Correlations between Knee Society Function Scores and Functional Force Measures

Cale A. Jacobs PhD, Christian P. Christensen MD

Received: 17 July 2008 / Accepted: 10 March 2009 / Published online: 3 April 2009
© The Association of Bone and Joint Surgeons 2009

Abstract Subjective evaluations often are used after knee arthroplasty to quantify function; however, these scores may be influenced by pain and/or function of the nonoperated limb. Multiple influences increase variability of these scores, which in turn may result in a greater change in score required to be considered clinically important. We determined the relationships among the Knee Society pain and function scores, range of motion (ROM), and functional force measures of the surgically treated and nonoperated limbs. Before and 3 months after total or unicompartmental knee arthroplasty, 36 patients answered questions necessary to calculate the Knee Society pain and function scores. A dual-force platform was used to record the lift-up force of each limb during a stepping task. Function scores were correlated to pain scores, lift-up force of the nonoperated limb, and ROM before surgery. After

surgery, function scores correlated with pain scores but not with objective functional measures or ROM. As patient-reported function scores and functional force measures of the surgically treated limb seem to provide distinctly different information, both measures may need to be collected after knee arthroplasty to fully understand a patient's functional recovery.

Introduction

The largest source of physical disability in the United States is osteoarthritis (OA) [9], with the knee most commonly affected [4]. An evaluation of outcomes after treatment of OA, whether operative or nonoperative, is necessary to fully understand the efficacy of any intervention. To do so, the Outcome Measures in Arthritis Clinical Trials conference (OMERACT) III recommended clinicians and researchers use tools that provide separate assessments of pain and physical function [1].

Although pain and physical function are inherently related, they are not the same. Patients undergoing TKA commonly report considerable improvements in pain ratings during the early postoperative period [7, 16–18]. In contrast, the use of functional testing in the clinical setting has identified functional deficits including decreased speed when walking or climbing stairs despite pain relief 1 year after TKA [25]. The relationships among physical activity, functional performance, and patient perceptions of pain and function have been examined. Thomas et al. [24] reported performance of functional tests by patients with end-stage knee OA yielded low correlations ($r = 0.31$ – 0.33) with patient-reported physical activity. Stratford and Kennedy [21] reported pain, and not performance of functional tests, was the principal determinant of the WOMAC Physical

One of the authors (CPC) has received funding from Biomet Inc, Warsaw, IN.

Each author certifies that his or her institution has approved the human protocol for this investigation, that all investigations were conducted in conformity with ethical principles of research, and that informed consent for participation in the study was obtained. This study was performed at ERMI, Inc, Atlanta, GA.

C. A. Jacobs (✉)
Research and Development, ERMI, Inc, 765 Sentry Ridge Xing,
Suwanee, Atlanta, GA 30024, USA
e-mail: calejacobs@hotmail.com

C. A. Jacobs
Division of Athletic Training, University of Kentucky,
Lexington, KY, USA

C. P. Christensen
Department of Orthopedics and Sports Medicine, Lexington
Clinic, Lexington, KY, USA

Function subscale and suggested patient-reported outcome and functional performance measures are necessary to better evaluate patients with OA.

Although it is becoming increasingly clear patient-reported outcome measures differ from clinical measures of functional performance, it remains unclear if the clinical tests used by previous authors truly provide information about the functional ability of the involved limb. Mizner and Snyder-Mackler [15] reported performance of functional tests after TKA was more related to strength of the uninvolved limb than of the surgical limb. With this in mind, it may be necessary to not only evaluate performance of functional tasks but also to do so using measurement techniques that allow individual evaluation of the involved and uninvolved limb. We presumed the Knee Society function score would be more related to pain and function of the nonoperative limb before and after knee arthroplasty and the Knee Society function score would show larger minimum detectable change than the functional force measure of the surgical limb.

To better evaluate functional recovery, we therefore (1) determined the relationships among the Knee Society pain and function scores, ROM, and individual functional force measures of the surgically treated and nonoperated limbs before and after knee arthroplasty; and (2) calculated the minimum detectable change for the patient-reported function scores and the functional force measures.

Materials and Methods

We recruited 73 patients with unilateral end-stage OA scheduled to undergo either unicompartmental knee arthroplasty (UKA) or TKA from among 316 patients having UKA or TKA from January 2007 to February 2008. Patients were not recruited to participate if they had sensory, neurologic, vestibular, or other health conditions that may affect balance, to protect against falls during functional testing. Patients who were unable to understand the questions used to obtain the pain or function score also were excluded. Of the 73 patients recruited, 43 chose to participate. Of these 43 patients, five had bilateral disease and were excluded from this analysis, and two had cancelled surgery after performing the preoperative data collection. The remaining 36 participants consisted of seven men and 29 women with a mean age of 67.4 ± 8.4 years and a mean body mass index of 29.1 ± 6.4 kg/m². Seven patients were treated with fixed-bearing UKA (Vanguard MTM; Biomet, Warsaw, IN), 18 with a mobile-bearing UKA (Oxford[®]; Biomet), and 11 with a fixed-bearing TKA (VanguardTM; Biomet). All surgeries were performed by one surgeon (CPC), and all patients were treated with identical perioperative pain control and rehabilitation protocols. We

obtained informed consent from all patients to participate in this Institutional Review Board-approved protocol.

We quantified patient-reported functional ability by calculating the pain and function subcomponents of the Knee Society score at each patient's routine preoperative and 3-month postoperative visit [5, 6]. The Knee Society score questionnaire used in this study included the 1993 modifications recommended by Dr. John Insall [6]. To calculate the pain score, patients were asked to rate their pain when walking, navigating stairs, and at rest as none, mild, moderate, or severe. The specific questions used to calculate the function score involved the patient's ability to walk without an assistive device, walking distance, and the ability to navigate stairs and rise from a seated position without upper extremity support.

Knee ROM was measured preoperatively and postoperatively using long-arm goniometry by one evaluator (CAJ). The intratester reliability of the ROM measurements were excellent, with intraclass correlation coefficients ($ICC_{2,1}$) = 0.92 (CAJ, unpublished data, 2006). The ICC statistic is commonly used to determine reliability of a measure [19], and Currier [2] suggested the following classification scheme for interpreting the magnitude of the reliability coefficients: 0.90 to 0.99 = excellent reliability, 0.80 to 0.89 = good reliability, 0.70 to 0.79 = fair reliability, and < 0.70 = poor reliability.

The functional testing required patients to perform two to three repetitions of the step-up- and-over tests on the Balance Master[®] long force plate (Neurocom, Inc, Clackamas, OR) (Fig. 1) per the manufacturer's instructions [12]. This test was included because it specifically mimics one of the questions used to calculate the patient-reported function score, which inquires about the patient's ability to ascend and descend stairs. We used the step-up-and-over test to quantify the patient's ability to ascend and descend a small step (Fig. 1). A box with a height of 10.2 cm was placed on the force plate, and patients were instructed to step up onto the box with the test leg as fast as they felt comfortable. They then were instructed to bring the other foot onto the box, momentarily resting with both feet on the box. We then instructed patients to step down with the nontest leg as gently as possible and then bring the test leg down onto the force plate. During the ascent onto the box, the force created by the test leg was measured and normalized to the patient's body weight (% BW). This value was termed the lift-up index and has been reliable in our facility with $ICC_{2,1}$ = 0.82. Greater lift-up indices have been associated with increased concentric ability of the test leg [12].

Participants were not asked to perform the functional test if they were unable to perform that test without upper extremity support, and four patients were unable to perform these tests at their postoperative followup. Patients were allowed to perform as many practice trials of the step-up-

Fig. 1A–B The images show a patient performing the step-up-and-over test. A 10.2-cm-high box is placed on a force plate. (A) The patient steps onto the box with the test leg, brings the other foot onto the box, (B) steps down with the nontest leg, and then brings the test leg down onto the force plate.

and-over test as they needed to become comfortable with the test. After completing the practice trials, patients performed a minimum of two test trials with each limb and the mean lift-up indices were calculated.

All variables, including the noncontinuous pain and function scores, passed Levine's test for homogeneity ($p > 0.05$), therefore allowing us to use parametric analyses. Pearson's product moment correlation coefficients (r) were performed using SPSS[®] Version 17.0 (SPSS Inc, Chicago, IL) to compare the relationships among the function score, pain score, ROM, and lift-up indices of the involved and uninvolved limbs. In addition, the minimum detectable change was calculated for the dependent variables using Microsoft[®] Excel[®] 2003 (Microsoft Corp, Redmond, WA). The minimum detectable change is useful to clinicians and researchers, as it allows one to interpret the clinical relevance of results in studies on the effectiveness of a given treatment [8]. Because of the partial dependence on sample size, statistical significance of a treatment effect does not always correspond to the clinical relevance of the effect [3]. Statistical significance is founded on the premise that the effect occurs beyond some level of chance. Clinical relevance, however, is related more to the benefits derived from a given intervention or treatment [3]. By computing the minimum detectable change of a dependent variable, the reader then may determine the amount of change of that variable that will be required to be considered clinically important. For example, take a patient whose Knee Society score increases from 40 (of 100) before arthroplasty to 50 1 year after the procedure. Although this difference may be

statistically significant when included as part of a larger data set, would we as clinicians consider this 10-point change clinically important? We chose to determine the minimum detectable change for the Knee Society function score and the lift-up index to provide insight regarding the amount of change to be considered clinically important after knee arthroplasty and also to determine whether patient-reported functional outcomes or objective functional testing were more sensitive to relatively smaller changes after surgery.

Results

Before and after surgery, the function score did not correlate with the lift-up index of the involved limb. Function scores were correlated to pain scores ($r = 0.49$), lift-up force of the nonoperated limb ($r = 0.52$), and ROM ($r = 0.34$) before surgery. After surgery, function scores correlated with pain scores ($r = 0.45$) but not with objective functional measures or ROM (Table 1). Lift-up index of the involved limb was not related to pain score either preoperatively ($r = 0.21$; $p = 0.29$) or postoperatively ($r = 0.11$; $p = 0.56$).

The minimum detectable change value for the patient-reported function score was 34.5 points, and the minimum detectable change for the lift-up index of the involved limb was 8.1% BW. The minimum detectable change is a function of the standard deviation and standard error of measurement of a variable, suggesting, based on the

Table 1. Preoperative and postoperative Pearson's product moment correlation coefficients (r)

Dependent variable	Preoperative r	p Value	Postoperative r	p Value
Knee Society pain score	0.49	0.003	0.45	0.01
Involved lift-up index	−0.05	0.82	0.14	0.45
Uninvolved lift-up index	0.52	0.01	0.25	0.17
Range of motion	0.34	0.04	0.22	0.21

Table 2. Preoperative and postoperative (mean ± standard deviation) results

Dependent variable	Preoperative	Postoperative	p Value
Knee function score (/100)	*42.8 ± 20.9	*66.1 ± 26.7	< .001
Knee pain score (/50)	*19.2 ± 17.8	*44.0 ± 26.7	< .001
†Involved lift-up index (%BW)	*19.6 ± 7.1	*21.6 ± 6.9	.01
†Uninvolved lift-up index (%BW)	*20.7 ± 7.2	21.7 ± 7.2	.03
Range of motion	120.0 ± 11.6	121.1 ± 9.0	.71

* Indicates statistical significance at $p \leq 0.05$; †lift-up indices expressed as a percent of body weight (% BW).

measured variability of these measures, changes of 34.5 points of 100 possible points in the function score and 8.1% BW of a possible 100% BW in lift-up index are necessary to be considered clinically important. However, despite statistically significant increases in function following surgery, the magnitude of change of neither the function score nor the lift-up index was large enough to meet the minimum detectable change. In other words, the statistically significant increases in these variables may not be clinically meaningful (Table 2).

Discussion

Subjective evaluations often are used after knee arthroplasty to quantify function; however, these scores may be influenced by pain and/or function of the nonoperated limb. Multiple influences increase variability of these scores, which in turn may result in a greater change in score required to be considered clinically important. We determined the relationships among the Knee Society pain and function scores, ROM, and individual functional force measures of the surgically treated and nonoperated limbs before and after knee arthroplasty and the minimum detectable change of the patient-reported and objective functional measures. We hypothesized the Knee Society function score would be more related to pain and function of the nonoperated limbs before and knee arthroplasty and the Knee Society function score would show larger minimum detectable change than the functional force measure of the surgical limb.

This study was not without limitations. We used the Knee Society scoring system that included Dr. Insall's 1993 modifications because it was the most current version

of the scoring system available when the study was initiated [5, 6]. However, at the time of writing, the Knee Society has formed a task force to develop and validate a new function score that would be compatible with the prior version of the Knee Society scoring system [23]. With more recent doubts of this scoring system, it may not be appropriate to generalize our results to other outcome scoring systems. However, by testing a specific task included in the patient-reported Knee Society score questionnaire, we believe our results suggest patient reports of function do not correlate to functional performance measures of the involved limb.

Previous studies have suggested patient-reported function scores are related to performance of functional tests, using correlations ranging from 0.4 to 0.6 to support this conclusion [10, 13, 20]. Other studies also have compared patient-reported functional outcome scores and performance of functional tests [11, 21]. Maly et al. [11] and Stratford and Kennedy [21] compared the physical function components of the SF-36 [11], WOMAC [11, 21], and Lower Extremity Function Score [21] with patients' abilities to perform the timed up-and-go test, stair test, and 6-minute walk test [11, 21]. Patients were timed during the timed up-and-go test as they rose from a standard arm chair, walked 3 m, turned around, walked back to the chair, and returned to a seated position [11, 21]. During the stair test, patients were timed as they ascended and descended five steps [11, 21]. For the 6-minute walk test, the distance a patient was able to walk in 6 minutes was recorded [11, 21]. Both groups of authors concluded patient-reported functional outcomes were more related to pain than to functional performance tests and suggested collection of patient-reported function scores and objective functional performance data may be necessary [11, 21]. Our study

showed similar results with the Knee Society function score, as function scores had greater correlation coefficients with pain and the lift-up index of the uninvolved limb during the step-up-and-over test.

Despite providing clinicians with objective measures of function, performance of clinical functional tests such as those used by Maly et al. [11] and Stratford and Kennedy [21] have been reported to be more related to strength of the uninvolved limb [15]. This potentially may limit the ability to generalize results to functional recovery of the surgically treated limb. This also suggests that to gain better understanding of functional recovery of the surgically treated limb, patient-reported functional outcome scores and unilateral functional performance tests may be required.

Our minimum detectable change results also suggest this. In calculating the minimum detectable change, increased variability and/or reduced reliability of a measure result in larger minimum detectable change values, meaning a greater amount of change will be necessary to be considered clinically important for measures with either high variability or poor reliability. Despite showing considerable improvement 3 months after surgery, the minimum detectable change of the function score was 34.5 points, suggesting only a change greater than this value should be considered clinically important. Our patients had considerably improved function scores, with mean function score increasing from 44.1 to 63.6 after surgery. It has been suggested patients tend to report higher function than is observed during functional testing, which subsequently may lead healthcare professionals to overestimate functional ability if they rely solely on outcome scoring [22]. Our patients showed only modest improvements in our measure of concentric ability of the involved limb, with the mean lift-up index increasing from 18.9% BW preoperatively to 21.4% BW postoperatively. The magnitude of postoperative changes in lift-up index we observed is consistent with the time of recovery of quadriceps strength and function reported by Mizner et al. [14]. Considering the smaller relative minimum detectable change values of the lift-up index and the fact that it more closely mimics the time of muscle and functional recovery without undue influence of pain, we suggest this single-limb measure may be better suited to detect changes of smaller magnitude that may still have clinical relevance.

Our data support the suggestion of Stratford and Kennedy [21] that better understanding of functional outcomes can be gained by collecting functional performance measures and patient-reported outcome measures. Our results also support the concept that patient-reported functional outcome measures may be more related to pain or function of the uninvolved limb, which necessitates the ability to individually measure function of each limb.

Whether evaluating the efficacy of perioperative pain control protocols, changes in surgical technique or implant design, or modifications to postoperative rehabilitation protocols, we believe studies should use patient-reported outcome questionnaires and objective measures of function, which seem better suited to detect changes of small magnitude that may be considered clinically important or relevant.

References

1. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, Altman R, Brandt K, Dougados M, Lequesne M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis: consensus development at OMERACT III. *J Rheumatol*. 1997; 24:799–802.
2. Currier D. *Elements of Research in Physical Therapy*. Baltimore, MD: Williams and Wilkins; 1990.
3. de Vet H, Terwee C, Ostelo R, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes*. 2006;5:54.
4. Felson D, Lawrence R, Hochberg M, McAlindon T, Dieppe PA, Minor MA, Blair SN, Berman BM, Fries JF, Weinberger M, Lorig KR, Jacobs JJ, Goldberg V. Osteoarthritis: new insights. Part 2. treatment approaches. *Ann Intern Med*. 2000;133:726–737.
5. Insall J, Dorr L, Scott R, Scott W. Rationale of the Knee Society clinical rating system. *Clin Orthop Relat Res*. 1989;248:13–14.
6. Insall J, Dorr L, Scott R, Scott W. Rationale of The Knee Society clinical rating system. Available at: http://www.kneesociety.org/pdfs/knee_society_rationale_article.pdf. Accessed January 27, 2009.
7. Jones CA, Voaklander DC, Suarez-Alma ME. Determinants of function after total knee arthroplasty. *Phys Ther*. 2003;83:696–706.
8. Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Tomás M, Mir MA, Cano A, Muriel A, Zamora J, Del Real MT, Gestoso M, Mufraggi N; Spanish Back Pain Research Network. Minimum detectable and minimally clinically important changes for pain in patients with nonspecific neck pain. *BMC Musculoskelet Disord*. 2008;9:43–44.
9. Lawrence R, Helmick C, Arnett F, Arnett FC, Deyo RA, Felson DT, Giannini EH, Heyse SP, Hirsch R, Hochberg MC, Hunder GG, Liang MH, Pillemer SR, Steen VD, Wolfe F. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. *Arthritis Rheum*. 1998;41:778–799.
10. Lin YC, Davey RC, Cochrane T. Tests for physical function of the elderly with knee and hip osteoarthritis. *Scand J Med Sci Sports*. 2001;11:280–286.
11. Maly MR, Costigan PA, Olney SJ. Determinants of self-report outcome measures in people with knee osteoarthritis. *Arch Phys Med Rehabil*. 2006;87:96–104.
12. Mattacola CG, Jacobs CA, Rund MA, Johnson DL. Functional assessment using the step-up-and-over test and forward lunge following ACL reconstruction. *Orthopedics*. 2004;27:602–608.
13. McCarthy CJ, Oldham JA. The reliability, validity and responsiveness of an aggregated locomotor function (ALF) score in patients with osteoarthritis of the knee. *Rheumatology (Oxford)*. 2004;43:514–517.

14. Mizner RL, Petterson SC, Snyder-Mackler L. Quadriceps strength and the time course of functional recovery after total knee arthroplasty. *J Orthop Sports Phys Ther.* 2005;35:424–436.
15. Mizner RL, Snyder-Mackler L. Altered loading during walking and sit-to-stand is affected by quadriceps weakness after total knee arthroplasty. *J Orthop Res.* 2005;23:1083–1090.
16. Ritter MA, Thong AE, Davis KE, Berend ME, Meding JB, Faris PM. Long-term deterioration of joint evaluation scores. *J Bone Joint Surg Br.* 2004;86:438–442.
17. Salmon P, Hall GM, Peerbhoy D, Shenkin A, Parker C. Recovery from hip and knee arthroplasty: patients' perspective on pain, function, quality of life, and well-being up to 6 months postoperatively. *Arch Phys Med Rehabil.* 2001;82:360–366.
18. Shields RK, Enloe LJ, Leo KC. Health related quality of life in patients with total hip or knee replacement. *Arch Phys Med Rehabil.* 1999;5:572–579.
19. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420–428.
20. Steultjens MP, Dekker J, van Baar ME, Oostendorp RA, Bijlsma JW. Internal consistency and validity of an observational method for assessing disability in mobility in patients with osteoarthritis. *Arthritis Care Res.* 1999;12:19–25.
21. Stratford PW, Kennedy DM. Performance measures were necessary to obtain a complete picture of osteoarthritic patients. *J Clin Epidemiol.* 2006;59:160–167.
22. Stratford PW, Kennedy DM, Woodhouse LJ. Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee. *Phys Ther.* 2006;86:1489–1496; discussion 1496–1500.
23. The Knee Society. Outcomes tools. Available at: <http://www.kneesociety.org/index.asp/fuseaction/site.outcomes>. Accessed January 27, 2009.
24. Thomas SG, Pagura SM, Kennedy D. Physical activity and its relationship to physical performance in patients with end stage knee osteoarthritis. *J Orthop Sports Phys Ther.* 2003;33:745–754.
25. Walsh M, Woodhouse LJ, Thomas SG, Finch E. Physical impairments and functional limitations: a comparison of individuals 1 year after total knee arthroplasty with control subjects. *Phys Ther.* 1998;78:248–258.