The Journal of Arthroplasty Vol. 22 No. 8 2007

Comparison of the Responsiveness of the SF-36 and WOMAC in Patients Undergoing Total Hip Arthroplasty

Nelson Fong SooHoo, MD, Raj M. Vyas, BS, David B. Samimi, BA, Ricardo Molina, MD, and Jay R. Lieberman, MD

Abstract: This study examines the responsiveness of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and Short Form-36 (SF-36) in patients undergoing total hip arthroplasty. Eighty-nine patients completed the WOMAC and SF-36 preoperatively and postoperatively. Standardized response means (SRMs) and effect sizes (ES) were used to measure responsiveness. Mean follow-up was 17 months. The SRMs for the WOMAC ranged from −0.93 to −1.49, and the ES ranged from −1.02 to −1.53. The SRMs for the SF-36 ranged from 0.22 to 1.64, and the ES ranged from 0.20 to 1.97. The highest values occurred with the physical functioning, bodily pain, and Physical Component Summary Scales. This study demonstrates a similar level of responsiveness of the WOMAC and several components of the SF-36. This suggests that the isolated use of the SF-36 may be adequate to monitor outcomes after total hip arthroplasty. There may still be a role for the WOMAC when comparing outcomes of specific designs or techniques of total hip arthroplasty.
Key words: outcomes, WOMAC, SF-36, total hip arthroplasty, responsiveness.

There has been significant uncertainty as to which outcomes tools should be used to report the results of treatment for patients after total hip arthroplasty. Several outcomes assessment instruments have been proposed for use in monitoring total joint patients. These include generic instruments, such as the Medical Outcomes Study Short Form-36 (SF-36), which are designed for broad use in a variety of medical conditions, as well as more specialized questionnaires, such as the Western Ontario and McMaster Universities Osteoarthritis Index [1-6]. The SF-36 is a 36-item questionnaire that generates subscale scores for 8 domains of health as well as summary scores for overall physical and mental functioning [7,8]. The WOMAC is a 24-item questionnaire that can be analyzed as a global score or as 3 subsection scores [9].

The criteria for selecting among these instruments are their validity, reliability, and responsiveness in evaluating the health of the targeted population [10-13]. The validity and reliability of the SF-36 and WOMAC have been addressed in previous studies. Despite being extensively validated and tested for reliability, the SF-36 is designed for use as a generic tool and has been shown to have lower levels of responsiveness to clinical change in patients with orthopaedic disorders than disease or region-specific tools [9,14-21]. More focused instruments such as the WOMAC have been designed to increase responsiveness and

From the Department of Orthopaedic Surgery, UCLA School of Medicine, Los Angeles, California.
Submitted December 29, 2005; accepted October 18, 2006.
No benefits or funds were received in support of the study.
Reprint requests: Nelson Fong SooHoo, MD, UCLA Department of Orthopaedic Surgery, 10945 Le Conte Ave., PVUB #3355, Los Angeles, CA 90095.

0883-5403/07/2208-0014$32.00/0
doi:10.1016/j.arth.2006.10.006

sensitivity to clinical change in patients with arthritic conditions [9]. However, previous studies have indicated that the relative responsiveness of the WOMAC in comparison with the SF-36 may not be consistently higher in patients undergoing total hip arthroplasty [3,21].

The purpose of this study is to compare the responsiveness to clinical change of the WOMAC with the SF-36 in a group of patients undergoing total hip arthroplasty by calculating the standardized response means (SRMs) and effect size (ES) for each questionnaire. The SRM and ES are commonly used measures of the sensitivity of an outcomes tool to clinical change [10-13]. Our hypothesis was that the WOMAC would have improved responsiveness as demonstrated by a higher SRM and ES than the SF-36 in patients after total hip arthroplasty. This would support the continued use of disease-specific questionnaires in conjunction with generic tools to follow the clinical outcomes of these patients. Conversely, similar levels of responsiveness of at least some of the SF-36 scales to the WOMAC would suggest that the isolated use of the SF-36 may be adequate to monitor outcomes in patients undergoing total hip arthroplasty. This would decrease the questionnaire load administered to patients and allow direct comparison to other health conditions without sacrificing adequate sensitivity to clinical change.

Materials and Methods

Human Subjects

This project was approved by the Institutional Review Board at the UCLA School of Medicine. Informed consent was obtained from all patients before participation.

Patient Sample

Patients were recruited through the joint replacement clinic at the UCLA School of Medicine. Patients were included if they presented with a chronic arthritic condition requiring total hip arthroplasty. Patients were excluded if their indication for surgery was acutely traumatic or due to fracture. In addition, revision arthroplasty patients were excluded. Other inclusion criteria included the ability to read English, age more than 18 years unless accompanied by a parent, and consent to participate. Patients unable to read English were excluded. Patients were not excluded on the basis of race or ethnicity.

Data Collection

Recruited patients were administered a packet including an Informed Consent Form approved by the Institutional Review Board, the Medical Outcomes Study Short Form-36 version 2 (SF-36), and the WOMAC. The patient packet was administered during the initial patient visit for all enrolled subjects. The same questionnaires were then administered again at subsequent postoperative visits for each patient. Questionnaires were checked for completion by one of the investigators, and patients were assisted in completing missing items.

Data Analysis

Raw data for the SF-36 and WOMAC were recorded using Microsoft Excel 2002 (Microsoft, Redmond, Wash, 2001). The WOMAC and SF-36 were scored using standard scoring techniques [7-9]. A global WOMAC score was calculated as well as scores for the 3 subscores relevant to pain, stiffness, and function [9]. The SF-36 data were scored using the SF-36 Health Outcomes Scoring Software Version 1.0 (QualityMetric Incorporated, Lincoln, RI, 2003). This software package generates output for all 8 subscales of the SF-36 in addition to the Physical (PCS) and Mental Component Summary (MCS) Scales.

Missing data were adjusted as appropriate for each outcomes tool. There is no established technique for missing data estimation with the WOMAC scales, and patients with incomplete data points for this tool were excluded from further study [9]. The SF-36 software scored missing data using standard algorithms. The SF-36 will allow complete scoring for patients with a limited number of missing items. More extensive missing data lead to incomplete scoring of some subscales or summary scores [7,8].

Statistical Analysis

The mean value for each of the scales was calculated both preoperatively and postoperatively. The mean change in score for each scale is also reported. The SRM and ES were calculated for each instrument including the global WOMAC score, the 3 WOMAC subscores, the 8 SF-36 subscales, and the 2 SF-36 summary scales. Microsoft Excel 2002 (Microsoft) was used for all statistical comparisons. The SRM and ES are accepted measures of responsiveness [10-13]. Higher absolute values indicate instruments more responsive to clinical change, whereas lower values reflect less sensitivity to underlying changes in health status. The SRM is

calculated as the mean change in scores divided by the standard deviation of these changes. The ES is calculated as the mean change in scores divided by the standard deviation of the preoperative scores. Small effects were considered higher than 0.20; moderate effects, higher than 0.50; and large effects, higher than 0.80 [10-13].

Results

Patient Sample

Eighty-nine patients were recruited and had adequate data to allow complete scoring of the WOMAC and SF-36 for the preoperative visit and minimum 5-month postoperative follow-up. The patients included in the study consisted of 48 women (54%) and 41 men (46%). The mean age of the patient sample was 60 years (range, 20-91 years).

The mean follow-up period between surgery and final questionnaire administration was 17 months (range, 5-43 months). Thirty-three patients completed follow-up questionnaires within 1 year of surgery. There were 33 patients who received follow-up during the second postoperative year, 16 during the third postoperative year, and the remaining 7 patients completed packets during the fourth postoperative year.

Baseline Scores

Mean baseline scores were calculated for the patient sample with results shown in Table 1 for the WOMAC scores, 8 SF-36 subscales, and 2 SF-36 summary scales.

Mean raw scores for each of the 8 SF-36 subscales are shown in Table 1. The 8 subscales include physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional (RE), and mental health (MH). These scales are scored from 0 to 100, with higher scores indicating better health status. The mean scores ranged from a high of 67 for the MH subscale to a low of 23 for the PF subscale.

Mean baseline scores are also reported for the PCS and MCS Scales of the SF-36 in Table 1. The summary scales are based on established norms, with scores calculated as variations from a reference value of 50 for healthy populations. Each 10 points higher or lower is consistent with one standard deviation from the normal population reference value of 50 [8]. The PCS Scale had a mean value of 31 for this patient sample. This is approximately 2 standard deviations below the mean for healthy subjects. In contrast, the mean MCS Scale score of 50 in this group of patients is equivalent to the reference value in healthy patients. The MCS and PCS scores seen at baseline in the present study are consistent with those in prior studies of the health effects of osteoarthritis and rheumatoid arthritis [22,23]. These findings are also similar to prior studies that indicate hip pathology has larger effects on the physical components than the mental components of health status measured by the SF-36 [24,25].

Table 1. Mean Baseline Scores, Mean Follow-Up Scores, and Mean Change in Scores for the WOMAC and SF-36 (n = 89)

Outcomes Scale	Mean Baseline Scores	Mean Follow-Up Scores	Mean Change in Scores
WOMAC global score	51	19	−32
WOMAC pain score	10	3	−8
WOMAC stiffness score	4	2	−2
WOMAC function score	36	15	−22
SF-36 subscales			
PF	23	61	37
RP	24	58	34
BP	30	61	31
GH	66	61	−6
VT	49	60	11
SF	58	77	20
RE	56	74	18
MH	67	76	9
SF-36 summary scales			
PCS Scale	31	42	12
MCS Scale	50	52	2

The mean global WOMAC score was 51 of a possible 96 points. Higher WOMAC scores are indicative of worse health. The mean WOMAC pain score was 10 of a possible 20, the mean WOMAC stiffness score was 4 of a possible 8, and the mean WOMAC function score was 36 of a possible 68 points. Normative values have not been established for the WOMAC scores.

Postoperative Scores

The mean postoperative scores and the mean change in scores are shown in Table 1. The mean change in score for the WOMAC global scale was a decrease of 32 points, which is consistent with an improvement in health status. Similar results were seen with improvement in the WOMAC scales for pain (mean change, −8), stiffness (mean change, −2), and function (mean change, −22).

Table 2. Standardized Response Means and ESs for the WOMAC and SF-36 (n = 89)

Outcomes Scale	SRM	ES
WOMAC global score	−1.49	−1.51
WOMAC pain score	−1.40	−1.53
WOMAC stiffness score	−0.93	−1.02
WOMAC function score	−1.37	−1.37
SF−36 subscales		
PF	1.64	1.97
RP	0.66	0.92
BP	1.12	1.48
GH	−0.34	−0.22
VT	0.53	0.49
SF	0.70	0.70
RE	0.35	0.39
MH	0.50	0.39
SF-36 summary scales		
PCS Scale	1.14	1.46
MCS Scale	0.22	0.20

Some of the SF-36 subscales showed similar improvement in scores when comparing preoperative and postoperative scores. The mean changes were highest for the PF (mean change, 37 points), RP (mean change, 34), and bodily pain (mean change, 31) subscales. The GH (mean change, −6), MH (mean change, 9), and VT (mean change, 11) subscales showed the least amount of change. The PCS Scale showed a higher mean change with improvement by 12 points as compared to a mean increase of 2 points for the MCS Scale.

Responsiveness Testing

The SRMs and ESs for each instrument are shown in Table 2. The WOMAC global score had an SRM of −1.49 and ES of −1.51. The WOMAC subscales also all had high levels of responsiveness including the pain (SRM, −1.40; ES, −1.53), stiffness (SRM, −0.93; ES, −1.02), and function (SRM, −1.37; ES, −1.37) components.

The SF-36 subscales and summary scales had a wider range of SRMs and ESs, with some subscales having high levels of responsiveness and other subscales demonstrating low levels of responsiveness. In general, higher SRMs and ES were seen in the scales that are overweighted in the PCS Scale. These include the PF (SRM 1.64/ES 1.97), bodily pain (SRM 1.12/ES 1.48), RP (SRM 0.66/ES 0.92), and GH (SRM −0.34/ES −0.22) subscales. Lower SRMs and ES were seen in most of the subscales overweighted in the MCS Scale. These include the MH (SRM 0.50/ES 0.39), VT (SRM 0.53/ES 0.49), RE (SRM 0.35/ES 0.39), and SF (SRM 0.70/ES 0.70) subscales. These differences resulted in a higher responsiveness of the PCS Scale (SRM 1.14/ES 1.46) compared with the MCS Scale (SRM 0.22/ES 0.20).

Discussion

This study demonstrated an adequate level of responsiveness to clinical change after total hip arthroplasty of several components of the SF-36 when compared with the disease-specific WOMAC. Specifically, there was evidence that the physical functioning subscale, PCS Scale, and bodily pain subscale of the SF-36 have a level of responsiveness similar to that seen with the WOMAC in patients undergoing total hip arthroplasty. This suggests that the isolated use of the SF-36 may be sufficient to follow outcomes in these patients while maintaining adequate sensitivity to clinical change. This would allow for a reduction in the questionnaire load administered to patients while using an outcomes instrument tested for validity and reliability in a variety of health conditions [7,8]. Therefore, the isolated use of the SF-36 may be a logical approach to simplify data collection and make monitoring of functional outcomes more feasible for practicing orthopedic surgeons.

Several studies have examined the responsiveness of generic tools such as the SF-36 and more specific instruments in patients with orthopaedic conditions. The SRMs seen in this study for the WOMAC global score, WOMAC pain scale, WOMAC function scale, SF-36 physical functioning and bodily pain subscales, and SF-36 PCS Scale were all within the range of 1.12 to 1.64, and the ESs were in the range of 1.37 to 1.97. These values are above the threshold of 0.7, which is commonly used for classifying a questionnaire as having adequate responsiveness [3,10]. In addition, these SRMs and ESs are of a similar magnitude to those seen with outcomes tools used for studies of patients undergoing treatment of shoulder disorders, orthopedic trauma, hip fractures, and radiculopathy [18,26-28].

The levels of sensitivity to clinical change reported in this study are similar to those seen in prior comparisons of the SF-36 and WOMAC over short-term follow-up of total hip arthroplasty [3,21]. Quintana et al [3] noted ESs ranging from 1.61 to 2.10 for the WOMAC subscales, with SRMs in the range of 1.39 to 1.80 after total hip arthroplasty. As in our study, adequate levels of responsiveness were seen in that study for the physical functioning (ES, 1.54; SRM, 1.10) and the bodily pain (ES, 1.15; SRM, 0.91) subscales of the SF-36 [3]. A separate study by Marx et al [21] also showed similar results with an SRM of 1.5 for the PCS Scale in patients after

total hip arthroplasty. Our study confirms and expands on these previous findings by demonstrating a high level of responsiveness of the WOMAC, the physical functioning, and bodily pain subscales of the SF-36 and the SF-36 PCS Scale in patients at a mean of 17 months after total hip arthroplasty. These results indicate that the WOMAC and selected components of the SF-36 all have an acceptable level of responsiveness to clinical change in patients undergoing total hip arthroplasty.

One weakness of the current study is that it does not address the responsiveness of the WOMAC and SF-36 in patients receiving treatments other than total hip arthroplasty for arthritic conditions of the hip. The isolated use of the SF-36 may not be adequate to detect the smaller changes in health status expected in patients receiving nonoperative management. In addition, this approach may not be adequate to compare small differences in health status between patients undergoing total hip arthroplasty using different indications (such as postfracture) or combined with different techniques (such as corrective osteotomies). The WOMAC may be more sensitive to these smaller changes in health status than the SF-36. Further studies may support the continued use of the WOMAC in clinical trials where the changes in health status are expected to be of lesser magnitude than that seen after total hip arthroplasty.

Outcomes assessment has become increasingly important in evaluating the efficacy of medical and surgical treatments. There has been ongoing uncertainty regarding the best tool or combination of tools to use in reporting the outcomes of patients after total hip arthroplasty. Outcomes tools should be valid, reliable, and responsive. Previous studies have demonstrated the validity and reliability of the SF-36, but there has been concern that it is not adequately responsive to clinical change in orthopedic patients [19]. The findings of this study indicate that the physical functioning subscale, bodily pain subscale, and PCS Scale of the SF-36 have a level of responsiveness similar to that seen with the WOMAC. This supports the isolated use of the SF-36 as a reasonable approach toward monitoring clinical outcomes after total hip arthroplasty. This would allow for a decreased burden of questions on patients being evaluated without sacrificing adequate responsiveness to clinical change. The PCS and MCS Scales may be particularly useful as a summation of patient outcome that incorporates the results of the more variable individual components of the SF-36. However, further study is required to determine whether the continued use of the WOMAC is advantageous in detecting the smaller clinical changes expected when evaluating the results of nonoperative management or when prospectively comparing different indications, designs, or techniques for total hip arthroplasty.

References

1. Bellamy N. The WOMAC knee and hip osteoarthritis indices: development, validation, globalization and influence on the development of the AUSCAN hand osteoarthritis indices. Clin Exp Rheum 2005; 23(Suppl 39):s-148.
2. Boardman DL, Dorey F, Thomas BJ, et al. The accuracy of assessing total hip arthroplasty outcomes. J Arthroplasty 2000;15:200.
3. Quintana JM, Escobar A, Bilbao A, Arostegui I, Lafuente I, Vidaurreta I, et al. Responsiveness and clinically important differences for the WOMAC and SF-36 after hip joint replacement. Osteoarthritis Cartilage 2005;13:1076-83 [Epub 2005 Sep 9 PMID:16154777].
4. Lieberman JR, Dorey F, Shekelle P, et al. Outcome after total hip arthroplasty. J Arthroplasty 1997; 12:639.
5. Soderman P, Malchau H. Validity and reliability of Swedish WOMAC osteoarthritis index. Acta Orthop Scand 2000;71:39.
6. Kosinski M, Keller S, Hatoum H, et al. The SF-36 health survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: tests of data quality, scaling assumptions, and score reliability. Med Care 1999;37:MS10.
7. Ware JE. SF-36 health survey manual and interpretation guide. Boston: The Health Institute, 1994.
8. Ware JE, Kosinski M, Keller SD. SF-36 Physical and mental health summary scales: a user's manual. Boston: The Health Institute, 1994.
9. Angst F, Aeschlimann A, Steiner W, et al. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Ann Rheum Dis 2001;60:834.
10. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. Clin Ther 1996; 18:979.
11. Liang MH, Evaluating measurement responsiveness. J Rheum 1995;22:1191.
12. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27:S178.
13. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. Spine 2000;25:3192.
14. McHorney CA, Ware JE, 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.
15. Tuttleman M, Pillemer SR, Tilley BC, et al. A cross sectional assessment of health status instruments in patients with rheumatoid arthritis participating in a clinical trial. J Rheumatol 1997;24:1910.
16. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. J Clin Epidemiology 1997;50:79.
17. Mangione CM, Goldman L, Orav EJ, et al. Health-related quality of life after elective surgery: measurement of longitudinal changes. J Gen Intern Med 1997;12:686.
18. Martin DP, Engelberg R, Agel J, et al. Comparison of the musculoskeletal function assessment questionnaire with the Short Form-36, the Western Ontario and McMaster Universities Osteoarthritis Index, and the sickness impact profile health-status measures. J Bone Joint Surg 1997;79-A:1323.
19. Liang MH, Fossel AH, Larson MG. Comparison of five health status instruments for orthopedic evaluation. Med Care 1990;28:632.
20. Lingard EA, Katz JN, Wright RJ, et al. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. J Bone Joint Surg 2001;83-A:1856.
21. Marx RG, Jones EC, Atwan NC, et al. Measuring improvement following total hip and knee arthroplasty using patient-based measures of outcome. J Bone Joint Surg 2005;87:1999.
22. Kosinski M, Keller S, Ware JE, et al. The SF-36 health survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: relative validity of scales in relation to clinical measures of arthritis severity. Med Care 1999;37:MS23.
23. Kosinski M, Zhao SZ, Dedhiya S, et al. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000;43:1478.
24. Croft P, Lewis M, Wynn Jones C, et al. Health status of patients awaiting hip replacement for osteoarthritis. Rheumatology 2002;41:1001.
25. Quintana JM, Escobar A, Arostegui I, et al. Health-related quality of life and appropriateness of knee or hip joint replacement. Arch Int Med 2006;166:220.
26. Beaton DE, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. J Shoulder Elbow Surg 1998;7:565.
27. Jaglow S, Lakhani Z, Schatzker J. Reliability, validity, and responsiveness of the lower extremity measure for patients with a hip fracture. J Bone Joint Surg 2000;82:955.
28. Walsh TL, Hanscom B, Lurie JD, et al. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary: the responsiveness of the Oswestry Disability Index MODEMS, and the SF-36. Spine 2003;28:607.