Comparison of Patient-Reported and Clinician-Assessed Outcomes Following Total Knee Arthroplasty

Gaurav Khanna, MD, Jasvinder A. Singh, MD, MPH, Donald L. Pomeroy, MD, and Terence J. Gioe, MD

Investigation performed at the Minneapolis Veterans Affairs Medical Center, Minneapolis, Minnesota, and the University of Louisville Medical College, Louisville, Kentucky

Background: Although the necessity of long-term follow-up after total knee arthroplasty is unquestioned, this task may become burdensome as greater numbers of total knee arthroplasties are performed. We sought to use comparisons with clinician-assessed values to determine whether patients could reliably assess their own outcome with use of a combination of American Knee Society Score and Oxford Knee Score questionnaires and self-reported knee motion. We hypothesized that patients would self-report worse pain and function and a similar range of knee motion than clinicians would.

Methods: One hundred and forty patients (181 knees) scheduled for routine follow-up at two centers after primary total knee arthroplasty were mailed American Knee Society Score and Oxford Knee Score questionnaires, a set of photographs illustrating knee motion in 5° increments for comparison with the patient’s range of knee motion, and a goniometer with instructions. The patient’s American Knee Society Score, Oxford Knee Score, and knee motion were then independently assessed within two weeks of the self-evaluation by one of three clinicians who had not been involved with the surgery. Patient-reported and clinician-assessed measures were compared with use of a paired-sample t test and the Spearman correlation coefficient.

Results: The mean patient-reported American Knee Society pain subscore was 4 points worse than the clinician-assessed score, and the function subscore was 10 points worse (p < 0.001 for both). The mean Oxford Knee Score did not differ significantly between the patient self-assessment and the clinician assessment (p = 0.05). The mean maximum flexion reported by the patient with use of the photographs differed by <1° from the mean value reported by the patient with use of the goniometer or the mean value measured by the clinician; these differences were not clinically important.

Conclusions: Patients’ self-reported American Knee Society pain and function subscores were worse than the corresponding clinician assessments, but the two Oxford Knee Scores were similar. Range of knee motion may reasonably be self-assessed by comparison with photographs. Long-term follow-up of patients after total knee arthroplasty may be possible with use of patient-reported measures, alleviating the burden of clinic visits yet maintaining contact, but further studies involving other validated instruments is warranted.

The long-term outcome of total knee arthroplasty is typically assessed on the basis of the improvement in measures of pain and function, including range of motion, stability, and ambulation. There is currently no widely accepted, comprehensive measure to assess the outcome after total knee arthroplasty, so orthopaedic surgeons generally use

Disclosure: None of the authors received payments or services, either directly or indirectly (i.e., via his or her institution), from a third party in support of any aspect of this work. One or more of the authors, or his or her institution, has had a financial relationship, in the thirty-six months prior to submission of this work, with an entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. Also, one or more of the authors has had another relationship, or has engaged in another activity, that could be perceived to influence or have the potential to influence what is written in this work. The complete Disclosures of Potential Conflicts of Interest submitted by authors are always provided with the online version of the article.

Disclaimer: This material is partially based upon work supported (or supported in part) by the Office of Research and Development, Department of Veterans Affairs. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.
various scoring systems combined with regular radiographic evaluation to monitor patient progress and long-term survival of the arthroplasty. Although the necessity of long-term follow-up of patients after total knee arthroplasty is unquestioned, this task may become burdensome as greater numbers of total knee arthroplasties are performed in younger populations. Self-reported patient questionnaires may lessen the burden on both clinicians and patients.

The Oxford Knee Score is a self-reported patient questionnaire that addresses the patient’s assessment of knee function and its effect on quality of life. The American Knee Society Score is another widely used scoring system that consists of two parts, the clinical score and the function score, and is typically administered in the clinician’s office. Medalla et al. noted a good correlation between the Oxford Knee Score and the American Knee Society Score at two years postoperatively, implying that self-reporting is a viable screening tool to identify patients who are in need of clinical follow-up. Our earlier study demonstrated an average 8 to 10-point difference between patient-reported and clinician-assessed American Knee Society Scores, with patients reporting poorer scores.

The range of knee motion is a key parameter for measuring the outcome of total knee arthroplasty. Edwards et al. showed that goniometer measurements of flexion made by experienced clinical evaluators after total knee arthroplasty were quite reproducible when the same technique was used on both occasions. Our earlier study showed that patients could reasonably record their own range of knee motion by following simple directions to compare their knee with a set of photographs illustrating specific knee flexion and extension angles, as the results had a moderate correlation coefficient of 0.42 relative to measurements made by a clinician.

For long-term follow-up of total knee arthroplasty to be efficient, practical, and valuable, the differences between patient-reported and clinician-assessed responses must be resolved. We studied a cohort of patients who had undergone total knee arthroplasty at two major medical centers to determine whether they could reliably assess their own outcomes with use of a combination of American Knee Society Score and Oxford Knee Score questionnaires and self-reported range of knee motion. We had the following three hypotheses: (1) patients can measure range of knee motion with a goniometer as accurately as the clinician can, and this measurement would be more accurate than comparison of the knee with photographic illustrations would be; (2) the patient-reported American Knee Society pain and function subscores and the Oxford Knee Score would be worse than those recorded by the clinician; and (3) the differences between patient-reported and clinician-assessed pain scores would have a greater impact on outcome scores than the difference in knee motion would.

Materials and Methods

One hundred and forty consecutive patients (181 knees) who had undergone primary total knee arthroplasty and were scheduled for routine follow-up as part of prospective studies of function following total knee arthroplasty by the surgeon authors (T.J.G. and D.L.P.) were recruited for the present study. The study was approved by the surgeons’ institutional review boards. All patients were assessed at a minimum of one year postoperatively (range, twelve to 228 months; mean, fifty-three months).

The most recent version of the Oxford Knee Score assessment consists of twelve equally weighted questions; the possible score on each question ranges from 0 to 4. The minimum total score of 0 indicates maximal disability, and the maximum score of 48 indicates normal function. Five of the questions relate specifically to pain. Seven of the questions could reasonably be considered to relate to function, and three, to range of motion of the knee. The American Knee Society clinical score assigns the maximum of 100 points to a well-aligned knee with no pain, a range of motion of 125°, and negligible anteroposterior or mediolateral instability. The function score uses self-reported stair-climbing ability and walking distance as the main factors, and use of walking aids results in a deduction of points. The pain score and the range of motion represent 50% and 25%, respectively, of the 100-point clinical score. The function score is a separate 100-point scale.

All patients were mailed the American Knee Society Score and Oxford Knee Score questionnaires, photographs illustrating knee flexion in 5° increments, and a plastic goniometer with simple instructions for measuring the active range of knee motion two weeks prior to the scheduled clinic visit. We used both the Oxford Knee Score and the American Knee Society Score because the Oxford Knee Score has relatively more validation data and the American Knee Society Score is one of most widely used knee outcome instruments. Patients were also asked to measure the range of knee motion once with use of the goniometer, since this mimics the usual examination by the clinician. Pat- ients were instructed to fill out and return the questionnaires promptly, and they were telephoned with reminders to do so if they did not respond promptly. Another mailing was sent to patients who had not yet returned the forms by one week prior to the appointment. If a patient’s forms contained incomplete data, the patient either was called on the telephone or met with a research assistant before the appointment. Within two weeks of completing the questionnaire, each patient was examined by the clinician, who asked the same questions that were contained in the American Knee Society Score and Oxford Knee Score instruments and measured the range of knee motion once, to the nearest degree, with use of a standard goniometer.

Standardization of Measurements

The three clinical examiners were physician assistants or nurse practitioners who were well experienced in performing these assessments. They had not been involved in the surgery and were blinded to the patient self-reports. Prior to initiation of the study, the clinicians assessed twenty knees, and the interobserver and intraobserver reliability of the measurements was determined. We required that all intraclass correlation coefficients (ICCs) be >0.80; the actual intraobserver values were 0.96, 0.93, and 0.94, and the interobserver values were 0.87, 0.92, and 0.91. The range of motion and clinical information for each patient in the study cohort were subsequently obtained by one of the same clinicians and entered into the patient’s electronic medical record, from which it was later abstracted into the study database by a research assistant.

Statistical Methods

We calculated that 160 knees would be required to detect a 5-point difference in the American Knee Society pain subscore (which has a possible range of 0 to 50) with an alpha of 0.05 and a power of 0.80. Assuming that 10% of patients would be lost to follow-up or provide incomplete information, we planned to include 180 knees. We calculated that this sample of 180 knees would also provide a power of 0.99 to detect a difference of 5° in flexion (within the usual range of 80° to 120°) and a 5° difference in extension (within the usual range of 0° to 30°) between the patient and clinician measurements. Our earlier study provided us with estimates of the expected difference between the measurement of the range of motion made by the patient with use of photographic illustrations of the knee and the measurement made by the clinician, and we assumed that the goniometer measurement by the patient would be at least as accurate as the measurement made with use of photographs. Our a priori assumption was that a difference of <5° in the range of motion would not be clinically meaningful, since the American Knee Society Score measures range of motion in 5° increments and the photographs were in 5° increments.
For the American Knee Society clinical and function subscores and the Oxford Knee Scores, we considered the clinic visit with the examiner to be the “gold standard,” and we compared the patient-reported scores with the clinician-assessed scores with use of the paired-sample t test. We also calculated a 95% confidence interval and a p value for each difference; a 95% confidence interval that does not include zero is approximately equivalent to a p value of <0.05.

Pain was reported by the patient both on the mailed questionnaire and during the clinic visit, and it was unclear which pain report should be considered to be the “gold standard.” Thus, we considered neither to represent the more accurate assessment of the patient’s pain. Since the American Knee Society pain subscore can range from 0 to 50 points (one-half of the 100-point range of the clinical subscore), we made the conservative priori assumption that a difference of <5 points between observers would not be clinically meaningful. This is also consistent with other studies in which a difference of 10 points on a pain scale of 0 to 100 was considered to be clinically meaningful11-12. Similarly, we made the a priori assumption that a 5-point difference in the Oxford Knee Score (roughly 10% of the total possible score of 48) might be clinically meaningful.

We assessed the “inaccuracy” of the range of motion measurements made by the patient with use of the photographic knee illustrations by subtracting the “gold standard” assessment by the clinician from the patient-reported knee flexion or extension. Similarly, we assessed the inaccuracy of the range of motion measurements made by the patient with use of the goniometer by subtracting the same “gold standard” clinician measurement from the goniometer measurement made by the patient. Since the range of motion measured by the patient could be either greater or less than the value measured by the clinician, averaging the differences could underestimate the magnitude of the typical difference; we therefore calculated the proportion of patients with a >10% discrepancy between the two flexion or extension measurements. Discrepancies in the patient-reported American Knee Society pain and function subscores were calculated by subtracting the clinician-assessed score from the patient-recorded score.

The correlation of patient-reported range of motion with that measured by the clinician was then assessed with use of the nonparametric Spearman correlation coefficient. This method was also used to assess the correlation of the American Knee Society Score and Oxford Knee Score values derived from the mailed survey with those obtained at the clinic visit. We used the Spearman correlation coefficient rather than the Pearson correlation coefficient because the latter can be more easily influenced by extreme observations. We considered the Spearman rank correlation to be high if it was >0.7, moderate if it was 0.4 to 0.7, and low if it was <0.4.

The results are reported on the basis of total knee arthroplasties rather than patients, since numerous patients had undergone bilateral primary total knee arthroplasty and the measurements were performed for each total knee arthroplasty. The effect of interdependence of observations in the forty-one patients who had had bilateral procedures was assessed by first performing an analysis that included all total knee arthroplasties and then performing a sensitivity analysis that was limited to one knee per patient. This simpler approach, rather than more complex methods such as generalized estimating equations, was used because clinicians represent the primary audience for this paper.

Source of Funding
There was no external source of funding for this study.

Results
One hundred and forty patients (181 knees) at the two centers (Minneapolis Veterans Affairs Medical Center and the University of Louisville Medical College) were included in the study. The majority (74%) of the patients were men, reflecting the predominance of men in the Veterans Affairs population. The mean age was 68.3 years, and the diagnosis was osteoarthritis in 94% of the patients. The other demographic information was characteristic of a population that had undergone total knee arthroplasty (see Appendix).

The mean extension (and standard deviation) measured by the patients with use of the set of 5°-increment photographs was 2.2° ± 4.1° (n = 154 knees). This differed from the clinician-assessed extension of 0.5° ± 2.5° by 1.7° ± 4.4° (95% confidence interval [CI], 1.0° to 2.4°; p < 0.001). The mean flexion measured by the patients with use of the photographs was 114.0° ± 13.1° (n = 154). This differed from the clinician-assessed flexion of 113.2° ± 10.3° by 0.7° ± 12.8° (95% CI, −1.3° to 2.8°; p = 0.48). The mean extension measured by the patients with use of the goniometer was 111.2° ± 20.8° (n = 161). This differed from the clinician-assessed goniometer measurement of 6.0° ± 2.7° by 2.6° ± 3.6° (95% CI, 1.8° to 3.5°; p < 0.001). The mean flexion measured by the patients with use of the goniometer was 111.2° ± 20.8°. This differed from the clinician-assessed goniometer measurement of 112.8° ± 11.6° by −1.7° ± 20.6° (95% CI, −4.9° to 1.6°; p = 0.32). The Spearman correlation

| TABLE I Correlation of Patient-Reported and Clinician-Assessed Range of Knee Motion |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Patient-Reported Flexion Using Photos** | **Patient-Measured Flexion** | **Clinician-Measured Flexion** | **Patient-Reported Extension Using Photos** | **Patient-Measured Extension** | **Clinician-Measured Extension** |
| Patient-reported flexion using photos | 1 | 0.93* | 0.35* | −0.22* | −0.20* | −0.27* |
| Patient-measured flexion | 1 | 0.37* | −0.20* | −0.23* | −0.28* |
| Clinician-measured flexion | 1 | −0.19* | −0.09 | −0.11 |
| Patient-reported extension using photos | 1 | −0.81* | 0.13 |
| Patient-measured extension | 1 | 0.15 |
| Clinician-measured extension | 1 |

*p ≤ 0.01.
coefficients for the range of motion measurements are shown in Table I.

Similar values were obtained in the sensitivity analyses that included only one knee per patient. The difference between extension reported by the patient with use of the photographs and extension measured by the clinician was 1.8° ± 4.5° (95% CI, 1.0° to 2.6°; p < 0.001). The difference between flexion reported by the patient with use of the photographs and flexion measured by the clinician was 0.5° ± 13.8° (95% CI, –2.0° to 3.0°; p = 0.70). The difference between extension reported by the patient with use of the goniometer and extension measured by the clinician was 2.8° ± 5.8° (95% CI, 1.8° to 3.8°; p < 0.001). The difference between flexion reported by the patient with use of the goniometer and flexion measured by the clinician was –2.5° ± 22.6° (95% CI, –6.4° to 1.5°; p = 0.23).

The mean patient-reported American Knee Society pain subscore was 40.1 ± 11.8 (n = 158). This differed from the clinician-assessed score of 44.0 ± 10.9 by –3.8 ± 10.8 (95% CI, –5.6 to –2.2; p < 0.001). The mean patient-reported American Knee Society function subscore was 61.6 ± 20.7 (n = 156). This differed from the mean clinician score of 70.9 ± 21.5 by –9.2 ± 22.0 (95% CI, –12.7 to –5.8; p < 0.001). The mean patient-reported Oxford Knee Score was 39.8 ± 7.7 (n = 128). This differed from the mean clinician score of 40.6 ± 7.7 by –0.8 ± 4.5 (95% CI, –1.6 to 0; p = 0.05) (Fig. 1). The Spearman correlation coefficients for the American Knee Society subscores are shown in Table II.

Again, sensitivity analyses that included only one knee per patient showed that interdependence of data from both knees of patients who had undergone bilateral total knee arthroplasty did not substantially influence any of our findings (Table III).

**Analyses of Clinically Meaningful Changes**

As previously noted, a 10° difference in flexion or extension was defined a priori as a clinically meaningful discrepancy. Such a discrepancy between the patient-reported and clinician-assessed values existed for 9% (15/158) of the extension measurements made by the patient with use of the goniometer, 4% (6/154) of the extension measurements made by the patient with use of the

![American Knee Society (AKS) pain and function subscores and Oxford Knee Score (OKS) as recorded by the patient and the clinician. Patients self-reported significantly greater pain and worse function on the AKS than the clinician recorded but a similar OKS.](image)

*p < 0.01
photographs, 30% (48/158) of the flexion measurements made by the patient with use of the goniometer, and 25% (38/151) of the flexion measurements made by the patient with use of the photographs. Sensitivity analyses limited to data from one knee per patient were also performed, and similar results were obtained, with corresponding values of 10% (13/126), 4% (5/123), 33% (41/126), and 28% (33/120), respectively.

A 5-point difference in the American Knee Society pain subscore or the Oxford Knee Score was defined as a clinically meaningful difference. Analyses revealed that 35% (56/158) of all knees had a clinically meaningful difference in the American Knee Society questionnaire consistently reported significantly less pain to the physician during the clinic visit than they did when self-reporting. In the present study, we noted a significant (p < 0.001) difference between the American Knee Society subscore recorded by the clinician and by the patient for both function and pain. In both studies, the pain and function subscore recorded by the clinician were higher (better) than those recorded by the patient. Clinicians in other disciplines have also been found to record less pain than the patient reported when filling out the same survey independently. The difference noted in the present study may have

### Discussion

The clinical follow-up of patients remains an important part of postoperative care after total knee arthroplasty. The follow-up interval and duration remain dependent on the physician’s anticipation of the clinical progress of the individual patient. However, a survey of active members of the American Association of Hip and Knee Surgeons demonstrated substantial variability in the typical frequency of clinical visits, with 45% of surgeons advising annual follow-up for the first five years and 37% recommending annual follow-up for more than ten years. Since the rate of total joint arthroplasty in the United States is expected to increase exponentially, the present study was designed to further explore the option for patients to self-report their outcome with use of validated outcome scales and an objective measurement of knee motion. Such self-reporting, perhaps coupled with periodic submission of radiographs via the Internet, could reasonably be expected to lower the burden on both the patient and the clinician without eliminating contact and thus sacrificing quality of care.

<table>
<thead>
<tr>
<th>TABLE III American Knee Society (AKS) Pain and Function Subscores and Oxford Knee Score as Reported by Patients or Recorded by the Clinician</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis including all total knee arthroplasties</td>
</tr>
<tr>
<td>AKS pain subscore (n = 158)</td>
</tr>
<tr>
<td>AKS function subscore (n = 156)</td>
</tr>
<tr>
<td>Oxford Knee Score (n = 128)</td>
</tr>
<tr>
<td>Analysis restricted to one total knee arthroplasty per patient</td>
</tr>
<tr>
<td>AKS pain subscore (n = 126)</td>
</tr>
<tr>
<td>AKS function subscore (n = 124)</td>
</tr>
<tr>
<td>Oxford Knee Score (n = 92)</td>
</tr>
</tbody>
</table>

*Values are given as the mean and standard deviation.
several explanations. Patients may be more reluctant to disappoint their surgeon by complaining about pain at the clinic visit than when filling out a survey at home. Expectations regarding what constitutes an acceptable outcome may differ between patients and physicians. Finally, the interval between completion of the survey and the clinic visit might allow for further improvement in the pain level, although this seems unlikely since the interval of two weeks was much shorter than the one year that had elapsed since the surgery. The fact that the patient-reported pain and function after total knee arthroplasty differed significantly from those recorded in the clinician’s office may have important implications for the use of these scoring instruments.

The reliability of the American Knee Society Score is not optimal; despite its widespread use by investigators after total knee arthroplasty, it has been shown to have high interobserver and intraobserver variability. Liow et al. noted that if a patient is evaluated by two different observers at two separate visits, a change of at least 23 points in the American Knee Society Score is needed in order for the observers to be confident that the change is real. Although we reported the proportion of patients with a difference between the patient-reported and the clinician-recorded American Knee Society Score of at least 5 points for pain or at least 10 points for function, these thresholds are less than the 23-point threshold based on interobserver variability. The American Knee Society Score has undergone revision since the study by Liow et al., and the new version may have improved reliability. Despite such concerns, however, the variability in the surveys completed by the patient at home appeared to be within the range of intraobserver and interobserver variability seen when the same current version of this outcome instrument was used in the clinician’s office.

We believe that the American Knee Society Score and the Oxford Knee Score offer independent benefits when evaluating patient outcome after total knee arthroplasty, and that using both instruments provides a more comprehensive picture of the patient’s true level of function and pain. However, a significant difference between patient-reported and clinician-recorded values was observed for the American Knee Society pain and function subscores but not for the Oxford Knee Score. In addition, the proportion of patients for which the difference between the patient-reported and clinician-recorded values was clinically meaningful was lower for the Oxford Knee Score (12% to 14%) than for the American Knee Society pain subscore (34% to 35%). This may reflect the subjective patient-reported nature of the Oxford Knee Score and the greater overall ease of use and greater interobserver reproducibility of the Oxford Knee Score compared with the American Knee Society Score. A recent publication by Medalla et al. compared the Oxford Knee Score reported by the patient and the American Knee Society Score recorded by the examiner at medium-term follow-up after total knee arthroplasty and reported that the correlation between the scores was good at two years postoperatively but diminished at five and ten years. Their conclusion was that the Oxford Knee Score was a good screening tool for the first two years after total knee arthroplasty but that it would need to be combined with clinical follow-up after that time. Our study indicated that patient reporting of pain and function with use of the American Knee Society Score questionnaire might reasonably be expected to fill that need for clinical follow-up.

Our study has several limitations. As noted previously, the interobserver reliability of the American Knee Society Score has been criticized. However, we believe that the difference between the consistently higher score recorded at the clinic visit and the lower score reported by the patient at home does represent a true difference. One possible explanation for such a systematic difference involves the assessment of knee stability; for instance, a patient with quadriceps weakness or spinal stenosis may record a feeling of instability on the survey completed at home but the medical practitioner may record the same knee as “stable” during the examination at the clinic. Because the function portion of the American Knee Society Score records walking distance, assistive devices used, and stair-climbing ability, we believe that the comparison between the function scores recorded by the patient and by the clinician is more valid, although patient exaggeration to please the surgeon in the office setting must again be considered. The duration of follow-up in our study ranged from twelve to 228 months, with a minimum of one year of follow-up after the arthroplasty. Since the range of motion and the scores on the two questionnaires typically show little variation with time beyond the one-year time point in well-functioning knees, the range in the duration of follow-up in our study likely had little impact on the study, and the timing of the follow-up was consistent with the ongoing follow-up advocated in clinical practice.

Our study has several implications for the long-term follow-up of total knee arthroplasties. Patients can be expected to report the range of knee motion with use of photographic illustrations with an accuracy that is reasonable for the purposes of clinical follow-up. In conjunction with our earlier study, we have shown that real differences may be observed between the pain and function reported by patients on questionnaires and reported to the clinician in the clinic setting. The use of patient-supplied outcome data, combined with periodic radiographs, may help to alleviate the logistical problems associated with the increasing burden of arthroplasty follow-up. The nature of such follow-up may need to change to accommodate the increasing number of arthroplasties. Total knee arthroplasties generally fail in a bimodal pattern, either early (within the first two years) or late (after more than ten years). The interim period between early and late follow-up may represent the best opportunity for the use of patient-reported outcome surveys. In our view, an ideal long-term, long-distance total knee arthroplasty follow-up program would involve a number of features: the use of validated and reliable patient outcome questionnaires encompassing the measurement of both joint-specific and global function and pain, feasible patient compliance, low cost, physician review of radiographs, use of the Internet for transmission and collection of questionnaires and radiographs, and wide acceptance by total knee arthroplasty investigators. Further studies to refine patient questionnaires...
and the use of telemedicine principles will assist us in reaching this goal.

Appendix

A table showing the clinical and demographic characteristics of the patients and knees is available with the online version of this article as a data supplement at jbjs.org.

References