A Validation Study of an Electronic Outcomes Tools in Orthopedics

Journal:	*Journal of Knee Surgery*
Manuscript ID:	JKS-13-Feb-11-OA
Manuscript Type:	Original Article
Specialty Area:	computer-based patient-reported outcomes, IKDC, responsiveness, reliability
Abstract:	Background: Patient reported outcomes 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 patient reported outcomes (PRO) measures are provided to a single patient. Additionally, the management and analysis of data collected in paper format involves labor intensive processes to score and render the data analyzeable. Computer-based outcomes systems have the potential to mitigate these problems by reformatting multiple outcomes tools into a single user friendly tool. Purpose: To determine whether the electronic outcomes system presented produces results comparable to the test-retest correlations reported for the corresponding orthopedic paper-based outcomes instruments. Study Design: Crossover based on consecutive orthopedic patients arriving at one of two designated orthopedic knee clinics Methods: Patients were assigned to complete either a paper or a computer-administered questionnaire based upon a similar set of questions (KOOS, IKDC, SF36v1, Lysholm). 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 to test-retest data from the original validation studies. Results: 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. Conclusions: Based on the correlation coefficients, the electromic appliation

	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 electronic outcomes system and validates comparison of scores between electronic and paper modes. OBERD Validity paper 2.12.13 with abstract revise for 325wc.doc

A Validation Study of an Electronic Method of Condensed Outcomes Tools Reporting in Orthopedics

ABSTRACT

Background: Patient reported outcomes 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 patient reported outcomes (PRO) measures are provided to a single patient. Additionally, the management and analysis of data collected in paper format involves labor intensive processes to score and render the data analyzeable. Computer-based outcomes systems have the potential to mitigate these problems by reformatting multiple outcomes tools into a single user friendly tool.

Purpose: To determine whether the electronic outcomes system presented produces results comparable to the test-retest correlations reported for the corresponding orthopedic paper-based outcomes instruments.

Study Design: Crossover based on consecutive orthopedic patients arriving at one of two designated orthopedic knee clinics

Methods: Patients were assigned to complete either a paper or a computer-administered questionnaire based upon a similar set of questions (KOOS, IKDC, SF36v1, Lysholm). 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 to test-retest data from the original validation studies.

Results: 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.

Conclusions: Based on the correlation coefficients, the electromic appliation 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 electronic outcomes system and validates comparison of scores between electronic and paper modes.

Introduction

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.[8,14] In fact, PROs are often considered to be more meaningful measures of outcomes than clinician reported outcomes (CROs).[1,2,9] More recently, PROs have a newly assigned role in healthcare reform with rewards attached for those physicians and entities who report them, as in the pay for performance movement.[6,11] Additionally, PROs address the need for transparency, meaningful use and quality of life measures. The value of PROs is therefore multi-faceted.

PROs may be categorized by disease (or in orthopaedics, by affected structure or pathology) or they may address broader quality of life (QOL) measures. As the terms suggest, disease-specific instruments evaluate change in the affected system only, whereas general health related outcomes tools (HRQoL) compare quality of life before and after treatment but can be impacted by other concurrent disease processes.[8,14] The advantage to having both disease-specific and HRQoL measures is that of determining outcome with regard to the specific joint, andpotentially 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.[14] In instances when a given disease process does not have a specific PRO, researchers will often use several overlapping PRO's to capture potential clinically relevant changes. The disadvantage, however, of more comprehensive measurements is that the acquisition and management of large amounts of data is 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 the present 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 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 is to evaluate a condensed electronic version of 4 PRO's, 3 knee-specific and one health related quality of life (HRQOL) and correlate to traditional paper ("expanded version") results. Our hypothesis is that no difference will exist in test-retest correlation between paper and electronics PRO's.

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 (whether local computer entry or web-based entry of outcomes) 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.[10] 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 the present study center 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

144 presentation utilizes visual elements designed to improve usability and reduce the cognitive load
145 for the patient) impact the overall validity of the responses. (Figure 1) These techniques are
146 especially important when the researcher wishes to combine several standard instruments. As an
147 example, if a question and its allowed answers are identical between two instruments the patient
148 only sees the question once.(Table 1) If only some of the answers are the same, then follow-up
149 information may be needed for the incompatible answers. If the questions are similar, but not
150 identical, then a compatible rewording may be presented. The final result is a separate score for
151 each of the distinct instruments. These scores are compared to results from the original paper
152 versions of the individual instruments.
153
154 The Knee injury and Osteoarthritis Outcome Score (KOOS), subjective International Knee
155 Documentation Committee form (IKDC), Short Form 36 version 1 (SF36v1) created by the
156 RAND Corporation and the Lysholm Knee Scoring Scale (Lysholm) were combined in this
157 OBERD application to construct a reformatted instrument, which generated separate scores for
158 the KOOS, IKDC, SF36v1 and Lysholm to be compared to their respective scores calculated on
159 each separate individual paper version of the four instruments involved. (Figure 2) The study
160 was approved by the institutational review board at each of the senior author's primary
161 institution Patients were informed of a $20 stipend, advised that their participation could be
162 withdrawn at any time, and verbal consent was obtained from those who qualified and agreed.
163 All new patients presenting to one of two orthopedic clinics (Institution A, Institution B) with a
164 primary knee problem were included in the study. Exclusion criteria were: 1) any condition that
165 severely limits the subject's ability to complete the paper-based or enhanced questionnaires (e.g.
166 blindness, physical disability); 2) under 18 years of age; 3) 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; 4) patients unwilling to participate and sign the informed consent form.

Patients were identified as likely to medt the inclusion criteria based on a review of the patient list one week before a clinic by research staff. Patients were contacted by phone prior to the office visit and offered entry to the study. Alternatively, patients were interviewed by the research staff on the day of visit and those that met the inclusion criteria were offered entry into the study.

Rush patients who verbally agreed to participate were emailed a link to the electronic questionnaire to complete approximately one week before their office visit. Such patients then completed each of the individual, paper versions of the KOOS, IKDC, SF36v1 and Lysholm onsite during the day of their appointment, upon the research coordinator formally verifying the participant's consent and completion of the computer-based version. Institution B patients who had, likewise, verbally agreed in advance to participate were sent the four paper-based individual instruments to their physical address approximately one week before their appointment to be completed and returned during their visit. Institution B patients were then requested to complete the electronic version on an iPad onsite during the day of their appointment, upon the research coordinator formally verifying the participant's consent and completion of each paper-based questionnaire.

Many patients could not be contacted to obtain verbal consent before their appointment for various reasons (e.g. disconnected phone, no answer, etc.). Such patients could not complete the

paper or electronic formatted version(s) beforehand as described and were therefore first introduced to the study on the day of their appointment by the research coordinators. Those who agreed to participate were consented like all others. Either the paper or computer-based version was completed before seeing the physician and the remaining version was completed after seeing the physician.

The research coordinators passed stipends to each participant after verifying all requirements were met and before participants left the clinic. 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 administration mode.

TESTING METHODOLOGY

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

Results

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 sufficienlty answer many of the questions. In this case, the specific instrument which was incomplete was removed from the analysis. (Table 2)

Because of the various subscales which they contain, the instruments used in this study provided a total of 17 different scores for comparison, resulting in a total of 1638 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 since the SF-36 is overtly multi-dimensional. However, it is included here, as it speaks to overall correlation.

The individual correlation percentage for each individual score is presented in Table 2. The range which is usually considered to show adequate correlation is a correlation coefficient > 0.70. The test-retest results found in the literature for these instruments are provided in Table 1. together with the electronic-paper correlation coefficients obtained in the present study. The overall correlation coefficient for all 1638 pairs of scores was 0.86.

Discussion

The results of this study demonstrate that 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 to 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 which cannot be eliminated without protocol or instrument changes. Thus the correlations reported argue that electronic methods cannot be distinguished from their paper-based ancestors, since they fall within the inherent error range of each instrument.

In response to the evolving need for development of electronic testing methods, the ISPOR created a task force to provide recommendations for validation of PRO measures transcribed to an electronic format. 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 which may involved differing cognitive processes. Both were applied in this study in providing a more user friendly interface for completion of outcome measures, and condensing the same or similar questions into one format. The task force further indicated that use of a randomized cross over design (such as that employed in this study) with reporting of correlation as an acceptable method of validation.[4]

There are multiple reasons by test scores may differ when the same test is re-administered 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 re-administering 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 to the

historical reports of within mode (paper v. paper) correlation indicating that the observed variablity is no different than if the same tested was re-administered using the same modality.

The acquisition and analysis of outcomes data is of significant interest to healthcare providers in today's quality driven healthcare industry. Paper-based outcomes collection has been the gold standard in research and clinical publications for decades. Additionally, 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 questions 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 strongly 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 PRO's 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 healthcare 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. 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 this is the first study to assess paper versus electronic correlation for orthopedic patients including knee specific PRO's.[5]

The limitations of the study include the fact that only specific PRO tools related to the knee and general quality of life were evaluated in this study. Second, patients were asked to completed both formats (paper and electronic) at variable time points, some within a matter of hours, others within days up to one 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.

1. 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.

2. Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, Appawu M, Iasonos A, Atkinson T, Goldfarb S, Culkin A, Kris MG, Schrag D. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. J Natl Cancer Inst 2009;101(23):1624-1632.

3. 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.

4. Coons SJ, Gwaltney CJ, Hays RD, Lundy J, Sloan JA, Revicki DA, Lenderking WR, Cella D, and Basch E. Recommendations of Evidence needed to support Measurement equivalence between electronic and paper-based patient-reported outcomes (PRO) measures: ISPOR ePRO Good Research Practice Task Force Report. Value in Health, 2009, 12(4): 419-429.

5. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of Electronic and paper-and-pencil administration of patient reported outcome measures: A meta-analytic review. Value in Health 2008, 11(2): 322-333.

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. Higgins LD, Taylor MK, Park D, Ghodadra N, Marchant M, Pietrobon R, Cook C, 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.

8. Value-based purchasing: A strategic overview for health care industry stakeholders [Internet]. [updated 2011; cited 2012]. Available from: http://www.deloitte.com/view/en_US/us/Insights/centers/center-for-health-solutions/82e4b2de53ebe210VgnVCM1000001a56f00aRCRD.htm

9. 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(1)-e117(7).

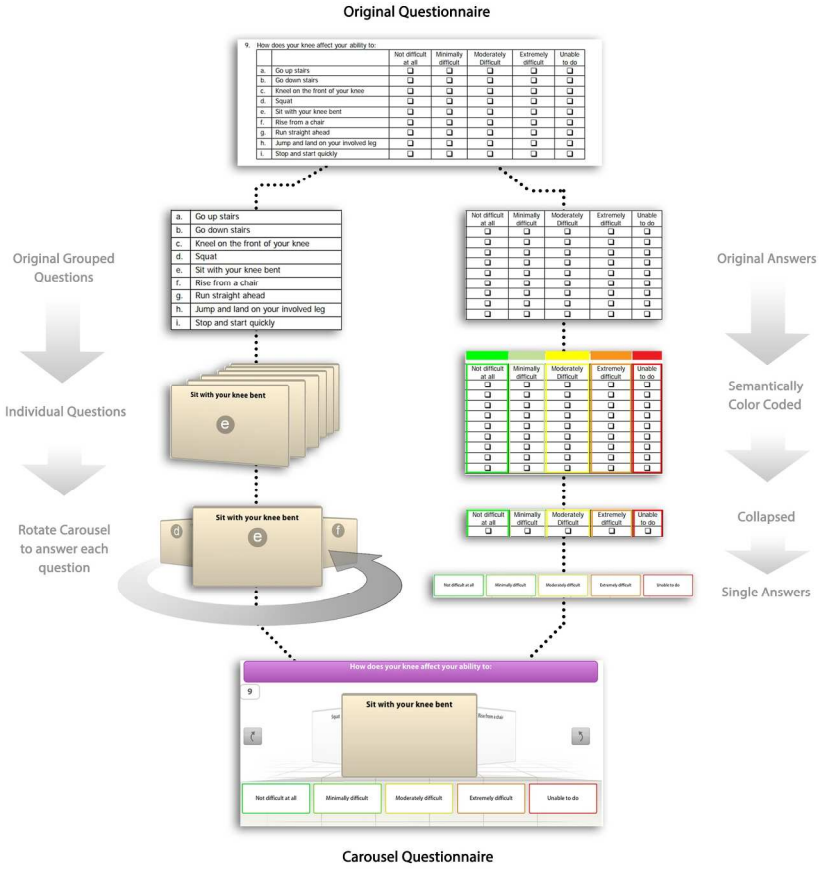
10. OBERD [Internet]. 2012]. Available from: http://www.oberd.com/

11. O'Connor PJ, Bodkin NL, Fradkin J, Glasgow RE, Greenfield S, Gregg E, Kerr EA, Pawlson LG, Selby JV, Sutherland JE, Taylor ML, Wysham CH. Diabetes performance measures: current status and future directions. Diabetes Care 2011;34(7):1651-1659.

Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998;28(2):88-96.

13. 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.

14. Suk M, Hanson B. Selection and assessment upper extremity. In: AO Foundation, ed. *Musculoskeletal Outcomes Measures and Instruments.* TAC; 2009:880.



OBERD strategies to condense questions and maintain outcomes tool integrity.
887x827mm (72 x 72 DPI)

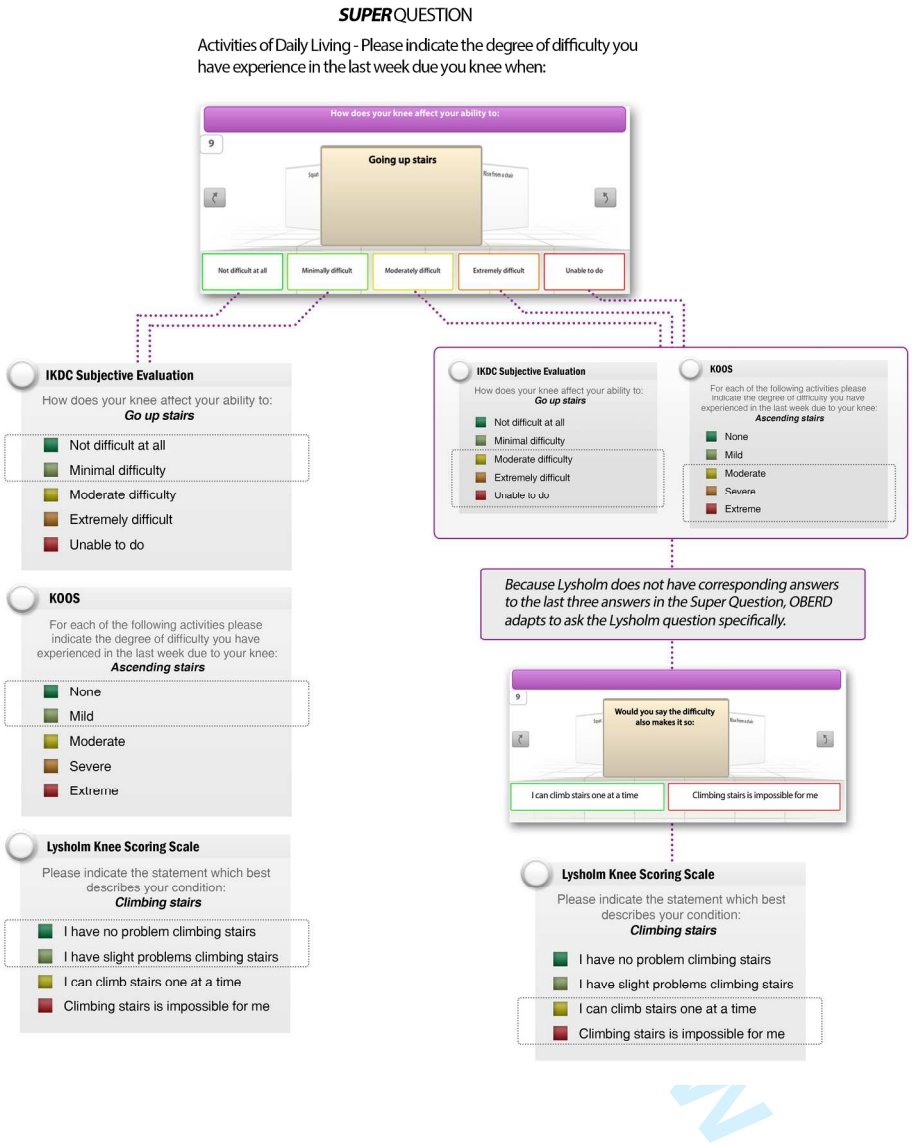

SUPER QUESTION
Activities of Daily Living - Please indicate the degree of difficulty you have experience in the last week due you knee when:
How does your knee affect your ability to:
Going up stairs
Not difficult at all
Minimally difficult
Moderately difficult
Extremely difficult
Unable to do
IKDC Subjective Evaluation
How does your knee affect your ability to:
Go up stairs
Not difficult at all
Minimal difficulty
Moderate difficulty
Extremely difficult
Unable to do
KOOS
For each of the following activities please indicate the degree of difficulty you have experienced in the last week due to your knee:
Ascending stairs
None
Mild
Moderate
Severe
Extreme
Lysholm Knee Scoring Scale
Please indicate the statement which best describes your condition:
Climbing stairs
I have no problem climbing stairs
I have slight problems climbing stairs
I can climb stairs one at a time
Climbing stairs is impossible for me
IKDC Subjective Evaluation
How does your knee affect your ability to:
Go up stairs
Not difficult at all
Minimal difficulty
Moderate difficulty
Extremely difficult
KOOS
For each of the following activities please indicate the degree of difficulty you have experienced in the last week due to your knee:
Ascending stairs
None
Mild
Moderate
Severe
Extreme
Because Lysholm does not have corresponding answers to the last three answers in the Super Question, OBERD adapts to ask the Lysholm question specifically.
Would you say the difficulty also makes it so:
I can climb stairs one at a time
Climbing stairs is impossible for me
Lysholm Knee Scoring Scale
Please indicate the statement which best describes your condition:
Climbing stairs
I have no problem climbing stairs
I have slight problems climbing stairs
I can climb stairs one at a time
Climbing stairs is impossible for me

Table 1: Condensation of questions achieved using electronic format.

IKDC+KOOS+Lysholm+SF-36v1

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)

Table 1. Cumulative effect of OBERD condensation methods. Adaptive methods are dynamic and and vary with individual answers

Table 2: Historical test-retest correlation values for each individucal PRO tool and observed test-retest correlation between paper and electronic versions.

STANDARD			STUDY			
Instrument	Correlation Coefficient	Reference	Correlation Coefficient	# Completed	OI Not Scoreable	Institution A Not Scoreable
IKDC score	0.95	Higgins5	0.87	93	3	6
KOOS – Daily Living	0.91	Roos10	0.88	90	7	5
KOOS - Pain	0.86	Roos10	0.92	91	4	7
KOOS – Quality of Life	0.83	Roos10	0.85	97	2	3
KOOS - Sports and Recreational Activities	0.78	Roos10	0.86	88	6	8
KOOS - Symptoms	0.84	Roos10	0.92	94	5	3
Lysholm	0.94	Briggs3	0.86	91	6	5
SF-36	*	*	0.95	86	9	7
SF-36 Body Pain	0.78	Ruta11	0.85	98	2	2
SF-36 General Health	0.86	Ruta11	0.94	98	2	2
SF-36 Mental Health	0.84	Ruta11	0.90	98	2	2
SF-36 Mental Health Composite (Dimension B)	**	Ruta11	0..82	100	2	2
SF-36 Physical Function	0.94	Ruta11	0.88	98	2	2
SF-36 Physical Health Composite (Dimension A)	**	Ruta11	0.84	100	0	2
SF-36 Role Emotional	0.83	Ruta11	0.73	98	2	2
SF-36 Role Physical	0.86	Ruta11	0.79	98	2	2
SF-36 Social Functioning	0.87	Ruta11	0.79	98	2	2

SF-36 Vitality	0.84	Ruta11	0.88	98	2	2
Overall Correlation of All Questions 0.86						
OI Sample Size 55						
Institution A Sample Size 47						

* Total SF-36 does not have a recognized clinical meaning. It only provides correlation data for the entire SF-36
see column 3.
** Composite scores are derived from the 8 other SF-36 cores, correlations are not reported in the original scores.