

Can patients help with long-term total knee arthroplasty surveillance? Comparison of the American Knee Society Score self-report and surgeon assessment

T. J. Gioe^{1,2}, D. Pomeroy³, K. Suthers⁴ and J. A. Singh^{2,5,6}

Objectives. To compare patient self-report of knee flexion, extension, range of motion (ROM) and American Knee Society (AKS) Pain, Knee and Functional scores with a clinician assessment.

Methods. A total of 239 consecutive total knee arthroplasty (TKA) patients (290 knees) were mailed surveys with an AKS questionnaire and lateral knee photographs that showed knee ROM in 10° increments to compare their operated knee(s) ROM. Patients were subsequently seen in clinic and their ROM, AKS Pain, Knee and Functional scores were measured. Patient- and physician-reported measures were compared using independent sample *t*-test and correlated using Spearman's correlation coefficient. *A priori* rules for comparisons were based on previously published reports.

Results. A total of 286 knees had both survey and clinic data available and constituted the analytic set. Patient-reported and physician-assessed extension, flexion and ROM were: $3 \pm 4.8^\circ$ vs $1.4 \pm 4.3^\circ$ ($P < 0.001$), $111.4 \pm 14.6^\circ$ vs $110 \pm 12.8^\circ$ ($P = 0.04$) and 108.6 ± 16.8 vs $108.6 \pm 14.3^\circ$ ($P = 0.98$). There was a moderate correlation between patient and physician assessments (extension = 0.31; flexion = 0.44; ROM = 0.42; $P \leq 0.001$ for all). Patient-reported and physician-assessed AKS Pain, Knee and Functional scores were: 35.8 ± 15.6 vs 43.9 ± 11.1 ($P < 0.001$), 79.8 ± 20 vs 88.9 ± 13.3 ($P < 0.001$) and 57.7 ± 23.1 vs 65.7 ± 26.4 ($P < 0.001$), respectively. Patient- and physician-assessed AKS Pain, Knee and Functional scores had moderate–high correlation ($r = 0.49, 0.49$ and 0.70 ; $P \leq 0.001$ for all).

Conclusion. Long-term surveillance of TKA patients may be possible using a self-report AKS, but the average 8- to 10-point difference between patient- and physician-reported AKS scores (patients reporting poorer scores) represents a substantial impact on this outcome instrument. Since patient-reported responses have clear value in global assessment, further evaluation with other validated outcome instruments is warranted.

KEY WORDS: Long-term surveillance, Total knee arthroplasty, Knee society score, Self-report, Surgeon assessment, Total knee replacement, Pain, Function.

Introduction

Outcomes after total knee arthroplasty (TKA) are typically assessed by improvement of pain and function in patients. There is increasing evidence that this assessment should include both a site- and joint-specific score and a more global function score, and that these scores should not be combined, since global function may deteriorate for reasons not related to the joint in question [1–6]. Orthopaedic surgeons have generally combined such scoring instruments with regular radiographic evaluation to monitor the patient's progress and the long-term survival of the arthroplasty.

There is little question that long-term surveillance of joint arthroplasty is necessary, but it has also become increasingly burdensome as greater numbers of TKAs are performed, and in younger populations [7]. Patient self-reported questionnaires are used by many investigators to decrease this burden on the surgeon or staff, and in combination with telemedicine radiographs might be a reasonable option to routine clinic visits. However, it is unclear how patient questionnaire responses correlate to an interview during a physician visit. There are some suggestions that patient-reported outcomes are more discriminatory and that

physician-recorded responses tend towards better scores than those self-reported by patients [8–12].

Many TKA investigators use some combination of the American Knee Society (AKS) score [13], the WOMAC [14], the Medical Outcomes Study short-form health survey (SF-36) [15] and a radiographic report [16] in routine follow-up of their patients and in reporting results. Although the WOMAC and the SF-36 are typically administered as patient self-reported surveys, the AKS has a knee-specific portion that documents range of motion (ROM), stability and extensor lag/flexion contracture as noted by the examiner. The reliability and validity of the AKS has been questioned [9, 16–21] and it is presently undergoing revision, but is still in widespread use.

Numerous studies in the physical therapy literature have documented reasonable intra- and interobserver agreement for ROM measurements when performed with a goniometer under strict testing conditions [22–25]. Other studies conducted under typical clinical conditions have shown wide variability in ROM measurements following TKA [19, 26–29]. It is not known if patients are capable of measuring ROM using simple directions with accuracy similar to that of health professionals.

If patients could reliably self-report at least one important aspect of the knee exam (ROM) and an outcome measurement (pain), it might allow future long-term and long-distance surveillance to occur more easily for both patient and physician. Additionally, if a patient self-reported version of the AKS is concordant with the physician-recorded exam, it may provide opportunities to conduct affordable long-term studies of knee arthroplasty outcomes, since the present AKS has been in long-term use in many centres. In a cohort of patients at two major medical centres, we compared the patient-assessed AKS Pain scores and ROM to those assessed by the physicians during the follow-up visit. We hypothesized that: (i) the ROM assessment by patients would be approximately similar to that measured by

¹Department of Orthopaedic Surgery, University of Minnesota, ²Minneapolis Veterans Affairs Medical Center, Minneapolis, MN, ³Department of Orthopaedic Surgery, University of Louisville Medical College, Louisville, KY, ⁴Clinical Research Department, DePuy®, Inc., Warsaw, IN, ⁵Department of Medicine (Rheumatology), University of Minnesota, Minneapolis and ⁶Mayo Medical School, Rochester, MN, USA.

Submitted 23 July 2008; revised version accepted 20 October 2008.

Correspondence to: T. J. Gioe, Department of Orthopaedic Surgery, University of Minnesota Medical School, Department of Veterans Affairs Medical Center, Section 112E, 1 Veterans Drive, Minneapolis, MN 55417, USA.
E-mail: tjgioe@gmail.com

physicians during follow-up clinic visits; and (ii) there would be a good agreement between patient and physician assessments of pain. Additionally, we aimed to compare and analyse the patient-reported and physician-assessed AKS scores.

Materials and methods

Two hundred and thirty-nine consecutive TKA patients (290 knees) that were routinely followed as part of prospective studies by the two surgeon authors (T.J.G., D.L.P.) were mailed an AKS outcome questionnaire and a knee ROM self-assessment form (See Appendix 1, available as supplementary data at *Rheumatology* Online). All patients were assessed at a minimum of 1 yr and a mean (\pm s.d.) of 50 ± 44 months from surgery. The AKS outcome questionnaire mimicked the AKS scoring instrument but phrased questions in an easily understandable fashion for patients. The questionnaire was submitted for readability (6th grade level) to our Education Department and then tested on 10 patients for readability and content. The AKS score is divided into a 100-point Knee Score (pain, ROM, stability, contracture, alignment) and a separate 100-point Function score (walking, stairs, assistive aids). For ROM, patients were asked to compare their operated knee(s) to lateral knee photographs that showed a knee flexion/extension arc in 10° increments from 0 to 125° with superimposed angle measurements. The questionnaires were part of a larger prospective total knee arthroplasty study that had IRB approval.

A letter of introduction and brief instructions for the ROM assessment were sent with each questionnaire (two questionnaires were sent to bilateral TKA patients). Patients were instructed to fill out and return the completed questionnaires promptly and were telephoned with reminders to do so if they had not responded. The completed questionnaires were not available to the physician assessor and were not a part of the permanent electronic medical record.

Patients were subsequently seen in clinic, their ROM was measured and the physician recorded the AKS score information based on that interview. The difference in time between the survey and clinic assessment was ≤ 30 days in almost all cases (93%), and 70% of the surveys were completed and returned within 14 days of the clinic visit. We obtained ethical approval for the study from all authors' institutions and all investigations were conducted in conformity with ethical principles of research. Informed consent for participation in the study was obtained.

Statistical analyses

We compared differences in patient-assessed vs physician recorded knee flexion, extension, ROM and AKS Pain subset, Knee and Functional scores using independent sample *t*-tests. The correlation of these measures was examined using non-parametric Spearman's correlation coefficients. For Spearman's rank correlation, we considered >0.7 as high, 0.3 – 0.7 as moderate and <0.3 as low. *P*-value < 0.05 was considered statistically significant.

A priori rules for comparison of ROM from patient vs physician assessment were made as follows: (i) since 10° incremental photographs were used, we considered differences $\leq 10^\circ$ as a 'good match'; (ii) as AKS score measures ROM in 5° increments, we considered a difference of $\leq 5^\circ$ a 'perfect match' [13]; and (iii) a difference of $>10^\circ$ was considered an 'imperfect match'.

A priori rule for comparison of AKS Pain scores from survey vs clinic visit was made based on the previously published study by Liow *et al.* [19], which showed that interobserver differences of 23 points in AKS Knee score (range 0–100) were accepted as within the limits of error. Since the pain score range is 0–50 (half the 100-point range of the AKS Knee score), we arbitrarily set the error cut-off at a conservative 10 points between observers. We present differences above and within these ranges for the AKS Pain scores in addition to means and absolute differences.

Since the unit of analysis was TKA, the presence of bilateral TKAs in 22% of the patients may potentially have violated the assumption of independence of observations, when all TKAs were analysed. We therefore conducted sensitivity analyses for all comparisons between patient-assessed and physician-recorded measures by restricting the analyses to unilateral TKAs and the first TKA in patients with bilateral TKAs.

Results

A total of 239 patients with 290 primary TKAs (188 patients with one TKA; 51 patients with bilateral TKAs, 13 of whom had simultaneous bilateral TKA procedure) were consecutively enrolled at two sites. A total of 210 TKAs were enrolled at Minneapolis (166 patients with one TKA and 22 with two TKAs) and 80 TKAs were enrolled at Louisville (22 with one TKA and 29 with two TKAs). The mean age was 70.1 ± 8.3 yrs (range 30–89 yrs) and 83% (198/239) were men. Of these, 286 knees provided both clinic and survey data, and constitute the analytic set and the denominator for most analyses. Two patients provided clinic but not survey data and two patients provided survey but not clinical data.

Table 1 shows the patient-reported and physician-assessed flexion, extension and ROM with very similar mean values. On average, patients over-estimated both flexion and extension by 1.5° each, resulting in almost no difference in ROM (Table 1). For ROM, a 'perfect match' ($\leq 5^\circ$ difference) was seen for 95/286 knees (33%), a 'good match' ($\leq 10^\circ$ difference) for 79/286 knees (28%) and an 'imperfect match' ($>10^\circ$ difference) for 112/286 knees (39%). Thus, good–perfect match was observed in 174/290 (61%) knees for ROM measurements. The correlation coefficients between patient and physician assessments for flexion and ROM were moderate (flexion = 0.44; ROM = 0.42; $P < 0.001$ for both) and low, but still significant, for extension (0.31; $P < 0.01$ for all; Table 2).

Table 3 shows patient-reported and physician-assessed AKS Pain subset, AKS Knee and AKS Functional scores. The average difference between patient and physician measurements for the AKS Pain scale was 8.0 ± 15.1 points. The correlation between the two measures was moderate (correlation coefficient = 0.49, $P \leq 0.001$; Table 4). In general, physicians' scores were 8 points higher (better or less pain) compared with patient survey reports. Using the arbitrary cut-off difference of >10 points in patient- and physician-reported pain scores, 70% (200/285) were within a 10-point difference, 3% (9/285) patients overestimated their scores by >10 points compared with physicians and 27% (76/285) underestimated it by >10 points. A discrepancy of >10 points in AKS Pain score difference was noted more frequently ($P = 0.003$) for patients aged ≥ 80 yrs (54%) compared with those aged 65–80 yrs (27%) or <65 yrs (25%). In general, patients aged ≥ 80 yrs both underestimated and overestimated their pain scores (compared with physicians) more frequently (44 and 10%) than patients aged 65–80 yrs (25 and 2%) and ≤ 65 yrs (22 and 2%), i.e. older patients' self-reported pain scores had much less agreement with physician pain scores than younger patients. There were no gender-related differences in AKS Pain score discrepancy ($P = 0.49$).

TABLE 1. Patient-reported and physician-measured knee flexion, extension and ROM (mean \pm s.d.)

	Patient-reported Mean \pm s.d. (n = 286)	Physician-reported Mean \pm s.d. (n = 286)	Difference (physician–patient)	
			Mean \pm s.d. [95% CI]	<i>P</i> -value ^a
Extension	3.0 \pm 4.8	1.4 \pm 4.3	–1.6 \pm 5.0 [–2.2, –1.0]	<0.001
Flexion	111.4 \pm 14.6	110.0 \pm 12.8	–1.5 \pm 12.4 [–3.0, –0.1]	0.04
ROM	108.6 \pm 16.8	108.6 \pm 14.3	0.02 \pm 14.3 [–1.6, 1.7]	0.98

^aComparisons were made using independent-sample *t*-test.

TABLE 2. Correlation of patient- vs physician-reported knee flexion, extension and ROM

	Physician-reported extension	Patient-reported extension	Physician-reported flexion	Patient-reported flexion	Physician-reported ROM	Patient-reported ROM
Physician-reported extension	1	0.31*	-0.09	-0.11**	-0.30*	-0.18*
Patient-reported extension		1	-0.17*	-0.32*	-0.24*	-0.52*
Physician-reported flexion			1	0.44*	0.89*	0.40*
Patient-reported flexion				1	0.42*	0.90*
Physician-reported ROM					1	0.42*
Patient-reported ROM						1

* $P \leq 0.001$, ** $P < 0.05$.

TABLE 3. AKS Pain, Knee and Functional scores as reported by patients or recorded by the physician

	Patient-reported score Mean \pm s.d. ($n=285$)	Physician-assessed score Mean \pm s.d. ($n=285$)	Difference (physician-patient) Mean \pm s.d. [95% CI] ($n=285$)	P -value*
AKS Pain score	35.8 \pm 15.6	43.9 \pm 11.1	8.0 \pm 15.1 [6.2, 9.8]	<0.001
AKS Knee score	79.8 \pm 20	88.9 \pm 13.3	9.1 \pm 17.8 [7.0, 11.2]	<0.001
AKS Functional score	57.7 \pm 23.1	65.7 \pm 26.4	8.5 \pm 19.5 [5.8, 10.1]	<0.001

*Comparisons were made using independent-sample t -test.

TABLE 4. Correlation of patient- vs physician-reported AKS Pain, Knee and Function scores

	Physician-reported AKS Pain score	Patient-reported AKS Pain score	Physician-reported AKS Knee score	Patient-reported AKS Knee score	Physician-reported AKS Functional score	Patient-reported AKS Functional score
Physician-reported AKS Pain score	1	0.49*	0.78*	-0.50*	-0.36*	0.36*
Patient-reported AKS Pain score		1	0.42*	0.90*	0.38*	0.43*
Physician-reported AKS Knee score			1	0.49*	0.37*	0.37*
Patient-reported AKS Knee score				1	0.38*	0.48*
Physician-reported AKS Functional score					1	0.70*
Patient-reported AKS Functional score						1

* $P \leq 0.001$.

The average difference between patient-reported and physician assessment for the AKS Knee score (which includes the Pain subset) was 9.1 ± 17.8 points and the average difference in the patient and physician AKS Functional score was 8.5 ± 19.5 points. The correlation between these measures was also moderate (correlation coefficients = 0.49 and 0.70, respectively; $P \leq 0.001$; Table 4). In both instances, patients reported lower or worse scores than those gathered during physician assessment.

Sensitivity analyses that restricted analyses to the unilateral knees and the first of the two bilateral staged (excluding simultaneous bilateral knees) did not change estimates or level of statistical significance for flexion, extension, ROM or AKS Pain, Knee or Functional scores.

Discussion

Numerous outcome instruments and patient-reported questionnaires are in use in arthroplasty research, but little consensus exists on which instrument to use routinely since all have some limitations [2, 4]. Patient-recorded questionnaires have become increasingly popular as a means to decrease the effort involved in seeing long-term follow-up TKA patients, especially those that travel long distances. In the absence of a truly objective measure of pain, we believe patients' self-report of pain is the true representation of the AKS Pain score. There is an increasing recognition that the inclusion of patient perceptions is an important part of global assessment and offers a more complete status estimation [30]. Patient compliance with follow-up is increased and costs are also lowered with such protocols [11, 31]. The finding in this study of significant differences between self-reported and physician-recorded AKS scores may have

substantial impact on how this outcome instrument in its present form is utilized.

The AKS is in widespread use among TKA investigators, despite ongoing concerns regarding its validity and reliability. The AKS has been subject to criticism for demographic bias [16], a confusing pain scale and poor discriminatory capacity [1], modest responsiveness to change [18], exclusion of psychological factors in assessment [5] and overall reliability as reflected in considerable intraobserver and interobserver variation [19, 20].

No study has attempted to correlate AKS Pain, Knee, Functional or total scores obtained from a patient questionnaire and that following a physician history and exam. McGrory *et al.* [11] showed significant correlation between Hospital for Special Surgery (HSS) knee scores calculated from patient questionnaire responses and those recorded during subsequent physician visits, but physicians gave significantly higher knee scores than patients. Various allowances were made for the ROM portion of the questionnaire, and the deformity and instability portions of the score were excluded from analysis. Mahomed *et al.* [32] showed excellent concordance between a self-reported and surgeon-assessed Harris Hip Score, but this score only has limp as a portion of the exam where a surgeon exam might be utilized. Lieberman *et al.* [12] used a visual analogue scale to compare patients' and physicians' evaluations of THA results, and found a disparity (patients rated the result lower than the physician) that typically increased when the patient was not satisfied with the outcome.

Our study sought to determine whether patients could reliably record their own knee ROM and self-report AKS Pain, Knee and Function scores. Most physicians or their assistants typically record knee ROM during clinic follow-ups with either a visual estimate or a goniometer. Unless the same observer is making the

measurement under strictly controlled conditions each time, goniometer measurement is prone to substantial interobserver error [22–29]. A number of studies assessing reliability of goniometric measurements have shown that increases in joint motion should exceed 4–10° to determine true improvement [22, 23, 25].

Of the patient-recorded visual estimates for total ROM arc based on lateral knee photographs, 61% were within 10° of that subsequently recorded by the physician examiner in our study. The mean difference in patient-reported *vs* physician-recorded overall ROM in our study was negligible (–0.02°). Although differences in flexion and extension measurements in our study were statistically significant, these were judged not to be clinically meaningful. McGrory *et al.* [11] sent patients an ROM ‘diagram’ with five flexion positions from 0 to 120° and showed that there were significant differences between the ROM described by patients and that measured by physicians, but did not elucidate further. This difference may reflect different patient demographics, or a more easily understandable pictorial representation of knee motion with more choices in our study.

Miner *et al.* [33] have called into question the use of ROM as an outcome measure, particularly for follow-up beyond 1 yr, showing that WOMAC pain and function scores at that point were correlates of patient satisfaction and perceived improvement in quality of life, but that knee flexion was not. They suggest that it may be sufficient to simply classify patients as having more or less than 95° of motion, an objective that our study supports can be determined by most patients with appropriate visual cues.

A number of studies that have compared patient *vs* physician reporting of function, or symptoms such as pain, have shown substantial differences between the two, typically with the physician reporting better scores than the patient [8–11]. The difference between patient and physician measurements for the AKS Pain scale in our study was 8 points with physicians consistently scoring the patient as having less pain than the patient self reported. That physicians tend to assign better (or lower) pain score ratings, or overestimate pain relief and satisfaction with surgery, has been previously reported following hip arthroplasty as well [12]. This difference might have a number of explanations. Patients and physicians often have different expectations of the surgery, and may reasonably differ as to what is an acceptable outcome. The difference between filling out a survey in private and responding to questions in a physician’s office or the mean 14-day gap between survey and exam may have impacted responses. Patients may be concerned about disappointing their surgeon, and prefer the anonymity of a survey.

Liow *et al.* [19], on the basis of their work on the reliability of the AKS, have stated that if a patient is seen by the same examiner on two separate occasions, a change in the AKS Knee score of <16 points cannot confidently said to be real. Our work has shown that patients and physicians may report significantly different AKS Pain scores in the same time frame. The variation between patient-reported and physician-recorded pain scores in our study was greater than the variation between patient- and physician-recorded ROM. Such variation in an important subset scale of the AKS has substantial impact on this outcome instrument. Similarly, the 8- to 10-point mean variations between patient-reported and physician-recorded AKS Knee and Functional scores are significant. Since 50% of the AKS Knee score is comprised of the Pain subset, the difference in the overall Knee score undoubtedly reflects the Pain subset difference. We caution that the AKS Knee score includes a stability subscale that is based on physician examination (as, of course, is the ROM). In our questionnaire, patients were asked to self-assess their stability, and we recognize this significant limitation to self-reporting the AKS Knee score. Patients may certainly perceive quadriceps weakness, loss of balance or any number of other factors as ‘knee instability.’ Patients may therefore describe a knee as ‘unstable’ in a questionnaire, thus incurring a deduction in their

AKS Knee score, when the physician examiner finds no evidence of true knee instability. Since the AKS Functional score records only walking distance, assistive devices used and how stairs are climbed, however, we believe the comparison here to be valid. Again, patients may overstate the distance walked, exaggerate their stair-climbing ability or minimize their need for assistive devices in an attempt to please their surgeon when asked these questions in his/her presence.

What are the implications of this study for long-term TKA surveillance? It would appear that patients might reasonably be expected to determine their own ROM within parameters that can be defined and illustrated. This can certainly be done with a degree of accuracy suitable for long-term joint health surveillance. We lend further support to the notion that patient-administered questionnaires to assess function after TKA may demonstrate substantial differences from the physician interview and that the former may reflect the ‘real’ situation better than the same questions posed in a surgeon’s office. Although questionnaires have certain limitations (stability testing is one example), they represent an important part of the armamentarium of the clinician-scientist interested in long-term, long-range total joint surveillance. An ideal long-term, long-range TKA surveillance programme would have the following features: (i) a reliable, validated patient questionnaire that was joint-specific; (ii) a reliable, validated patient questionnaire that examined overall function, global health and psychological attributes; (iii) modest ‘responder burden’ for the patient; (iv) physician review of regular radiographs; (v) low cost; (vi) internet capable for both transmission/collection of questionnaires and radiographs; and (vii) widespread use and acceptance by TKA investigators. Although at present this programme does not exist, we remain optimistic that studies such as this one will help us as we move forward to refine the questionnaires and self-measurement tools that are needed.

Rheumatology key messages

- Patients can accurately record their knee ROM using pictures of knee flexion and extension.
- Patient-reported pain scores in home setting and physician-recorded pain scores in office setting differ significantly, leading to differences in AKS scores.
- Long-term surveillance of knee arthroplasty patients using tele-medicine tools and a mailed self-report AKS may be possible.

Acknowledgements

Funding: This material is based upon work supported (or supported in part) by the Office of Research and Development, Department of Veterans Affairs. The effort of J.A.S. was supported by the NIH CTSA Award 1 KL2 RR024151-01 (Mayo Clinic Center for Clinical and Translational Research). The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

Disclosure statement: D.L.P. has received grants for clinical research on joint replacement and honoraria for speaking engagements from DePuy/Johnson & Johnson, Inc. All other authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

- 1 Bach CM, Nogler M, Steingruber IE *et al.* Scoring systems in total knee arthroplasty. *Clin Orthop Relat Res* 2002;399:184–96.
- 2 Davies AP. Rating systems for total knee replacement. *Knee* 2002;9:261–6.

- 3 Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br* 1998;80:63–9.
- 4 Drake BG, Callahan CM, Dittus RS, Wright JG. Global rating systems used in assessing knee arthroplasty outcomes. *J Arthroplasty* 1994;9:409–17.
- 5 Kantz ME, Harris WJ, Levitsky K, Ware JE Jr, Davies AR. Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. *Med Care* 1992;30:MS240–52.
- 6 Konig A, Scheidler M, Rader C, Eulert J. The need for a dual rating system in total knee arthroplasty. *Clin Orthop Relat Res* 1997;161–7.
- 7 Bankes MJ, Coull R, Ferris BD. How long should patients be followed-up after total hip replacement? Current practice in the UK. *Ann R Coll Surg Engl* 1999;81:348–51.
- 8 Basch E, Iasonos A, McDonough T *et al.* Patient versus clinician symptom reporting using the national cancer institute common terminology criteria for adverse events: results of a questionnaire-based study. *Lancet Oncol* 2006;7:903–9.
- 9 Bullens PH, van Loon CJ, de Waal Malefijt MC, Laan RF, Veth RP. Patient satisfaction after total knee arthroplasty: a comparison between subjective and objective outcome assessments. *J Arthroplasty* 2001;16:740–7.
- 10 Cohen SB, Strand V, Aguilar D, Ofman JJ. Patient- versus physician-reported outcomes in rheumatoid arthritis patients treated with recombinant interleukin-1 receptor antagonist (anakinra) therapy. *Rheumatology* 2004;43:704–11.
- 11 McGrory BJ, Freiberg AA, Shinar AA, Harris WH. Correlation of measured range of hip motion following total hip arthroplasty and responses to a questionnaire. *J Arthroplasty* 1996;11:565–71.
- 12 Lieberman JR, Dorey F, Shekelle P *et al.* Differences between patients' and physicians' evaluations of outcome after total hip arthroplasty. *J Bone Joint Surg Am* 1996;78:835–8.
- 13 Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the knee society clinical rating system. *Clin Orthop Relat Res* 1989;248:13–4.
- 14 Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833–40.
- 15 Stewart AL, Hays RD, Ware JE Jr. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care* 1988;26:724–35.
- 16 Ewald FC. The Knee Society total knee arthroplasty roentgenographic evaluation and scoring system. *Clin Orthop Relat Res* 1989;248:9–12.
- 17 Brinker MR, Lund PJ, Barrack RL. Demographic biases of scoring instruments for the results of total knee arthroplasty. *J Bone Joint Surg Am* 1997;79:858–65.
- 18 Lingard EA, Katz JN, Wright RJ, Wright EA, Sledge CB. Validity and responsiveness of the Knee Society clinical rating system in comparison with the SF-36 and WOMAC. *J Bone Joint Surg Am* 2001;83-A:1856–64.
- 19 Liow RY, Walker K, Wajid MA, Bedi G, Lennox CM. The reliability of the American Knee Society score. *Acta Orthop Scand* 2000;71:603–8.
- 20 Liow RY, Walker K, Wajid MA, Bedi G, Lennox CM. Functional rating for knee arthroplasty: comparison of three scoring systems. *Orthopedics* 2003;26:143–9.
- 21 Marx RG, Jones EC, Atwan NC, Closkey RF, Salvati EA, Sculco TP. Measuring improvement following total hip and knee arthroplasty using patient-based measures of outcome. *J Bone Joint Surg Am* 2005;87:1999–2005.
- 22 Boone DC, Azen SP, Lin CM, Spence C, Baron C, Lee L. Reliability of goniometric measurements. *Phys Ther* 1978;58:1355–90.
- 23 Bovens AM, van Baak MA, Vrencken JG, Wijnen JA, Verstappen FT. Variability and reliability of joint measurements. *Am J Sports Med* 1990;18:58–63.
- 24 Brosseau L, Balmer S, Tousignant M *et al.* Intra- and intertester reliability and criterion validity of the parallelogram and universal goniometers for measuring maximum active knee flexion and extension of patients with knee restrictions. *Arch Phys Med Rehabil* 2001;82:396–402.
- 25 Mayerson NH, Milano RA. Goniometric measurement reliability in physical medicine. *Arch Phys Med Rehabil* 1984;65:92–4.
- 26 Edwards JZ, Greene KA, Davis RS, Kovacic MW, Noe DA, Askew MJ. Measuring flexion in knee arthroplasty patients. *J Arthroplasty* 2004;19:369–72.
- 27 Kafer W, Fraitzl CR, Kinkel S, Clessienne CB, Puhl W, Kessler S. [Outcome assessment in total knee arthroplasty: is the clinical measurement of range of motion a reliable measurable outcome variable?]. *Z Orthop Ihre Grenzgeb* 2005;143:25–9.
- 28 Shultz SJ, Nguyen AD, Windley TC, Kulas AS, Botic TL, Beynon BD. Intratester and intertester reliability of clinical measures of lower extremity anatomic characteristics: implications for multicenter studies. *Clin J Sport Med* 2006;16:155–61.
- 29 Walkins MA, Riddle DL, Lamb RL, Personius WJ. Reliability of goniometric measurements and visual estimates of knee range of motion obtained in a clinical setting. *Phys Ther* 1991;71:90–6; discussion 6–7.
- 30 Cella D, Yount S, Rothrock N *et al.* The patient-reported outcomes measurement information system (promis): progress of an NIH roadmap cooperative group during its first two years. *Med Care* 2007;45:S3–11.
- 31 Wasson J, Gaudette C, Whaley F, Sauvigne A, Baribeau P, Welch HG. Telephone care as a substitute for routine clinic follow-up. *J Am Med Assoc* 1992;267:1788–93.
- 32 Mahomed NN, Arndt DC, McGrory BJ, Harris WH. The Harris Hip Score: comparison of patient self-report with surgeon assessment. *J Arthroplasty* 2001;16:575–80.
- 33 Miner AL, Lingard EA, Wright EA, Sledge CB, Katz JN. Knee range of motion after total knee arthroplasty: how important is this as an outcome measure? *J Arthroplasty* 2003;18:286–94.