

# How to assess a survey report: a guide for readers and peer reviewers

Karen E.A. Burns MD MSc, Michelle E. Kho PT PhD

**Competing interests:** None declared.

This article has been peer reviewed.

**Correspondence to:**

Karen Burns,  
burnsk@smh.ca

*CMAJ* 2015. DOI:10.1503/  
cmaj.140545

Although designing and conducting surveys may appear straightforward, there are important factors to consider when reading and reviewing survey research. Several guides exist on how to design and report surveys, but few guides exist to assist readers and peer reviewers in appraising survey methods.<sup>1-9</sup> We have developed a guide to aid readers and reviewers to discern whether the information gathered from a survey is reliable, unbiased and from a representative sample of the population. In our guide, we pose seven broad questions and specific subquestions to assist in assessing the quality of articles reporting on self-administered surveys (Box 1). We explain the rationale for each question posed and cite literature addressing its relevance in appraising the methodologic and reporting quality of survey research. Throughout the guide, we use the term “questionnaire” to refer to the instrument administered to respondents and “survey” to define the process of administering the questionnaire. We use “readers” to encompass both readers and peer reviewers.

## Questions to ask about a survey report

### Was a clear objective posed?

Every questionnaire should be guided by a simple, clearly articulated objective that highlights the topic of interest, the type of respondents and

the primary research question to be addressed.<sup>2</sup> Readers should use their judgment to determine whether the survey report answered the research question posed.

### Was the target population defined, and was the sample representative of the population?

The population of interest should be clearly defined in the research report. Because administering a questionnaire to an entire population is usually infeasible, researchers must survey a sample of the population. Readers should assess whether the sample of potential respondents was representative of the population. Representativeness refers to how closely the sample reflects the attributes of the population. The “sampling frame” is the target population from which the sample will be drawn.<sup>2</sup>

How the sampling frame and potential respondents were identified should be specified in the report. For example, what sources did the authors use to identify the population (e.g., membership lists, registries), and how did they identify individuals within the population as potential respondents? Sample selection can be random (probability design [simple random, systematic random, stratified random or cluster sampling]) or deliberate (nonprobability design),<sup>2</sup> with advantages and disadvantages associated with each sampling strategy (Table 1).<sup>10</sup> Except for cluster sampling, investigators rely on lists of individuals with accurate and up-to-date contact information (e.g., postal or email addresses, telephone numbers) to conduct probability sampling.

Readers should consider how similar the individuals invited to participate in the survey were to the target population based on the data sources and sampling strategy used, the technique used to administer the survey questionnaire and the respondents’ demographic data.

### Was a systematic approach used to develop the questionnaire?

Questionnaire development includes four phases: item generation, item reduction, formatting and pretesting. Readers should discern

## KEY POINTS

- Seven broad questions and several specific subquestions are posed to guide readers and peer reviewers in systematically assessing the quality of an article reporting on survey research.
- Questions probe whether investigators posed a clear research question, gathered information from an unbiased and representative sample of the population, used a systematic approach to develop the questionnaire and tested the questionnaire.
- Additional questions ask whether authors reported the response rate (or used techniques to assess nonresponse bias), defined the response rate, adopted strategies to enhance the response rate and justified the sample size.
- Finally, the guide prompts readers and peer reviewers to assess the clarity and transparency of the reporting of the survey results.

whether a systematic approach was used to develop the questionnaire and understand the potential consequences of not using a methodical approach. Use of a systematic approach to questionnaire development reassures readers that key domains were not missed and that the authors carefully considered the phrasing of individual questions to limit the chance of misunderstanding. When evaluating questionnaire development, readers should ask themselves whether the methods used by the authors allowed them to address the research question posed.

First, readers should assess how items for inclusion in the questionnaire were generated (e.g., literature reviews,<sup>11</sup> in-depth interviews or focus group sessions, or the Delphi technique involving experts or potential respondents<sup>2,12,13</sup>). In item generation, investigators identify all potential constructs or items (ideas, concepts) that could be included in the questionnaire with the goal of tapping into important domains (categories or themes) of the research question.<sup>14</sup> This step helps investigators define the constructs they wish to explore, group items into domains and start formulating questions within domains.<sup>2</sup> Item generation continues until they cannot identify new items.

Second, readers should assess how the authors identified redundant items and constrained the number of questions posed within domains, without removing important constructs or entire domains, with the goal of limiting respondent burden. Most research questions can be addressed with at least five domains and 25 or fewer items.<sup>12,15</sup> To determine how the items were reduced in number, readers should examine the process used (e.g., investigators may have conducted interviews or focus groups), who was involved (e.g., external appraisers or experts) and how items were identified for inclusion or exclusion (e.g., use of binary responses [in/out], ranking [ordinal scales], rating [Likert scales]<sup>16</sup> or statistical methods).

Next, readers should review the survey methods and the appended questionnaire to determine whether measures were taken by the authors to limit ambiguity while formatting question stems and response options. Each question should have addressed a single construct, and the perspective from which each question was posed should be clear. Readers should examine individual question stems and response formats to ensure that the phrasing was simple and easily understood, each question addressed a single item, and the response formats were mutually exclusive and exhaustive.<sup>2,17</sup> Closed-ended response formats, whereby respondents are restricted to specific responses (e.g., yes or no) or a limited number of categories, are the most frequently used and the

easiest to aggregate and analyze. If necessary, authors may have included indeterminate responses and “other” options to provide a comprehensive list of response options. Readers should note whether the authors avoided using terminology that could be perceived as judgmental, biased or absolute (e.g., “always,” “never”)<sup>15</sup> and avoided using negative and double-barrelled items<sup>17</sup> that may bias responses or make them difficult to interpret.

#### Box 1: A guide for appraising survey reports

1. Was a clear research question posed?
  - 1a. Does the research question or objective specify clearly the type of respondents, the topic of interest, and the primary and secondary research questions to be addressed?
2. Was the target population defined, and was the sample representative of the population?
  - 2a. Was the population of interest specified?
  - 2b. Was the sampling frame specified?
3. Was a systematic approach used to develop the questionnaire?
  - 3a. *Item generation and reduction*: Did the authors report how items were generated and ultimately reduced?
  - 3b. *Questionnaire formatting*: Did the authors specify how questionnaires were formatted?
  - 3c. *Pretesting*: Were individual questions within the questionnaire pretested?
4. Was the questionnaire tested?
  - 4a. *Pilot testing*: Was the entire questionnaire pilot tested?
  - 4b. *Clinimetric testing*: Were any clinimetric properties (face validity or clinical sensibility testing, content validity, inter- or intra-rater reliability) evaluated and reported?
5. Were questionnaires administered in a manner that limited both response and nonresponse bias?
  - 5a. Was the method of questionnaire administration appropriate for the research objective or question posed?
  - 5b. Were additional details regarding prenotification, use of a cover letter and an incentive for questionnaire completion provided?
6. Was the response rate reported, and were strategies used to optimize the response rate?
  - 6a. Was the response rate reported (alternatively, were techniques used to assess nonresponse bias)?
  - 6b. Was the response rate defined?
  - 6c. Were strategies used to enhance the response rate (including sending of reminders)?
  - 6d. Was the sample size justified?
7. Were the results clearly and transparently reported?
  - 7a. Does the survey report address the research question(s) posed or the survey objectives?
  - 7b. Were methods for handling missing data reported?
  - 7c. Were demographic data of the survey respondents provided?
  - 7d. Were the analytical methods clear?
  - 7e. Were the results succinctly summarized?
  - 7f. Did the authors' interpretation of the results align with the data presented?
  - 7g. Were the implications of the results stated?
  - 7h. Was the questionnaire provided in its entirety (as an electronic appendix or in print)?

Lastly, readers should note whether pretesting was conducted. Pretesting ensures that different respondents interpret the same question similarly and that questions fulfill the authors' intended purpose and address the research question posed. In pretesting, the authors obtain feedback (e.g., written or verbal) from individuals who are similar to prospective respondents on whether to accept, reject or revise individual questions.

### Was the questionnaire tested?

Several types of questionnaire testing can be performed, including pilot, clinical sensibility, reliability and validity testing. Readers should assess whether the investigators conducted formal testing to identify problems that may affect how respondents interpret and respond to individual questions and to the questionnaire as a whole. At a minimum, each questionnaire should have undergone pilot testing. Readers should evaluate what process was used for pilot testing the questionnaire (e.g., investigators sought feedback in a semi-structured format), the number and type of people involved (e.g., individuals similar to those in the sampling frame) and what features (e.g., the flow, salience and acceptability of the questionnaire) were assessed. Both pretesting and pilot testing minimize the chance that respondents will misinterpret questions. Whereas pretesting focuses on the wording of the questionnaire, pilot testing assesses the flow and relevance of the entire questionnaire, as well as individual questions, to identify unusual, irrelevant,

poorly worded or redundant questions and responses.<sup>18</sup> Through testing, the authors identify problems with questions and response formats so that modifications can be made to enhance questionnaire reliability, validity and responsiveness.

Readers should determine whether additional testing (e.g., clinical sensibility, inter- or intra-rater reliability, and validity testing) was conducted and, if so, the number and type of participants involved in each assessment.

Clinical sensibility testing addresses the comprehensiveness, clarity and face validity of the questionnaire.<sup>2</sup> Readers should assess whether such an assessment was made and how it was done (e.g., use of a structured assessment sheet with either a Likert scale<sup>16</sup> or nominal response format). Clinical sensibility testing reassures readers that the investigators took steps to identify missing or redundant items, evaluated how well the questionnaire addressed the research question posed and assessed whether existing questions and responses were easily understood.

Reliability testing determines whether differences in respondents' answers were due to poorly designed questions or to true differences within or between respondents. Readers should assess whether any reliability testing was conducted. In intra-rater (or test-retest) reliability testing, investigators assess whether the same respondent answered the same questions consistently when administered at different times, in the absence of any expected change. With internal consistency, they determine whether items

**Table 1:** Commonly used strategies for probability sampling

Sampling design	Description	Advantages	Disadvantages
Simple random	<ul style="list-style-type: none"> <li>• Every individual in population has equal chance of being included in the sample</li> <li>• Potential respondents are selected using various techniques (e.g., lottery process or random-number generator)</li> </ul>	<ul style="list-style-type: none"> <li>• Requires little advance knowledge of population</li> </ul>	<ul style="list-style-type: none"> <li>• May not capture specific groups</li> <li>• May not be efficient</li> </ul>
Systematic random	<ul style="list-style-type: none"> <li>• Starting point on a list is randomly chosen, and individuals are selected at prespecified intervals</li> <li>• Starting point and sampling interval are determined by required sample size</li> </ul>	<ul style="list-style-type: none"> <li>• High precision</li> <li>• Easy to analyze data and compute sampling errors</li> </ul>	<ul style="list-style-type: none"> <li>• Ordering of elements in sampling frame may create biases</li> <li>• May not capture specific groups</li> <li>• May not be efficient</li> </ul>
Stratified random	<ul style="list-style-type: none"> <li>• Potential respondents are organized into strata or categories and sampled using simple or systematic sampling within strata to ensure possible representation of specific groups</li> <li>• Sampled proportion can be proportionate or disproportionate across strata</li> </ul>	<ul style="list-style-type: none"> <li>• Captures specific groups</li> <li>• Disproportionate sampling possible</li> <li>• Highest precision</li> </ul>	<ul style="list-style-type: none"> <li>• Requires advance knowledge of population</li> <li>• More complex to analyze data and compute sampling errors</li> </ul>
Cluster	<ul style="list-style-type: none"> <li>• Population is divided into clusters that are mutually exclusive, heterogeneous and exhaustive</li> <li>• Clusters are sampled in a stepwise manner</li> </ul>	<ul style="list-style-type: none"> <li>• Lower field costs</li> <li>• Enables sampling of groups if individuals not available</li> </ul>	<ul style="list-style-type: none"> <li>• More complex to analyze data and compute sampling errors</li> <li>• Lowest precision</li> </ul>

Adapted, with permission, from Aday and Cornelius.<sup>10</sup>

within a construct are associated with one another. A variety of statistical tests can be used to assess test–retest reliability and internal consistency.<sup>2</sup> Test–retest reliability is commonly reported in survey articles. Substantial or near-perfect reliability scores (e.g., intraclass correlation coefficient > 0.61 and 0.80, respectively) should reassure readers that the respondents, when presented with the same question on two separate occasions, answered it similarly.<sup>19</sup>

Types of validity assessments include face, content, construct and criterion validity. Readers should assess whether any validity testing was conducted. Although the number of validity assessments depends on current or future use of the questionnaire, investigators should have assessed at a minimum the face validity of their questionnaire during clinical sensibility testing.<sup>2</sup> In face validity, experts in the field or a sample of respondents similar to the target population determine whether the questionnaire measures what it aims to measure.<sup>20</sup> In content validity, experts assess whether the content of the questionnaire includes all aspects considered essential to the construct or topic. Investigators evaluate construct validity when specific criteria to define the concept of interest are unknown; they verify whether key constructs were included using content validity assessments made by experts in the field or using statistical methods (e.g., factor analysis).<sup>2</sup> In criterion validity, investigators compare responses to items with a gold standard.<sup>2</sup>

### **Was the questionnaire administered in a manner that limited both response and nonresponse bias?**

Although a variety of techniques (e.g., postal, telephone, electronic) can be used to administer questionnaires, postal and electronic (email or Internet) methods are the most common. The technique chosen depends on several factors, such as the research question, the amount and type of information desired, the sample size, available personnel and financial resources. The selected administration technique may result in response bias and ultimately influence how well survey respondents represent the target population. For example, telephone surveys will tend to identify respondents at home, and electronic surveys are more apt to be answered by those with Internet access and computer skills. Moreover, although electronic questionnaires are less labour intensive to conduct than postal questionnaires, their response rates may be lower.<sup>21</sup>

Readers should assess the alignment between the administration technique used and the research question posed, the potential for the

administration technique to have influenced the survey results and whether the investigators made efforts to contact nonrespondents to limit nonresponse bias.

### **Was the response rate reported, and were strategies used to optimize the response rate?**

Readers should examine the reported survey methods and results to determine whether the response rates (numerators and denominators) align with the definition of the response rates provided, whether a target sample was provided and whether the investigators used specific strategies to enhance the response rate.<sup>22</sup> A high response rate (i.e., > 80%) minimizes the potential for bias due to absent responses, ensures precision of estimates and generalizability of survey findings to the target population and enhances the validity of the questionnaire.<sup>1,2,23</sup>

The “sampling element” refers to the respondents for whom information is collected and analyzed. To compute accurate response rates, readers need information on the number of surveys sent (denominator) and the number of surveys received (numerator). They should then examine characteristics of the surveys returned and identify reasons for nonresponse. For example, returned questionnaires may be classified as eligible (completed) or ineligible (e.g., returned and opted out because it did not meet eligibility criteria or self-reported as ineligible). Questionnaires that were not returned may represent individuals who were eligible but who did not wish to respond or participate in the survey or individuals with indeterminate eligibility.<sup>8</sup> Readers should also determine how specific eligibility circumstances (e.g., return-to-sender questionnaires, questionnaires completed in part or in full) were managed and analyzed by the investigators. Transparent and replicable formulas for calculating response rates are continually being developed and updated by the American Association for Public Opinion Research.<sup>8</sup> The use of standardized formulas enables comparison of response rates across surveys. Investigators should define and report the response rates (overall and for prespecified subgroups) and provide sufficient detail to understand how they were computed. This information will help readers to verify the computations and determine how closely actual and target response rates align.

To optimize the number of valid responses, authors should report a sample size estimate based on anticipated response rates.<sup>10,24</sup> An a priori computation of the sample size helps guide the number of respondents sought and, if realized, increases readers’ confidence in the survey results.<sup>25</sup> Response

rates below 60%, between 60% and 70%, and 70% or higher have all traditionally been considered acceptable,<sup>12,14,16</sup> with lower mean rates reported among physicians (54%–61%)<sup>26,27</sup> than among nonphysicians (68%).<sup>26</sup> A recent meta-analysis of 48 studies identified an overall survey response rate of 53% among health professionals.<sup>28</sup> A rate of 60% or higher among physicians is reasonable and has face validity. Notwithstanding, some authors feel that there is no scientifically established minimum acceptable response rate and assert that response rates may not be associated with survey representativeness or quality.<sup>29</sup> In such instances, the more important consideration in determining representativeness is the degree to which sampled respondents differ from the target population (or nonresponse bias), which can be assessed using a variety of techniques.<sup>29</sup>

Relevant or topical research questions are more likely to garner interest and enthusiasm

from potential respondents and favourably influence response rates.<sup>22</sup> Readers should assess whether potential respondents received a pre-notification statement and an incentive (monetary or nonmonetary) for completing the questionnaire, and what type and number of reminder questionnaires were issued to nonrespondents. All of these factors can enhance response rates. In Tables 2 and 3, we summarize strategies shown to substantially influence response rates to postal and electronic questionnaires in a large meta-analysis.<sup>22</sup> Reminder strategies may involve multiple follow-up waves.<sup>32</sup> For postal surveys, each additional mailed reminder could increase initial response rates by 30% to 50%.<sup>33</sup> The ordering of survey administration techniques can also be important. In one study, the response rate was higher when an initial postal survey was followed by an online survey to nonrespondents than when the reverse was done.<sup>34</sup>

**Table 2:** Strategies that enhance or reduce the response rate to mailed questionnaires<sup>22</sup>

Strategy	No. of studies	No. of participants	Odds ratio (95% CI)	Heterogeneity*	
				I <sup>2</sup> value, %	p value
<b>Enhances response rate</b>					
Monetary incentive	94	160 004	1.87 (1.73–2.03)	84	< 0.001
Recorded delivery	15	18 931	1.76 (1.43–2.18)	71	< 0.001
Teaser on envelope†	1	190	3.08 (1.27–7.44)	NA	NA
More interesting topic	3	2 711	2.00 (1.32–3.04)	80	0.01
Prenotification	47	79 651	1.45 (1.29–1.63)	89	< 0.001
Follow-up contact	19	32 778	1.35 (1.18–1.55)	76	< 0.001
Unconditional incentive	24	27 569	1.61 (1.36–1.89)	88	< 0.001
Shorter questionnaire	56	60 119	1.64 (1.43–1.87)	91	< 0.001
Second copy of questionnaire at follow-up	11	8 619	1.46 (1.13–1.90)	82	< 0.001
Mention of obligation to respond	3	600	1.61(1.16–2.22)	0	0.98
University sponsorship	14	21 628	1.32 (1.13–1.54)	83	< 0.001
Nonmonetary incentive	94	135 934	1.15 (1.08–1.22)	79	< 0.001
Personalized questionnaire	58	60 184	1.14 (1.07–1.22)	63	< 0.001
Handwritten address	7	5 091	1.25 (1.08–1.45)	14	0.3
Stamped return envelope‡	27	48 612	1.24 (1.14–1.35)	69	< 0.001
Assurance of confidentiality	1	25 000	1.33 (1.24–1.42)	NA	NA
First class outward mailing	2	8 300	1.11 (1.02–1.21)	0	0.8
<b>Reduces response rate</b>					
Questions of sensitive nature	10	21 393	0.94 (0.88–1.00)	0	0.5

Note: CI = confidence interval, NA = not applicable.  
 \*The I<sup>2</sup> statistic describes the percentage of total variance across studies that can be attributed to heterogeneity (differences within and between studies) rather than to chance.<sup>30</sup> Typically, I<sup>2</sup> statistic thresholds of 0% to 40%, 30% to 60%, 50% to 90% and > 75% represent between-study heterogeneity that might not be important or that might be moderate, substantial or considerable, respectively.<sup>31</sup> A  $\chi^2$  test is typically used to assess heterogeneity. Because this test has low power in most meta-analyses (e.g., when studies have few patients and studies are few in number), a significant result may indicate a problem with heterogeneity; however, a nonsignificant result may not represent the absence of heterogeneity. Consequently, a p value < 0.1, rather than < 0.05, is usually used to determine statistical significance.  
 †Comment that implies participants may benefit from opening envelope.  
 ‡Versus franked return envelope.



### Were the results reported clearly and transparently?

The reported results should directly address the primary and secondary research questions posed. Survey findings should be reported with sufficient detail to be clear and transparent. Readers should assess whether the methods used to handle missing data and to conduct analyses have been reported. They should also determine if the authors' conclusions align with the data presented and whether the authors have discussed the implications of their findings. Readers should review the demographic data on the survey respondents to determine how similar or different the sampling frame is from their own population. Finally, they may wish to review the questionnaire, which ideally should be appended as an electronic supplement, to align specific questions with reported results.

Although several guides have been published for reporting survey findings,<sup>3-6</sup> survey methods are often poorly reported, leaving readers to question whether specific steps in questionnaire design were conducted or simply not reported.<sup>9,35</sup> In a review of 117 published surveys, few provided the questionnaire or core questions (35%), defined the response rate (25%), reported the validity or reliability of the instrument (19%), discussed the representativeness of the sample (11%) or identified how missing data were han-

dled (11%).<sup>9</sup> In another review of survey reporting in critical care, Duffett and colleagues<sup>35</sup> identified five key features of high-quality survey reporting: a stated question (or objective); identification of the respondent group; the number of questionnaires distributed; the number of returned questionnaires completed in part or in full (or a response rate); and the methods used to deal with incomplete surveys. They identified additional methodologic features that should be reported to aid in interpretation and to limit bias: the unit of analysis (e.g., individual, hospital, city); the number of potential respondents who could not be contacted; the rationale for excluding entire questionnaires from the analysis; and the type of analysis conducted (e.g., descriptive, inferential or higher level analysis).<sup>35</sup> Finally, they highlighted the need for investigators to provide demographic data about respondents and to append a copy of the questionnaire to the published survey report.<sup>35</sup> With transparent reporting, readers will be able to identify the strengths and limitations of survey studies more easily and to determine how applicable the results are to their setting.

### Discussion

The seven broad questions and specific subquestions posed in our guide are designed to help

**Table 3:** Strategies that enhance or reduce the response rate to electronic questionnaires<sup>22</sup>

Strategy	No. of studies	No. of participants	Odds ratio (95% CI)	Heterogeneity*	
				I <sup>2</sup> value, %	p value
<b>Enhances response rate</b>					
Nonmonetary incentive	6	17 493	1.72 (1.09–2.72)	96	< 0.001
Shorter questionnaire	2	7 589	1.73 (1.40–2.13)	68	0.08
Statement that others had responded	1	8 586	1.52 (1.36–1.70)	NA	NA
More interesting topic	1	2 176	1.85 (1.52–2.26)	NA	NA
Lottery with immediate notification of results	1	2 233	1.37 (1.13–1.65)	NA	NA
Offer of survey results	1	2 332	1.36 (1.15–1.61)	NA	NA
Use of a white background	1	6 090	1.31 (1.10–1.56)	NA	NA
Personalized questionnaire	12	48 910	1.24 (1.17–1.32)	41	0.07
Simple header	1	5 075	1.23 (1.03–1.48)	NA	NA
Textual representation of response categories	1	5 413	1.19 (1.05–1.36)	NA	NA
Deadline given	1	8 586	1.18 (1.03–1.34)	NA	NA
Picture included in email	2	720	3.05 (1.84–5.06)	19	0.3
<b>Reduces response rate</b>					
Survey mentioned in email subject line	2	3 845	0.81 (0.67–0.97)	0	0.3
Male signature in email message	2	720	0.55 (0.38–0.80)	0	0.96

Note: CI = confidence interval, NA = not applicable.  
\*See Table 2 for explanation of heterogeneity.

readers assess the quality of survey reports systematically. They assess whether authors have posed a clear research question, gathered information from an unbiased and representative sample of the population and used a systematic approach to develop and test their questionnaire. They also probe whether authors defined and reported response rates (or assessed nonresponse bias), adopted strategies to enhance the response rate and justified the sample size. Finally, our guide prompts readers to assess the clarity and transparency with which survey results are reported. Rigorous design and testing, a high response rate, and appropriate interpretation and reporting of analyses enhance the credibility and trustworthiness of survey findings.

Although surveys may be categorized under the umbrella of observational research, there are important distinctions between surveys and other observational study designs (e.g., case-control, cross-sectional and cohort) that are not captured by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) framework.<sup>36,37</sup> For example, with surveys, authors question respondents to understand their attitudes, knowledge and beliefs, whereas with observational studies, investigators typically observe participants for outcomes. Reasons for non-response and nonparticipation in surveys are often unknown, whereas individuals can be accounted for at various stages of observational studies. Survey researchers typically do not have detailed information about the characteristics of nonrespondents, whereas in observational studies, investigators may have demographic data for eligible participants who were not ultimately included in the cohort. Although they bear some similarities to observational studies, surveys have unique features of development, testing, administration and reporting that may justify the development and use of a separate reporting framework.

Our guide has limitations. First, we did not conduct a systematic review with the goal of critically appraising, comparing and synthesizing guidelines and checklists for reporting survey research. As a result, our guide may not include some items thought to be important by other researchers. Second, our guide is better suited to aiding readers in evaluating reports of self-administered questionnaires as opposed to interviewer-administered questionnaires. Third, although it does not bear directly on the methodologic quality of questionnaires, we did not address issues pertaining to research ethics approval and informed consent. Although ethics approval is required in Canada, it is not mandatory for conducting survey research in some jurisdictions. Completion and return of questionnaires typically implies consent to participate; however, written consent may be required to allow ongoing prospective follow-up, continued engagement and future contact. Finally, our intent was to develop a simple guide to assist readers in assessing and appraising a survey report, not a guidance document on how to conduct and report a survey. Although our guide does not supplant the need for a well-developed guideline on survey reporting,<sup>9</sup> it provides a needed framework to assist readers in appraising survey methods and reporting, and it lays the foundation for future work in this area.

naires typically implies consent to participate; however, written consent may be required to allow ongoing prospective follow-up, continued engagement and future contact. Finally, our intent was to develop a simple guide to assist readers in assessing and appraising a survey report, not a guidance document on how to conduct and report a survey. Although our guide does not supplant the need for a well-developed guideline on survey reporting,<sup>9</sup> it provides a needed framework to assist readers in appraising survey methods and reporting, and it lays the foundation for future work in this area.

## References

1. Thoma A, Cornacchi SD, Farrokhyar F, et al.; Evidence-Based Surgery Working Group. How to assess a survey in surgery. *Can J Surg* 2011;54:394-402.
2. Burns KE, Duffet M, Kho M, et al.; ACCADEMY Group. A guide for the design and conduct of self-administered surveys of clinicians. *CMAJ* 2008;179:245-52.
3. 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.
4. Huston P. Reporting on surveys: information for authors and peer reviewers. *CMAJ* 1996;154:1695-704.
5. Boynton PM. Administering, analyzing and reporting your questionnaire. *BMJ* 2004;328:1372-5.
6. 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.
7. Draugalis JR, Coons SJ, Plaza CM. Best practices for survey research reports: a synopsis for authors and reviewers. *Am J Pharm Educ* 2008;72:11.
8. *Standard definitions: final dispositions of case codes and outcome rates for surveys*. 7th ed. Deerfield (IL): American Association for Public Opinion Research; 2011. Available: [www.aapor.org/Standard\\_Definitions2.htm](http://www.aapor.org/Standard_Definitions2.htm) (accessed 2014 Nov. 14).
9. Bennett C, Khangura S, Brehaut JC, et al. Reporting guidelines for survey research: an analysis of published guidance and reporting practices. *PLoS Med* 2010;8:e1001069.
10. Aday LA, Cornelius LJ. *Designing and conducting health surveys: a comprehensive guide*. 3rd ed. San Francisco: Jossey-Bass; 2006.
11. Birch DW, Eady A, Robertson D, et al.; Evidence-Based Surgery Working Group. Users' guide to the surgical literature: how to perform a literature search [published erratum in *Can J Surg* 2003;46:250]. *Can J Surg* 2003;46:136-41.
12. Passmore C, Dobbie AE, Parchman M, et al. Guidelines for constructing a survey. *Fam Med* 2002;34:281-6.
13. 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.e1-14.
14. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985;38:27-36.
15. Fox J. Designing research: basics of survey construction. *Minim Invasive Surg Nurs* 1994;8:77-9.
16. Elder JP, Artz LM, Beaudin P, et al. Multivariate evaluation of health attitudes and behaviors: development and validation of a method for health promotion research. *Prev Med* 1985;14:34-54.
17. Babbie E. *Survey research methods*. 2nd ed. Belmont (CA): Wadsworth; 1998.
18. Collins D. Pre-testing survey instruments: an overview of cognitive methods. *Qual Life Res* 2003;12:229-38.
19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
20. Turocy PS. Survey research in athletic training: the scientific method of development and implementation. *J Athl Train* 2002;37:S174-9.
21. 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.
22. Edwards PJ, Roberts I, Clarke MJ, et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev* 2009;(3):MR000008.
23. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th ed. Oxford (UK): Oxford University Press; 2008.

24. Lemeshow S, Hosmer DW Jr, Klar J, et al. *Adequacy of sample size in health studies*. Chichester (UK): John Wiley & Sons; 1990.
25. Al-Subaihi AA. Sample size determination. Influencing factors and calculation strategies for survey research. *Saudi Med J* 2003; 24:323-30.
26. Asch DA, Jedrzwieski MK, Christakis NA. Response rates to mail surveys published in medical journals. *J Clin Epidemiol* 1997;50:1129-36.
27. Cummings SM, Savitz LA, Konrad TR. Reported response rates to mailed physician questionnaires. *Health Serv Res* 2001;35: 1347-55.
28. Cho YI, Johnson TP, Van Geest JB. Enhancing surveys of health care professionals: a meta-analysis of techniques to improve response. *Eval Health Prof* 2013;36:382-407.
29. Johnson TP. Response rates and nonresponse errors in surveys. *JAMA* 2012;307:1805-6.
30. Higgins JPT, Altman DG, Sterne JAC; Cochrane Statistical Methods Group; Cochrane Bias Methods Group. Assessing risk of bias in included studies [chapter 8]. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Version 5.0.1 [updated Sept. 2008]. Oxford (UK): Cochrane Collaboration; 2011. Available: [www.cochrane-handbook.org](http://www.cochrane-handbook.org) (accessed 2014 Nov. 14).
31. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.
32. Dillman DA. *Mail and telephone surveys: the total design method*. New York: John Wiley & Sons; 1978.
33. Sierles FS. How to do research with self-administered surveys. *Acad Psychiatry* 2003;27:104-13.
34. Beebe TJ, Locke GR III, Barnes SA, et al. Mixing web and mail methods in a survey of physicians. *Health Serv Res* 2007 42(3 Pt 1): 1219-34.
35. Duffett M, Burns KE, Adhikari NK, et al. Quality of reporting of surveys in critical care journals: a methodologic review. *Crit Care Med* 2012;40:441-9.
36. von Elm E, Altman DG, Egger M, et al. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344-9.
37. Vandembroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007;4:e297.

**Affiliations:** Divisions of Respiriology and Critical Care (Burns), and Li Ka Shing Knowledge Institute (Burns), St. Michael's Hospital, Toronto, Ont.; Department of Clinical Epidemiology and Biostatistics (Burns) and School of

Rehabilitation Science (Kho), McMaster University, Hamilton, Ont.; Department of Physiotherapy (Kho), St. Joseph's Healthcare, Hamilton, Ont.; Department of Physical Medicine and Rehabilitation (Kho), Johns Hopkins University, Baltimore, Md.

**Contributors:** Karen Burns performed the literature search, and Michelle Kho provided scientific guidance. Both authors drafted and revised the manuscript, approved the final version submitted for publication and agreed to act as guarantors of the work.

**Acknowledgements:** Karen Burns holds a Clinician Scientist Award from the Canadian Institutes of Health Research (CIHR) and an Ontario Ministry of Research and Innovation Early Researcher Award. Michelle Kho holds a CIHR Canada Research Chair in Critical Care Rehabilitation and Knowledge Translation.

#### Resources

- American Association for Public Opinion Research ([www.aapor.org](http://www.aapor.org))
- Best practices for survey research and public opinion ([www.aapor.org/AAPORKentico/Standards-Ethics/Best-Practices.aspx](http://www.aapor.org/AAPORKentico/Standards-Ethics/Best-Practices.aspx))
- 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 ([www.jmir.org/2004/3/e34/](http://www.jmir.org/2004/3/e34/))
- Burns KE, Duffett M, Kho M, et al.; ACCADEMY Group. A guide for the design and conduct of self-administered surveys of clinicians. *CMAJ* 2008;179:245-52 ([www.cmaj.ca/content/179/3/245.long](http://www.cmaj.ca/content/179/3/245.long))
- 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 (<http://intqhc.oxfordjournals.org/content/15/3/261.long>)