampling within strata to ensure that specific subgroups of interest are represented. Stratified sampling can be proportionate (sampling the same proportion of cases in each stratum) or disproportionate (sampling fraction varies across strata).⁶
- Cluster sampling: Investigators divide the population into clusters and sample clusters (or individuals within clusters) in a stepwise manner. Clusters should be mutually exclusive and exhaustive and, unlike strata, heterogeneous.

With the exception of cluster sampling, investigators require lists of individuals in the sampling frame, with contact information, to conduct probability sampling. It is important to ensure that each member of the sampling frame can be contacted. Table 1 presents the advantages and disadvantages of different approaches to probability sampling.⁸

From the Interdepartmental Division of Critical Care (Burns, Adhikari, Sinuff), University of Toronto, Toronto, Ont.; Keenan Research Centre and the Li Ka Shing Knowledge Institute (Burns), St. Michael's Hospital, Toronto, Ont.; Pediatric Critical Care (Duffett), McMaster Children's Hospital, Hamilton, Ont.; Department of Clinical Epidemiology and Biostatistics (Duffett, Kho, Meade, Cook), McMaster University, Hamilton, Ont.; Department of Critical Care Medicine and Sunnybrook Research Institute (Adhikari, Sinuff), Sunnybrook Health Sciences Centre, Toronto, Ont. Members of the ACCADEMY (Academy of Critical Care: Development, Evaluation and Methodology) Group are listed at the end of the article.

A nonprobability sampling design is chosen when investigators cannot estimate the chance of a given individual being included in the sample. Such designs enable investigators to study groups that may be challenging to identify. Nonprobability designs include purposive sampling, quota sampling, chunk sampling and snowball sampling.⁶

- Purposive sampling: Individuals are selected because they meet specific criteria (e.g., they are physiotherapists).
- Quota sampling: Investigators target a specific number of respondents with particular qualities (e.g., female physicians between the ages of 40 and 60 who are being promoted).
- Chunk sampling: Individuals are selected based on their availability (e.g., patients in the radiology department’s waiting room).
- Snowball sampling: Investigators identify individuals meeting specific criteria, who in turn identify other potential respondents meeting the same criteria.⁶

The extent to which the results of a questionnaire can be generalized from respondents to a target population depends on the extent to which respondents are similar to nonrespondents. It is rarely possible to know whether respondents differ from nonrespondents in important ways (e.g., demographic characteristics, answers) unless additional data are obtained from nonrespondents. The best safeguard against poor generalizability is a high response rate.

Development

Item generation

The purpose of item generation is to consider all potential items (ideas, concepts) for inclusion in the questionnaire, with the goal of tapping into important domains (categories or themes) suggested by the research question.⁹ Items may be generated through literature reviews, in-depth interviews, focus-group sessions, or a combination of these methods with potential respondents or experts. Item generation continues until no new items emerge, often called

“sampling to redundancy.” The Delphi process, wherein items are nominated and rated by experts until consensus is achieved, can also be used to generate items.^{5,10} Following item generation, investigators should define the constructs (ideas, concepts) that they wish to explore,⁵ group the generated items into domains and begin formulating questions within the domains.

By creating a “table of specifications,” investigators can ensure that sufficient items have been generated to address the research question and can identify superfluous items.² Investigators list research questions on the vertical axis and either the domains of interest or the type of information sought (knowledge, attitudes and practice) on the horizontal axis. Subtopics or concepts can be added within identified domains.¹⁰ This table is revisited as questions are eliminated or altered and to establish validity.¹⁰

Item reduction

In this step, investigators limit the large number of potentially relevant questions within domains to a manageable number without eliminating entire domains or important constructs. The requirement for information must be balanced against the need to minimize respondent burden, since lengthy questionnaires are less likely to be completed.^{11,12} In general, most research questions are addressed with 25 or fewer items⁵ and at least 5 items in each domain.¹¹

Item reduction is an iterative process that can be achieved using one of several methods, some of which require respondent data. Redundant items can be eliminated in interviews or focus-group sessions with content experts or external appraisers. Participants are asked to evaluate the relative merit of included items by ranking (e.g., ordinal scales) or rating (e.g., Likert scales) items or by providing binary responses (e.g., include/exclude). Alternatively, investigators may reduce items using statistical methods that examine the relation between and among items within domains; this method requires data obtained through pilot testing.

Questionnaire formatting

Question stems

The question stem is the statement or question to which a response is sought. Each question should focus on a single construct. Question stems should contain fewer than 20 words and be easy to understand and interpret,^{5,13} nonjudgmental and unbiased.¹³ Investigators should phrase questions in a socially and culturally sensitive manner. They should avoid absolute terms (e.g., “always,” “none” or “never”),¹¹ abbreviations and complex terminology.² Investigators should specify the perspective from which questions should be addressed, particularly for questions about attitudes that may elicit different responses depending on how they are

Table 1: Different approaches to random sample selection (probability sampling)

Sampling design	Advantages	Disadvantages
Simple random	<ul style="list-style-type: none"> • Requires little advance knowledge of population 	<ul style="list-style-type: none"> • May not capture specific groups • May not be efficient
Systematic random	<ul style="list-style-type: none"> • High precision • Easy to analyze data and compute sampling errors 	<ul style="list-style-type: none"> • Ordering of elements in sampling frame may create biases • May not capture specific groups • May not be efficient
Stratified random	<ul style="list-style-type: none"> • Captures specific groups • Disproportionate sampling possible • Highest precision 	<ul style="list-style-type: none"> • Requires advance knowledge of population • More complex to analyze data and compute sampling errors
Cluster	<ul style="list-style-type: none"> • Lower field costs • Enables sampling of groups if individuals not available 	<ul style="list-style-type: none"> • More complex to analyze data and compute sampling errors • Lowest precision

Adapted, with permission, from Aday and Cornelius.⁸

worded.¹⁴ The language used influences the response formats used, which may affect the response rate. Demonstrative questions are often followed by binary responses, whereas question stems requesting respondents to rank items or elicit their opinions should adopt a neutral tone. The wording of the question and the order of response categories can influence the responses obtained.^{3,15} Moreover, the manner in which question stems and responses are synthesized and presented can influence potential respondents' decisions to initiate and complete a questionnaire.³

Response formats

Response formats provide a framework for answering the question posed.⁵ As with question stems, investigators should develop succinct and unbiased response formats, either "open" (free text) or "closed" (structured). Closed response formats include binary (yes/no), nominal, ordinal, and interval and ratio measurements.

- **Nominal responses:** This response option consists of a list of mutually exclusive, but unordered, names or labels (e.g., administrators, physicians, nurses) that typically reflect qualitative differences in the construct being measured.
- **Ordinal responses:** Although ordinal responses (e.g., Likert scales) imply a ranked order, they do not reflect a quantity or magnitude of the variable of interest.¹⁶ Likert scales can be used to elicit respondents' agreement (ranging from strongly disagree to strongly agree) with a statement.
- **Interval and ratio measurements:** These response options depict continuous responses. Both formats demonstrate a constant relation between points. However, only ratio measurements have a true zero and exhibit constant proportionality (proportions of scores reflect the magnitude of the variable of interest).

Collaboration with a biostatistician is helpful during questionnaire development to ensure that data required for analyses are obtained in a usable format.

When deciding on the response options, investigators should consider whether to include indeterminate response options, to avoid "floor and ceiling" effects and to include "other" response options.

- **Indeterminate response options:** Although the inclusion of indecisive response options (e.g., "I don't know," "I have no opinion") may let respondents "off the hook" too easily,¹⁷ they acknowledge uncertainty.¹³ These response options may be suitable when binary responses are sought or when respondent knowledge, as opposed to attitudes or opinions, is being probed.²
- **Floor and ceiling effects:** These effects reflect responses that cluster at the top or bottom of scales.⁵ During item reduction, investigators should consider removing questions that demonstrate floor or ceiling effects, or using another response format to increase the range of responses. Providing more response options may increase data dispersion, and may increase discrimination among responses.⁵ Floor and ceiling effects sometimes remain after response options are modified; in such cases they reflect true respondent views.
- **"Other" response options:** Providing an "other" response

option or requesting "any other comments" allows for unanticipated answers, alters the power balance between investigators and respondents,¹⁸ and may enhance response rates to self-administered questionnaires.³ During questionnaire testing, "other" response options can help to identify new issues or elaborate on closed response formats.¹⁸

Questionnaire composition

Cover letter

The cover letter creates the first impression. The letter should state the objective of the survey and highlight why potential respondents were selected.¹⁹ To enhance credibility, academic investigators should print cover letters on departmental stationery with their signatures. To increase the response rate, investigators should personalize the cover letter to recipients known to them, provide an estimate of the time required to complete the questionnaire and affirm that the recipient's participation is imperative to the success of the survey.²⁰

Questionnaire

Some investigators recommend highlighting the rationale for the survey directly on the questionnaire.¹³ Presenting simple questions or demographic questions first may ease respondents into questionnaire completion. Alternatively, investigators may reserve demographic questions for the end if the questions posed are sensitive. The font style and size should be easy to read (e.g., Arial 10–12 point). The use of bold type, shading and broad lines can help direct respondents' attention and enhance visual appeal. McColl and colleagues³ highlighted the importance of spatial arrangement, colour, brightness and consistency in the visual presentation of questionnaires.

Questionnaires should fit neatly inside the selected envelope along with the cover letter, a return (stamped or metered) envelope and an incentive, if provided. Often longer questionnaires are formatted into booklets made from larger sheets of paper (28 × 36 cm [11 × 14 inches]) printed on both sides that are folded in half and either stapled or sutured along the seam. Investigators planning to send reminders to nonrespondents should code questionnaires before administration. "Opt out" responses identify respondents who do not wish to complete the questionnaire or were incorrectly identified and can limit additional correspondence.²

For Internet-based surveys, questions are presented in a single scrolling page (single-item screen) or on a series of linked pages (multiple-item screens) often with accompanying electronic instructions and links to facilitate questionnaire flow. Although the use of progress indicators can increase questionnaire completion time, multiple-item screens significantly decrease completion time and the number of "uncertain" or "not applicable" responses.²¹ Respondents may be more likely to enter invalid responses in long- versus short-entry boxes, and the use of radio buttons may decrease the likelihood of missing data compared with entry boxes.²¹ [Radio buttons, or option buttons, are graphic interface objects used in electronic surveys that allow users to choose only one option from a predefined set of alternatives.]

Questions should be numbered and organized. Every question stem should include a clear request for either single or multiple responses and indicate the desired notation (e.g., check, circle). Response options should appear on separate lines. Tables can be used to present ordinal responses of several constructs within a single question. The organization of the questionnaire should assist respondents' thought processes and facilitate questionnaire flow.⁵ Questions can be ordered on the basis of content (e.g., broad questions preceding specific ones),^{3,13} permutations in content (scenario-based questions) or structure (questions presented within domains or based on the similarity of response formats when a single domain is being explored).⁵ Operational definitions are helpful before potentially ambiguous questions,⁵ as are clear instructions to skip nonapplicable questions.¹⁷

In a systematic review, Edwards and colleagues²² identified 292 randomized trials and reviewed the influence of 75 strategies on responses to postal questionnaires. They found that specific formatting strategies (e.g., the use of coloured ink, the placement of more interesting questions first, and shorter length) enhanced response rates (Table 2).

Pre-testing

The quality of questionnaire data depends on how well respondents understand the items. Their comprehension may be affected by language skills, education and culture.⁵ Pre-testing initiates the process of reviewing and revising questions. Its purpose is to evaluate whether respondents interpret questions

in a consistent manner, as intended by the investigator,²³ and to judge the appropriateness of each included question. Investigators ask colleagues who are similar to prospective respondents¹⁴ to evaluate each question through interviews (individual or group) or written feedback. They also ask them to determine a course of action, including whether to accept the original question and meaning, to change the question but keep the meaning, to eliminate the question or to write a new question).²⁴

Testing

Pilot testing

During pilot testing, investigators present questions as they will appear in the penultimate draft of the questionnaire to test respondents who are similar to the sampling frame.²⁴ The purpose is to assess the dynamics of the questionnaire in a semi-structured interaction. The respondents are asked to examine the questionnaire with regard to its flow, salience, acceptability and administrative ease,²³ identifying unusual, redundant, irrelevant or poorly worded question stems and responses. They are also asked to record the time required to complete the questionnaire. Pre-testing and pilot testing minimize the chance that respondents will misinterpret questions, fail to recall what is requested or misrepresent their true responses.²³ The information obtained through pre-testing and pilot testing is used to improve the questionnaire.

Following pilot testing, investigators can reduce items further

Table 2: Strategies shown in a meta-analysis to enhance response rates to postal questionnaires²²

Strategy	No. of trials (no. of participants)	Odds ratio (95% CI)	p value for heterogeneity
Questionnaire development			
More interesting v. less interesting questionnaire	2 (2 151)	2.44 (1.99–3.01)	
User friendly questionnaire v. standard	1 (3 540)	1.46 (1.21–1.75)	
Factual questions v. factual and attitudinal	1 (1 280)	1.34 (1.01–1.77)	
More relevant questions first v. other items first	1 (5 817)	1.23 (1.10–1.37)	
Shorter v. longer questionnaire	40 (40 669)	1.86 (1.55–2.24)	< 0.001
Coloured ink v. standard	1 (3 540)	1.39 (1.16–1.67)	
Questionnaire administration			
Precontact v. no precontact	28 (28 793)	1.54 (1.24–1.92)	< 0.001
More personalized v. less personalized	38 (39 210)	1.16 (1.06–2.28)	< 0.001
Stamped return envelope v. business reply/franked	14 (38 259)	1.26 (1.13–1.41)	< 0.001
First class mailing v. other class	1 (7 370)	1.12 (1.02–1.23)	
Recorded delivery v. standard	6 (2 127)	2.21 (1.51–3.25)	< 0.01
Monetary incentive v. none	49 (46 474)	2.02 (1.79–2.27)	< 0.001
Incentive with questionnaire v. on return	10 (13 713)	1.71 (1.29–2.26)	< 0.001
Nonmonetary incentive v. no incentive	45 (44 708)	1.19 (1.11–1.28)	< 0.001
Follow-up v. no follow-up	12 (16 740)	1.44 (1.22–1.70)	< 0.001
Postal follow-up with questionnaire v. without	6 (6 310)	1.41 (1.02–1.94)	< 0.001
University sponsorship/source v. other source	13 (20 428)	1.31 (1.11–1.54)	< 0.001
Explanation for nonparticipation requested v. not requested	1 (1 240)	1.32 (1.05–1.66)	

Note: CI = confidence interval.

through factor analysis by examining mathematical relations among items and seeing how items cluster into specific domains.²⁵ Measures of internal consistency (see “Reliability”) can assess the extent to which candidate items are related to selected items and not to other items within a domain. Correlations between 0.70 and 0.90 are optimal;²⁶ correlations below 0.70 suggest that different concepts are being measured, and those above 0.90 suggest redundant items.²⁶ At least 5 respondents per candidate item (i.e., 100 respondents for a 20-item questionnaire) are required for factor analysis. Factor analysis can highlight items that require revision or removal from a domain.²⁶

Clinical sensibility testing

The goals of clinical sensibility testing are to assess the comprehensiveness, clarity and face validity of the questionnaire. The testing addresses important issues such as whether response formats are simple and easily understood, whether any items are inappropriate or redundant or missing, and how likely the questionnaire is to address the survey objective. During clinical sensibility testing, investigators administer a 1-page assessment sheet to respondents with the aforementioned items presented as questions, with either Likert scale (e.g., very unlikely, unlikely, neutral, likely, very likely) or nominal (e.g., yes/no/don’t know/unclear) response formats. An example of a clinical sensibility testing tool is shown in Appendix 1 (available at www.cmaj.ca/cgi/content/full/179/3/245/DC1). Following pre-testing, pilot testing and clinical sensibility testing, questionnaires may need to be modified to an extent that additional testing is required.

Although some overlap exists among pre-testing, pilot testing and clinical sensibility testing, each is distinct. Pre-testing focuses on the clarity and interpretation of individual questions and ensures that questions meet their intended purpose. Pilot testing focuses on the relevance, flow and arrangement of the questionnaire, in addition to the wording of the questionnaire. Although pilot testing can detect overt problems with the questionnaire, it rarely identifies their origins, which are generally unveiled during pre-testing.²³ Clinical sensibility testing focuses on how well the questionnaire addresses the topic of interest and the survey objective.

Reliability

Ideally, questions discriminate among respondents such that respondents who think similarly about a question choose similar responses, whereas those who think differently choose diverse responses.⁵ Reliability assessment is part of rigorous evaluation of a new questionnaire.²⁷

- Test–retest reliability: With this method, investigators assess whether the same question posed to the same individuals yields consistent results at different times (typically spanning 2–4 weeks).
- Interrater reliability: Investigators assess whether different respondents provide similar responses where expected.
- Internal consistency: Investigators appraise whether different items tapping into the same construct are correlated.⁶ Three tests can be used to assess internal consistency: the corrected item-total correlation (assesses the correlation of an item with the sum of all other items), split-half reliability (assesses correlation between scores derived by splitting a set of questions in half) and the α reliability coefficients (derived by determining key dimensions and assessing items that tap into specific dimensions).

The reliability assessment required depends on the objective of the survey and the type of data collected (Table 3).²⁷

Validity

There are 4 types of validity that can be assessed in questionnaires: face, content, construct and criterion validity.

- Face validity: This is the most subjective aspect of validity testing. Experts and sample participants evaluate whether the questionnaire measures what it intends to measure during clinical sensibility testing.²⁰
- Content validity: This assessment is best performed by experts (in content or instrument development) who evaluate whether questionnaire content accurately assesses all fundamental aspects of the topic.
- Construct validity: This is the most abstract validity assessment. It should be evaluated if specific criteria cannot be identified that adequately define the construct being measured. Expert determination of content validity or factor analysis can substantiate that key constructs underpinning the content are included.

Table 3: Statistical tests used in reliability assessments*

Type of reliability testing	Description	Type of data; correlation coefficients			
		Nominal	Ordinal	Interval	Other
Test–retest	To assess whether the same question posed to the same individuals yields consistent results at different times	Chi-square based	Spearman rho	Pearson <i>r</i>	
Interrater	To assess whether different respondents provide similar results	Kappa	Spearman rho	Pearson <i>r</i>	Intra-class (several raters, continuous data)
Internal consistency	To assess whether different items tapping into the same construct are correlated		Spearman–Brown Cronbach alpha	Corrected item-total <i>r</i>	Kuder–Richardson (dichotomous data)

*Adapted, with permission, from Carmines and Zeller.²⁷

- Criterion validity: In this assessment, responses to survey items are compared to a “gold standard.”

Investigators may engage in one or more assessments of instrument validity depending on current and anticipated uses of the questionnaire. At a minimum, they should assess the questionnaire’s face validity.

Administration

Advanced notices, for example in professional newsletters or a premailed letter, should announce the impending administration of a questionnaire.¹⁹ Self-administered questionnaires can be distributed by mail or electronically via email or the Internet. The administration technique chosen depends on the amount and type of information desired, the target sample size, investigator time, financial constraints and whether test properties were established.² In a survey of orthopedic surgeons, Leece and colleagues²⁸ compared Internet ($n = 221$) and postal ($n = 221$) administration techniques using alternating assignment. Nonrespondents to the mailed questionnaire were sent up to 3 additional copies of the questionnaire; nonrespondents to the Internet questionnaire received up to 3 electronic requests to complete the questionnaire and, if necessary, were mailed a copy of the questionnaire. Compared with the postal arm, Internet recipients had a lower response rate (45% [99/221] v. 58% [128/221]; absolute difference 13%, 95% confidence interval 4%–22%; $p < 0.01$). Other studies^{29,30} also showed a lower response rate with electronic than with postal administration techniques, which suggests that a trade-off may exist with electronic administration between cost (less investigator time required for questionnaire administration) and response rate. A systematic review of Internet-based surveys of health professionals identified 17 publications of sampling from e-directories and Web postings or electronic discussion groups; 12 reported variable response rates ranging from 9% to 94%.³¹

Internet-based surveys pose unique technical challenges and methodologic concerns.³¹ Before choosing this administration technique, investigators must have the support of skilled information technologists and the required server space. They must also ensure that potential respondents have access to electronic mail or the Internet. Electronic software is needed for questionnaire development and analysis; otherwise commercial electronic survey services can be used (Vovici [formerly WebSurveyor], SurveyMonkey and QuestionPro). As with postal surveys, an advanced notice by email should closely precede administration of the electronic questionnaire. Potential respondents can be sent an electronic cover letter either with the initial or reminder questionnaires attached or with a link to the an Internet-based questionnaire. Alternatively, the cover letter and questionnaire can be posted on the Web. Incentives can also be provided electronically (e.g., online coupons, entry into a lottery).

Response rate and estimation of sample size

High response rates increase the precision of parameter estimates, reduce the risk of selection bias³ and enhance valid-

ity.²⁸ The lower the response rate, the higher the likelihood that respondents differ from those of nonrespondents, which casts doubt on whether the results of the questionnaire reflect those of the target population.⁵ Investigators may report the *actual response rate*, which reflects the sampling element (including respondents who provide partially or fully completed questionnaires and opt-out responses), or the *analyzable response rate*, which reflects information obtained from partially or fully completed questionnaires as a proportion of the sampling frame (all potential respondents contacted). Although response rates of at least 70% are desirable for external validity,^{2,4,5,17} response rates between 60% and 70%, and sometimes less than 60% (e.g., for controversial topics), may be acceptable.¹⁷ Mean response rates of 54%³² to 61%³³ for physicians and 68%³² for nonphysicians have been reported in recent systematic reviews of postal questionnaires.

In another systematic review of response rates to postal questionnaires, Nakash and colleagues³⁴ identified 15 randomized trials in health care research in patient populations. Similar to Edwards and colleagues,²² whose systematic review of 292 randomized trials was not limited to medical surveys, Nakash and colleagues found that reminder letters and telephone contact had a favourable impact on response rates (odds ratio [OR] 3.71, 95% CI 2.30–5.97); shorter versus longer questionnaires also had an influence, although to a lesser extent (OR 1.35, 95% CI 1.19–1.54). However, unlike Edwards and colleagues, Nakash and coworkers found no evidence that providing an incentive increased the response rate (OR 1.09, 95% CI 0.94–1.27) (see Appendix 2, available at www.cmaj.ca/cgi/content/full/179/3/245/DC1).

Reminders have a powerful and positive influence on response rates. For postal surveys, each additional mailed reminder yields about 30%–50% of the initial responses.¹⁷ If the initial response rate to a questionnaire is 40%, the response rate to a second mailing is anticipated to be between 12% and 20%. In this circumstance, a third mailing would be expected to achieve an overall response rate of 70%. Dillman and colleagues³⁵ proposed the use of 3 follow-up “waves”: an initial reminder postcard sent 1 week after the initial mailing of the questionnaire to the entire sample, and 2 reminders (a letter plus replacement questionnaire) sent at 3 and 7 weeks to nonrespondents, with the final letter and replacement questionnaire sent by certified mail. As with postal surveys, the use of reminders with electronic surveys of health professionals have been found by several authors^{36–38} to increase response rates substantively.

The survey objective, hypotheses and design inform the approach to estimating the sample size. Appendix 3 (available at www.cmaj.ca/cgi/content/full/179/3/245/DC1) outlines the steps involved in estimating the sample size for descriptive survey designs (synthesizing and reporting factual data with the goal of estimating a parameter) and explanatory or experimental survey designs (drawing inferences between constructs to test a hypothesis).⁶ In Appendices 4 and 5 (available at www.cmaj.ca/cgi/content/full/179/3/245/DC1), we provide commonly used formulas for estimating sample sizes in descriptive and experimental study designs, respectively.³⁹

Survey reporting

Complete and transparent reporting is essential for a survey to provide meaningful information for clinicians and researchers. Although infrequently adopted, several recommendations have been published for reporting findings from postal⁴⁰⁻⁴² and electronic surveys.⁴³ One set of recommended questions to consider when writing a report of findings from

Table 4: Questions to consider when preparing a report of findings from postal surveys*

Section	Question
Abstract	<ul style="list-style-type: none"> • Is the objective clearly stated? • Is the design of the study stated? • Is the study setting well described? • Is the survey population described? • Is the response rate reported? • Are the outcome measures identified? • Are the main results clearly reported? • Are the conclusions appropriate?
Introduction	<ul style="list-style-type: none"> • Is the problem clearly stated? • Is the pertinent literature cited and critically appraised? • Is the relevance of the research question explained? • Is the objective clearly stated?
Methods	<ul style="list-style-type: none"> • Is the study design appropriate to the objective? • Is the setting clearly described? • Are the methods described clearly enough to permit other researchers to duplicate the study? • Is the survey sample likely to be representative of the population? • Is the questionnaire described adequately? • Have the validity and reliability of the questionnaire been established? • Was the questionnaire administered in a satisfactory way? • Are the statistical methods used appropriately?
Results	<ul style="list-style-type: none"> • Do the results address the objective? • Are all respondents accounted for? • Are the results clearly and logically presented? • Are the tables and figures appropriate? • Are the numbers consistent in the text and the tables?
Discussion	<ul style="list-style-type: none"> • Are the results succinctly summarized? • Are the implications of the results stated? • Are other interpretations considered and refuted? • Are the limitations of the study and its results explained? • Are appropriate conclusions drawn?

*Modified from Huston.⁴⁰

postal surveys appears in Table 4.⁴⁰ Reviews of the quality of survey reports showed that only 51% included a response rate,⁴⁴ 8%–16% provided access to the questionnaire,^{45,46} and 67% reported validation of the questions.⁴⁵ Only with sufficient detail and transparent reporting of the survey's methods and results can readers appraise the survey's validity.

Conclusion

In this guide for the design and conduct of self-administered surveys of clinician's knowledge, attitudes and practice, we have outlined methods to identify the sampling frame, generate items for inclusion in the questionnaire and reduce these items to a manageable list. We have also described how to further test and administer questionnaires, maximize response rates and ensure transparent reporting of results. Using this systematic approach (summarized in Appendix 6, available at www.cmaj.ca/cgi/content/full/179/3/245/DC1), investigators should be able to design and conduct valid, useful surveys, and readers should be better equipped to appraise published surveys.

This article has been peer reviewed.

Competing interests: None declared.

Contributors: Karen Burns, Neill Adhikari, Maureen Meade, Tasnim Sinuff and Deborah Cook conceived the idea. Karen Burns drafted a template of the guide, reviewed and synthesized pertinent literature and prepared the initial and subsequent drafts of the manuscript. Mark Duffett and Michelle Kho reviewed and synthesized pertinent literature and drafted sections of the manuscript. Maureen Meade and Neill Adhikari contributed to the organization of the guide. Tasnim Sinuff contributed to the organization of the guide and synthesized information in the guide into a summary appendix. Deborah Cook aided in drafting the layout of the guide and provided scientific and methodologic guidance on drafting the guide. All of the authors revised the manuscript critically for important intellectual content and approved the final version submitted for publication.

Acknowledgements: We thank Dr. Donnie Arnold, McMaster University, Hamilton, Ont., for his review of this manuscript.

Karen Burns and Tasnim Sinuff hold a Clinician–Scientist Award from the Canadian Institutes of Health Research (CIHR). Michelle Kho holds a CIHR Fellowship Award (Clinical Research Initiative). Deborah Cook is a Canada Research Chair of the CIHR.

REFERENCES

1. Howie JG. Research in general practice. Anatomy of a research project. *BMJ* 1982;285:266-8.
2. Henry RC, Zivick JD. Principles of survey research. *Fam Pract Res J* 1986;5:145-57.
3. McColl E, Jacoby A, Thomas L, et al. Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. *Health Technol Assess* 2001;5:1-256
4. Rubenfeld GD. Surveys: an introduction. *Respir Care* 2004;49:1181-5.
5. Passmore C, Dobbie AE, Parchman M, et al. Guidelines for constructing a survey. *Fam Med* 2002;34:281-6.
6. Aday LA, Cornelijs LJ. *Designing and conducting health surveys. A comprehensive guide*. 3rd ed. San Francisco: Jossey-Bass; 2006.
7. Sudman S. Applied sampling. In: Rossi PH, Wright JD, Anderson AB, editors. *Handbook of survey research*. San Diego: Academic Press; 1983:145-94.
8. Aday LA, Cornelijs LJ. Advantages and disadvantages of different probability sample designs. In: *Designing and conducting health surveys. A comprehensive guide*. San Francisco (CA): Jossey-Bass; 2006. p. 135.
9. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985;38:27-36.
10. Ehrlich A, Koch T, Amin B, et al. Development and reliability testing of a standardized questionnaire to assess psoriasis phenotype. *J Am Acad Dermatol* 2006; 54:987-91.
11. Fox J. Designing research: basics of survey construction. *Minim Invasive Surg Nurs* 1994;8:77-9.

12. Dillman DA. Mail and other self-administered questionnaires. In: Rossi PH, Wright JD, Anderson AB, editors. *Handbook of survey research*. San Diego (CA): Academic Press; 1983. p. 359-76.
13. Stone DH. Design a questionnaire. *BMJ* 1993;307:1264-6.
14. Woodward CA. Questionnaire construction and question writing for research in medical education. *Med Educ* 1988;22:345-63.
15. Guyatt GH, Cook DJ, King D, et al. The framing of questionnaire items regarding satisfaction with training and its effects on residents' responses. *Acad Med* 1999; 74:192-4.
16. Horvath T. *Basic Statistics for behavioral sciences*. Glenview (IL): Scott, Foresman and Company; 1985. p. 9-21.
17. Sierles FS. How to do research with self-administered surveys. *Acad Psychiatry* 2003;27:104-13.
18. 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.
19. Dillman DA. *Mail and Internet surveys: the tailored design method*. 2nd ed. Hoboken (NJ): John Wiley & Sons; 2000.
20. Turocy PS. Survey research in athletic training: the scientific method of development and implementation. *J Athl Train* 2002;37:S174-9.
21. Couper MP, Traugott MW, Lamias MJ. Web survey design and administration. *Public Opin Q* 2001;65:230-53.
22. Edwards P, Roberts I, Clarke M, et al. Increasing response rates to postal questionnaires: systematic review. *BMJ* 2002;324:1183-91.
23. Collins D. Pre-testing survey instruments: an overview of cognitive methods. *Qual Life Res* 2003;12:229-83.
24. Bowden A, Fox-Rushby JA, Nyandieka L, et al. Methods for pre-testing and piloting survey questions: illustrations from the KENQOL survey of health-related quality of life. *Health Policy Plan* 2002;17:322-30.
25. Juniper EF, Guyatt GH, Streiner DL, et al. Clinical impact versus factor analysis for quality of life questionnaire construction. *J Clin Epidemiol* 1997;50:233-8.
26. Norman GR, Streiner DL. *BioStatistics: the bare essentials*. 2nd ed. Hamilton (ON): BC Decker; 2000.
27. Carmines EG, Zeller RA. Correlation matrix of self-esteem items. In: *Reliability and validity assessment* [Quantitative Applications in the Social Sciences series]. Newbury Park (CA): Sage Publications; 1979. p. 64.
28. Leece P, Bhandari M, Sprague S, et al. Internet versus mailed questionnaires: a randomized comparison (2) [published erratum appears in *J Med Internet Res* 2004;6:e38; corrected and republished in *J Med Internet Res* 2004;6:e39]. *J Med Internet Res* 2004;6:e30.
29. Kim HL, Hollowell CM, Patel RV, et al. Use of new technology in endourology and laparoscopy by American urologists: Internet and postal survey. *Urology* 2000;56:760-5.
30. Raziano DB, Jayadevappa R, Valenzuela D, et al. E-mail versus conventional postal mail survey of geriatric chiefs. *Gerontologist* 2001;41:799-804.
31. Braithwaite D, Emery J, de Lusignan S, et al. Using the Internet to conduct surveys of health professionals: A valid alternative? *Fam Pract* 2003;20:545-51.
32. Asch DA, Jedrzewski MK, Christakis NA. Response rates to mail surveys published in medical journals. *J Clin Epidemiol* 1997;50:1129-36.
33. Cummings SM, Savitz LA, Konrad TR. Reported response rates to mailed physician questionnaires. *Health Serv Res* 2001;35:1347-55.
34. Nakash RA, Hutton JL, Jørstad-Stein EC, et al. Maximising response to postal questionnaires: a systematic review of randomised trials in health research. *BMC Med Res Methodol* 2006;6:5.
35. Dillman DA. *Mail and telephone surveys. The total design method*. Hoboken (NJ): John Wiley & Sons; 1978.
36. Fischbacher C, Chappel D, Edwards R, et al. Health surveys via the Internet: Quick and dirty or rapid and robust? *J R Soc Med* 2000;93:356-9.
37. McLean SA, Feldman JA. The impact of changes in HCFA documentation requirements on academic emergency medicine: results of a physician survey. *Acad Emerg Med* 2001;8:880-5.
38. Schleyer TKL, Forrest JL. Methods for the design and administration of Web-based surveys. *J Am Med Inform Assoc* 2000;7:416-25.
39. Lemeshow S, Hosmer DW Jr, Klar J, et al. *Adequacy of sample size in health studies*. Chichester (UK): John Wiley & Sons; 1990.
40. Huston P. Reporting on surveys: information for authors and peer reviewers. *CMAJ* 1996;154:1695-704.
41. Boynton PM. Administering, analyzing and reporting your questionnaire. *BMJ* 2004;328:1372-5.
42. Kelley K, Clark B, Brown V, et al. Good practice in the conduct and reporting of survey research. *Int J Qual Health Care* 2003;15:261-6.
43. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res* 2004;6:e34.
44. Badger F, Werrett J. Room for improvement? Reporting response rates and recruitment in nursing research in the past decade. *J Adv Nurs* 2005;51:502-10.
45. Rosen T, Olsen J. The art of making questionnaires better. *Am J Epidemiol* 2006; 164:1145-9. Epub 2006 Oct 13.
46. Schilling LM, Kozak K, Lundahl K, et al. Inaccessible novel questionnaires in published medical research: hidden methods, hidden costs. *Am J Epidemiol* 2006; 164:1141-4. Epub 2006 Oct 13.

Correspondence to: Dr. Deborah J. Cook, Professor, Department of Clinical Epidemiology and Biostatistics, McMaster University, Rm. 2C11, 1200 Main St. W, Hamilton ON L8N 3Z5; fax 905 521-6068; debcook@mcmaster.ca

Members of the ACCADEMY (Academy of Critical Care: Development, Evaluation and Methodology) Group: Dr. Neill K.J. Adhikari, Sunnybrook Health Sciences Centre, Sunnybrook Research Institute and Interdepartmental Division of Critical Care, Toronto, Ont.; Donald Arnold, Department of Medicine, McMaster University, Hamilton, Ont.; Dr. Karen E.A. Burns, St. Michael's Hospital, the Interdepartmental Division of Critical Care, Keenan Research Centre and the Li Ka Shing Knowledge Institute, Toronto, Ont.; Dr. Karen Choong, Department of Pediatrics and Division of Critical Care, McMaster Children's Hospital, Hamilton, Ont.; Dr. Deborah J. Cook, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont.; Dr. Cynthia Cupido, Department of Pediatrics and Division of Critical Care, McMaster Children's Hospital, Hamilton, Ont.; Ines De Campos RN, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont., and St. Michael's Hospital, Toronto, Ont.; Dr. Mark Duffett, Department of Pharmacy and Division of Critical Care, McMaster Children's Hospital, and Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont.; Dr. François Lamontagne, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont.; Dr. Wendy Lim, Department of Medicine, McMaster University, Hamilton, Ont.; and Dr. Maureen O. Meade, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont.