

THE JOURNAL OF BONE & JOINT SURGERY

# JB&JS

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

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

---

## Measuring Improvement Following Total Hip and Knee Arthroplasty Using Patient-Based Measures of Outcome

Robert G. Marx, Edward C. Jones, Nawal C. Atwan, Robert F. Closkey, Eduardo A. Salvati and Thomas P. Sculco  
*J Bone Joint Surg Am.* 87:1999-2005, 2005. doi:10.2106/JBJS.D.02286

---

**This information is current as of September 22, 2005**

### Reprints and Permissions

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

### Publisher Information

The Journal of Bone and Joint Surgery  
20 Pickering Street, Needham, MA 02492-3157  
[www.jbjs.org](http://www.jbjs.org)

# MEASURING IMPROVEMENT FOLLOWING TOTAL HIP AND KNEE ARTHROPLASTY USING PATIENT-BASED MEASURES OF OUTCOME

BY ROBERT G. MARX, MD, MSc, FRCSC, EDWARD C. JONES, MD, MA, NAWAL C. ATWAN, AB,  
ROBERT F. CLOSKEY, MD, EDUARDO A. SALVATI, MD, AND THOMAS P. SCULCO, MD

*Investigation performed at The Hospital for Special Surgery, New York, New York*

**Background:** Patient-derived outcome scales have become increasingly important to physicians and clinical researchers for measuring improvement in function after surgery. The goal of the present study was to evaluate the ability of health-status instruments to measure early functional recovery after total hip and total knee arthroplasty.

**Methods:** Four hundred and six patients undergoing total hip arthroplasty and 266 patients undergoing total knee arthroplasty completed health-status questionnaires preoperatively and six months postoperatively to determine the standardized response mean. In the second phase of the study, a group of patients undergoing knee and hip arthroplasty were evaluated with several instruments before and after surgery to test for postoperative ceiling effects.

**Results:** The standardized response mean at six months was 1.7 for the MODEMS Hip Core, 1.2 for the MODEMS Knee Core, and 1.5 and 1.1 for the Physical Component Summary of the SF-36 for patients managed with hip and knee replacement, respectively. A standardized response mean of 1.0 is generally satisfactory for measuring improvement in orthopaedic surgery. In Phase 2 of the study, the vast majority of patients who had a score of 95 to 100 (that is, a maximum or near-maximum score) on the joint-specific scales generally believed that the hip or knee was normal and could not be better.

**Conclusions:** The MODEMS, Oxford, and WOMAC scales all demonstrated a ceiling effect following total knee and total hip arthroplasty. These scores likely reflected the patients' perception of the status of the knee or hip rather than an inability to measure their improvement beyond the highest possible score. The Physical Component Summary score of the SF-36 had similar standardized response means when compared with hip and knee-specific instruments, and, therefore, consideration should be given to using this scale without a joint-specific scale for the measurement of improvement following total knee and total hip replacement, as a way to decrease responder burden (that is, the time required to complete the questionnaires).

Patient-derived outcome scales have become increasingly important to physicians and clinical researchers for measuring improvement in function after surgery<sup>1,2</sup>. Hunsaker et al. collected population-based normative data on the eleven American Academy of Orthopaedic Surgeons musculoskeletal outcomes scales, including the MODEMS (Musculoskeletal Outcomes Data Evaluation and Management System) Hip/Knee Core scale<sup>3</sup>. In that study, the sensitivity of the MODEMS Hip/Knee Core scale to meaningful clinical change following the treatment of hip and knee disorders was not determined<sup>3</sup>. This information is important for the comparison of different clinical interventions for the hip and knee and the long-term evaluation of patients.

Among the many hip and knee outcome-rating scales that have been developed to evaluate patients, the MODEMS Hip/Knee Core instrument is commonly used for patients undergoing total hip and total knee replacement<sup>3</sup>. This scale was validated by the American Academy of Orthopaedic Surgeons<sup>4</sup>; however, we are aware of no published studies that have assessed the effectiveness of the MODEMS Hip/Knee Core scale in detecting clinically meaningful improvement following total hip or knee arthroplasty<sup>5</sup>.

The purpose of the first phase of the present study was to evaluate the responsiveness of the Hip/Knee Core scale to clinical change at six and twelve months after total hip and total knee arthroplasty. A ceiling effect was noted in this first phase; that is, many patients received the maximum score (or

close to the maximum score) on the scale. The purposes of the second phase of the study were to determine (1) whether this ceiling effect occurred with other rating scales following total hip and knee arthroplasty and (2) whether these elevated scores accurately reflected the way the patients felt about the status of the involved hip or knee.

## Materials and Methods

### Phase 1

This study included preoperatively collected data on 406 patients undergoing total hip arthroplasty and 266 patients undergoing total knee arthroplasty. Osteoarthritis was the primary diagnosis in all but eight patients, who had a primary diagnosis of rheumatoid arthritis based on International Classification of Diseases, Ninth Revision (ICD-9) coding. The mean age of the patients undergoing knee replacement was seventy years (range, twenty-eight to ninety years) at the time of surgery, and 62% (165) of these patients were women. The mean age of the patients undergoing hip replacement was sixty-six years (range, twenty-one to ninety years) at the time of surgery, and 53% (214) of these patients were women. The MODEMS Hip/Knee Core scale and the SF-36 general health-status instrument were administered to all patients before surgery and six months after surgery. Consecutive patients managed by two surgeons (E.A.S., T.P.S.) were evaluated, covering the range of symptoms and disability seen in our institution. Two hundred and sixteen of the patients undergoing hip replacement and 108 of the patients undergoing knee replacement also completed the questionnaires twelve months after surgery.

The MODEMS Hip/Knee Core scale was developed to measure hip and knee pain and the overall impact of a pathological condition on daily function. The instrument comprises seven patient-reported multiple-choice questions. Patients receive a score of 0 to 100 points on the MODEMS Hip/Knee Core scale, with 100 being the best possible score<sup>6</sup>. The SF-36 comprises thirty-six items that measure general health in eight subscales. This instrument is widely used in conjunction with region-specific questionnaires in orthopaedics<sup>7-11</sup>. A Physical Component Summary score and a Mental Component Summary score can be derived from the SF-36<sup>12-15</sup>. The standardized response mean (the observed change divided by the standard deviation of change) was used to calculate responsiveness. The standardized response means for validated orthopaedic instruments generally have ranged from 0.9 to 1.9<sup>16</sup>.

The data from the questionnaires were entered by manually scanning the forms with use of a digital scanner. The data were analyzed with use of SPSS software (version 11.0; SPSS, Chicago, Illinois) for personal computers.

Construct validity analysis was conducted with use of the baseline MODEMS Hip/Knee Core scale data for both the patients undergoing hip replacement and those undergoing knee replacement. The term validity defines whether the instrument actually measures what it is intended to measure<sup>17</sup>. The eight subscales of the SF-36 scale were used to determine

construct validity. It was hypothesized that the MODEMS Hip/Knee Core scale would correlate better with the physical function, role-physical, and bodily pain subscales than with the general health, vitality, social function, role-emotional, and mental health subscales according to the Spearman correlation coefficient.

### Phase 2

In Phase 1, a ceiling effect was noted. While medical data often have a skewed distribution, Phase 2 of the study was performed to determine (1) whether such scores are also seen with other instruments following total hip arthroplasty and total knee arthroplasty and (2) whether these elevated scores are valid.

To ensure that we had a sufficient number of patients who had a score of at least 95 points on the scales, we planned to recruit at least fifty patients undergoing knee arthroplasty and at least fifty undergoing hip arthroplasty. A sample-size calculation was not required because the analysis was purely descriptive. Two additional cohorts of patients undergoing total hip arthroplasty or total knee arthroplasty were recruited, and these patients were evaluated at baseline (sixty-six and fifty-eight patients, respectively), at the six-month follow-up (sixty-six and fifty-eight patients, respectively), and at the twelve-month follow-up (forty-eight and forty-six patients, respectively). The mean age of the additional patients undergoing hip arthroplasty was sixty-eight years (range, forty-three to ninety years), and 56% (thirty-seven) of these patients were women. The mean age of the additional patients undergoing knee arthroplasty was sixty-nine years (range, forty-two to ninety years), and 62% (thirty-six) of these patients were women. As in Phase 1, consecutive patients were evaluated. The patients undergoing total knee arthroplasty completed the MODEMS outcome tool as well as the Oxford and WOMAC instruments. The patients undergoing total hip arthroplasty completed the MODEMS and Oxford scales. All patients completed the SF-36 at all follow-up points.

In order to determine if the ceiling effect was valid, patients in Phase 2 of the study were asked, "Is there room for improvement in your hip/knee function?," "Could your hip/knee be better?," and "Does your hip/knee feel normal?" at the time of the twelve-month follow-up. As the concept of ceiling is essentially binary in the context of a continuous measure, we employed these three binary questions in an attempt to determine if a perfect (or nearly perfect) score accurately reflected the way the patient felt about the status of the involved hip or knee.

Patients who received the maximum score as well as those who received a score of between 95 and 100 points on a 100-point scale (or the equivalent on the Oxford scale) were studied to determine how they answered the three additional questions. The hypothesis was that patients who received the maximum score (or close to it) would state that the involved hip or knee was normal, that it could not be any better, and that there was no room for improvement and function. It was thought that, if these patients answered in the hypothesized

TABLE I Rating Scale Scores for Hip and Knee Arthroplasty Patients

| Instrument                       | Mean Score | Standard Deviation | Lowest Score | Highest Score |
|----------------------------------|------------|--------------------|--------------|---------------|
| <b>Total hip arthroplasty</b>    |            |                    |              |               |
| MODEMS Hip Core                  |            |                    |              |               |
| Baseline (n = 406)               | 59.2       | 16.3               | 8.0          | 100.0         |
| 6 Months (n = 406)               | 88.9       | 10.3               | 43.0         | 100.0         |
| 12 Months (n = 216)              | 88.9       | 11.4               | 38.0         | 100.0         |
| SF-36 Physical Component Summary |            |                    |              |               |
| Baseline (n = 406)               | 30.1       | 8.3                | 13.0         | 54.0          |
| 6 Months (n = 406)               | 45.5       | 10.3               | 17.0         | 65.0          |
| 12 Months (n = 216)              | 45.6       | 10.0               | 16.0         | 64.0          |
| <b>Total knee arthroplasty</b>   |            |                    |              |               |
| MODEMS Knee Core                 |            |                    |              |               |
| Baseline (n = 266)               | 58.3       | 15.7               | 6.0          | 95.0          |
| 6 Months (n = 266)               | 79.9       | 15.2               | 25.0         | 100.0         |
| 12 Months (n = 108)              | 83.1       | 14.4               | 39.0         | 100.0         |
| SF-36 Physical Component Summary |            |                    |              |               |
| Baseline (n = 266)               | 28.8       | 8.2                | 10.0         | 54.0          |
| 6 Months (n = 266)               | 41.7       | 10.8               | 15.0         | 62.0          |
| 12 Months (n = 108)              | 43.7       | 10.6               | 17.0         | 59.0          |

fashion, then the ceiling effect noted in our study was valid for the scale in this patient population.

## Results

### Phase 1

The MODEMS Hip/Knee Core instrument detected substantial improvement in health-related quality of life at

six months postoperatively. The patients managed with hip arthroplasty demonstrated greater improvement on this scale compared with those managed with knee arthroplasty. The standardized response means for the patients managed with hip arthroplasty were 1.7 for the MODEMS Hip Core, 1.5 for the SF-36 Physical Component Summary score, and 0.3 for the SF-36 Mental Component Summary score. The standard-

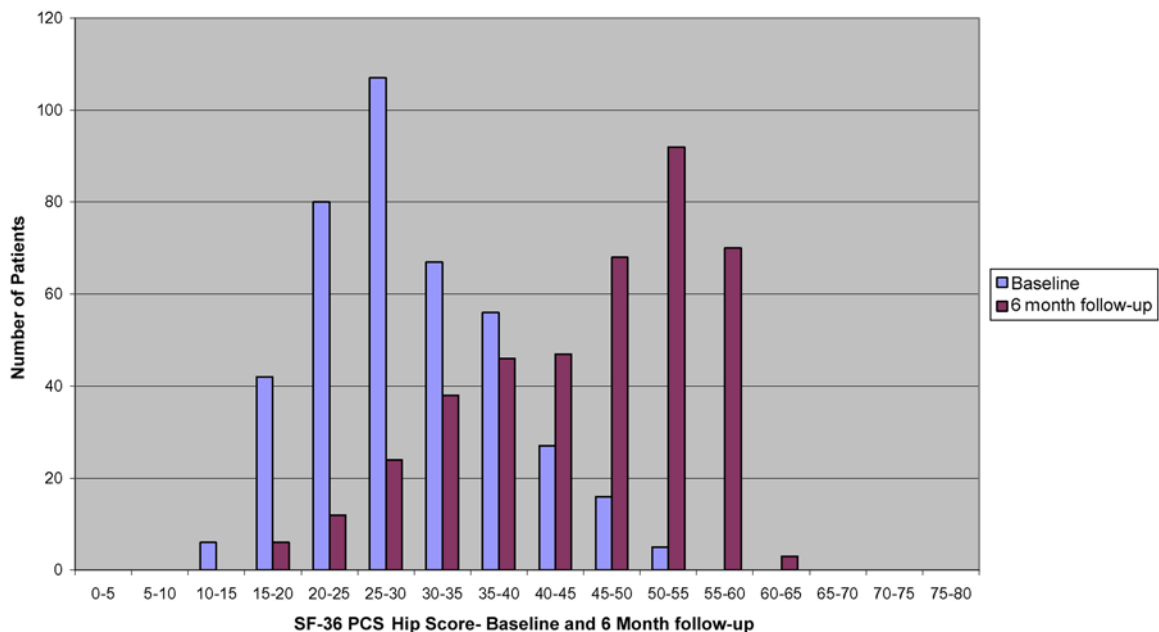


Fig. 1

Bar graph showing the baseline and six-month follow-up SF-36 Physical Component Summary (PCS) scores for the 406 patients managed with total hip arthroplasty.

**TABLE II Percentage of Patients Receiving Maximum or Near-Maximum Scores on Condition-Specific Instruments at Six and Twelve Months Following Total Hip Arthroplasty or Total Knee Arthroplasty**

| Instrument              | Percentage of Patients Receiving a Score of 100 (or Equivalent) |               | Percentage of Patients Receiving a Score of 95-100 (or Equivalent) |                |
|-------------------------|-----------------------------------------------------------------|---------------|--------------------------------------------------------------------|----------------|
|                         | 6 months                                                        | 12 months     | 6 months                                                           | 12 months      |
| Total hip arthroplasty  |                                                                 |               |                                                                    |                |
| MODEMS                  | 12% (8 of 66)                                                   | 19% (9 of 48) | 30% (20 of 66)                                                     | 38% (18 of 48) |
| Oxford                  | 9% (6 of 66)                                                    | 17% (8 of 48) | 39% (26 of 66)                                                     | 52% (25 of 48) |
| Total knee arthroplasty |                                                                 |               |                                                                    |                |
| MODEMS                  | 5% (3 of 58)                                                    | 15% (7 of 46) | 14% (8 of 58)                                                      | 28% (13 of 46) |
| Oxford                  | 5% (3 of 58)                                                    | 7% (3 of 46)  | 14% (8 of 58)                                                      | 22% (10 of 46) |
| WOMAC                   | 7% (4 of 58)                                                    | 4% (2 of 46)  | 17% (10 of 58)                                                     | 20% (9 of 46)  |

ized response means for the patients managed with knee arthroplasty were 1.2 for the MODEMS Knee Core, 1.1 for the SF-36 Physical Component Summary score, and 0.2 for the SF-36 Mental Component Summary score. A standardized response mean of 1.0 is generally satisfactory for measuring improvement in orthopaedic surgery.

The MODEMS Hip/Knee Core scale exhibited a ceiling effect in the evaluation of patients managed with hip replacement. Of the patients managed with hip replacement, 17% (sixty-seven) received a score of 100 points (the maximum score) and a total of 35% (144) received a score of 95 points at the time of the six-month follow-up (Fig. 1).

The ceiling effect was less evident for the patients managed with knee replacement, although the distribution of the six-month scores was also severely skewed to the right (Fig. 2). Of the patients managed with knee replacement, 9% (twenty-four) received a score of 100 points and 19% (fifty) received a

score of 95 at the time of the six-month follow-up.

The skew of data was even greater at twelve months postoperatively both for the patients managed with hip replacement and for those managed with knee replacement. A total of 24% (fifty-one) of the patients managed with hip replacement and 16% (seventeen) of the those managed with knee replacement who completed the questionnaires at twelve months received a score of 100 points on the MODEMS Hip/Knee Core scale. A total of 38% (eighty-two) of the patients managed with hip replacement and 27% (twenty-nine) of those managed with knee replacement who completed the questionnaires at twelve months postoperatively received a score of 95 points on the MODEMS Hip/Knee Core scale.

The change in the mean scores on the MODEMS Hip/Knee Core instrument from six months to twelve months postoperatively for the subset of patients who completed the questionnaires at twelve months was very small, which is con-

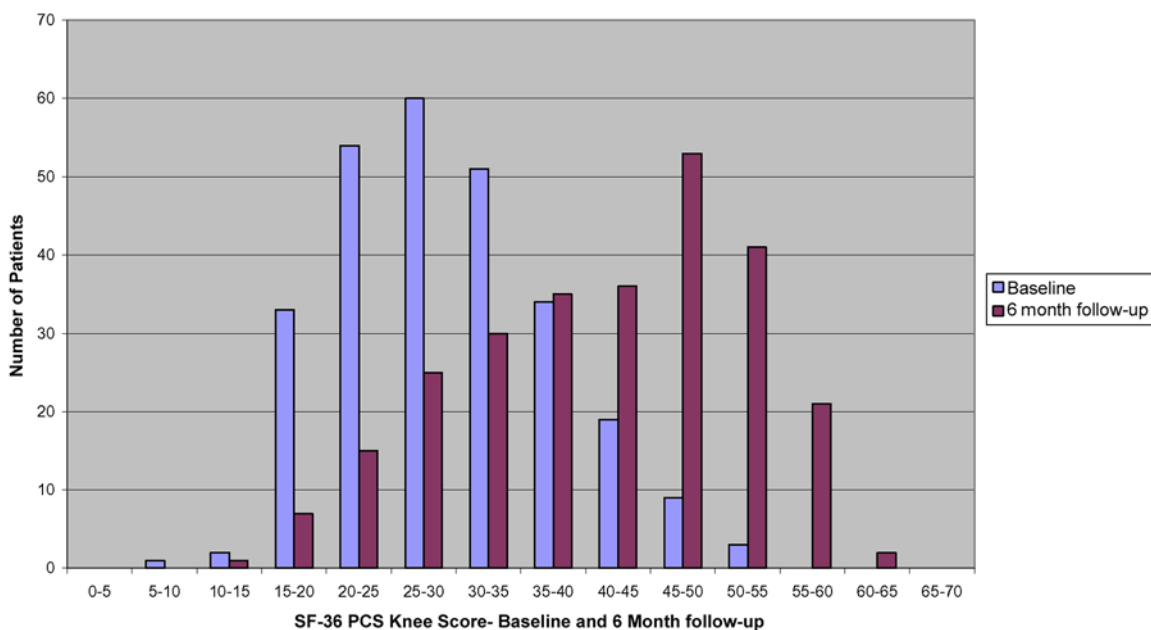


Fig. 2

Bar graph showing the baseline and six-month follow-up SF-36 Physical Component Summary (PCS) scores for the 266 patients managed with total knee arthroplasty.

**TABLE III Patient Scoring on Questions 2, 3, 4 at Twelve-Month Follow-up**

| Procedure, Instrument,<br>and Score | Question*                          |                                               |                                   |
|-------------------------------------|------------------------------------|-----------------------------------------------|-----------------------------------|
|                                     | Does Your Hip/Knee<br>Feel Normal? | Is There Room for<br>Improvement in Function? | Could Your Hip/Knee<br>be Better? |
| <b>Total hip arthroplasty</b>       |                                    |                                               |                                   |
| Maximum score                       |                                    |                                               |                                   |
| MODEMS (100 points)                 | 9/9                                | 1/9                                           | 0/9                               |
| Oxford (100 points)                 | 7/8                                | 2/8                                           | 1/8                               |
| Maximum or near-maximum score       |                                    |                                               |                                   |
| MODEMS (95-100 points)              | 17/18                              | 2/17                                          | 0/18                              |
| Oxford (95-100 points)              | 22/25                              | 4/25                                          | 2/25                              |
| <b>Total knee arthroplasty</b>      |                                    |                                               |                                   |
| Maximum score                       |                                    |                                               |                                   |
| MODEMS (100 points)                 | 5/7                                | 1/7                                           | 1/7                               |
| Oxford (100 points)                 | 3/3                                | 0/3                                           | 0/3                               |
| WOMAC (100 points)                  | 2/2                                | 0/2                                           | 0/2                               |
| Maximum or near-maximum score       |                                    |                                               |                                   |
| MODEMS (95-100 points)              | 10/13                              | 1/13                                          | 1/13                              |
| Oxford (95-100 points)              | 8/10                               | 1/10                                          | 1/10                              |
| WOMAC (95-100 points)               | 8/9                                | 0/9                                           | 0/9                               |

\*The data are given as the number of patients who answered yes, followed by the total number of patients who answered the question.

sistent with the findings reported in other studies following hip surgery<sup>18</sup>. The mean score for the patients managed with hip replacement remained the same, whereas the mean score for the patients managed with knee replacement improved by 3.2 points (Table I). The Physical Component Summary score of the SF-36 increased by 0.1 point for the patients managed with hip replacement and by 2.0 points for those managed with knee replacement (Table I). As expected, the magnitude of the change in the mean outcome scores between baseline and six months postoperatively was much greater than the magnitude of the change in the mean outcome scores between six and twelve months postoperatively (Table I).

All eight subscales of the SF-36 were positively correlated with the MODEMS Hip/Knee Core scale for both the patients managed with hip replacement and those managed with knee replacement. The physical function and bodily pain subscales correlated better with the MODEMS Hip/Knee Core scale than any of the other subscales did (physical function,  $r = 0.64$  for patients managed with hip replacement and  $r = 0.48$  for those managed with knee replacement; bodily pain,  $r = 0.67$  for patients managed with hip replacement and  $r = 0.63$  for those managed with knee replacement). The role-physical subscale correlated well with the MODEMS Hip/Knee Core scale ( $r = 0.38$  for patients managed with hip replacement and  $r = 0.35$  for those managed with knee replacement), but the social function and vitality subscale scores correlated as well, or nearly as well, for both groups (social function,  $r = 0.51$  for patients managed with hip replacement and  $r = 0.43$  for those managed with knee replacement; vitality,  $r = 0.42$  for patients managed with hip replacement and  $r = 0.34$  for those managed with knee replacement).

### Phase 2

For all scales, at both six and twelve months after surgery, the percentage of patients who received maximal scores and the percentage of patients who received scores of between 95 and 100 points (normalized for all scales to a 100-point scale) were similar on the basis of the small numbers available (Table II).

Patients also were asked, at twelve months postoperatively, if the involved hip or knee “feels normal.” The vast majority of patients who received either a score of 100 points or of between 95 and 100 points on all scales following both total hip arthroplasty and total knee arthroplasty responded “yes” (Table III). Similarly, a small minority responded “yes” to either “Is there room for improvement in function of your hip/knee?” or “Could your hip/knee be better?” (Table III).

### Discussion

Responsive outcome instruments for total hip and knee replacement are important for clinical research and quality-of-care studies. They are also required because prospective orthopaedic clinical research often involves differentiating small improvements in patient outcomes. For example, the difference in outcome between patients who undergo posterior cruciate ligament-sacrificing compared with posterior cruciate ligament-retaining total knee replacement may be relatively difficult to detect.

While condition or joint-specific instruments generally have been found to be more responsive than generic health-status tools<sup>19,21</sup>, the MODEMS Hip/Knee Core Scales did not demonstrate significantly greater improvement for patients managed with total hip or knee arthroplasty, with the numbers

available, at six months of follow-up, compared with the Physical Component Summary score of the SF-36. At six months postoperatively, the standardized response mean for the MODEMS Knee Core was 1.2 and the Physical Component Summary Score of the SF-36 was 1.1 for the patients managed with knee replacement, whereas it was 1.7 for the Hip Core and 1.5 for the Physical Component Summary score of the SF-36 for the patients managed with hip replacement, findings that are similar to those in other published reports<sup>22</sup>. The MODEMS Hip/Knee Core Scale fulfilled our hypotheses for construct validity as it correlated well with overall health as measured by the SF-36. The role-physical subscale of the SF-36 correlated less well with hip and knee-specific disability ( $r = 0.38$  and  $0.35$ , respectively) than did the physical function subscale ( $r = 0.64$  and  $0.48$ , respectively) and the bodily pain subscale ( $r = 0.67$  and  $0.63$ , respectively) as the latter two subscales are more relevant to patients with hip and knee disorders.

The distribution of the MODEMS Hip/Knee Core scores was skewed to the right at six months postoperatively for patients managed with total hip and total knee replacement. This postoperatively skewed distribution has not been described previously, to our knowledge. This "ceiling effect," coupled with the standardized response mean, which was equivalent to that of the Physical Component Summary score of the SF-36, was considered to possibly indicate that the MODEMS instruments were not more responsive to detect improvement following lower extremity joint replacement surgery. These two observations (the lack of responsiveness compared with the Physical Component Summary of the SF-36 and the ceiling effect) are probably related because more than one-third of the patients managed with total hip arthroplasty had perfect, or nearly perfect, scores at six months postoperatively.

After Phase 1 of this study, it was not clear whether a perfect (or nearly perfect) score on the MODEMS Hip/Knee Core Instrument following joint replacement was truly indicative of full functional recovery. Many patients are generally believed to have continued improvement beyond six months, which may not be detected by this scale if the patients have already achieved the highest possible rating. Although the mean scores for patients managed with hip and knee replacement increased very little from six to twelve months postoperatively, the ceiling effect was even greater at twelve months for both groups. Therefore, we wanted to investigate whether this skewed distribution was indeed valid, that is, whether patients with high scores indeed believed that their outcome was excellent.

To determine whether this ceiling effect was present with other scales, and to assess whether the high scores noted at one year postoperatively were valid, Phase 2 of this study was carried out prospectively in two additional cohorts of patients undergoing total hip arthroplasty and total knee arthroplasty.

In Phase 2 of the study, the Oxford and WOMAC scales demonstrated ceiling effects similar to the MODEMS. At six and twelve months, the ceiling effect for patients managed with knee replacement was less than that for patients managed

with hip replacement. This is consistent with findings that early functional recovery is slower for patients managed with knee replacement than for those managed with hip replacement<sup>23</sup>. The increase in the magnitude of the ceiling effect from six to twelve months following surgery was greater for the patients managed with knee replacement than it was for those managed with hip replacement. This is also consistent with findings that greater improvement is generally realized for patients managed with knee arthroplasty during the six to twelve-month postoperative interval<sup>12,24,25</sup>.

While the hip and knee-specific scales were all found to have substantial ceiling effects, this finding probably was a reflection of the patients' outcome and not a reflection of the tools themselves. Although the sample size was small, the patients' responses to the three additional questions in Phase 2 of the study suggest that there is a subset of patients who attain what they believe to be a perfect or nearly perfect hip or knee-related health status at twelve months following surgery. For these individuals, further improvement with regard to their hip or knee symptoms and disability is not thought to be possible, either because of their low demands and expectations or the relief from their severe preoperative symptoms. Therefore, the skewed distribution of the MODEMS scores postoperatively and the finding that the standardized response mean of the MODEMS was similar to that of the Physical Component Summary of the SF-36 in Phase 1 may be a reflection of the way that the patients truly feel rather than of an inability to measure improvement beyond a certain level of function and pain.

The Physical Component Summary score of the SF-36 had a similar measure of responsiveness (standardized response mean) as the disease-specific measures that were studied and also was correlated with all joint-specific scales for patients managed with both hip and knee arthroplasty. Additionally, the distribution of the postoperative Physical Component Summary scores was less skewed. Therefore, the Physical Component Summary of the SF-36 may be sufficient to evaluate patients following total hip arthroplasty and total knee arthroplasty without a disease-specific instrument. This is due to the fact that the SF-36 has many items that are relevant to low-demand physical activities that involve the lower extremities, such as walking and other daily activities. However, there is no downside to using a knee or hip-specific instrument in addition to the SF-36, aside from increased responder burden (that is, the time required to complete the questionnaires). ■

Robert G. Marx, MD, MSc, FRCSC

Edward C. Jones, MD, MA

Nawal C. Atwan, AB

Robert F. Closkey, MD

Eduardo A. Salvati, MD

Thomas P. Sculco, MD

Department of Orthopedic Surgery, The Hospital for Special Surgery,  
535 East 70th Street, New York, NY 10021. E-mail address for R.G. Marx:  
marxr@hss.edu

In support of their research or preparation of this manuscript, one or more of the authors received grants or outside funding from the Cornell Center for Aging Research and Clinical Care. None of the authors received payments or other benefits or a commitment or agreement to provide such benefits from a commercial entity. No commercial entity paid or directed, or agreed to pay or direct, any benefits to any research

fund, foundation, educational institution, or other charitable or nonprofit organization with which the authors are affiliated or associated.

doi:10.2106/JBJS.D.02286

## References

1. Keller RB. Outcomes research in orthopaedics. *J Am Acad Orthop Surg.* 1993;1:122-9.
2. Bellamy N. Outcome measurement in osteoarthritis clinical trials. *J Rheumatol Suppl.* 1995;43:49-51.
3. Hunsaker FG, Cioffi DA, Amadio PC, Wright JG, Caughlin B. The American academy of orthopaedic surgeons outcomes instruments: normative values from the general population. *J Bone Joint Surg Am.* 2002;84:208-15.
4. Johanson NA, Liang MH, Daltroy L, Rudicel S, Richmond J. American Academy of Orthopaedic Surgeons lower limb outcomes assessment instruments. Reliability, validity, and sensitivity to change. *J Bone Joint Surg Am.* 2004;86:902-9.
5. American Academy of Orthopaedic Surgeons. AAOS Outcomes Studies Instruments: normative data and hip and knee scoring documentation. 2001.
6. American Academy of Orthopaedic Surgeons. Scoring algorithms for the lower limb outcomes data collection instrument, version 2.0. Rosemont, IL: American Academy of Orthopaedic Surgeons; 1998.
7. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br.* 1996;78:185-90.
8. 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.
9. Hawker G, Melfi C, Paul J, Green R, Bombardier C. Comparison of a generic (SF-36) and a disease specific (WOMAC) (Western Ontario and McMaster Universities Osteoarthritis Index) instrument in the measurement of outcomes after knee replacement surgery. *J Rheumatol.* 1995;22:1193-6.
10. Lieberman JR, Dorey F, Shekelle P, Schumacher L, Kilgus DJ, Thomas BJ, Finerman GA. Outcome after total hip arthroplasty. Comparison of a traditional disease-specific and a quality-of-life measurement of outcome. *J Arthroplasty.* 1997;12:639-45.
11. Naughton MJ, Anderson RT. Outcomes research in orthopaedics: health-related quality of life and the SF-36. *Arthroscopy.* 1998;14:127-9.
12. Kiebzak GM, Vain PA, Gregory AM, Mokris JG, Mauerhan DR. SF-36 general health status survey to determine patient satisfaction at short-term follow-up after total hip and knee arthroplasty. *J South Orthop Assoc.* 1997;6:169-72.
13. Ware JE Jr, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston: The Health Institute, New England Medical Center; 1993.
14. Benroth R, Gawande S. Patient-reported health status in total joint replacement. *J Arthroplasty.* 1999;14:576-80.
15. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care.* 1993;31:247-63.
16. Marx RG, Jones EC, Allen AA, Altchek DW, O'Brien SJ, Rodeo SA, Williams RJ, Warren RF, Wickiewicz TL. Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. *J Bone Joint Surg Am.* 2001;83:1459-69.
17. Feinstein AR. *Clinimetrics.* New Haven: Yale University Press; 1987.
18. Peterson MG, Allegrante JP, Cornell CN, MacKenzie CR, Robbins L, Horton R, Ganz SB, Augurt A. Measuring recovery after a hip fracture using the SF-36 and Cummings scales. *Osteoporos Int.* 2002;13:296-302.
19. McGuigan FX, Hozack WJ, Moriarty L, Eng K, Rothman RH. Predicting quality-of-life outcomes following total joint arthroplasty. Limitations of the SF-36 Health Status Questionnaire. *J Arthroplasty.* 1995;10:742-7.
20. Bergner M, Rothman ML. Health status measures: an overview and guide for selection. *Annu Rev Public Health.* 1987;8:191-210.
21. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987;40:171-8.
22. Mangione CM, Goldman L, Orav EJ, Marcantonio ER, Pedan A, Ludwig LE, Donaldson MC, Sugarbaker DJ, Poss R, Lee TH. Health-related quality of life after elective surgery: measurement of longitudinal changes. *J Gen Intern Med.* 1997;12:686-97.
23. Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. *J Rheumatol.* 2000;27:1745-52.
24. Bachmeier CJ, March LM, Cross MJ, Lapsley HM, Tribe KL, Courtenay BG, Brooks PM; Arthritis Cost and Outcome Project Group. A comparison of outcomes in osteoarthritis patients undergoing total hip and knee replacement surgery. *Osteoarthritis Cartilage.* 2001;9:137-46.
25. Anderson JG, Wixson RL, Tsai D, Stulberg SD, Chang RW. Functional outcome and patient satisfaction in total knee patients over the age of 75. *J Arthroplasty.* 1996;11:831-40.