

THE JOURNAL OF BONE & JOINT SURGERY

JB&JS

This is an enhanced PDF from The Journal of Bone and Joint Surgery

The PDF of the article you requested follows this cover page.

Patients Can Provide a Valid Assessment of Quality of Life, Functional Status, and General Health on the Day They Undergo Knee Surgery

Dianne Bryant, Paul Stratford, Robert Marx, Stephen Walter and Gordon Guyatt
J Bone Joint Surg Am. 2008;90:264-270. doi:10.2106/JBJS.G.00336

This information is current as of August 4, 2008

Supplementary material

Commentary and Perspective, data tables, additional images, video clips and/or translated abstracts are available for this article. This information can be accessed at <http://www.ejbjs.org/cgi/content/full/90/2/264/DC1>

Letters to The Editor are available at
<http://www.ejbjs.org/cgi/content/full/90/2/264#responses>

Reprints and Permissions

Click here to [order reprints or request permission](#) to use material from this article, or locate the article citation on jbjs.org and click on the [Reprints and Permissions] link.

Publisher Information

The Journal of Bone and Joint Surgery
20 Pickering Street, Needham, MA 02492-3157
www.jbjs.org

Patients Can Provide a Valid Assessment of Quality of Life, Functional Status, and General Health on the Day They Undergo Knee Surgery

By Dianne Bryant, MSc, PhD, Paul Stratford, PT, MSc, Robert Marx, MD, FRCSC, Stephen Walter, PhD, and Gordon Guyatt, MD, MSc

Investigation performed at McMaster University, Hamilton, and The University of Western Ontario, London, Ontario, Canada

Background: In the interest of efficiency, investigators often offer participants in surgical trials the option of completing baseline assessments on the day of surgery. The emotional affects of this day may, however, increase bias or random error. We studied the validity and reliability of collecting subjective ratings of health on the day of surgery.

Methods: One hundred and seventy-seven patients undergoing anterior cruciate ligament reconstruction and/or knee arthroscopy completed quality-of-life, functional status, and general health instruments at four weeks preoperatively, on the day of surgery, and one year postoperatively. We evaluated results with use of three conceptual frameworks: (1) that ratings provided four weeks preoperatively provide a gold standard for preoperative ratings, (2) that there is no gold standard for preoperative ratings and that, if valid, ratings on the day of surgery should be highly correlated with ratings at four weeks preoperatively and moderately and similarly correlated with ratings at one year postoperatively, and (3) that ratings provided four weeks preoperatively and on the day of surgery are measuring identical constructs and should therefore show high reliability.

Results: Most patients (97%) had a chronic injury as the interval between the injury and surgery was more than ninety days. Data collected on the day of surgery demonstrated high predictive validity with data collected within one month before surgery. There was no significant heterogeneity between variances for data collected four weeks preoperatively and on the day of surgery. The correlation between data collected on the day of surgery and four weeks preoperatively was moderate to high (range, 0.64 to 0.93), and the correlation between preoperative ratings and the one-year postoperative ratings was moderate (range, 0.40 to 0.59) across all instruments. Agreement between the ratings provided four weeks preoperatively and on the day of surgery was excellent (intraclass correlation coefficient, 0.64 to 0.91), and the standard error of measurement was small across instruments.

Conclusions: In the treatment of chronic knee injuries, patients can accurately rate their quality of life, general health, and functional status on the day on which they undergo surgery.

Patient self-ratings of quality of life, general health, and functional status have gained popularity as outcome measures for the evaluation of the effects of surgical interventions. In the context of a clinical trial, investigators often collect pretreatment or baseline measurements to establish health status before treatment in order to allow for a

more powerful statistical comparison of treatment effect between groups. In the interest of efficiency and for the sake of convenience, investigators may offer trial participants the option of completing baseline assessments on the day of surgery instead of making a separate appointment prior to surgery. The advantage of doing so is that the patient is available and

Disclosure: In support of their research for or preparation of this work, one or more of the authors received, in any one year, outside funding or grants in excess of \$10,000 from Lawson Health Research Institute and Smith and Nephew. Neither they nor a member of their immediate families received payments or other benefits or a commitment or agreement to provide such benefits from a commercial entity. No commercial entity paid or directed, or agreed to pay or direct, any benefits to any research fund, foundation, division, center, clinical practice, or other charitable or nonprofit organization with which the authors, or a member of their immediate families, are affiliated or associated.



A commentary is available with the electronic versions of this article, on our web site (www.jbjs.org) and on our quarterly CD-ROM (call our subscription department, at 781-449-9780, to order the CD-ROM).

has time to complete assessments, usually while waiting in the preoperative area for one to two hours before surgery with few competing issues (e.g., physicians waiting to see them or tests needing to be completed).

The question remains, however, regarding the validity and reliability of data collected on the day of surgery, which for many patients may be filled with anxiety that could compromise their ability to provide accurate responses. There are three possibilities. First, it is possible that the emotional affects of the day of surgery have no effect on the validity or reliability of the measurement. Second, the emotional affects of the day of surgery could contribute to an overall biased measurement of health status if patients within a treatment group similarly overestimate or underestimate their disability. Third, the emotional affects of surgery may vary greatly among patients, affecting some much more than others; some patients may overestimate their status, whereas others may underestimate their status, contributing to greater overall random error. The purpose of the present study was to determine the validity and reliability of collecting self-rated baseline measurements of quality of life, general health, and functional status on the day of surgery in patients undergoing a surgical intervention for the treatment of a variety of knee conditions.

Materials and Methods

Study Design

The present study focused on additional objectives of a larger, multicenter randomized clinical trial to determine patients' ability to recall preoperative quality of life, general health, and function. A description of the design, methods, and results for the randomized trial have been reported elsewhere¹.

The present study included consecutive patients undergoing knee arthroscopy and/or anterior cruciate ligament reconstruction. All patients completed quality-of-life, general health, and functional status instruments four weeks preoperatively, on the day of surgery (before the procedure), and two weeks and one year postoperatively. For the purposes of this study, our analyses included ratings provided at four weeks preoperatively, on the day of surgery, and at one year postoperatively only. All questionnaires were standardized to inquire about the patient's health status over the past two weeks.

We conducted the analysis of the validity and reliability of ratings collected on the day of surgery with use of three conceptual approaches. First, we assumed that the ratings collected at four weeks preoperatively provide a gold or criterion standard of patients' preoperative quality of life, functional status, and general health and that, if valid, ratings collected on the day of surgery should accurately predict ratings provided at four weeks preoperatively. Second, we assumed that both the four-week preoperative rating and the day-of-surgery rating measure the same construct and will therefore have excellent agreement, thus we determined reliability. Third, using a construct validity approach, we postulated that if the ratings from the day of surgery were valid, they would correlate highly with the ratings from four weeks preopera-

tively, would have similar variances, and would demonstrate a similar and moderate correlation with outcome at one year postoperatively.

Instruments

Instruments included the Subjective Knee Form of the International Knee Documentation Committee (IKDC)², the Knee Injury and Osteoarthritis Outcome Score (KOOS)³, and the Short-Form Health Survey (SF-36)⁴. In addition, patients undergoing an isolated arthroscopy completed the Western Ontario Meniscal Evaluation Tool (WOMET)⁵ and patients undergoing anterior cruciate ligament reconstruction completed the Quality of Life Outcome Measure (Questionnaire) for Chronic Anterior Cruciate Ligament Deficiency (ACL-QOL)⁶. All questionnaires were self-administered under the supervision of a research assistant.

The IKDC was designed to measure the functional ability of patients with knee problems. It includes eighteen items related to symptoms (seven items), limitations in recreational activities (ten items), and function before the injury (one item). The number of response options per question varies from two options (one item) to five options (fourteen items) to eleven options (three items). The majority of response options are presented as ordered categories (e.g., not at all, mildly, moderately, very, extremely) or require patients to rate their symptom on a scale from 0 to 10, with 10 being the worst possible score. A patient's total score is determined by calculating the raw score (the sum of all responses), calculating the difference between the raw score and the lowest possible score (18), dividing this difference by the range of scores (87), and multiplying by 100. The resulting number represents a total score out of 100 possible points, with 100 points representing perfect knee function. This instrument has face validity, has demonstrated construct validity, has excellent test-retest reliability (intraclass correlation coefficient, 0.94; 95% confidence interval, 0.88 to 0.97), and is responsive to change^{2,7}.

The KOOS⁸ is a forty-two-item knee-specific questionnaire developed for patients with anterior cruciate ligament injury, meniscus injury, or osteoarthritis. It has five separately reported domains of pain (nine items), symptoms (seven items), limitations in activities of daily living (seventeen items), recreation (five items), and quality of life (four items). Each item has five ordered response options (always, often, sometimes, rarely, never). In the present study, each domain yielded similar results, and thus, for ease of reporting, we computed an overall aggregate score only (the sum of all items divided by the total number of items converted to a total score out of 100%, with 100% representing the best possible score) and presented results with use of this aggregate score. This instrument has face validity, has demonstrated construct validity, has excellent test-retest reliability for each domain (range, 0.75 to 0.93), and is responsive to change^{8,9}.

The ACL-QOL⁶ is a thirty-two-item questionnaire designed to measure the quality of life of patients with anterior cruciate ligament deficiency. It has five domains related to physical symptoms (five items), limitations in occupation (four items), recreational activities (twelve items), lifestyle (six

TABLE I Demographic Characteristics of the 177 Patients Who Were Available for Analysis

Male gender	62%
Age* (yr)	39 ± 13
Height* (in [cm])	70 ± 14 (178 ± 36)
Weight* (lb [kg])	182 ± 38 (83 ± 17)
Third-party compensation†	16%
Smoking status‡	
Never smoked	51%
Smoked, but quit	26% (11 ± 12)
Current smoker	23% (12 ± 11)
Right knee affected	56%
Arthroscopy only (no anterior cruciate ligament reconstruction)	65%
Previous surgery	42%
Time from injury to surgery	
<3 months	3%
≥3 months to 1 year	30%
>1 year to 3 years	24%
>3 years to 5 years	5%
>5 years to 10 years	6%
>10 years	32%

*The values are given as the average and the standard deviation. †Includes Workman's Safety and Insurance Board, disability, and litigation. ‡The values are expressed as the proportion of patients who smoked, with the number of pack-years (the number of years of smoking multiplied by the number of packs per day) in parentheses. The number of pack-years is expressed as the average and the standard deviation.

items), and social and emotional concerns (five items). Each item has one 100-mm visual analog scale with labeled anchors at 0 mm (extremely concerned) and 100 mm (not concerned at all). A patient's score is calculated by converting the average of each of the five domain scores to a total average score out of 100%, with 100% representing the highest quality of life. This instrument has face validity, has demonstrated content and construct validity, has excellent test-retest reliability (standard error of measurement, 6%), and is responsive to change⁶.

The WOMET⁵ is a sixteen-item questionnaire designed to measure quality of life in patients with meniscal pathology. There are three domains related to physical symptoms (nine items); restrictions in recreational, occupational, and lifestyle activities (four items); and emotional well-being (three items). Each item has one 100-mm visual analog scale response option with anchors at 0 mm (not at all worried) and 100 mm (extremely worried). A patient's score is calculated by adding the scores for all items, subtracting the raw score from 1600 (the worst possible score), and dividing by 16 to obtain a score out of 100%, with 100% representing the highest possible quality of life. This instrument has face validity, has content and construct validity, has excellent test-retest reliability (intraclass

correlation coefficient, 0.79; 95% confidence interval, 0.59 to 0.87), and is responsive to change⁵.

The SF-36⁴ is a thirty-six-item generic general health instrument that evaluates eight domains of health, including restrictions or limitations on physical and social activities, activities and responsibilities of daily living, pain, mental health and well-being, and perceptions of health. The SF-36 can be reported as eight domain scores (physical functioning, role physical, bodily health, social functioning, role social, mental health, general health, and vitality) or as two component scores (the Physical Component Score [PCS] or the Mental Component Score [MCS]). Component scores are computed aggregates of all eight domain scores, with the weight of the contribution of each domain score being unique for each component score. The SF-36 has been used extensively and has been shown to be valid, reliable, and responsive in a wide variety of populations and contexts^{10,11}, including patients with orthopaedic conditions^{9,12,13}. For the purpose of the present study, we present the PCS and the MCS only.

Statistical Analysis

To address the first conceptualization, we assumed that ratings provided at four weeks preoperatively represent a valid assessment of preoperative status (the so-called gold standard). We used linear regression to determine the extent to which ratings given on the day of surgery (independent variable) predict ratings given four weeks prior to surgery (dependent variable). If ratings obtained on the day of surgery are valid, we expect the coefficient (β) for the independent variable to approach 1.0 with a significant ($p < 0.01$) proportion of the variance being explained by the model (i.e., with the value of R^2 approaching 1.0). In contrast with a Pearson correlation coefficient (r), a linear regression implies directionality and is

TABLE II Descriptive Statistics of Ratings Provided at Four Weeks Preoperatively, on the Day of Surgery, and at One Year Postoperatively*

Questionnaire†	4 Weeks Preop.	Day of Surgery	1 Year Postop.
WOMET	34.3 ± 17.4	35.8 ± 18.3	56.1 ± 26.7
ACL-QOL	32.2 ± 17.5	33.4 ± 17.5	61.7 ± 20.9
IKDC	46.2 ± 17.1	45.5 ± 17.7	62.1 ± 22.4
KOOS	57.2 ± 17.1	57.4 ± 17.3	71.4 ± 19.2
SF-36 PCS	38.6 ± 10.1	39.9 ± 9.9	45.1 ± 10.8
SF-36 MCS	51.7 ± 9.2	52.1 ± 9.2	52.8 ± 9.1

*The values are expressed, in points, as the average and the standard deviation. †WOMET = Western Ontario Meniscal Evaluation Tool, ACL-QOL = Quality of Life Outcome Measure (Questionnaire) for Chronic Anterior Cruciate Ligament Deficiency, IKDC = International Knee Documentation Committee, KOOS = Knee Injury and Osteoarthritis Outcome Score, SF-36 PCS = Short-Form Health Survey Physical Component Score, and SF-36 MCS = Short-Form Health Survey Mental Component Score.

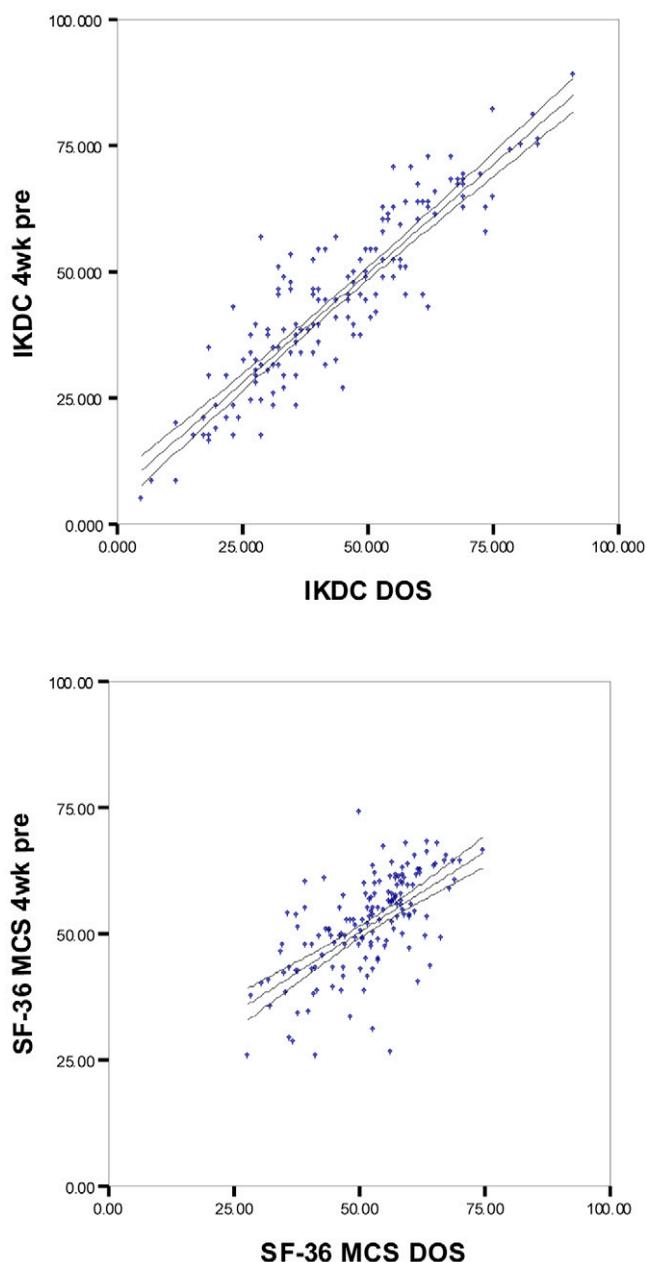


Fig. 1
Scatterplots illustrating the agreement between patients' ratings on the day of surgery (DOS) and at four weeks preoperatively (4wk pre) for (a) the Subjective Form of the International Knee Documentation Committee (IKDC) and (b) the Mental Component Score of the Short-Form Health Survey (SF-36 MCS).

therefore more appropriate when assessing the strength of the association between the independent predictor variable (preoperative ratings measured on the day of surgery) and the dependent variable (preoperative ratings measured prior to the day of surgery). We constructed 95% confidence intervals around the prediction line for group and individual ratings.

To address the second conceptualization (that the instruments are measuring the same construct at both time-

points), we assessed reliability with use of repeated-measures analysis of variance where the between-subject variance was a random variable (patient) and the within-subject variable (time) was fixed and was defined by the ratings given four weeks preoperatively and those given on the day of surgery. Using the variances from the analysis of variance, we calculated an intraclass correlation coefficient (two-way mixed model with measures of absolute agreement) to estimate the level of agreement between ratings and the standard error of measurement to estimate the error associated with the reproducibility of individual measurements¹⁴. In contrast with a Pearson correlation coefficient (r), an intraclass correlation coefficient takes into account systematic differences as well as random differences between variables and is therefore more appropriate when assessing the level of agreement.

To address the third conceptualization, we hypothesized (1) that the magnitude of the correlation between ratings provided four weeks preoperatively and on the day of surgery would be high (>0.80), (2) that the variance associated with ratings at each time-point would be similar, and (3) that the correlation between preoperative ratings and one-year post-operative outcome ratings would be moderate in magnitude and similar, regardless of whether preoperative ratings were provided four weeks preoperatively or on the day of surgery, and tested for significant differences between correlations.

To test the first hypothesis, we calculated the Pearson correlation coefficient (r) between the four-week preoperative rating and the day-of-surgery rating across all instruments. To test the second hypothesis, we assumed that, if the condition of the patients remained stable between time-points, the overall variance would also remain stable. Although total variance includes between and within-subject variability and random error, we assumed that the condition of the patients was stable (the majority

TABLE III Predictive Validity of Data Collected on the Day of Surgery, with Ratings Provided One Month Before Surgery Used as a Gold Standard

Questionnaire*	R ²	Coefficient (β)†	P Value
WOMET	0.74	0.82 (0.73 to 0.91)	<0.001
ACL-QOL	0.73	0.75 (0.63 to 0.86)	<0.001
KOOS	0.87	0.92 (0.87 to 0.98)	<0.001
IKDC	0.83	0.88 (0.82 to 0.94)	<0.001
SF-36 PCS	0.64	0.81 (0.72 to 0.90)	<0.001
SF-36 MCS	0.41	0.64 (0.52 to 0.76)	<0.001

*WOMET = Western Ontario Meniscal Evaluation Tool, ACL-QOL = Quality of Life Outcome Measure (Questionnaire) for Chronic Anterior Cruciate Ligament Deficiency, IKDC = International Knee Documentation Committee, KOOS = Knee Injury and Osteoarthritis Outcome Score, SF-36 PCS = Short-Form Health Survey Physical Component Score, and SF-36 MCS = Short-Form Health Survey Mental Component Score. †The 95% confidence interval is shown in parentheses.

TABLE IV Reliability Between Ratings Provided Four Weeks Preoperatively and on the Day of Surgery

Questionnaire*	Intraclass Correlation Coefficient†	Standard Error of Measurement†
WOMET	0.86 (0.80 to 0.90)	6.65 (5.88 to 7.65)
ACL-QOL	0.84 (0.76 to 0.91)	5.74 (4.88 to 6.98)
KOOS	0.93 (0.91 to 0.95)	4.43 (4.01 to 4.94)
IKDC	0.91 (0.88 to 0.93)	5.22 (4.73 to 5.84)
SF-36 PCS	0.80 (0.74 to 0.85)	4.48 (4.05 to 5.00)
SF-36 MCS	0.64 (0.54 to 0.72)	5.66 (5.10 to 6.35)

*WOMET = Western Ontario Meniscal Evaluation Tool, ACL-QOL = Quality of Life Outcome Measure (Questionnaire) for Chronic Anterior Cruciate Ligament Deficiency, IKDC = International Knee Documentation Committee, KOOS = Knee Injury and Osteoarthritis Outcome Score, SF-36 PCS = Short-Form Health Survey Physical Component Score, and SF-36 MCS = Short-Form Health Survey Mental Component Score. †The 95% confidence interval is shown in parentheses.

of patients had sustained an injury more than six months previously) and interpreted evidence of heterogeneity of variance as an increase in random error only. Thus, we compared the overall variance of ratings provided four weeks preoperatively with the variance of ratings provided on the day of surgery by conducting a test for heterogeneity of nonindependent variances^{15,16}, where a positive test ($p < 0.05$) indicates unequal variances.

Finally, we calculated the Pearson correlation coefficient (r) to describe the association between the four-week preoperative ratings and the one-year postoperative ratings and compared it with the Pearson correlation coefficient describing the association between the day-of-surgery ratings and the one-year postoperative ratings with use of the Fisher z test¹⁷. Similar correlations suggest similar ratings at four weeks preoperatively and on the day of surgery.

Results

Of the 211 patients who began these assessments, thirty-four did not complete the study; specifically, sixteen patients withdrew their consent, six patients were lost to follow-up, ten patients did not undergo the scheduled procedure because it was cancelled by either the physician or the patient, and two patients were excluded during surgery because they required an ineligible procedure. Table I describes the characteristics of the remaining 177 patients who were available for analysis. For most patients (97%), the time from the injury to the operation was more than ninety days. Descriptive statistics of ratings provided at four weeks preoperatively, on the day of surgery, and one year postoperatively are provided in Table II.

Ratings Collected on the Day of Surgery Can Accurately Predict Earlier Ratings

Scatterplots with regression lines and 95% confidence intervals are suggestive of excellent predictive validity for both disease-

specific and knee-specific instruments and the PCS of the SF-36. Figure 1 shows scatterplots with the prediction line and 95% group and individual prediction intervals for the IKDC (an example with large between-subject variability) and the MCS of the SF-36 (an example with small between-subject variability). Across all instruments, the ratings obtained on the day of surgery were a significant predictor of the ratings obtained four weeks preoperatively ($p < 0.001$) (Table III). Residual analysis verified that the data were consistent with the assumptions of linear regression.

Ratings Collected on the Day of Surgery and Four Weeks Preoperatively Show Excellent Reliability

The reliability of the recalled and actual ratings was excellent for the WOMET (intraclass correlation coefficient, 0.86; 95% confidence interval, 0.80 to 0.90) and ACL-QOL (intraclass correlation coefficient, 0.84; 95% confidence interval, 0.76 to 0.91) disease-specific questionnaires, the IKDC (intraclass correlation coefficient, 0.91; 95% confidence interval, 0.88 to

TABLE V Assessment of the Heterogeneity of Variance and Comparison of the Magnitude of the Association with One-Year Postoperative Data Between Ratings Provided on the Day of Surgery and at Four Weeks Preoperatively

Questionnaire*	Variance	P Value	Pearson's r	P Value
WOMET		0.31		1.0
4 weeks preop.	302.8		0.53	
Day of surgery	334.9		0.53	
ACL-QOL		0.94		1.0
4 weeks preop.	309.8		0.40	
Day of surgery	306.2		0.40	
IKDC		0.30		0.49
4 weeks preop.	292.4		0.55	
Day of surgery	313.3		0.50	
KOOS		0.68		0.91
4 weeks preop.	292.4		0.58	
Day of surgery	299.3		0.59	
SF-36 PCS		0.67		0.77
4 weeks preop.	102.0		0.56	
Day of surgery	98.0		0.54	
SF-36 MCS		0.71		0.57
4 weeks preop.	88.4		0.53	
Day of surgery	84.6		0.50	

*WOMET = Western Ontario Meniscal Evaluation Tool, ACL-QOL = Quality of Life Outcome Measure (Questionnaire) for Chronic Anterior Cruciate Ligament Deficiency, IKDC = International Knee Documentation Committee, KOOS = Knee Injury and Osteoarthritis Outcome Score, SF-36 PCS = Short-Form Health Survey Physical Component Score, and SF-36 MCS = Short-Form Health Survey Mental Component Score.

0.93) and KOOS (intraclass correlation coefficient, 0.93; 95% confidence interval, 0.91 to 0.95) knee-specific questionnaires, and the SF-36 PCS (intraclass correlation coefficient, 0.80; 95% confidence interval, 0.74 to 0.85), whereas the reliability of the SF-36 MCS was moderate (intraclass correlation coefficient, 0.64; 95% confidence interval, 0.54 to 0.72) (Table IV).

The standard error of measurement was small for all questionnaires, including the SF-36 MCS, suggesting that the lower level of agreement between the four-week preoperative ratings and the day-of-surgery ratings for the MCS was not attributable to greater measurement error but rather to less heterogeneity in scores between patients (smaller between-subject variability). In other words, the scores for the patients in our study population did not represent the entire range of scores possible for the SF-36 MCS (i.e., the scores of all patients fell within the middle part of the range) but did represent a greater proportion of the range of possible scores for the disease or region-specific measures (greater between-subject variability).

Ratings Collected on the Day of Surgery Demonstrate Construct Validity

The ratings of quality of life, functional status, and physical health (as reflected by the ACL-QOL, WOMET, IKDC, KOOS, and the PCS of the SF-36) that were provided at four weeks preoperatively and on the day of surgery were highly correlated (range, 0.80 to 0.93), whereas the ratings of mental health (as reflected by the MCS of the SF-36) that were provided at four weeks preoperatively and on the day of surgery were moderately correlated ($r = 0.64$). Across all instruments, the variances were similar between four weeks preoperatively and the day of surgery, and no difference approached significance (Table V).

The correlation between the one-year postoperative and four-week preoperative ratings ranged from 0.40 to 0.58. The magnitude of the association between the one-year postoperative and day-of-surgery ratings ranged from 0.40 to 0.59. The magnitude of the difference between correlations between time-points for each instrument ranged from 0.00 to 0.05, and none of the differences approached significance (Table V).

Discussion

The present study provides evidence of the validity and reliability of data collected on the day of surgery. First, if one regards ratings one month prior to surgery as a gold standard, then ratings on the day of surgery are valid in that they accurately predict the gold standard. Second, ratings obtained on the day of surgery and those obtained four weeks prior to surgery show a high level of agreement and a low and consistent level of within-subject error.

Finally, when we assumed that there is no gold standard for the measurement of preoperative status and applied principles of construct validation to the ratings obtained on the day of surgery, we found that day-of-surgery ratings demonstrated the anticipated high correlation with ratings obtained four weeks prior to surgery and that there was no difference in the overall variance between time-points (Table IV). The finding

that ratings provided on the day of surgery and those provided within one month before surgery showed a similar moderate correlation with the one-year postoperative outcome provides further evidence of construct validity.

The validity of the day-of-surgery rating has important implications for the planning of randomized trials of interventions for patients undergoing surgery of the knee and, to the extent that our results are generalizable, to other surgical procedures. When conducting a randomized trial in which one wishes to make comparisons between groups, statistical approaches include a t test of the postoperative end point scores, a change score (postoperative end point minus preoperative measurement), or an analysis of covariance in which the baseline preoperative measure provides a covariate and the postoperative end point is used as the dependent variable. Using only postoperative end point scores increases the likelihood of a false-negative result due to random error¹⁸⁻²³ unless the association between preoperative baseline and postoperative end point scores is zero. When conducting an analysis involving the use of either change scores or an analysis of covariance, the magnitude of the association between the preoperative baseline and postoperative end point scores becomes extremely important. For an analysis involving the use of change scores, the magnitude of this correlation needs to be at least 0.50 or the analysis becomes inefficient (i.e., loss of power)²⁰⁻²². For an analysis of covariance, the consequences of a low correlation are less serious (power will not be lost if the correlation is low) but the gain in power by considering the pretest value still increases with an increasing strength of the association^{18,20,22}.

The present study offers two important pieces of information related to statistical planning. The first is that data collected on the day of surgery are similar to data collected four weeks preoperatively with respect to the magnitude of the association with postoperative data. Thus, collecting data on the day of surgery will not interfere with the power of the analysis. The second is that across all questionnaires, including disease-specific, joint-specific, and generic health instruments, the magnitude of the correlation was never >60% and most often hovered around 50%, suggesting that the use of change scores may be inefficient and will certainly be less efficient than the use of an analysis of covariance.

The strengths of the present study include its large sample size, its use of different types of self-assessment instruments (from disease-specific to generic), and the broad spectrum of knee injuries and conditions represented in our population. This study included patients with a wide range of ages (from fifteen to seventy-eight years) who had knee conditions spanning minor patellofemoral disease, meniscal injuries, anterior cruciate ligament deficiencies, and differing severities of osteoarthritis.

The weaknesses of the present study may include limitations related to the generalizability of the results. It is possible that patients undergoing urgent or emergent surgical procedures or elective procedures with an associated risk of death and serious physical or cognitive impairment experience a greater degree of anxiety that may impair their ability to provide an accurate self-assessment of their health status.

In conclusion, patients undergoing surgery for the treatment of a chronic knee injury can provide an accurate self-assessment of their quality of life, general health, and functional status on the day of surgery. The results of the present study suggest that investigators can improve the efficiency of data collection for clinical studies for this patient population, with no expected loss of statistical power, by obtaining baseline self-assessments on the day of surgery. ■

NOTE: The authors wish to acknowledge the late Dr. Alexandra Kirkley for her contribution to the conceptualization of this project and for her insight into the importance of this work; Dr. Peter Fowler, Dr. Robert Giffin, Dr. Robert Litchfield, Dr. Rick Ogilvie, Dr. Devin Peterson, and Dr. Kevin Willits, who allowed us to recruit their patients; Peggy Austin, who managed the data; and Diane Heels-Ansdell, Lauren Griffith, and Bruce Weaver for lending their expertise to numerous statistical discussions.

Dianne Bryant, MSc, PhD
Faculty of Health Sciences, Elborn College, Room 1438, The University of

Western Ontario, London, ON N6G 1H1, Canada. E-mail address: dianne.bryant@uwo.ca

Paul Stratford, PT, MSc
School of Rehabilitation Sciences, Institute of Applied Sciences, Room 430, McMaster University, 1200 Main Street West, Hamilton, ON L8N 3Z5, Canada. E-mail address: stratfor@mcmaster.ca

Robert Marx, MD, FRCSC
Sports Medicine and Shoulder Service, Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021. E-mail address: marxr@hss.edu

Stephen Walter, PhD
Gordon Guyatt, MD, MSc
Clinical Epidemiology and Biostatistics, Hamilton Health Sciences Center, Rooms 2C16 (S.W.) and 2C12 (G.G.), McMaster University, 1200 Main Street West, Hamilton, ON L8N 3Z5, Canada. E-mail address for S. Walter: walter@mcmaster.ca. E-mail address for G. Guyatt: guyatt@mcmaster.ca

References

- Bryant D, Norman G, Stratford P, Marx RG, Walter SD, Guyatt G. Patients undergoing knee surgery provided accurate ratings of preoperative quality of life and function 2 weeks after surgery. *J Clin Epidemiol.* 2006;59:984-93.
- Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, Richmond JC, Shelborne KD. Development and validation of the international knee documentation committee subjective knee form. *Am J Sports Med.* 2001;29:600-13.
- Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynonn BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28:88-96.
- 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.
- Kirkley A, Griffin S, Whelan D. The development and validation of a quality of life measurement tool for patients with meniscal pathology: The Western Ontario Meniscal Evaluation Tool (WOMET). *Clin J Sport Med.* 2007;17:349-56.
- Mohtadi N. Development and validation of the quality of life outcome measure (questionnaire) for chronic anterior cruciate ligament deficiency. *Am J Sports Med.* 1998;26:350-9.
- Irrgang JJ, Ho H, Harner CD, Fu FH. Use of the International Knee Documentation Committee guidelines to assess outcome following anterior cruciate ligament reconstruction. *Knee Surg Sports Traumatol Arthrosc.* 1998;6:107-14.
- Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS)—validation of a Swedish version. *Scand J Med Sci Sports.* 1998;8:439-48.
- Di Fabio RP, Boissonnault W. Physical therapy and health-related outcomes for patients with common orthopaedic diagnoses. *J Orthop Sports Phys Ther.* 1998;27:219-30.
- McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care.* 1994;32:40-66.
- McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31:247-63.
- Jette DU, Jette AM. Physical therapy and health outcomes in patients with knee impairments. *Phys Ther.* 1996;76:1178-87.
- Shapiro ET, Richmond JC, Rockett SE, McGrath MM, Donaldson WR. The use of a generic, patient-based health assessment (SF-36) for evaluation of patients with anterior cruciate ligament injuries. *Am J Sports Med.* 1996;24:196-200.
- Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther.* 1997;77:745-50.
- Pitman EJG. A note on normal correlation. *Biometrika.* 1939;31:9-12.
- Snedecor GW, Cochran WG. *Statistical methods.* 6th ed. Ames, Iowa: Iowa State University Press; 1967.
- Fisher RA. On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron.* 1921;1:3-32.
- Cronbach LJ, Furby L. How should we measure 'change'—or should we? *Psychol Bull.* 1970;74:68-80.
- Knapp TR. The (un)reliability of change scores in counseling research. *Meas Eval Guid.* 1980;13:149-57.
- Lee J. A note on the comparison of group means based on repeated measurements of the same subject. *J Chron Dis.* 1980;33:673-5.
- Norman GR. Issues in the use of change scores in randomized trials. *J Clin Epidemiol.* 1989;42:1097-105.
- Oldham PD. A note on the analysis of repeated measurements of the same subjects. *J Chronic Dis.* 1962;15:969-77.
- Stanek EJ. Choosing a pretest-posttest analysis. *Am Stat.* 1988;42:178-83.