Validation Study of an Electronic Method of Condensed Outcomes Tools Reporting in Orthopaedics

Jack Farr, MD[1] Nikhil Verma, MD[2] Brian J. Cole, MD MBA[2]

[1] Department of Orthopaedics, Indiana Orthopaedic Hospital, Indianapolis, Indiana
[2] Department of Orthopaedics, Rush University Medical Center, Chicago, Illinois

Address for correspondence Jack Farr MD, OrthoIndy, 1260 Innovation Pkwy, Greenwood, IN 46143 (e-mail: indyknee@hotmail.com).

J Knee Surg

Abstract

Patient-reported outcomes (PRO) instruments are a vital source of data for evaluating the efficacy of medical treatments. Historically, outcomes instruments have been designed, validated, and implemented as paper-based questionnaires. The collection of paper-based outcomes information may result in patients becoming fatigued as they respond to redundant questions. This problem is exacerbated when multiple PRO measures are provided to a single patient. In addition, the management and analysis of data collected in paper format involves labor-intensive processes to score and render the data analyzable. Computer-based outcomes systems have the potential to mitigate these problems by reformatting multiple outcomes tools into a single, user-friendly tool. The study aimed to determine whether the electronic outcomes system presented produces results comparable with the test–retest correlations reported for the corresponding orthopedic paper-based outcomes instruments.

The study is designed as a crossover study based on consecutive orthopaedic patients arriving at one of two designated orthopedic knee clinics.

Patients were assigned to complete either a paper or a computer-administered questionnaire based on a similar set of questions (Knee injury and Osteoarthritis Outcome Score, International Knee Documentation Committee form, 36-Item Short Form survey, version 1, Lysholm Knee Scoring Scale). Each patient completed the same surveys using the other instrument, so that all patients had completed both paper and electronic versions. Correlations between the results from the two modes were studied and compared with test–retest data from the original validation studies.

The original validation studies established test–retest reliability by computing correlation coefficients for two administrations of the paper instrument. Those correlation coefficients were all in the range of 0.7 to 0.9, which was deemed satisfactory. The present study computed correlation coefficients between the paper and electronic modes of administration. These correlation coefficients demonstrated similar results with an overall value of 0.86.

On the basis of the correlation coefficients, the electronic application of commonly used knee outcome scores compare variably to the traditional paper variants with a high rate of test–retest correlation. This equivalence supports the use of the condensed

Keywords

- computer-based
- patient-reported outcomes
- IKDC
- responsiveness
- reliability

received
February 13, 2013
accepted
April 7, 2013

DOI http://dx.doi.org/10.1055/s-0033-1347361.
ISSN 1538-8506.

electronic outcomes system and validates comparison of scores between electronic and paper modes.

A growing need for patient-reported outcomes (PROs) has emerged within the last two decades, providing both complementary and independent objective counterparts to physical examination and physician-reported results.[1,2] In fact, PROs are often considered to be more meaningful measures of outcomes than clinician-reported outcomes (CROs).[3–5] More recently, PROs have a newly assigned role in health care reform with rewards attached for those physicians and entities who report them, as in the pay for performance movement.[6,7] In addition, PROs address the need for transparency, meaningful use, and quality-of-life (QOL) measures. The value of PROs is therefore multifaceted.

PROs may be categorized by disease (or in orthopaedics, by affected structure or pathology) or they may address broader QOL measures. As the terms suggest, disease-specific instruments evaluate change only in the affected system, whereas general health-related QOL outcomes tools (HRQOL) compare QOL before and after treatment but can be impacted by other concurrent disease processes.[1,2] The advantage to having both disease-specific and HRQOL measures is that of determining outcome with regard to the specific joint, and potentially identifying the impact of disease improvement on the patients' quality of life. Furthermore, there is an added benefit of increased specificity with the use of additional disease-specific PRO measures.[2] In instances when a given disease process does not have a specific PRO, researchers will often use several overlapping PROs to capture potential clinically relevant changes. The disadvantage, however, of more comprehensive measurements is that the acquisition and management of large amounts of data are potentially a burden for patients, clinicians, and researchers.

The collection of multiple PRO assessments requires high resource utilization. The paper-based process may include mailings of multiple instruments to patients, verification of patient compliance when the patient arrives for their appointment (or loss of data points if verification is not concomitant with the visit), and a manual scoring and documentation of the instruments. This process is cumbersome for both the researcher and patient. In addition, as many PRO assessments have the same or similar questions, patients may answer the same question multiple times. A patient completing numerous instruments may experience cognitive overload, or "form fatigue," which may result in less accurate data collection and an increased perception of dissatisfaction with their clinical experience. An alternative is to have PRO-related initiatives begin with a patient-centric approach.

An electronic PRO electronic data system reduces this burdensome increase in questionnaires and requires fewer resources for delivery and verification. The assumption that a standard instrument can be transformed to a more efficient form without altering its psychometric characteristics requires careful consideration and scrutiny. The design of this investigation is to evaluate a set of subjects who are administered both the standard (paper-based) forms and the condensed electronic outcomes questionnaire. These results will be compared directly to test–retest findings from the original validation studies of the paper instruments (which have been expressed as a correlation coefficient between two paper administrations for the questionnaires studied here). If the condensed electronic and the paper versions correlate similarly (no significant differences), then the condensed electronic administration may be deemed to be as reliable as the paper method. Moreover, the condensed electronic version thereby inherits the other validity evidence collected for the standard "expanded" paper version. This purpose of this study was to evaluate a condensed electronic version of four PROs, three knee-specific and one HRQOL and correlate to traditional paper (*expanded version*) results. Our hypothesis is that no difference will exist in the test–retest correlation between paper and electronics PROs.

Methods

This study was designed to address the question of whether a system of reformatted (condensed) electronic capture of PRO data from standard PRO instruments could reliably be used interchangeably with paper-based collection (each PRO presented in its entirety). The specific electronic capture database system under evaluation is OBERD (Outcomes Based Electronic Research Database) (OBERD, Columbia, MO). It was selected for this study specifically because of its capability to simultaneously populate multiple outcomes tools data fields to achieve the goals of improved patient and clinician satisfaction with the PRO process.[8] The presentational methods of OBERD included psychometrically optimized screen colors, screen formatting, reminders to complete skipped questions (all questions have to be answered to complete the PRO), answered questions rolling off the screen, and the condensed question format. These factors and the reformatting represented a significant format departure from the paper mode.

The issues of greatest concern for this study revolve around whether modification of the original format of the outcomes tools in both adaptive features (i.e., question presentation is dynamically modified in light of patient answers.) and cognitive features (i.e., question presentation uses visual elements designed to improve usability and reduce the cognitive load for the patient.) impact the overall validity of the responses (►**Fig. 1**). These techniques are especially important when the researcher wishes to combine several standard instruments. As an example, if a question and its allowed answers are identical between two instruments, the patient only sees the question once (►**Table 1**). If only some of the answers are the same, then follow-up information may be needed for the incompatible answers. If the questions are similar, but not identical, then a compatible rewording may be presented. The final result is a separate score for each of the distinct

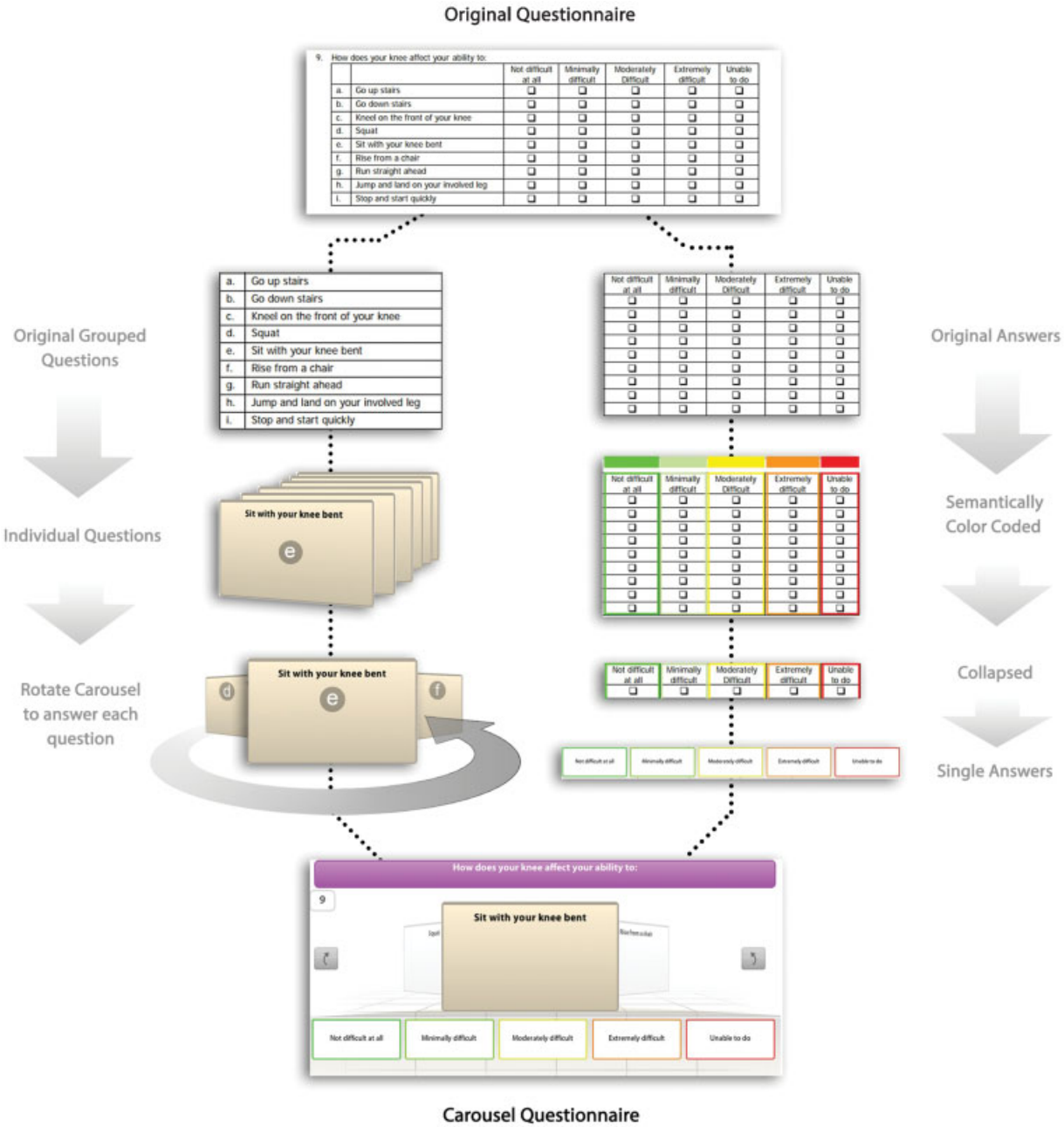

Fig. 1 OBERD strategies to condense questions while maintaining outcomes tool integrity. OBERD, Outcomes Based Electronic Research Database.

Table 1 Cumulative Effect of OBERD Condensation Methods[a]

Presentation method	Number of questions
Paper	105
OBERD: Elimination of repetitions	92
OBERD: Adaptive methods applied	87–92
OBERD: Cognitive load reduction applied	61–66 (user perception)

Abbreviation: OBERD, Outcomes Based Electronic Research Database.
[a]Adaptive methods are dynamic and vary with individual answers.

instruments. These scores are compared with results from the original paper versions of the individual instruments.

The Knee injury and Osteoarthritis Outcome Score (KOOS), subjective International Knee Documentation Committee form (IKDC), 36-Item Short Form, version 1 (SF-36 v1), and the Lysholm Knee Scoring Scale (Lysholm) outcomes tools were combined in this OBERD application to construct a reformatted instrument, which generated separate scores for the KOOS, IKDC, SF-36 v1, and Lysholm to be compared with their respective paper scores.[9–12] (►**Fig. 2**). The study was approved by the Institutional Review Board at each institution (Indianapolis Orthopaedic Hospital and Rush University Medical Center). Patients were informed of a stipend given upon completion of both forms, advised that their participation could be withdrawn at any time, and gave verbal consent if they agreed and wished to participate. All new

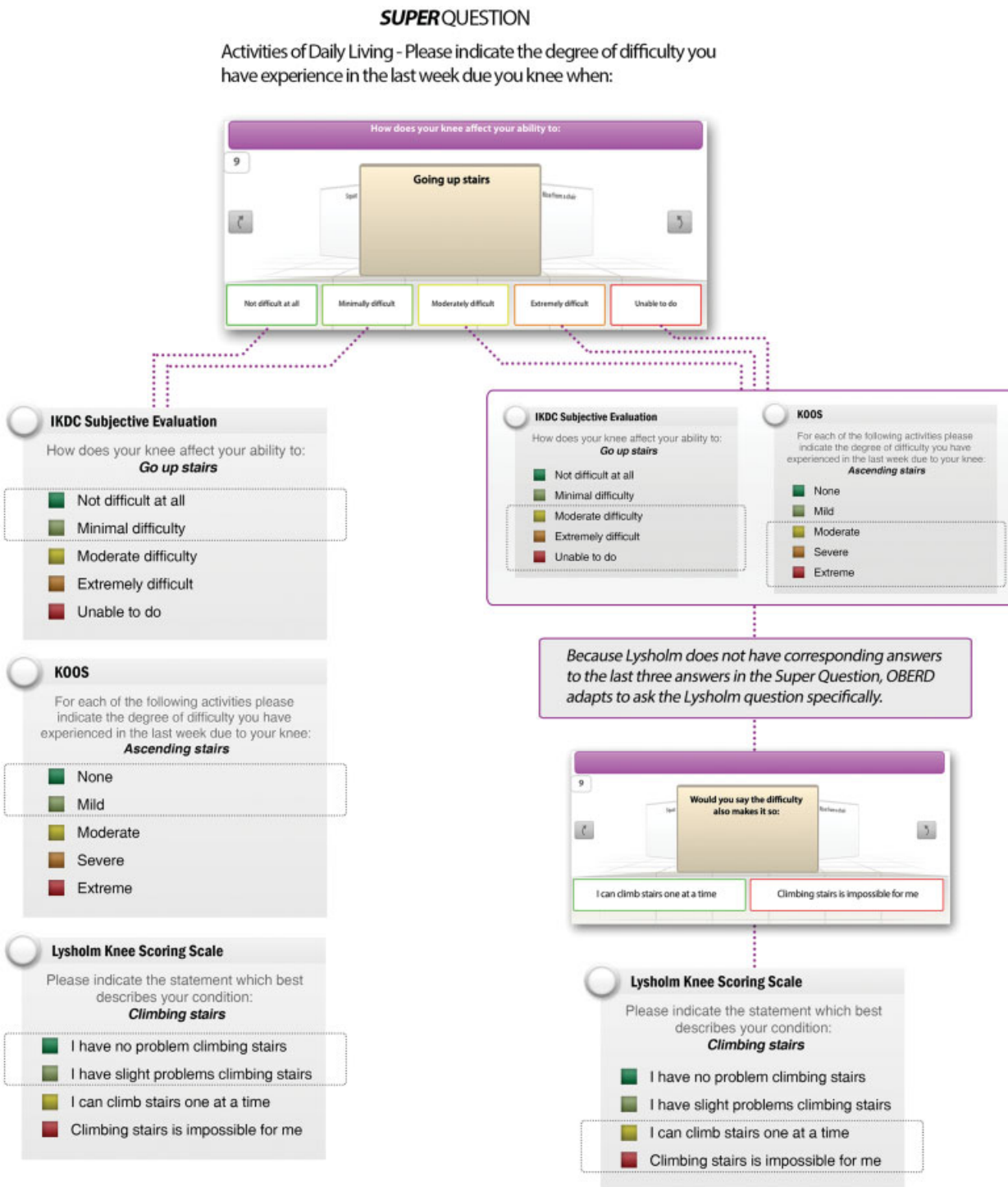

Fig. 2 Condensation of questions achieved using electronic format: IKDC + KOOS + Lysholm + SF-36 v1. IKDC, International Knee Documentation Committee form; KOOS, Knee injury and Osteoarthritis Outcome Score; Lysholm, Lysholm Knee Scoring Scale; SF-36 v1, 36-Item Short Form, version 1.

patients presenting to the orthopedic clinics (Institution A, Institution B) with a primary knee problem were included in the study. Exclusion criteria included (1) any condition that severely limited the subject's ability to complete the paper-based or enhanced questionnaires (e.g., blindness, physical disability, etc.), (2) the patient being younger than 18 years of age, (3) those patients scheduled to receive some type of treatment or medical intervention (therapy, injection, etc.) between the first and second tests or any patient who sustained an additional injury between taking the first and second tests, and (4) patients unwilling to participate.

Patients were identified as likely to meet the inclusion criteria after the research staff reviewed the patient list prior to the patient's clinic visit. Patients were contacted over the phone before the office visit and offered entry into the study. Alternatively, patients were interviewed by the research staff on the day of their visit and those who qualified were offered entry into the study.

Patients who could not be contacted prior to their office visit were first introduced to the study on the day of their office visit by the research coordinators. Those who agreed to participate were consented and completed either the paper-based version or the computer-based version before seeing the physician. The remaining version was completed after their office visit.

Once both version were completed, stipends were given to the participants. Paper-based instruments were then manually entered into the OBERD system for automatic scoring and comparison to the individual scores extracted from the electronic version, which were calculated in real time when the patient completed the electronic version.

Testing Methodology

This procedure constitutes a crossover experimental design. The paired observations should exhibit a high correlation, comparable with the test–retest correlations reported for the paper instrument if the research question posed is answered in the affirmative. Correlation coefficients were the only measure previously reported for all of the instruments of interest that had bearing on their inherent repeatability. Hence, the correlation coefficients were the only measure available for quantitative comparison in this study. Percentage correlation between electronic and paper versions were calculated and reported for each individual PRO measure.

Results

A total of 47 participants from Institution A and 55 participants from Institution B filled out the instruments. Some individual instruments could not be scored because of a violation of the protocol specified for the particular instrument, typically failure to sufficiently answer many of the questions. In this case, the specific instrument that was incomplete was removed from the analysis (►**Table 2**).

Because of the various subscales that they contain, the instruments used in this study provided a total of 17 different scores for comparison, resulting in a total of 1,638 sets of scores for comparing electronic and paper-based results. One of these scores, the SF-36 overall average, is not recommended for clinical use by the instrument's creators because the SF-36 is overtly multidimensional. However, it is included here because it speaks to overall correlation.

The individual correlation percentage for each individual score is presented in ►**Table 2**. The range that is usually considered to show adequate correlation is a correlation coefficient of more than 0.70. The test–retest results found in the literature for these instruments are provided in ►**Table 2**. Altogether, with the electronic-paper correlation coefficients obtained in the present study, the overall correlation coefficient for all 1,638 pairs of scores was 0.86.

Discussion

The results of this study demonstrate that the instruments tested through the OBERD electronic outcomes questionnaire system achieved levels of correlation that would be considered adequate on test–retest measurement, and overall the results are comparable with the correlation coefficients historically reported for the individual scores administered via paper. Test–retest correlation coefficient is generally considered to represent the inherent reliability of an instrument in validation studies. Its difference from 1.0 is attributed to random factors that cannot be eliminated without protocol or instrument changes. Thus, the correlations reported argue that electronic methods cannot be distinguished from their paper-based ancestors because they fall within the inherent error range of each instrument.

In response to the evolving need for development of electronic testing methods, the ISPOR (International Society for Pharmacoeconomics and Outcomes Research) created a task force to provide recommendations for validation of PRO measures transcribed to an electronic format.[13] The task force has recommended equivalency testing when moderate modifications have been made to the written format. This includes changes in item wordings or changes in mode of administration that may involve differing cognitive processes. Both were applied in this study in providing a more user-friendly interface for completion of outcome measures and for condensing the same or similar questions into one format. The task force further indicated that the use of a randomized crossover design (such as that employed in this study) with reporting of correlation as an acceptable method of validation.[13]

There are multiple reasons why test scores may differ when the same test is readministered on more than one occasion. Patients may change their mind regarding an answer between tests, their condition may change resulting in a different answer even over a short period of time, or random variation may occur. To this end, a correlation coefficient of 1.0 is rarely, if ever, achieved even when readministering the same test using the same modality. A correlation coefficient of 0.70 or above has been generally accepted when validating a patient outcome measure over repeated administration. The results of this study compare favorably with the historical reports of within-mode (paper vs. paper) correlation

Table 2 Historical Test–retest Correlation Values for Each Individual PRO Tool and Observed Test–retest Correlation between Paper and Electronic Versions

Standard		Study			
Instrument	Historical correlation coefficient	No. Completed	Institution A-Not scorable	Institution B-Not scorable	Correlation coefficient
IKDC score[10]	0.95	93	6	3	0.87
KOOS – Daily Living[9]	0.91	90	5	7	0.88
KOOS - Pain[9]	0.86	91	7	4	0.92
KOOS – Quality of Life[9]	0.83	97	3	2	0.85
KOOS - Sports and Recreational Activities[9]	0.78	88	8	6	0.86
KOOS - Symptoms[9]	0.84	94	3	5	0.92
Lysholm[12]	0.94	91	5	6	0.86
SF-36	[a]	86	7	9	0.95
SF-36 Body Pain[11]	0.78	98	2	2	0.85
SF-36 General Health[11]	0.86	98	2	2	0.94
SF-36 Mental Health[11]	0.84	98	2	2	0.90
SF-36 Mental Health Composite11 (Dimension B)	[b]	100	2	2	0..82
SF-36 Physical Function[11]	0.94	98	2	2	0.88
SF-36 Physical Health Composite11 (Dimension A)	[b]	100	2	0	0.84
SF-36 Role Emotional[11]	0.83	98	2	2	0.73
SF-36 Role Physical[11]	0.86	98	2	2	0.79
SF-36 Social Functioning[11]	0.87	98	2	2	0.79
SF-36 Vitality[11]	0.84	98	2	2	0.88

Abbreviations: PRO, patient-reported outcomes; IKDC, International Knee Documentation Committee form; KOOS, Knee injury and Osteoarthritis Outcome Score; SF-36, 36-Item Short Form survey.
Note: Overall correlation of all questions = 0.86. Institution B sample size = 55. Institution A sample size = 47.
[a]Total SF-36 does not have a recognized clinical meaning. It only provides correlation data for the entire SF-36 (see column 3).
[b]Composite scores are derived from the eight other SF-36 scores, correlations are not reported in the original scores.

indicating that the observed variability is no different than if the same test was readministered using the same modality.

The acquisition and analysis of outcomes data are of significant interest to health care providers in today's quality-driven health care industry. Paper-based outcomes collection has been the gold standard in research and clinical publications for decades. In addition, part of that historical standard was that outcomes tools be used in only the recommended formats for data to be considered "valid." This study evaluated whether or not the electronic administration of a condensed outcomes–based questionnaire was a reliable way to collect outcomes information that could be translated into accepted outcomes reporting tools and maintain acceptable correlation coefficients.

The study results presented here indicate that the reformatting of the selected orthopedic outcomes questions in an efficient (condensed) electronic collection tool maintains the historical correlation coefficients of the original tools and can reliably be used for outcomes data collection by orthopedic providers for the specific PROs tested. The march toward evidence-based medicine will likely push clinicians increasingly toward data collection on their own patients, even if they do not intend to publish. The results of this study provide evidence that a properly designed computer-based methodology is valid. The use of such a system may provide significant benefit by reducing the burden on patients, physicians, and health care budgets.

Correlation between paper and electronic outcomes tools has been reported previously by other authors. A 2008 meta-analysis reported on 46 studies including a total of 278 scales providing correlation coefficients between electronic and paper methods.[14] The authors noted that the average correlation coefficient was 0.90, and 94% of correlation coefficients were above 0.75. The authors also reported that within-mode comparison (paper vs. papers) correlation coefficients were nearly identical to cross-mode (paper vs. electronic) correlation. A further review of the literature, however, indicates that the study presented in this paper is the first study to assess paper versus electronic correlation for orthopedic patients including knee specific PROs.

The limitations of the study include that only specific PRO tools related to the knee and general QOL were evaluated. Secondly, patients were asked to complete both formats (paper and electronic) at variable time points, some within a matter of hours, others within days up to 1 week. In addition, patient diagnosis was not standardized; however, the lack of standardization is similar to the clinical use of outcomes tools. Finally, the question of whether or not electronic capture of outcomes information is more efficient for clinicians or more satisfying for patients was not addressed within the scope of this study and is of obvious interest for future study as electronic outcomes reporting becomes more prevalent.

Acknowledgments

The authors thank Vicki Snodgrass-Miller and Andrea Clifford at OrthoIndy and Vasili Karas at Midwest Orthopaedics at Rush for their contributions in supporting this study. The authors also thank the staff of Universal Research Solutions, LLC, for providing the technical and related assistance required for this specific investigation.

References

1 Value-based purchasing: a strategic overview for health care industry stakeholders. . Updated 2011, Accessed May 2013. Available at: http://www.deloitte.com/us/valuebasedpurchasing
2 Suk M, Hanson BP, Norvell D, Helfet D. The purpose of patient-reported outcomes. In: Musculoskeletal Outcomes Measurements and Instruments. AO Foundation, ed. Switzerland Thieme: 2009:880.
3 Basch E, Bennett A, Pietanza MC. Use of patient-reported outcomes to improve the predictive accuracy of clinician-reported adverse events. J Natl Cancer Inst 2011;103(24):1808–1810
4 Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst 2009;101(23):1624–1632
5 Khanna G, Singh JA, Pomeroy DL, Gioe TJ. Comparison of patient-reported and clinician-assessed outcomes following total knee arthroplasty. J Bone Joint Surg Am 2011;93(20):e117–e117, 1, 7
6 Haffer SC, Bowen SE. Measuring and improving health outcomes in Medicare: the Medicare HOS program. Health Care Financ Rev 2004;25(4):1–3
7 O'Connor PJ, Bodkin NL, Fradkin J, et al. Diabetes performance measures: current status and future directions. Diabetes Care 2011;34(7):1651–1659
8 OBERD 2012. Accessed May 2013. Available at: http://www.oberd.com/
9 Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998;28(2):88–96
10 Higgins LD, Taylor MK, Park D, et al; International Knee Documentation Committee. Reliability and validity of the International Knee Documentation Committee (IKDC) Subjective Knee Form. Joint Bone Spine 2007;74(6):594–599
11 Ruta DA, Abdalla MI, Garratt AM, Coutts A, Russell IT. SF 36 health survey questionnaire: I. Reliability in two patient based studies. Qual Health Care 1994;3(4):180–185
12 Briggs KK, Lysholm J, Tegner Y, Rodkey WG, Kocher MS, Steadman JR. The reliability, validity, and responsiveness of the Lysholm score and Tegner activity scale for anterior cruciate ligament injuries of the knee: 25 years later. Am J Sports Med 2009; 37(5):890–897
13 Coons SJ, Gwaltney CJ, Hays RD, et al; ISPOR ePRO Task Force. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009;12(4):419–429
14 Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2): 322–333