

Computerized Questionnaires and the Quality of Survey Data

Brett Hanscom, MS,* Jon D. Lurie, MD, MS,*†‡ Karen Homa, MS,† and James N. Weinstein, DO, MS*†‡

Study Design. A retrospective data quality analysis was conducted.

Objective. To compare missing response rates and internal consistency between computerized and paper surveys administered to spine patients.

Summary of Background Data. Computerized patient surveys have been shown to offer numerous advantages over traditional paper surveys. It has been assumed that computerized surveys also improve data quality, but quantitative comparisons have not been made.

Methods. Between January 1998 and December 2000, approximately 3500 computerized questionnaires and 15,000 paper questionnaires containing the MOS 36-Item Short-Form Health Survey (SF-36) and the Oswestry Low Back Pain Disability Questionnaire were administered in the National Spine Network. Missing response rates and the Response Consistency Index (RCI) were compared between computerized and paper questionnaire data.

Results. Computer surveys had approximately half the missing response rate of paper surveys. For the SF-36, the

computer survey had 1.7% missing, as compared with 3.3% missing on paper ($P < 0.001$). For the Oswestry, the computer survey had 2.9% missing, as compared with 6% missing on paper ($P < 0.001$). Whereas 84% of the SF-36 surveys and 85% of the Oswestry surveys collected by computer were completely filled out (no missing responses), only 68% of the SF-36 surveys ($P < 0.001$) and 77% of the Oswestry surveys ($P < 0.001$) collected on paper were completely filled out. The SF-36 data collected by computer had better internal consistency than the paper-form data, with average Response Consistency Index scores of 0.12 and 0.16, respectively ($P = 0.001$).

Conclusions. Superior response rates and higher internal consistency suggest that computerized survey systems improve data quality, and may enhance instrument validity for commonly used measures of spine patient health. [Key words: computerized surveys, data quality, NSN, Oswestry, SF-36] **Spine 2002;27:1797-1801**

From *Dartmouth Medical School, Hanover, †The Spine Center at Dartmouth Hitchcock Medical Center, Lebanon, and ‡The Center for Evaluative Clinical Sciences, Hanover, New Hampshire. Supported in part by NIAMS #AR45444-01 and the National Spine Network.

Acknowledgment date: August 31, 2001.

First revision date: December 6, 2001.

Acceptance date: February 4, 2002.

*Members of the National Spine Network: Alegant Health System/Nebraska Spine Surgeons, Omaha, NE; Alliant Health System/Kenton D. Leatherman Spine Center, Louisville, KY; Dartmouth Hitchcock Medical Center, Hanover, NH; Emory HEALTHCARE/Emory Spine Center, Atlanta, GA; Georgetown University Medical Center, Washington, DC; Hospital for Joint Diseases Spine Institute, New York, NY; Lakewood Orthopaedics, Denver, CO; Medical College of Wisconsin, Milwaukee, WI; Medical University of South Carolina, Charleston, SC; New England Baptist Hospital, Boston, MA; North Carolina Spine Center, Chapel Hill, NC; North Kansas City Hospital/Spine and Seeliosis Surgery, Ltd., Kansas City, MO; Rush-Presbyterian/St. Luke's Medical Center, Chicago, IL; University Hospital Spine Institute, Cleveland, OH; University of Iowa Hospitals and Clinics, Iowa City, IA; University of Missouri Health Sciences Center, Columbia, MO; The Hospital for Special Surgery, New York, NY; Thomas Jefferson University Hospital/Rothman Institute, Philadelphia, PA; University of California, San Francisco, San Francisco, CA; University of Pittsburgh Medical Center/University Orthopaedics, Inc., Pittsburgh, PA; University of Utah School of Medicine, Salt Lake City, UT; Washington University School of Medicine; Barnes-Jewish and St. Louis Children's Hospitals, St. Louis, MO; Vanderbilt University Spine Center, Nashville, TN; SUNY Health Science Center, Syracuse, NY; Tulane University Medical Center, New Orleans, LA; William Beaumont Hospital, Detroit, MI.

Device/drug statement: The submitted manuscript does not contain information about medical devices or drugs.

Conflict of interest: No funds were received in support of this work. One or more of the authors have received or will receive benefits for personal or professional use from a commercial party related directly or indirectly to the subject of this manuscript (e.g., honoraria, gifts, consultancies).

DOI: 10.1097/01.BRS.0000020303.08783.20

The advent of the microcomputer in the 1960s inspired psychologists and medical practitioners to explore the possibility of using computer-assisted patient interviews.^{7,12-16,18} Numerous advantages were foreseen, including increased efficiency of administration, reduction of data entry errors, improved ability to solicit sensitive or confidential information, elimination of paper storage, and the potential for automatic scoring and report generation.

Computerized surveys have now been installed and tested in a variety of clinical settings and have generally met expectations. As hoped, computerized systems have been found to reduce the burden of paper handling and data entry.^{3,9,22} Both paper storage and the time-consuming, error-prone process of data entry are eliminated, and the cost of data entry technicians is removed. Computer survey software also can immediately score patient responses, retrieve and compare information from previous encounters and external databases, and provide clinicians with detailed summary information and comparisons that were not previously available in a timely manner.²²

In addition, numerous reports suggest that patient acceptance of computer-based questionnaires is better than expected. Although not all patients appreciate computer-based surveys, the majority of patients find them easy to use, private, and generally comparable with paper-and-pencil surveys or face-to-face interviews.^{1,6,10,11,17,23} In fact, most patients report that they prefer computer-based surveys over paper surveys.^{3,5,8,9,23} Taenzer et al¹⁷ reported that patients' attitudes toward computer surveys improve after they complete a computer survey for

the first time. Contrary to some speculation, Yarnold et al²⁴ reported that geriatric patients are quite accepting of computer-aided assessment.

Not only can computerized surveys simplify the data collection process and provide an agreeable medium for patients, but they should also be able to improve data quality. In addition to eliminating data entry, computer-based surveys can, in theory, improve data quality in a variety of ways. Software can be programmed to prevent patients from marking multiple responses to the same question, or from marking inconsistent or illogical responses, both common errors with paper forms. Also, some authors speculate that displaying questions one at a time on a computer screen, as opposed to printing numerous questions on a single paper page, allows patients to focus more easily on individual questions and reduces the chance that they might skip items or become confused.^{5,22,23}

In effect, computerized survey technology can provide an efficient and desirable method for collecting data directly from patients. Therefore, it might be assumed that the resulting data should be of better quality than data collected with paper. The purpose of this study was to quantify any differences in data quality that may exist between a computerized and a paper-based data collection process.

■ Methods

The National Spine Network Health Status Survey. The National Spine Network is a consortium of spine care centers across the United States that collaborate in collecting outcomes data on their patients.* The NSN Health Status Survey consists of both a patient and a clinician component. The patient component contains the MOS 36-Item Short-Form Health Survey (SF-36) and the Oswestry Low Back Pain Disability Questionnaire, as well as questions regarding demographics, symptom history, and comorbid illness. Clinicians report the clinical symptoms, the diagnosis, and a treatment plan. Participating clinicians administer the NSN Health Status Survey to patients when they arrive at a clinic for the first time, and subsequently at follow-up visits. The NSN was formed in 1995, and as of January 2001, the outcomes database contained nearly 70,000 records. Data collection is ongoing.

Data Collection Methods. The NSN Initial Visit Health Status Survey is an eight-page questionnaire consisting mostly of multiple choice questions with “bubble” responses, and a limited number of write-in responses for dates, height, weight, and the like. Once completed, the paper surveys are mailed to a central data repository, where each survey is double entered by hand into a database. Questions with no visible response and those with more than one response are recorded as “missing.”

The DHMC Spine Center, an NSN member in Lebanon, New Hampshire, began using a computerized version of the same questionnaire in 1998.²² The questions are identical, and they are administered in the same order as the questions on the paper survey. Patients use a portable touch-pad computer (small PC laptop with a touch-sensitive monitor and no keyboard) to complete the questionnaire in the clinic waiting room. The questions are presented one at a time on the screen

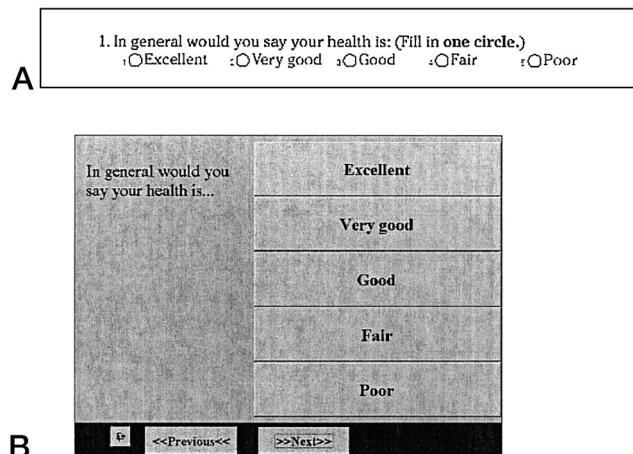


Figure 1. Question format on paper survey (A) and computer questionnaire (B).

in a large font with large, push-button responses (Figure 1.) For each question, the patient has the option either to choose a response from the list of available responses or to press the “next” button and proceed to the next question. No questions are required. Completed surveys are loaded into a local database. A summary report is printed for the clinician and patient to review during the encounter, and the survey records are sent electronically to the NSN data repository.

Study Cohort and Missing Value Rates. Between January 1998 and December 2000, 14,815 paper-survey records and 3574 computer-survey records were added to the NSN database. All the computer-survey records came from the DHMC Spine Center, and all the paper-survey records came from other NSN centers. All the available surveys, both computer and paper versions, were included in the analysis. Although the NSN questionnaires contain a variety of additional questions, missing value data only from the well-known SF-36 and the Oswestry questionnaires were evaluated. Missing response rates (*i.e.*, the number of missed questions divided by the total number of questions) were calculated for each of the two instruments.

The SF-36 and the Oswestry. The SF-36 is a general health status instrument designed to measure overall health.^{20,21} The 36 questions contained in the survey are scored and reduced to eight scales, each measuring a different domain of physical or mental health: bodily pain, physical function, role physical, vitality, mental health, social function, role emotional, and general health. If questions are skipped, the scales still can be computed as long as a majority of the questions that make up that scale have been answered. These eight scales can be further reduced into two summary scales, the physical component summary (PCS) and the mental component summary (MCS), that measure overall physical and mental health.¹⁹

The Oswestry is a low back-specific questionnaire designed to assess the disabling effects of low back pain.⁴ All the questions on the Oswestry may be scored and reduced to a single scale: the Oswestry Disability Index (ODI). This scale, like those of the SF-36, may be computed as long as the patient has responded to a majority of the survey questions.

Table 1. Patient Characteristics

	Computer Survey (n = 3574)	Paper Survey (n = 14,815)	P Value*
Age (mean)	50.3	49.2	0.000
Gender (% female)	48.7	53.0	0.000
Education (% high school graduate)	54.2	62.0	0.000
Work status (% working)	47.2	45.8	0.144
Compensation (% receiving)	8.2	8.4	0.719

* Small P values likely reflect the large sample size more than any clinically meaningful differences.

The Response Consistency Index. The authors of the SF-36 Health Survey have developed a measure of internal consistency known as the Response Consistency Index (RCI) (Ware JE: Response Consistency Checks of SF-36 Items, 1993, unpublished data). Fifteen pairs of questions from the SF-36 survey were chosen, and “inconsistent” responses to these questions were defined. The RCI is simply the number of inconsistent responses to the selected question pairs. For example, an inconsistent pair of responses would be “no limitation walking more than a mile” and “severe limitation walking one block.” The RCI ranges from 0 (no inconsistencies) to 15 (15 pairs of inconsistent responses).

Statistical Analysis. All statistical analyses were performed using Stata 6.0 (Stata, College Station, TX). Rates and proportions were compared using Pearson's χ^2 test, and means were compared using Student's *t* test. Multivariate logistic regression was used to identify patient characteristics that related to missing responses.

■ Results

Patients completing computer surveys were similar in age to patients completing paper surveys (50 years old *vs.* 49 years old). They also were similar in terms of gender (49% female *vs.* 51% female) (Table 1). (The small differences in age and gender between the two study groups were statistically significant primarily because of the large sample size.) The patient's completing computer surveys were somewhat less likely to have completed high school (54% *vs.* 62%; $P < 0.001$), but were just as likely to be working (47% *vs.* 46%; $P = 0.144$) and to be receiving worker's compensation (both approximately 8%; $P = 0.719$).

The overall missing value rate for the computer surveys was approximately half the missing value rate for the paper surveys (Table 2). The computerized SF-36 had 1.7% missing item responses, as compared with 3.3% missing responses on paper ($P < 0.001$), and the computerized Oswestry had 2.9% missing item responses, as compared to 6% missing responses on paper ($P < 0.001$). Approximately 85% of both the SF-36 and Oswestry surveys were completely filled out by patients using the computer, whereas only 77% of the Oswestry surveys and 68% of the SF-36 surveys were completely filled out by patients using the paper version.

The SF-36 scales and the ODI may be computed only if a survey has responses for more than one half of all the

items that make up the scale. Approximately 1% of the eight standard SF-36 scales and the ODI could not be calculated for the computer surveys (Table 2), as compared with approximately 2% to 3% of the scales for the paper surveys ($P < 0.001$). The SF-36 summary scales, PCS and MCS (which require scores for each of the eight standard scales), could not be calculated for 3% of the computer surveys and nearly 8% of the paper surveys ($P < 0.001$).

For both the paper and computer surveys, the Oswestry showed notably higher missing response rates than the SF-36. Much of this difference was caused by the high number of missing responses on the final question of the Oswestry survey: “How has pain affected your sex life?” This was the only question related to potentially sensitive and confidential information, and it was missing on 9% of computer surveys and 13.5% of paper surveys. Excluding this question, overall missing response rates for the Oswestry would have been 2.1% for the computer survey and 5.1% for paper survey.

According to the RCI, fewer inconsistent responses were encountered in the SF-36 data collected *via* computerized survey than in the SF-36 data collected *via* paper. The average RCI score for the computerized SF-36 was 0.12, as compared with 0.16 for the paper version ($P = 0.001$). Slightly more than 91% of the computer surveys had no inconsistent responses (RCI = 0), as compared with 90% of paper surveys. It should again be emphasized that these differences are statistically significant, primarily because of the large sample size. The computer survey appears to produce only slightly better results with respect to internal consistency than the paper survey.

Because differences were observed between patients surveyed by computer and those surveyed by paper (Table 1), multivariate logistic regression modeling was used

Table 2. Overall Missing Value Rates and Missing Score Rates*

	Computer Survey (n = 3574) Percent Missing	Paper Survey (n = 14,815) Percent Missing
Overall missing		
SF-36	1.66	3.34
Oswestry	2.89	6.01
Missing scores		
Bodily pain	0.73	2.15
Physical function	0.64	1.49
Role physical	0.81	2.59
General health	0.95	2.54
Vitality	0.81	1.87
Role emotional	1.93	3.94
Social function	0.67	1.36
Mental health	0.98	1.93
Physical component summary	3.13	7.94
Mental component summary	3.13	7.94
Oswestry disability index†	1.01	3.18

* Missing response rates are all significantly larger for paper than for computer ($P \leq 0.001$).

† Assuming that at least 5 of 9 Oswestry questions must be answered in order to calculate the Oswestry disability index.

to adjust for these potentially confounding factors. Older patients, women, and patients with less than a high school education were more likely to have missing and inconsistent data. Because patients completing the computer survey were older and less educated, adjustment for age and education actually enhanced the relation between the survey media and the missing response rates. The same was true for the relation between the survey media and response consistency. Patient gender contributed only slightly to the models. The differences between adjusted and unadjusted rates were very small. Therefore, only the unadjusted rates are presented in Table 2.

■ Discussion

The results of this study demonstrate that missing and inconsistent data are relatively rare in both paper and computer survey systems, but computerized surveys can improve on data completeness and consistency. There are several obvious differences between the two data collection systems that may explain the computerized survey's superior performance. The computer software displays only one question at a time, using the full computer screen, and the patient must respond to the question or press the "next" button to move on. Therefore, it is harder to skip a question accidentally than it would be with a paper page containing 20 printed questions.

Also, computer survey questions are presented in a sizable font, with large, clearly delineated buttons for each response, making questions easier to read for persons with visual difficulties. Patients using paper surveys may skip a question because of the small font size, or because it is unclear which "bubbles" represent the appropriate responses. If this were the only important difference between paper and computer surveys, using a paper survey with one question printed on each page in a large font would generate the same results as the computer survey. Such an approach certainly is not practical. A variety of more reasonable strategies for optimizing paper survey design are discussed by Dillman,² and if used appropriately, may improve the performance of the paper survey.

Finally, and perhaps most importantly, computer survey systems eliminate the need for data entry technicians. On a paper form, patients can easily mark more than one response to a question, or make a mark halfway between two response bubbles. Data entry technicians cannot interpret such responses, so they must record the data as "missing." Also, despite the use of a double-entry system (two technicians independently entering the same survey, with errors reconciled), there still is a chance that data entry errors will be made. Using the computer survey system, on the other hand, the patient serves, in effect, as the data entry technician, and is allowed to mark only one response per question, unless more than one response is appropriate. Thus it may be that patients do not skip more questions on a paper survey, but rather, the computer prevents patients from marking multiple responses or "creative" responses and eliminates discrep-

ancies between patients' intended responses and the data entry technicians' key entries.

Two common strategies for increasing response rates in computerized surveys were not used in the current study: requiring responses to all questions and re-asking skipped questions. Requiring patients to respond to each and every question may be undesirable because some patients may not feel comfortable with specific questions, or may not understand them. However, these obstacles seem to be rare in the assessment of general health status and pain. Discomfort with survey questions is more likely to occur when patients are surveyed about sensitive topics such as sexual history or substance abuse, as exemplified by the low response rates on the Oswestry sex-life question.

Computerized questionnaires that re-ask missed questions at the end of a survey may also be problematic. Although this method has proved to be effective in reducing missing data, it is unclear how this distinctive presentation of questions might affect patient responses, and subsequently the final scoring of the instrument.^{1,5} It seems unlikely that the effect would be large, but such issues should be explored.

It is important and encouraging to note that although paper did not serve as well as the computer in this study, missing response rates and inconsistent responses were not particularly high for either media. With missing response rates between 3% and 6%, it is unlikely that missing information compromises the validity of these instruments. In addition, with not a single inconsistency in 90% of all the SF-36 surveys, there does not seem to be any reason for concern regarding survey validity in this population. Statistical significance levels for the results presented in Tables 1 and 2 should be interpreted with caution given the large sample size in this study population.

Incremental gains in data quality may be possible by further enhancing the survey software. After marking an inconsistent pair of responses on the SF-36, for example, a patient could be notified and given the chance to respond to these questions again. Such a restructuring of the survey would require reevaluation of the survey's psychometric properties, but it could potentially eliminate many patient errors.

The results of this study do not necessarily suggest that it is either important or cost effective to use computer-based surveys in all circumstances. The absolute gains in terms of missing data are fairly small, only 1% to 4%, whereas the cost of computer hardware, software, and technical support can be quite high. The cost effectiveness of implementing a computerized survey likely depends on a variety of factors, including patient volume, project scale, and system reliability. Investing in a computer and an electronic survey for a small study of 50 patients, for example, may not be worth the added gains in data quality. However, a computer survey intended for permanent clinical use or a large-scale clinical trial

may be cost effective and may offer substantial data quality and data management advantages.

In summary, the results of this study suggest that although both paper and computer surveys generate good data with few missing responses, computerized surveys produce data with fewer missing values and better internal consistency than data collected from paper. This benefit, along with the variety of advantages associated with computer survey systems, makes computer surveys an attractive alternative to traditional paper-and-pencil surveys.

■ Key Points

- Computerized surveys have been used effectively to collect data from patients, and they offer many advantages over traditional paper and pencil surveys.
- To compare data quality between computer and paper surveys, approximately 3500 computer surveys and 15,000 paper surveys from the National Spine Network (NSN) database were evaluated.
- For two commonly used measures of health status and functioning among spine patients (the SF-36 and the Oswestry), computerized survey systems generated data with significantly fewer missing values and resulted in significantly more surveys with no missing data than traditional paper surveys.
- The SF-36 data collected *via* computer had better internal consistency, as measured by the Response Consistency Index, than SF-36 data collected *via* paper and pencil.

References

1. Buxton J, White M, Osoba D. Patients' experiences using a computerized program with a touch-sensitive video monitor for the assessment of health-related quality of life. *Qual Life Res* 1998;7:513-19.
2. Dillman DA. *Mail and Internet Surveys; The Tailored Design Method*. New York: Wiley, 2000.
3. Drummond HE, Ghosh S, Ferguson A, et al. Electronic quality-of-life questionnaires: A comparison of pen-based electronic questionnaires with conventional paper in a gastrointestinal study. *Qual Life Res* 1995;4:21-6.
4. Fairbank JCT, Davies JB, Couper J, et al. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271-3.
5. Honaker L, Harrell T, Buffalo J. Equivalency of microtest computer MMPI administration for standard and special cases. *Comput Hum Behav* 1988;4:323-37.
6. Locke SE, Kowaloff HB, Hoff RG, et al. Computer-based interview for screening blood donors for risk of HIV transmission [see comments]. *JAMA* 1992;268:1301-5.
7. Maultsby MC Jr, Slack WV. A computer-based psychiatry history system. *Arch Gen Psychiatry* 1971;25:570-2.
8. Millstein SG, Irwin CE Jr. Acceptability of computer-acquired sexual histories in adolescent girls. *J Pediatrics* 1983;103:815-19.
9. O'Connor KP, Hallam RS, Hinchcliffe R. Evaluation of a computer interview system for use with neuro-otology patients. *Clin Otolaryngol Allied Sci* 1989;14:3-9.
10. Pouwer F, Snoek FJ, van der Ploeg HM, et al. A comparison of the standard and the computerized versions of the Well-Being Questionnaire (WBQ) and the Diabetes Treatment Satisfaction Questionnaire (DTSQ). *Qual Life Res* 1998;7:33-8.
11. Skinner HA, Allen BA. Does the computer make a difference? Computerized *versus* face-to-face *versus* self-report assessment of alcohol, drug, and tobacco use. *J Consult Clin Psychol* 1983;51:267-75.
12. Slack WV, Hicks GP, Reed CE, et al. A computer-based medical history system. *N Engl J Med* 1966;274:194-8.
13. Slack WV, Peckham BM, van Cura LJ, et al. A computer-based physical examination system. *JAMA* 1967;200:224-8.
14. Slack WV, van Cura LJ. Computer-based patient interviewing. *Postgrad Med* 1968;43:115-120.
15. Slack WV, van Cura LJ. Computer-based patient interviewing. 1. *Postgrad Med* 1968;43:68-74.
16. Slack WV, van Cura LJ. Patient reaction to computer-based medical interviewing. *Comput Biomed Res* 1968;1:527-31.
17. Taenzer PA, Specca M, Atkinson MJ, et al. Computerized quality-of-life screening in an oncology clinic. *Cancer Pract* 1997;5:168-75.
18. van Cura LJ, Slack WV, Frey SR. Elements of a computer medical interview system. *Biomed Sci Instrum* 1971;8:33-42.
19. Ware J, Kosinski M, Keller S. *SF-36 Physical and Mental Health Summary Scales: A User's Manual*. Boston: Health Assessment Lab, 1994.
20. Ware J, Snow K, Kosinski M, et al. *SF-36 Health Survey: Manual and Interpretation Guide*. Boston: The Health Institute, 1993.
21. Ware JE Jr, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;30:473-83.
22. Weinstein JN, Brown PW, Hanscom B, et al. *Designing an Ambulatory Clinical Practice for Outcomes Improvement: From Vision to Reality*. The Spine Center at Dartmouth-Hitchcock, Year One. Quality Management in Health Care 8. Frederick, MD: Aspen Publishers, 2000.
23. White D, Clements C, Fowler R. A comparison of computer administration with standard administration of the MMPI. *Comput Hum Behav* 1985;1:153-62.
24. Yarnold PR, Stewart MJ, Stille FC, et al. Assessing functional status of elderly adults *via* microcomputer. *Percept Mot Skills* 1996;82:689-90.

Address reprint requests to

Brett Hanscom, MS
The Spine Center at DHMC
One Medical Center Drive
Lebanon, NH 03756

E-mail: Brett.Hanscom@Dartmouth.Edu