

USERS' GUIDE TO THE SURGICAL LITERATURE

How to assess a survey in surgery

Achilleas Thoma, MD, MSc
Sylvie D. Cornacchi, MSc
Forough Farrokhyar, MPhil, PhD
Mohit Bhandari, MD, MSc
Charlie H. Goldsmith, PhD
for the Evidence-Based
Surgery Working Group*

From the Surgical Outcomes Research
Centre, Department of Surgery,
McMaster University and St. Joseph's
Healthcare, Hamilton, Ont.

*The Evidence-Based Surgery Working
Group members comprises Drs. S.
Archibald, M. Bhandari, M. Cadeddu,
S. Cornacchi, F. Farrokhyar, C.H. Goldsmith,
T. Haines, R. Hansebout, R. Jaeschke,
C. Levis, P. Lovrics, M. Simunovic,
V. Tandan, A. Thoma

Accepted for publication
July 8, 2011

Correspondence to:
Dr. A. Thoma
101-206 James St. S
Hamilton ON L8P 3A9
athoma@mcmaster.ca

DOI: 10.1503/cjs.025910

The use of surveys is expanding in all domains of society. Frequently surgeons are presented with hard-copy or email surveys asking for information about their knowledge, beliefs, attitudes and practice patterns. The purpose of these questionnaires may be to obtain an accurate picture of what is going on in their surgical practices, and results are used by local, regional or national organizations to effect changes in surgical practice. Questionnaires can collect descriptive (reporting actual data) or explanatory (drawing inferences between constructs or concepts) data and can explore several constructs at a time.^{1,2}

There are 2 basic types of surveys: cross-sectional and longitudinal surveys. Some cross-sectional surveys gather descriptive information on a population at a single time (e.g., survey of orthopedic trauma surgeons to explore the influence of physician and practice characteristics on referral for physical therapy in patients with traumatic lower-extremity injuries³). A different cross-sectional survey questionnaire might be designed to determine the relation between 2 factors on a representative sample at a particular time. For example, a population-based cross-sectional survey was conducted to explore geographic and sociodemographic factors associated with variation in the accessibility of total hip and knee replacement surgery in England.⁴ The authors found evidence of unequal access based on age, sex, rurality and race. Longitudinal surveys are conducted to determine changes in a population over a period of time.⁵ An example is a prospective longitudinal survey on quality of life among 558 women with breast cancer who underwent surgical treatment and were compared according to whether or not they received chemotherapy.⁶ The authors reported that the quality of life of both groups improved significantly in the year after primary treatment ended, but adjuvant chemotherapy was associated with more severe physical symptoms. Note that prevalence rather than incidence is normally determined in a cross-sectional survey. On the other hand, the temporal sequence of a cause and effect relation can be assessed using longitudinal surveys.

The aim of a survey is to obtain reliable and unbiased data from a representative sample.⁷ Surveys can have a major impact if surgical organizations act on the results. If the surveys have sound methodology, most likely their inferences are correct and will be helpful. However, if proper methodology was not considered and inferences are adopted, surveys can have undesired consequences. High response rates are needed to ensure validity and reduce nonresponse bias.^{1,2} Response rates to mail and email surveys are particularly low among surgeons, some as low as 9%.⁸⁻¹¹ Response rates as high as 80% have also been reported.^{12,13} The purpose of this article is to help surgeons critically appraise survey results reported in the surgical literature.

CLINICAL SCENARIO

You are a urologist, and at the most recent weekly urologic rounds there was a heated exchange between 2 colleagues after the presentation of a clinical

case: the younger one was criticizing his senior colleague for not adopting evidence-based principles in his practice. The head of the urology division became concerned that his senior faculty members may be resistant to adopting evidence-based clinical practice, which *BMJ* hailed as one of the most important 15 medical breakthroughs in the last 166 years.¹⁴ He asks you, the most junior member of the division, to find out to what extent North American urologists have adopted evidence-based practices. He asks you to present your findings during next week's rounds.

LITERATURE SEARCH

To obtain the most specific and up-to-date information about urologists' opinions and adoption of evidence-based practices, you access PubMed (www.ncbi.nlm.nih.gov/PubMed) on one of the hospital library computers. Keywords to use in your search are derived from your clinical question (refer to "Users' guide to the surgical literature: How to perform a literature search"¹⁵ for detailed information on how to develop a clinical question and conduct a successful literature search). You use the search terms "surveys," "urology" and "evidence-based medicine," which yields 68 results. To further restrict the search, you use the "limits" function in PubMed to obtain only English-language studies carried out on human subjects published in the last 5 years. The results of this search yields 42 articles. You scan through the titles and find 4 articles that describe surveys of urologists on the topic of evidence-based medicine.^{11,16-18} A review of the article abstracts reveals that they are surveys conducted by a research group at Duke University; 1 article discusses a survey of urology program directors concerning evidence-based surgery training in residency,¹⁶ and the 3 other articles

describe the results of 2 surveys of the American Urological Association (AUA) membership on the topic of perceptions and competence in evidence-based medicine in 2005-2006.^{11,17,18} The article by Sur and colleagues¹¹ appears to be particularly relevant to your clinical question, so you obtain an online version of this article to review.

We will use a framework (Box 1) similar to that in previous articles,¹⁹⁻²² to appraise the validity of the study by Sur and colleagues,¹¹ interpret the results and apply them to our scenario. The key characteristics of the study by Sur and colleagues are summarized in Table 1.

ARE THE RESULTS VALID?

Primary guides

Did the investigators have a clear objective for this survey? Did they ask a clear question?

The validity of a survey will depend to a great extent on whether the investigators have asked a pertinent research question that can be answered by the survey data. In general, good questions are clear, simple, clinically relevant, interesting and overall answerable. Important and interesting questions are more likely to attract the attention of the target audience and are more likely to be answered. It is important that questions are specific and precise. The refinement of the question requires an operational definition of the terms involved; detailed operational definitions minimize vagueness and permit the investigator to state refined research questions.²³ Sur and colleagues¹¹ had a clear objective for their survey. They wanted to determine the attitudes of urologists toward evidence-based medicine

Box 1. Guidelines for how to use surgical survey articles

Are the results valid?

Primary guides

- Did the investigators have a clear objective for this survey? Did they ask a clear question?
- Was there an explicit sampling frame?
- Was the development of the questionnaire appropriate (item generation, item reduction, item formatting, composition, pretesting)?

Secondary guides

- Did the investigators perform clinical sensitivity testing?
- Was there reliability and validity testing?
- Was the administration of the questionnaire appropriate?

What were the results?

- What was the magnitude of the response rate?
- Were appropriate statistical methods used?
- Was the reporting transparent?
- Were the conclusions appropriate?

Will the results help me alter my practice?

- Are the results generalizable to my practice?
- Will the information/conclusions from this survey help me change or improve my practice/behaviour?

Table 1. Key characteristics of the survey by Sur and colleagues¹¹

Characteristic	Survey
Survey development	Questions modelled on 2 previously tested surveys on similar topic
Pilot testing: validity, reliability, clinical sensibility	Not reported
Sample	All members of AUA with a listed email address
Method of administration	Email/Web-based survey
No. of contacts	Email invitation sent once, no reminders
Response rate, %	8.8
Undeliverable email invitations, %	13.0
Responder characteristics, %	
Male sex	88.4
Age > 46 yr	61.5
Community practice	71.4
Practice location in North America	71.3
Incentives for completing survey	None
AUA = American Urological Association.	

and their familiarity with technical terms and resources related to evidence-based medicine.

Was there an explicit sampling frame?

Ideally investigators would like to administer their questionnaire among all potential responders in their target population. If this is impossible owing to the size of the population or financial constraints, investigators may choose to survey a sample of the target population.²⁴ The “sampling frame” is the target population from which the sample will be drawn.^{1,24,25} The “sampling element” refers to the respondents for whom information is collected and analyzed.²⁵ It is essential that the sampling frame should represent the population of interest. If an inappropriate sampling method is used it may limit the generalizability of the survey results. Different sampling methods and their advantages and disadvantages are listed in Table 2.

Sur and colleagues¹¹ contacted all members of the AUA with a listed email address to participate in the survey. They indicated that an introductory letter was sent by the Chair of the AUA Practice Guidelines Committee. The survey was Web-based and administered from an AUA server. The sampling frame was appropriate, as the investigators sought to capture all American urologists, although they did not justify their sampling strategy. The authors did not mention the percentage of urologists who had email access and whether the email addresses of those who did have access were updated at the time they sent out their survey.

Was the development of the questionnaire appropriate?

The development of a survey questionnaire is a well-defined process that requires item generation, item reduction, questionnaire formatting, composition and pretesting.¹

The purpose of item generation is to consider all potential items (ideas, concepts) for inclusion in the question-

naire, with the goal of tapping into important domains (categories or themes) suggested by the research question.²⁶ Items may be generated with potential responders or experts through literature reviews, in-depth interviews, focus-group sessions, or a combination of these methods.¹ This process ends only when “sampling to redundancy” has been achieved.¹ At this stage, no new items emerge. One such technique is the Delphi process, in which items are nominated and rated by experts until consensus is achieved.^{23,27}

In the item-reduction stage, investigators prune the large number of potentially relevant questions within each domain to a manageable number, as a lengthy questionnaire is unlikely to be completed. It is important that in this process one does not eliminate entire domains or important constructs. A number of methods of item reduction exist: use of interviews or focus groups with content experts, external appraisal, participant input (e.g., ranking or rating) and statistical methods.¹

Questionnaire formatting includes “question stems,” which are statements or questions to which responses are sought. Each question should focus on a single construct. Question stems should contain fewer than 20 words and be easy to understand and interpret.^{23,28} Furthermore, the questions should be nonjudgmental and unbiased.²⁸ They should be socially and culturally sensitive. Dillman²⁹ recommends using simple rather than specialized words, vertical rather than horizontal layout for scalar categories, and using equal numbers of positive and negative categories for scalar questions. Absolute terms, such as “always,” “none” or “never,” and abbreviations should be avoided.^{30,31} It is important to include a biostatistician early in the survey questionnaire development to ensure that the data required for analysis are obtained in a usable format. It is important to consider whether the responses will be nominal or ordinal or whether they will express intervals or ratios.

Table 2. Different approaches to random sample selection*

Sampling design	Definition	Advantages	Disadvantages
Simple random	Every individual in the population of interest has an equal chance of being included in the sample (e.g., using random number generator)	<ul style="list-style-type: none"> • Requires little advance knowledge of the population 	<ul style="list-style-type: none"> • May not capture specific groups • May not be efficient
Systematic random	Investigator randomly selects a starting point on a list and then selects individuals at a prespecified sampling interval (e.g., every tenth individual)	<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	Potential respondents are organized into distinct categories and randomly sampled using simple or systematic sampling within categories to ensure that specific subgroups of interest are represented	<ul style="list-style-type: none"> • Captures specific groups • Enables disproportionate sampling and optimal allocation within strata • Highest precision 	<ul style="list-style-type: none"> • Requires advanced knowledge of population • More complex to analyze data and compute sampling errors
Cluster	Researchers divide the population into mutually exclusive, heterogeneous clusters and then randomly sample the clusters (e.g., clusters of housing units or telephone numbers)	<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 LA and Cornelius LJ. *Designing and conducting health surveys: a comprehensive guide*. 3rd ed. San Francisco (CA): Jossey-Bass (John Wiley & Sons Inc.); 2006.²⁸

Investigators also need to consider whether “indeterminate” response options will be allowed for uncertainty.²⁸ If “ceiling or floor” questions (i.e., responses that tend to cluster at the top or bottom of scales) are identified, investigators may consider removing them during item reduction.¹ See Dillman²⁹ for a more detailed discussion of question design.

Questionnaire developers should consider asking the simple questions and those requesting demographic information early and the difficult, complex questions later in the questionnaire. 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 responders’ attention and enhance visual appeal.¹ Questions need to 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) for responses.¹ Specific formatting strategies (e.g., the use of coloured ink), the placement of more interesting questions first, shorter survey length and the use of symbols (e.g., arrows in combination with larger and darker fonts to indicate skip patterns) can influence respondents’ answers and enhance response rates.^{32–34} For Internet-based surveys,

there are a number of options. The investigators may use a single scrolling page or a series of linked pages often accompanied by electronic instructions and links to facilitate questionnaire flow.¹ Table 3 lists additional considerations for Internet surveys.

The questionnaire should include a cover letter that creates a positive first impression. This cover letter should state the purpose of the survey and highlight why potential respondents were selected.^{2,29} To optimize response rate the following strategies have been suggested:

- print cover letters on departmental stationery and include the signatures of the investigators,
- personalize cover letters to recipients who are known to the investigators,
- 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.^{1,32,35}

Questionnaires sent by mail should include the cover letter, a return (stamped or metered) envelope and incentive (if provided).¹ Email cover letters may have the survey embedded within the email or may provide a link to a website from which to access the survey (Table 3).^{36–41}

Table 3. Summary of Internet/email survey design considerations and advantages/disadvantages^{29,32,34,36–41,47,49}

Design consideration	Internet/email surveys
Potential sampling frames	<ul style="list-style-type: none"> • Electronic directories of specified target population (e.g., list of all AMA members) • Commercial organizations’ email directories • Targeted volunteers recruited via websites or electronic discussion groups
Questionnaire formatting	<ul style="list-style-type: none"> • Formatting may need to be different from mail surveys to simplify data entry and clarify possible responses • Variations in respondents’ computer hardware/software must be considered when planning visual design • A single scrolling page or multiple linked pages with questions can be used • The order in which respondents see the questions can be controlled • Respondents can be prevented from going back and changing previous answers • The number of responses can be limited to 1 per question
Administration and cover letter	<ul style="list-style-type: none"> • Surveys can be embedded into the body of an email; respondents are instructed to hit “reply,” fill out the survey and use “send” to return the survey • An email invitation/cover letter can be sent with a link to the survey website; a username and password can be used to ensure confidentiality and restricted access • Personalization of the email cover letter has been shown to increase response rates in some studies^{32,36} • Using the word “survey” in the email subject line has been shown to decrease response rates, whereas a blank subject line reportedly yielded higher response rates³² • Increases in response rates have been found with multiple reminders or mixed methods (i.e., follow up mailings or telephone interview)^{37,38}
Advantages	<ul style="list-style-type: none"> • Ability to conduct large-scale survey with lower costs than mail survey • Faster return time • Convenient (e.g., easy to administer surveys and send multiple reminders) • Responses can be immediately downloaded into a database, leading to reduced data entry errors and costs • Lower numbers of unanswered questions have been reported³⁷ • Steps can be taken to ensure confidentiality and anonymity (e.g., encryption methods) • Longer responses to open-ended questions have been reported³⁹
Disadvantages	<ul style="list-style-type: none"> • Decreased response rates compared with mail surveys leads to nonresponse bias and issues with representativeness of results⁴⁹ • Recipients may view surveys as junk mail, or institutions may have spam detectors or firewalls in place to filter or prevent unsolicited emails • Out-of-date or incomplete email lists • Various Internet browsers may display formatting differently • Some potential respondents may not be comfortable with Internet or may not have access • Technical difficulties (e.g., server failure) could result in losing responses⁴⁰

AMA = American Medical Association.

Pretesting is an important component in the development of a survey. This step involves the investigators presenting the questions as they would appear in the final draft of the questionnaire to a group of respondents representative of the sampling frame.⁴² The purpose of pretesting is to ensure that the survey meets certain criteria relevant to acceptability and administrative ease,⁴¹ avoidance of redundancy and poorly worded question stems and responses, and reasonable amount of time to complete the questionnaire.¹ Pretesting minimizes the chance that the respondents will misinterpret certain questions, fail to remember what was requested of them or in general answer questions in a way that misrepresents their intentions.⁴³

Sur and colleagues¹¹ used an Internet-based survey. The composition of the survey included an introductory letter from the Chair of the AUA Practice Guidelines Committee, which added credibility to the process. Some studies have shown that response rates can be affected by the “connectedness” of the respondent to the surveying organization; for example, higher response rates have been reported with the use of university envelopes or when the cover letter was signed by a well-known or senior person.³² However, having the Chair of the AUA Practice Guidelines Committee did not increase the response rate to the survey by Sur and colleagues, which was 8.8%. Bhandari and colleagues⁴⁴ found that the addition of a letter listing expert surgeons who endorsed their survey of orthopedic surgeon practice patterns led to a lower primary response rate than when no endorsement was used. These results highlight that using opinion leaders to endorse a survey may not always have a positive effect on response rate.⁴⁴

Sur and colleagues¹¹ did not provide detail on questionnaire development; there is no information concerning item generation, item reduction or pretesting. They stated “the instrument was developed based on previously described surveys of attitudes toward [evidence-based medicine], initially tested at a local continuing medical education event in print and subsequently adapted to a Web-based format.”^{45,46} We reviewed both the articles (Stapleton and colleagues⁴⁵ and McColl and colleagues⁴⁶) on which the authors modelled their survey and found that neither study describes how their survey questions were developed or mentions whether any pretesting was performed. Sur and colleagues¹¹ provide a copy of the survey they designed as an appendix, which allows the reader to review the wording and formatting of their questions. Sample questions from their survey using a Likert scale from 1 (strongly disagree) to 10 (strongly agree) included: “Practising evidence-based medicine improves patient care in urology,” “All of your surgical therapy decisions incorporate evidence-based medicine” and “Every urologist should be familiar with techniques for critical appraisal of studies.”¹¹

Secondary guides

Did the investigators perform clinical sensibility testing?

The goals of clinical sensibility testing are to assess the comprehensiveness, clarity and face validity of the questionnaire.¹ Sensibility testing ensures that survey questions are simple, easily understood and appropriate; it also identifies questions that are redundant or missing and determines how likely the questionnaire is to address the survey objective.¹ The testing is typically conducted independently by a number of assessors who rate the “sensibility” of the survey based on a series of direct questions. Sur and colleagues¹¹ did not explicitly report the conduct of clinical sensibility testing in their survey development. However, they were careful to report that their survey was adapted from previously conducted surveys, which may have included such analyses in their development.

Was there reliability and validity testing?

The essence of reliability testing in surveys is to ensure that questions discriminate among respondents. In other words, respondents’ answers to a given question are similar to those of respondents who feel the same or dissimilar to those of respondents who feel differently.²³ Reliability is related to the reproducibility of the results or test scores and is an “interaction among the instruments, the specific group of people taking the test and the test situation.”⁴⁷ Reliability is usually expressed as a ratio of the variability among individuals compared to the total variability in the scores.⁴⁷ There are a number of ways to measure reliability. The reproducibility of results can be measured across different times. With test–retest reliability, the same survey is given to the same respondent on 2 different occasions to see if results from the first test correlate with those from the second test. The same individual should provide consistent answers at different times.⁴⁷ Interrater reliability assesses the degree of agreement among different observers, whereas intrarater reliability measures the agreement between observations made by the same rater at 2 different times.⁴⁷ Internal consistency is the extent to which the results of different items correlate with each other. In other words, several different questions that propose to measure the same general construct should produce similar answers.⁴⁷ It is important to note that consistent results obtained from repeated administrations do not ensure that survey questions are measuring what we intend them to measure.

Validity is the extent to which an instrument/survey is measuring what was intended, and empirical evidence must be produced to determine validity.⁴⁷ Early theories of validity focused on showing that “a scale is valid,” but new conceptualizations of validity emphasize “the process whereby we determine the degree of confidence we can place on the inferences we make about people based on their scores from that scale.”⁴⁷ An instrument can be shown to be valid

in a specific group of people in the context in which it was tested. If a survey or instrument is being used with a different group of individuals in a different context, then the original validation study may not apply, and further piloting and validation is recommended.⁴⁷ Different approaches to validating an instrument or scale are outlined in Table 4. These can be separated into situations when other similar scales are available for comparison or when no other measure exists.⁴⁷ Validity also involves the process of specifying and evaluating proposed interpretations and uses of scores (see Kane⁴⁸ for a detailed discussion). For an in-depth discussion of reliability and validity, see Streiner and Norman,⁴⁷ as a full discussion is beyond the scope of this paper.

The article by Sur and colleagues¹¹ did not mention whether they assessed reliability or validity. Their survey was adapted from previously published surveys,^{45,46} and a review of these articles showed that no reliability or validity testing was reported. Caution should be exercised when interpreting the findings of Sur and colleagues because the reliability and validity of the instrument have not been evaluated.

Was the administration of the questionnaire appropriate?

There are a number of strategies to enhance the response rate of a survey questionnaire and these need to be considered a priori (see Sprague and colleagues⁴⁹ and Edwards and colleagues³² for more in-depth discussions of strategies to increase response rates to mail and Internet surveys). Advanced notices in professional newsletters or a mailed letter should announce the impending survey.²⁹ Survey questionnaires can be distributed by mail, email, Internet or fax. The administration method chosen will depend on the type of information desired, target sample size, investigator time and financial constraints.⁷ There is evidence from orthopedic surveys that

Internet respondents had a lower response rate than mail respondents (45% v. 58%).⁸ This evidence contradicts other studies in which the reverse has been observed.^{50,51} It is possible that a trade-off exists between cost and response rate (Internet administration is less costly than mail administration but has a lower response rate). Table 3 lists important considerations for implementing or evaluating Internet-based surveys.

The key considerations in determining whether the administration process was appropriate include the method of administration (email, telephone, Internet, fax, mail), the rigour in follow-up and decreasing nonresponse rates. The survey by Sur and colleagues¹¹ was administered on a customized page on the AUA website. The chair of the AUA sent an invitation letter on behalf of the association to all members informing them of the survey. The sampling frame was appropriate, and the Internet survey, in principle, was an efficient way to access a large body of urologists. Sur and colleagues¹¹ reported that only AUA members with a listed email address were contacted, therefore part of the population of interest was excluded from the survey (the percentage excluded was not reported). The choice of Internet-based survey administration was a good one, assuming that the surgeons contacted were responsive to email and that sufficient absolute numbers of surveys could be returned using the chosen method. Whereas Sur and colleagues¹¹ reported a poor response rate (8.8%), the absolute numbers returned (714 surveys) likely provided adequate data for analysis. The 13% email error rate could have been reduced if the invitation letter to AUA surgeons had been sent by email ahead of time; that way, necessary updates to email addresses could have been made before the survey was administered. Questionnaires administered by mail should also be pretested to confirm addresses and reduce the cost of administration. Interpretation of the survey results must include a consideration of whether nonresponders were uniquely different from responders (nonresponse bias).

What were the results?

What was the magnitude of the response rate?

High response rates increase the precision of parameter estimates, reduce selection bias and enhance validity.¹ As the response rate decreases, the likelihood that the characteristics of the respondents differ from those of nonresponders increases.²³ Therefore, the findings from a survey with a low response rate are less likely to be generalizable to the target population. The “actual response rate” reflects the sampling element (fully and partially completed questionnaires and opt-out responses), whereas the “analyzable response rate” (fully and partially completed questionnaires) reflects the percentage of the sampling frame.¹ Many investigators consider a response rate of 70% adequate for generalization to the target population,

Table 4. Different approaches to the validation of a survey or instrument⁴⁷

Types of validity testing	Process description
Content validation	The systematic examination, preferably by experts in the field, of the content of the survey questions to determine if they accurately assess all the different domains of the construct being measured. The higher the content validity, the broader the inferences that can validly be drawn about a person under a variety of conditions.
Construct validation	Abstract constructs, such as intelligence, cannot be directly observed and need to be operationally defined. Construct validation is an ongoing process of hypothesis testing whereby one develops and tests new predictions based on a construct.
Criterion validation	Survey or scale results are correlated with some other measure of the construct, ideally a “gold standard” that has been used and accepted in the field. One assesses how a person who scores at a certain level on a new scale does on some criterion measure.

though this may vary according to the purpose and nature of the study.²³ Some investigators consider a response rate between 60% and 70% (or less than 60% for controversial topics) acceptable.¹ The response rate for electronic (email or Internet) questionnaires has been shown to be lower than that for surveys administered by mail.^{50,51} Different methods are proposed to increase the response rate for electronically administered surveys.^{32,52}

Sur and colleagues¹¹ emailed 9319 members of the AUA asking them to complete a Web-based survey. Of 9319 emails, 1213 (13%) were returned undeliverable with an incorrect address. Of the 8100 delivered emails, 724 surveys were completed and 714 (8.8%) contained analyzable data. The fact that the survey was online for only 4 weeks and that no reminders were sent out for technical reasons might partly explain the low response rate. The investigators mentioned the low response rate of 8.8% as a limitation to their survey. The very low response rate raises the question of how representative the survey sample is of the AUA members and to what extent their findings are generalizable to that population. The investigators claimed that the characteristics of the survey responders, although selective, were similar to the AUA profile, which might indicate unbiased results and gives some legitimacy to the internal and external validity of the study. The investigators could have estimated sample size based on a defined sampling frame, eligibility criteria and objectives and then randomly drawn the desired sample size (with a conservative correction for nonresponders) from the AUA members list. They could have ensured their eligibility and email addresses before conducting the survey. In fact, in a subsequent study by the same investigators,¹⁸ an almost identical survey was conducted and dealt with some of these limitations. A similar evidence-based medicine survey with additional questions was sent out by mail to a random sample of 2000 AUA members. Weekly email reminders were sent to nonresponders, and a second copy of the survey was sent after 6 weeks. Nonresponders were also offered the opportunity to complete an Internet-based version of the survey. The response rate for this survey was 44.5%, demonstrating that certain techniques (i.e., using mixed methods and sending reminders and replacement surveys) can help to increase response rates.^{18,32,38}

Were appropriate statistical methods used?

Surveys can be descriptive or explanatory. Descriptive surveys synthesize and report the factual data with the goal of estimating a parameter (e.g., surgical residents' satisfaction with their residency programs). Explanatory surveys draw inferences between constructs and concepts to test a hypothesis and can explore several constructs at a time. Surveys can address 1 or more underlying constructs, such as an idea, attitude or measure. Surveys with 1 construct, such as an instrument to measure residents' knowledge, are unidimensional scales, and those measuring more than

1 construct, such as residents' knowledge and attitudes, are multidimensional scales.²³ Like other designs, surveys require sample size estimation (power analysis or precision analysis) a priori, and the research question, objectives, hypotheses and design inform the method of sample size estimation. Burns and colleagues¹ have provided some useful formulas of sample size estimation for descriptive and explanatory survey designs. The statistical methods used for data analysis must be based on the objectives of the survey and the characteristics used for power estimation, and they should be planned a priori. Sur and colleagues¹¹ conducted an explanatory survey and appropriately listed 2 hypotheses on the understanding of technical terms related to evidence-based medicine. They used univariate logistic regression to test their hypotheses, but they did not report performing an a priori power calculation. They might have decided that with a target of 9319 members, they would obtain more than enough completed questionnaires to answer the research question.

Was the reporting transparent?

The findings of a survey should address its objectives. They should be clearly and logically presented with appropriate tables and figures. The results should account for all respondents and represent information obtained from partially or fully completed questionnaires as a proportion of the sampling frame.¹ Sur and colleagues¹¹ presented the results of analyzable data accounting for all respondents. They appropriately presented the summary data on demographics and question responses in tables and figures with adequate explanation in the text. They found that the surgeons who were full-time academics or who completed training less than 10 years previous to the survey administration had a better understanding of technical terms related to evidence-based medicine.

Were the conclusions appropriate?

The impact of the nonresponse rate on the validity of the findings should be discussed in detail. Appropriate methods, such as multiple imputation, should be used to handle the missing data.^{53,54} The results from both the imputed data set and actual data set should be reported to measure the impact of the missing data. The discussion should succinctly summarize the results and state their implications. The findings from other similar studies should be interpreted, considered or refuted. The limitations of the study and their implications on the findings should be explained, and appropriate conclusions should be drawn accordingly. Sur and colleagues¹¹ appropriately summarized and interpreted their findings and discussed other interpretations with respect to their findings. They listed a number of study limitations but gave no explanation of the study power and its implication on their results. They did not discuss the reliability and validity of the tool used for their survey. They indicated that their conclusions must be

interpreted with caution and that further studies would be needed to answer their research question.

Will the results help me alter my practice?

Are the results generalizable to my practice?

The generalizability of the study is the extent to which we can apply the results from our sample to the whole population.¹ The survey by Sur and colleagues¹¹ had some serious limitations, including a low response rate (8.8%), which was most likely not representative of the AUA membership as nonresponders and members without listed email addresses (> 90% of sample) may have had different attitudes than responders. Furthermore, there was concern about the reliability and validity of the survey instrument as reported. Considering that the investigators themselves indicated in their conclusion that their survey results must be interpreted with caution, we question whether the findings of the survey are generalizable.

Will the information/conclusions from this survey help me change or improve my practice/behaviour?

Notwithstanding its methodologic limitations, the survey by Sur and colleagues found that surgeons who were full-time academic practitioners or who had completed their training less than 10 years previously were more likely to have better understanding of technical terms related to evidence-based medicine. Based on this finding we are more likely to adopt a prudent approach and familiarize ourselves with evidence-based medicine principles.

RESOLUTION OF THE CLINICAL SCENARIO

At the next urology rounds, you report your findings to the head of your division. You inform him that there were methodologic weaknesses to the survey, but you could not ignore that there was some evidence to support the view of the junior staff person. The professor, however, is not yet ready to chastise the senior surgeon; he wants better evidence. He therefore assigns you a new research project: carry out a new survey with better methodology and a higher response rate.

CONCLUSION

Results from physician surveys can provide useful information about knowledge, practice patterns, beliefs and attitudes that may help highlight research needs and inform the adoption of practice guidelines or resources/program implementation.⁸ Poorly designed surveys can produce inaccurate and misleading results. The present article provides surgeons with useful guides to critically assess the quality of a survey's design and the validity and generalizability of its results. We outlined the importance of survey question design, reliability and validity testing,

methods for maximizing response rates, and appropriate data analysis and interpretation.

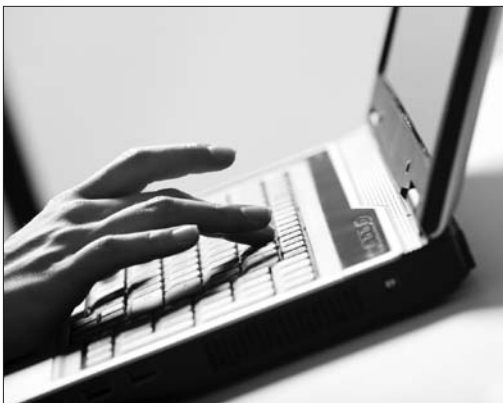
Competing interests: None declared.

Contributors: All authors designed the article, reviewed it and approved its publication. Drs. Thoma, Farrokhyar and Cornacchi acquired the data and wrote the article together with Dr. Goldsmith.

References

1. Burns KE, Duffet M, Kho M, et al. A guide for the design and conduct of self-administered surveys of clinicians. *CMAJ* 2008;179:245-52.
2. 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.
3. Archer KR, MacKenzie EJ, Bosse MJ, et al. Factors associated with surgeon referral for physical therapy in patients with traumatic lower-extremity injury: results of a national survey of orthopedic trauma surgeons. *Phys Ther* 2009;89:893-905.
4. Judge A, Welton NJ, Sandhu J, et al. Equity in access to total joint replacement of the hip and knee in England: cross sectional study. *BMJ* 2010;341:c4092.
5. Rabbie ER. *Survey research methods*. 2nd ed. Belmont (CA): Wadsworth Publishing Co.; 1990.
6. Ganz PA, Kwan L, Stanton AL, et al. Physical and psychosocial recovery in the year after primary treatment of breast cancer. *J Clin Oncol* 2011;29:1101-9.
7. 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.
8. Leece P, Bhandari M, Sprague S, et al. Internet versus mailed questionnaires: a controlled comparison (2). Corrected and republished from: *J Med Internet Res* 2004; 6:e39.
9. Mathes DW, Schlenker R, Ploplys E, et al. A survey of North American hand surgeons on their current attitudes toward hand transplantation. *J Hand Surg Am* 2009;34:808-14.
10. Chokshi NK, Simeone DM, Chari RS, et al. A survey of academic surgeons: work, stress, and research. *Surgery* 2009;146:462-8.
11. Sur RL, Scales CD, Preminger G, et al. Evidence-Based Medicine: a survey of American Urological Association Members. *J Urol* 2006; 176:1127-34.
12. Alderman AK, Hawler ST, Waljee J, et al. Correlates of referral practices of general surgeons to plastic surgeons for mastectomy reconstruction. *Cancer* 2007;109:1715-20.
13. Wang TS, Roman SA, Cox H, et al. The management of thyroid nodules in patients with primary hyperthyroidism. *J Surg Res* 2009; 154:317-23.
14. Godlee F. Milestones on the long road to knowledge. *BMJ* 2007;334 Suppl 1:s2-3.
15. Birch DW, Eady A, Robertson D, et al. Users' guide to the surgical literature: How to perform a literature search. *Can J Surg* 2003;46:136-41.
16. Dahm P, Preminger GM, Scales CD Jr, et al. Evidence-based medicine training in residency: a survey of urology programme directors. *BJU Int* 2009;103:290-3.
17. Scales CD Jr, Voils CI, Fesperman SF, et al. Barriers to the practice of evidence-based urology. *J Urol* 2008;179:2345-9.
18. Dahm P, Poolman RW, Bhandari M, et al. American Urological Association Membership. Perceptions and competence in evidence-based medicine: a survey of the American Medical Association Membership. *J Urol* 2009;181:767-77.
19. Thoma A, Tandan V, Sprague S. Users' guide to the surgical literature: How to use an article on economic analysis. *Can J Surg* 2001;44:347-54.
20. Hong D, Tandan V, Goldsmith CH, et al. Users' guide to the surgical literature: How to use an article reporting population-based volume-outcome relationships in surgery. *Can J Surg* 2002;45:109-15.

21. Thoma A, Farrokhyar F, Bhandari M, et al. Users' guide to the surgical literature: How to assess a randomized controlled trial (RCT) in surgery. *Can J Surg* 2004;47:200-8.
22. Huston P. Reporting on surveys: information for authors and peer reviewers. *CMAJ* 1996;154:1695-704.
23. Passmore C, Dobbie AE, Parchman M, et al. Guidelines for constructing a survey. *Fam Med* 2002;34:281-6.
24. Williams B. *A sampler on sampling*. New York (NY): John Wiley and Sons, Inc.; 1978.
25. Aday LA, Cornelius LJ. *Designing and conducting health surveys. A comprehensive guide*. 3rd ed. San Francisco (CA): Jossey-Bass; 2006.
26. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985;38:27-36.
27. Ehrlich A, Koch T, Amin B, et al. Development and reliability testing of a standardized questionnaire of assess psoriasis phenotype. *J Am Acad Dermatol* 2006;54:987-91.
28. Stone DH. Design a questionnaire. *BMJ* 1993;307:1264-6.
29. Dillman DA. *Mail and Internet surveys: the tailored design method*. 2nd ed. Hoboken (NJ): John Wiley & Sons; 2000.
30. Fox J. Designing research: basics of survey construction. *Minim Invasive Surg Nurs* 1994;8:77-9.
31. Henry RC, Zivick JD. Principles of survey research. *Fam Pract Res J* 1986;5:145-57.
32. Edwards PJ, Roberts I, Clarke MJ, et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev* 2009;8:MR000008.
33. Christian LM, Dillman DA. The influence of graphical and symbolic language manipulations on responses to self-administered questions. *Public Opin Q* 2004;68:57-80.
34. Dillman DA, Smythe JD. Design effects in the transition to web-based surveys. *Am J Prev Med* 2007;32:S90-6.
35. Turocy PS. Survey research in athletic training: the scientific method of development and implementation. *J Athl Train* 2002;37:S174-9.
36. Heerwegh D. Effects of personal salutations in e-mail invitations to participate in a web survey. *Public Opin Q* 2005;69:588-98.
37. McMahon SR, Iwamoto M, Massoudi MS, et al. Comparison of e-mail, fax and postal surveys of pediatricians. *Pediatrics* 2003;111:e299.
38. Beebe TJ, Locke GR, Barnes SA, et al. Mixing web and mail methods in a survey of physicians. *Health Serv Res* 2007;42:1219-34.
39. Seguin R, Godwin M, MacDonald S, et al. Email or snail mail? Randomized controlled trial on which works better for surveys. *Can Fam Physician* 2004;50:414-9.
40. 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.
41. Eysenbach G. Improving the quality of Web surveys: the checklist for reporting results of Internet surveys (CHERRIES). *J Med Internet Res* 2004;6:e34.
42. Bowden A, Fox-Rushby JA, Nyandieka L, et al. Methods for pre-testing and pilot-testing survey questions: illustrations from the KENQOL survey of health-related quality of life. *Health Policy Plan* 2002;17:322-30.
43. Collins D. Pre-testing survey instruments: an overview of cognitive methods. *Qual Life Res* 2003;12:229-38.
44. Bhandari M, Devreux PJ, Swiontkowski MF, et al. A randomized trial of opinion leader endorsement in a survey of orthopaedic surgeons: effect on primary response rates. *Int J Epidemiol* 2003;32:634-6.
45. Stapleton AMF, Cuncins-Hearn A, Pinnock C. Attitudes to evidence-based practice in urology: results of a survey. *ANZ J Surg* 2001;71:297-300.
46. McColl A, Smith H, White P, et al. General practitioners' perceptions of the route to evidence-based medicine: a questionnaire survey. *BMJ* 1998;316:361-5.
47. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th ed. Oxford: Oxford University Press; 2008.
48. Kane M. Validation. In: Brennan R, editor. *Educational measurement*. 4th ed. Westport (CT): American Council on Education and Praeger; 2006. p. 17-64.
49. Sprague S, Quigley L, Bhandari M. Survey design in orthopaedic surgery: getting surgeons to respond. *J Bone Joint Surg Am* 2009;91 (Suppl 3):27-34.
50. Kim HL, Hollowell CM, Pater RV, et al. Use of new technology in endourology and laparoscopy by American urologists: Internet and postal survey. *Urology* 2000;56:760-5.
51. Raziano DB, Jayadevappa R, Valenzuela D, et al. E-mail versus conventional postal mail survey of geriatric chiefs. *Gerontologist* 2001;41:799-804.
52. 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.
53. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York (NY): J. Wiley & Sons Inc.; 1987.
54. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods* 2002;7:147-77.



Canadian Journal of Surgery

**Online manuscript
submission and peer review**

AVAILABLE at

<http://mc.manuscriptcentral.com/cjs>