

## The SF-36 General Health Status Survey Documents the Burden of Osteoarthritis and the Benefits of Total Joint Arthroplasty: But Why Should We Use It?

Gary M. Kiebzak, PhD; Meredith Campbell, BS; and David R. Mauerhan, MD

**Objectives:** To document the disease burden of osteoarthritis and the benefits of total joint replacement by using the Short Form Health Survey (SF-36) general health status survey and evaluate other factors that could affect scores.

**Study Design:** Prospective study.

**Patients and Methods:** All patients scheduled for primary total hip arthroplasty (THA) or total knee arthroplasty (TKA) (n = 622 preoperatively) in 2 years were surveyed using the SF-36, which assesses health-related quality of life (HRQOL) in patients' physical and social functioning and mental health. Follow-up surveys were administered 12 months after surgery to all patients and 3 and 24 months after surgery to a subset of patients.

**Results:** Preoperatively, patient scores were significantly lower than normative scores in the physical functioning, bodily pain, and social functioning domains. Preoperative scores were not different between THA and TKA patients. Women scored lower than men. Comorbid conditions were weakly associated with low SF-36 scores. Postoperatively, the largest incremental improvement in scores was seen at 3-month follow-up. Scores improved sooner and more substantially in THA vs TKA patients and in men vs women, paralleling improvement in clinical and subjective ratings of postoperative physical function and pain.

**Conclusions:** The SF-36 has the sensitivity to document improvement in HRQOL after surgery and to reveal differences in THA vs TKA and in men vs women. However, routine use of outcome assessment instruments to monitor this patient population is costly and unjustified in our current healthcare environment.

(*Am J Manag Care* 2002;8:463-474)

It is well established that total joint replacement is one of the most successful modern surgical procedures.<sup>1</sup> However, the recent pattern of intense focus on healthcare cost containment, com-

binated with heightened efforts to objectively identify which medical interventions are effective and necessary, mandates continued documentation of the success of total joint replacement. One way to show the effectiveness and success of surgical procedures is to periodically assess health-related quality-of-life (HRQOL) variables using patient-completed survey instruments. A variety of instruments are available for patient outcome assessment.<sup>2-11</sup>

The ideal solution for monitoring patient satisfaction and change in HRQOL after surgery is to use a readily available validated survey instrument that is easy to complete, process, and score yet sensitive enough to detect changes in a wide spectrum of health categories. Furthermore, use of standardized surveys allows comparisons between studies, cohorts, normative data, etc.

The results of many recent studies<sup>8,12-19</sup> show the usefulness of the Short Form Health Survey (SF-36) general health status survey for documenting change after surgery. Controversy persists, however, regarding whether a general health status survey is an adequate outcome tool for assessing site-specific treatments

From Miller Orthopaedic Clinic, Charlotte, NC. Dr Kiebzak is now with the Center for Orthopaedic Research and Education, St Luke's Episcopal Hospital, Houston, TX.

Presented in part as a poster at the Annual Meeting of the Association for Health Services Research, Washington, DC, 1998, and as a lecture at the Annual Meeting of the American Academy of Orthopaedic Surgeons, San Francisco, CA, June 21-23, 2001.

The study was internally funded by Miller Orthopaedic Clinic Inc, Charlotte, NC.

Address correspondence to: Gary M. Kiebzak, PhD, Chief Research Scientist, Center for Orthopaedic Research and Education, St Luke's Episcopal Hospital, 6720 Bertner Avenue, Houston, TX 77063. E-mail: gkiebzak@sleh.com.

such as total hip arthroplasty (THA) and total knee arthroplasty (TKA).<sup>5,7-10,18,20</sup> Also, further study of differences between surgical sites and between men and women and the effect of age, preexisting comorbid conditions, etc, is necessary to clarify which survey instrument should be routinely used.<sup>15,21-30</sup>

The primary objective of this study was to document the burden of osteoarthritis and the benefits of THA and TKA by using the SF-36 general health status survey. The secondary objective was to evaluate the time course of change after THA vs TKA, differences in outcome between men and women in the postoperative recovery period, and the effect of age and comorbid conditions on HRQOL variables.

On completion of this study, we assessed (1) the practicality of routine administration of the SF-36 to patients undergoing total joint replacement; (2) whether the information derived was truly novel (ie, could not be obtained from patient charts); and (3) whether SF-36 scores could have an impact on the management of an individual.

... METHODS ...

For 2 years, the SF-36 was given to all patients at the time of THA or TKA at Miller Orthopaedic Clinic, Charlotte, NC. Surveys were either complet-

ed in the clinic or were taken home and mailed back to our research staff. Twelve-month follow-up surveys were mailed to patients with a self-addressed, stamped envelope in which the completed survey could be returned. Patients enrolled in the first 3 months of the study were asked to complete an SF-36 at 3- and 24-month follow-up in addition to 12-month follow-up to establish the time course of change in HRQOL variables.<sup>18</sup>

Patients of all ages were included. Only patients scheduled for elective primary THA or TKA were assessed in this study. American Society of Anesthesiologists (ASA) scores (ASA surgical risk assessment system: the greater the score, the greater the surgical risk) and the number of comorbid conditions were obtained from medical chart review. At our institution, a standard form listing comorbid conditions is completed during presurgery certification. No attempt was made to assign weighted values to comorbid conditions. Patients from 5 surgeons were pooled to complete this study. There was no surgeon effect with respect to preoperative scores (ie, results were not biased based on surgeon). Most femoral prosthetic components and knee components were cemented (as opposed to cementless or press-fit); we did not stratify results by prosthetic system used.

The SF-36 consists of 36 multiple-choice questions sorted into 8 domains that describe overall HRQOL (Table 1).<sup>31,32</sup> Low numeric scores reflect a perception of poor health, loss of function, and presence of pain. High numeric scores reflect a perception of good health, no functional deficits, and absence of pain. The survey instrument was scanned using Quick-Scan (Response Technologies Inc, East Greenwich, RI), and scores were calculated using the Starting Line system (Response Technologies Inc).

In a subset of patients from one surgeon, SF-36 scores for physical functioning and bodily pain were correlated with Harris hip scores (n = 54 men and women)<sup>33</sup> and knee function scores (n = 58 men and women),<sup>34</sup> which are clinical

**Table 1.** Domains of the SF-36 That Describe Overall Health Status

Domain	Description	Meaning of Score Improvement
Physical functioning	Physical limitations in performance of daily living	Performs all types of physical activities including the most vigorous, without limitations due to health
Role physical	Problems encountered with daily activities or work due to physical health	Fewer problems with work or other daily activities as a result of physical health
Bodily pain	Overall pain severity	Less pain or limitations due to pain
General health	Overall general health	Evaluates personal health as excellent
Vitality	Frequency of feeling full of energy vs tired	Feels full of pep and energy more of the time
Social functioning	Performance of social activities in lieu of health problem (eg, osteoarthritis)	Performs usual social activities without interference due to physical or emotional problems
Role emotional	Problems encountered with daily activities or work due to emotional health	Fewer problems with work or other daily activities as a result of emotional problems
Mental health	Degree of nervousness or	Feels peaceful, happy, and calm more

ratings of pain and function. This was done to evaluate how well SF-36 scores compare with other, less subjective, measures. We calculated the follow-up response rate (ie, patient compliance) at 12-month follow-up. We also calculated descriptive statistics to subjectively assess potential technical limitations of the SF-36 due to floor and ceiling effects (ie, the percentage of individual surveys with domains scored 0 or 100). Retrospectively, we estimated the time devoted to the project and the cost of survey forms, software, mailing, etc.

The ordinal scaling of the SF-36 data mandated use of nonparametric statistics. The Wilcoxon matched pairs test was used for matched comparisons (eg, preoperative vs postoperative data). Pairwise comparisons (eg, men vs women for a particular domain) were performed using the Mann-Whitney test. The Friedman test for nonparametric repeated-measures analysis of variance was used to compare scores over time (eg, the change in physical functioning for men from before surgery to 3 months, 12 months, and 24 months after surgery). Correlation analysis (Pearson *r*) was used to assess relationships between SF-36 scores and age, number of preexisting comorbidities, etc. We considered *P* < .05 to be statistically significant.

To show the disease burden of osteoarthritis, we compared patient data with normative data previously published by Ware et al<sup>32</sup> using the unpaired *t* test.

Nonparametric statistics could not be used because only mean values were available for normative data.

... RESULTS ...

**Preoperative SF-36 Scores**

*THA vs TKA.* Preoperative SF-36 surveys were completed by 622 patients: 118 women and 89 men undergoing THA and 234 women and 181 men undergoing TKA. Preoperative scores were not significantly different between patients undergoing THA vs TKA (Table 2). Consequently, in some cases, data for THA and TKA were pooled for further analysis.

*Effect of Sex.* Although the difference in scores between men and women was generally not large, women consistently scored statistically significantly lower than men for all domains of the SF-36 (Table 2). Consequently, for subsequent data analysis, scores for men and women were generally kept separate.

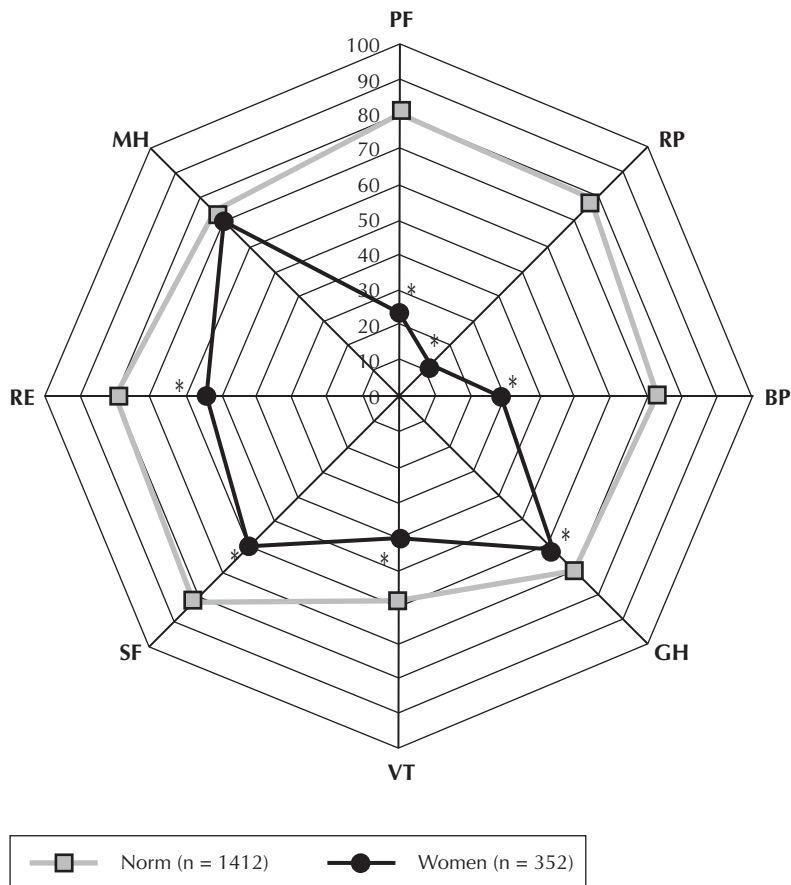
*Effect of Age.* Age did not have a substantial effect on SF-36 scores in patients scheduled for THA (both men and women) (data not shown). In contrast, for patients scheduled for TKA, there was a statistically significant correlation between age and SF-36 scores for men (bodily pain, general health, vitality, social functioning, and mental health) and women (bodily pain, general health, and mental health). The rela-

**Table 2.** Preoperative SF-36 Scores for Men and Women Undergoing Total Hip Arthroplasty (THA) or Total Knee Arthroplasty (TKA)

Domain	THA			TKA			THA + TKA		
	Women (n = 118)	Men (n = 89)	<i>P</i>	Women (n = 234)	Men (n = 181)	<i>P</i>	Women (N = 352)	Men (N = 270)	<i>P</i>
Physical functioning	22.2 ± 20.0	34.8 ± 23.3	<.001	24.4 ± 19.2	33.2±19.6	<.0001	23.7 ± 19.5	33.7 ± 20.8	<.0001
Role physical	10.2 ± 20.5	21.0 ± 30.8	.0207	14.3 ± 26.0	22.5 ± 32.4	.0182	13.0 ± 24.5	22.0 ± 31.8	.001
Bodily pain	26.1 ± 14.5	33.9 ± 15.4	.0003	29.2 ± 17.1	33.1 ± 17.3	.0260	28.2 ± 16.3	33.3 ± 16.7	.0001
General health	62.3 ± 23.9	70.0 ± 18.4	.0222	61.1 ± 20.0	68.2 ± 18.6	.0002	61.5 ± 21.4	68.8 ± 18.5	<.0001
Vitality	40.6 ± 22.0	51.0 ± 20.8	0.0008	39.3 ± 20.7	49.6 ± 22.1	<.0001	39.8 ± 21.1	50.1 ± 21.6	<.0001
Social functioning	55.3 ± 26.6	67.8 ± 25.2	0.0013	61.8 ± 27.3	69.2 ± 26.8	.0043	59.6 ± 27.2	68.8 ± 26.3	<.0001
Role emotional	53.8 ± 44.8	62.8 ± 43.5	0.1375	55.2 ± 42.8	66.6 ± 42.5	.0084	54.8 ± 43.4	65.4 ± 42.8	.0027
Mental health	69.7 ± 19.7	76.1 ± 19.2	0.0119	70.9 ± 19.0	75.3 ± 17.7	.0211	70.5 ± 19.2	75.5 ± 18.2	.0007
Age, y	67.6 ± 11.2	63.9 ± 12.0	.0028	67.8 ± 9.8	66.0 ± 9.7	.0631	—	—	—

\*Data are given as mean ± SD.

**Figure 1.** Radar Plot Comparing Mean Preoperative SF-36 Scores From Women Undergoing Total Joint Replacement (Hip or Knee) with Published Normative Scores (Age Range, 18 to >75 years; data from reference 32.)



Each spoke represents a domain of the SF-36. Plots are read from the center outward along each spoke. Scores are shown on concentric circles, beginning with 0 (at the center) and increasing to 100 (outer line).  
 PF = physical functioning; RP = role physical; BP = bodily pain; GH = general health; VT = vitality; SF = social functioning; RE = role emotional; MH = mental health.  
 \*Statistically significant at  $P < .05$ .

tionship, however, was not strong; the greatest  $r^2$  value was 12%. Furthermore, in all cases in which there was a significant correlation, scores increased slightly with increasing age.

**Comparison with Normative Data.** Scores for THA and TKA were pooled for all men and for all women and were compared with previously published normative data,<sup>32</sup> ignoring age as a covariable. Preoperative SF-36 scores for patients with degenerative osteoarthritis were significantly lower than normative scores for all domains except mental health for women and men undergoing THA or TKA

(Figure 1 and Figure 2). Because normative scores decrease with increasing age, we sorted patients by age (45-54, 55-64, and >65 years) and compared preoperative SF-36 scores with age-matched normative scores. All comparisons for physical functioning, role physical, and bodily pain showed significantly lower scores for THA and TKA. Scores for general health and mental health, however, were generally not different between age-matched patients and norms (data not shown).

**Effect of Comorbid Conditions.** The number of preexisting comorbid conditions had a significant but inconsistent effect on SF-36 scores. Statistically significant correlations were found between SF-36 scores and the number of comorbid conditions, but the range of  $r^2$  values was 18% to 37%, suggesting only a weak relationship. The relationship was stronger for women than for men on the basis of a higher number of significant correlations than for men. A similar finding was recorded with ASA ratings (data not shown) in which the general relationship was the higher the ASA rating, the lower the SF-36 score.

**Correlation with Clinical Hip and Knee Ratings.** For patients undergoing THA, there was a statistically significant correlation between preoperative SF-36 bodily pain scores and Harris hip scores ( $r = 0.53$ ;  $P < .001$ ). For patients undergoing TKA, there was a statistical-

ly significant correlation between preoperative SF-36 physical functioning scores and knee function scores ( $r = 0.39$ ;  $P = .0026$ ).

**Change Over Time in SF-36 Scores After Surgery**

A complete dataset of preoperative and 3-, 12-, and 24-month follow-up SF-36 scores was obtained from 13 men and 29 women undergoing THA and 27 men and 43 women undergoing TKA. (A total of 126 patients were asked to complete the 3-month post-operative survey with the intent to also complete 12- and 24-month surveys; 89% completed all 3

postoperative surveys.) Scores for men and women were pooled for this analysis to increase sample size. **Figure 3** and **Figure 4** show the change in SF-36 scores after surgery. Although the greatest absolute score was most often found at 12 months, the greatest change from one time to another was from the presurgery evaluation to the 3-month evaluation. In no case did scores significantly increase from 12- to 24-month follow-up. Consequently, with the remainder of our study cohort, follow-up data were collected only at the 12-month visit.

**Assessment of Change After Surgery: Presurgery vs 12-Month Postsurgery SF-36 Scores**

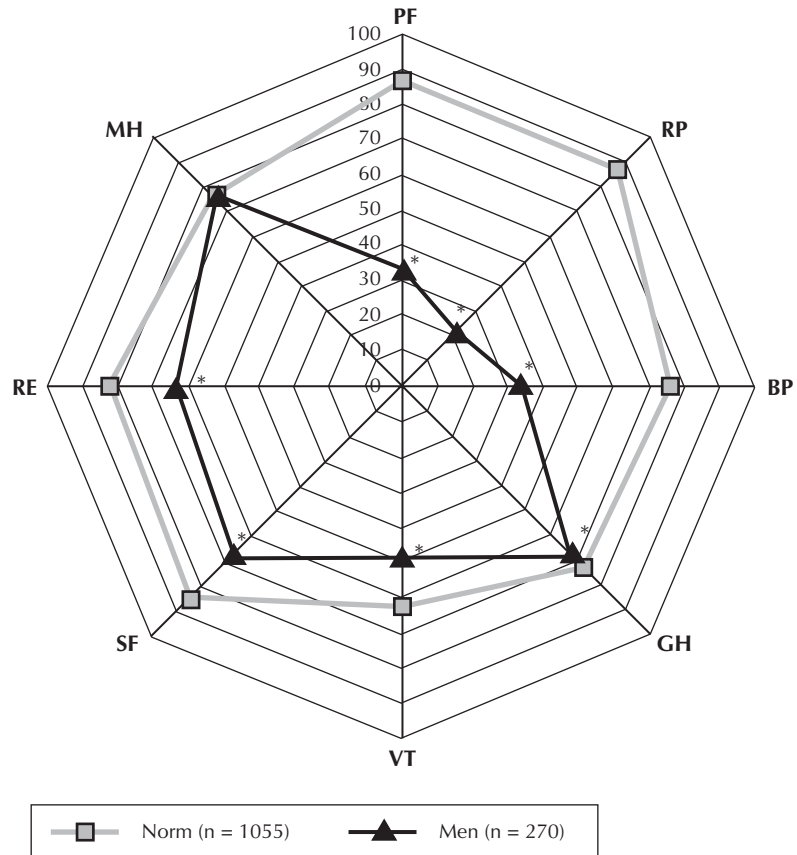
*Benefits of Surgery: Change at 12-Month Follow-up.* There was an 81.2% response rate for receipt of the 12-month follow-up SF-36. **Figures 5-8** show 12-month follow-up SF-36 scores compared with presurgery scores. In most cases, except general health for men who underwent THA, general health and mental health for women who underwent TKA, and role emotional for men who underwent TKA, improvements in scores were highly statistically significant. General health scores for men who underwent TKA decreased slightly but significantly.

*THA vs TKA.* For all domains of the SF-36 for both men and women (except general health for men who underwent THA), patients who underwent THA had statistically significantly higher scores at 12-month follow-up than did patients who underwent TKA ( $P < .04$ ).

*Sex Effect: Women vs Men.* Scores on the SF-36 at 12-month follow-up were consistently slightly greater for men than for women; differences were significant for physical functioning for patients who underwent THA ( $P = .0123$ ) ( $P = .09$  to  $.054$  for bodily pain, vitality, social functioning, and mental health) and for all domains ( $P < .04$ ) except role emotional for patients who underwent TKA.

*Effect of Age.* Age  $>65$  years did not preclude significant improvement in SF-36 scores; in fact, for

**Figure 2.** Radar Plot Comparing Mean Preoperative SF-36 Scores From Men Undergoing Total Joint Replacement (Hip or Knee) with Published Normative Scores (Age Range, 18 to  $>75$  years; data from reference 32.)



Each spoke represents a domain of the SF-36. Plots are read from the center outward along each spoke. Scores are shown on concentric circles, beginning with 0 (at the center) and increasing to 100 (outer line).

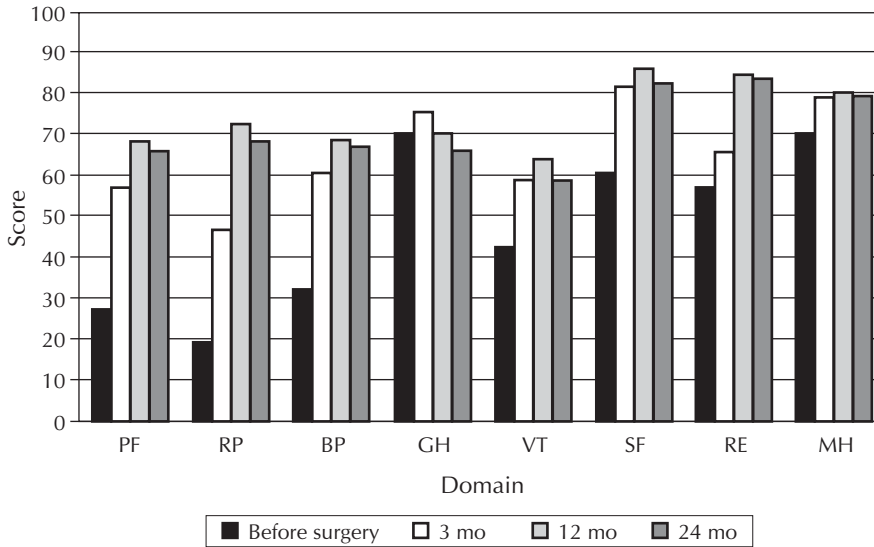
PF = physical functioning; RP = role physical; BP = bodily pain; GH = general health; VT = vitality; SF = social functioning; RE = role emotional; MH = mental health.

\*Statistically significant at  $P < .05$ .

men  $>65$  years who underwent TKA, 12-month scores were higher than for men aged 55 to 64 years for bodily pain, general health, vitality, social functioning, and mental health ( $P < .04$ ) (data not shown).

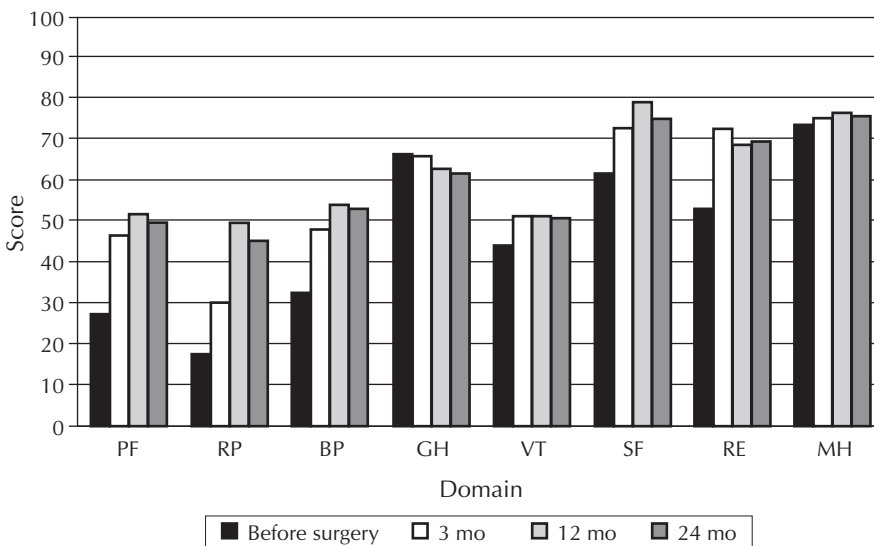
*Comparison to Normative Scores.* Although nearly all scores significantly increased from baseline to 12-month follow-up, differences still remained between patient scores and normative scores for the general population. Specifically, significant differences remained for physical functioning, role physical, bodily pain, and mental health for women who

**Figure 3.** Change Over Time in Pooled SF-36 Scores for Men and Women Who Underwent Total Hip Arthroplasty (n = 42)



PF = physical functioning; RP = role physical; BP = bodily pain; GH = general health; VT = vitality; SF = social functioning; RE = role emotional; MH = mental health.

**Figure 4.** Change Over Time in Pooled SF-36 Scores for Men and Women Who Underwent Total Knee Arthroplasty (n = 70)



PF = physical functioning; RP = role physical; BP = bodily pain; GH = general health; VT = vitality; SF = social functioning; RE = role emotional; MH = mental health.

underwent THA; all domains except mental health for women who underwent TKA; physical functioning and role physical for men who underwent THA; and all categories except mental health for men who underwent TKA (data not shown).

**Potential Technical Limitations and Cost of Using the SF-36**

Table 3 summarizes the percentage floor and ceiling effects for the preoperative SF-36 data. We estimated that the cost per patient to administer and score the SF-36, track follow-up, and readminister and score the SF-36 was approximately \$10 per patient. This retrospective estimate is for 1 preoperative and 1 postoperative SF-36 bubble form, postage, and the cost of personnel. About 5% to 10% of the time of the research staff (2 full-time equivalents) was devoted to this project. However, the initial purchase of hardware and software to scan and score the SF-36 was costly (several thousand dollars). Data tabulation, statistical evaluation, and interpretation required many hours on the part of the research team, including the director. A crude estimate of the cost per patient to complete this study is approximately \$50.

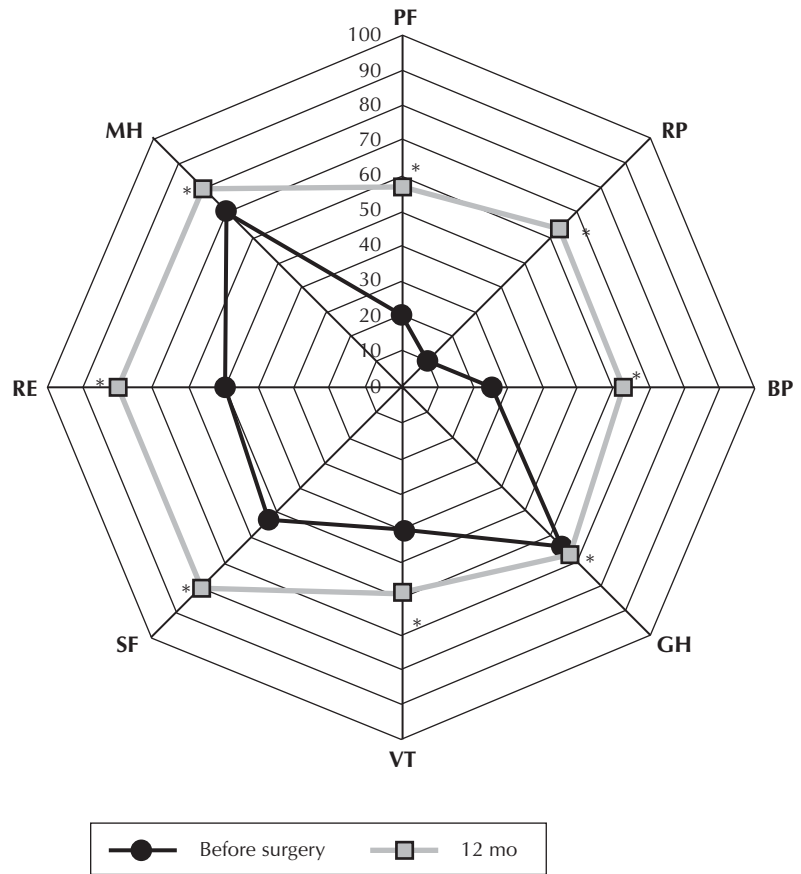
... DISCUSSION ...

The SF-36 effectively documented the dramatic difference in physical functioning and pain variables between patients with

degenerative osteoarthritis of the hip and knee and healthy men and women in the general US population. Degenerative osteoarthritis is a chronic disabling disease that affects all aspects of an individual's life.<sup>3,35-42</sup> Nearly all patients experience pain. Physical discomfort ranges from stiffness of afflicted joints to frank pain at rest and during the night while in bed trying to sleep. Because of pain, patients may experience functional limitations, including difficulty in doing housework, working, and participating in recreational activities. The inability to work results in earning losses and financial problems.<sup>39,43,44</sup> Clearly, the predominant reason patients eventually elect to undergo total joint replacement is to relieve pain and restore physical function.

For many patients, pain and physical function limitations may contribute to decreased self-esteem and self-confidence.<sup>42</sup> Several studies<sup>38,39,41,42</sup> have shown that individuals with osteoarthritis who have negative attitudes and low self-esteem score low on HRQOL measures. We observed significant differences in scores for the social functioning and mental health domains between patients with degenerative osteoarthritis and healthy individuals; however, these differences were generally smaller than those observed for scores for the physical functioning and bodily pain domains between patients with osteoarthritis and healthy individuals. This is not necessarily surprising because pain and physical function limitations are what typically drive the decision to have surgery. Patients who have developed coping strategies, who have psychological and social support to help manage the stress of osteoarthritis, and who can afford total joint replacement may rate their HRQOL higher than those who lack these support systems. We did not attempt to sort patients into groups based on level of social support. An alternative explanation is that patients with unilateral joint disease may not

**Figure 5.** Radar Plot Comparing Mean 12-Month Follow-up SF-36 Scores with Baseline (Preoperative) Scores for Women Who Underwent Total Hip Arthroplasty (n = 80)



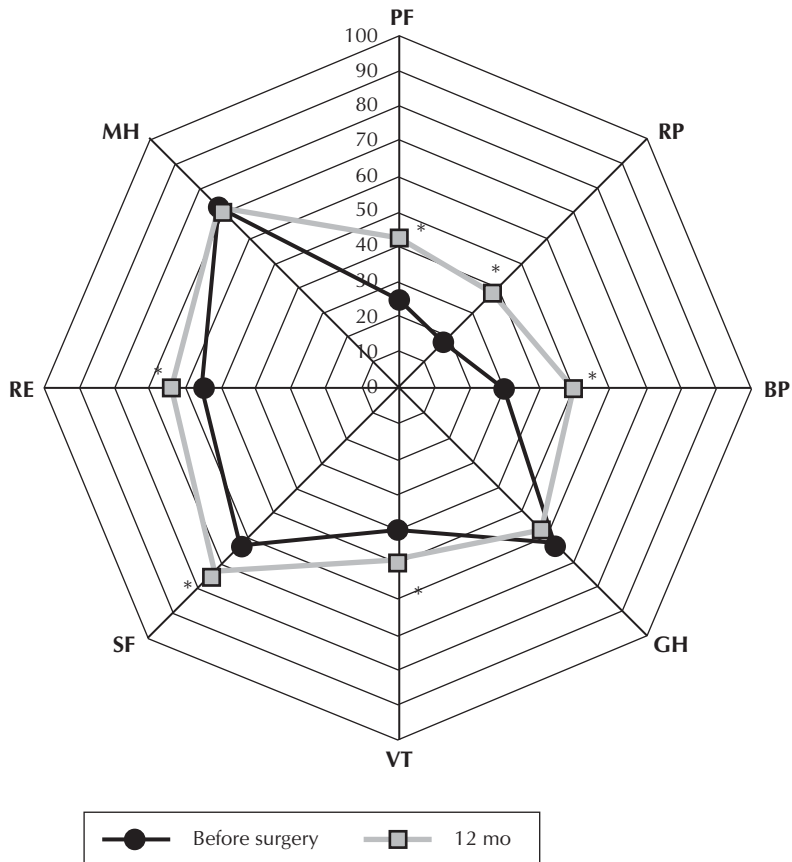
Each spoke represents a domain of the SF-36. Plots are read from the center outward along each spoke. Scores are shown on concentric circles, beginning with 0 (at the center) and increasing to 100 (outer line).

PF = physical functioning; RP = role physical; BP = bodily pain; GH = general health; VT = vitality; SF = social functioning; RE = role emotional; MH = mental health. Asterisks indicate a statistically significant difference between 12-mo follow-up and baseline at  $P < .05$  using the Wilcoxon matched pairs test. Rounded 95% confidence intervals for baseline scores were as follows: PF, 17-26; RP, 7-16; BP, 23-30; GH, 59-69; VT, 37-47; SF, 49-60; RE, 41-61; and MH, 66-75. Rounded 95% confidence intervals for 12-month scores were as follows: PF, 51-65; RP, 55-73; BP, 58-70; GH, 62-74; VT, 53-63; SF, 75-87; RE, 72-88; and MH, 74-82.

have the same degree of social and emotional problems as patients with multiple joint involvement. Finally, the SF-36 may not effectively document emotional, social, and mental health decrements in patients with osteoarthritis. Our study was not designed to clarify these issues.

There was no significant difference in preoperative scores between patients scheduled for THA or TKA. As a group, however, patients who underwent

**Figure 6.** Radar Plot Comparing Mean 12-Month Follow-up SF-36 Scores with Baseline (Preoperative) Scores for Women Who Underwent Total Knee Arthroplasty (n = 130)



Each spoke represents a domain of the SF-36. Plots are read from the center outward along each spoke. Scores are shown on concentric circles, beginning with 0 (at the center) and increasing to 100 (outer line).

PF = physical functioning; RP = role physical; BP = bodily pain; GH = general health; VT = vitality; SF = social functioning; RE = role emotional; and MH = mental health. Asterisks indicate a statistically significant difference between 12-month follow-up and baseline at  $P \leq .05$  using the Wilcoxon matched pairs test. Rounded 95% confidence intervals for baseline scores were as follows: PF, 22-28; RP, 13-23; BP, 27-33; GH, 58-65; VT, 37-44; SF, 58-67; RE, 48-63; and MH, 69-75. Rounded 95% confidence intervals for 12-month scores were as follows: PF, 38-47; RP, 30-44; BP, 45-54; GH, 55-63; VT, 44-51; SF, 68-78; RE, 58-71; and MH, 70-76.

THA recorded greater numerical increases in SF-36 scores for all domains at 3-month follow-up except role emotional. In general, as a group, patients who underwent TKA showed slower improvement and less dramatic increases in scores compared with those who underwent THA. Other researchers<sup>15,45</sup> also showed differences in the recovery rate of patients who underwent THA vs TKA: TKA is much

more painful, with a longer duration of pain after surgery.<sup>46</sup> Nonetheless, the explanation for the difference in response between patients who underwent THA vs TKA is not straightforward and warrants additional study.

Both preoperative and postoperative scores were generally lower for women than for men. This finding is not widely recognized because scores for men and women are often pooled when analyzing data. In the future, we must always stratify data by site of surgery (THA vs TKA) and sex. The explanation for the sex difference may relate to the slightly older age of women at the time of surgery and the fact that women tend to have a greater number of comorbidities at the time of surgery. Social support may be more of a problem for women than for men after surgery. These are factors that should be studied further.<sup>22</sup>

Normative data for HRQOL variables measured using the SF-36 showed a significant and consistent decrease with increasing age. In contrast, preoperative data for patients scheduled for THA or TKA did not show age-related differences. The lack of an age-related effect in patients may be a function of the floor effect, ie, scores are universally low for all patients with end-stage osteoarthritis, and additive effects of age-related decrements in function may not be well discriminated because the scoring algorithm does not allow for even lower scores to be recorded. In contrast, scores for older patients scheduled for THA and TKA compared with US population normative scores may reflect

the impact of demographic variables, such as greater affluence or social support systems, that enable these individuals to proceed with a major surgical procedure. The lack of an effect of age on HRQOL variables related to THA and TKA warrants further investigation.<sup>16,26</sup>

The greatest change in scores (compared with preoperative baseline measures) was seen at 3-

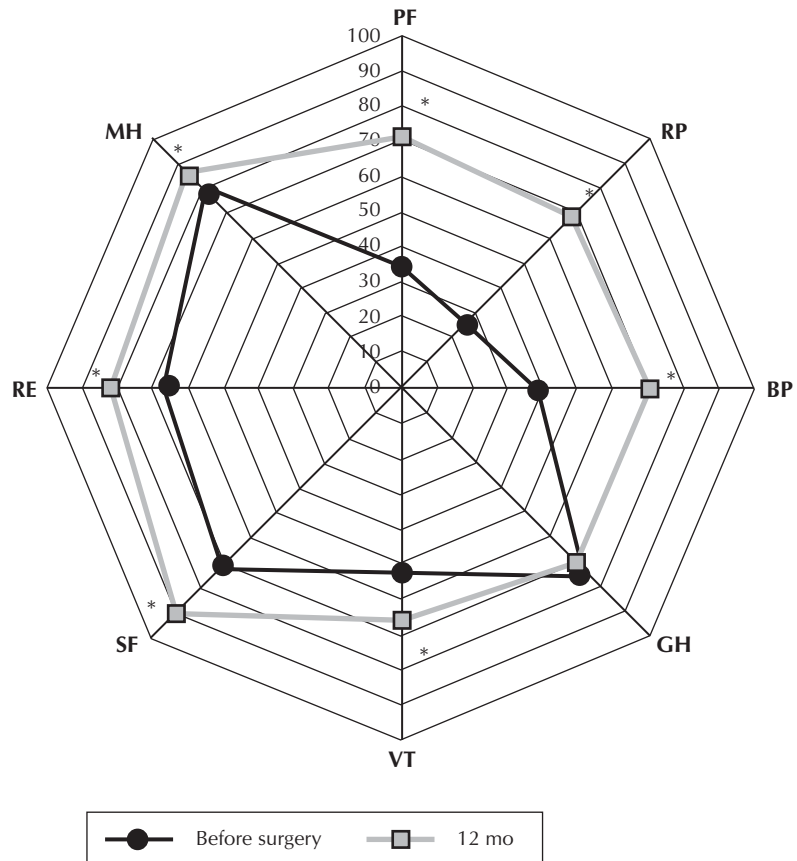


month follow-up.<sup>18</sup> After that time, scores improved less dramatically. In fact, 12-month general health scores for men who underwent TKA were slightly but significantly lower than baseline scores. No significant difference was seen between 12- and 24-month scores. At 24-month follow-up, patients may have new medical or social problems, such as the death of a spouse or a new disease process. Other researchers<sup>8,9</sup> have shown that scores for general health status surveys are more responsive to the effects of comorbidities than are scores for condition-specific surveys. The time course of change in SF-36 scores impacts how we administer HRQOL instruments and how we design studies that have the goal of monitoring change after surgery or the goal of validating the benefit of a new procedure.

The SF-36 is easy to administer and score, thus precluding use of more complicated and intricate condition-specific survey measures for the purpose of documenting quality of care and internal quality control. Improvement in SF-36 scores for the physical functioning and bodily pain domains paralleled improvement in clinical ratings for physical function and pain, thus validating these domains for use with patients undergoing total joint replacement (and, as mentioned previously, scores in the social functioning and mental health domains were lower than normative scores reported by others using a variety of assessment instruments.<sup>35-44</sup>) There are drawbacks, however, to use of the SF-36, and any type of survey instrument for that matter. First, there are operational costs involved with administration and processing of data obtained with any survey measure. There is no reimbursement for administration of these surveys. Collection of HRQOL data after surgery can become a costly endeavor for a clinic.

A potential technical limitation of the SF-36 relates to the scoring algorithm, with resultant ceil-

**Figure 7.** Radar Plot Comparing Mean 12-Month Follow-up SF-36 Scores with Baseline (Preoperative) Scores for Men Who Underwent Total Hip Arthroplasty (n = 61)

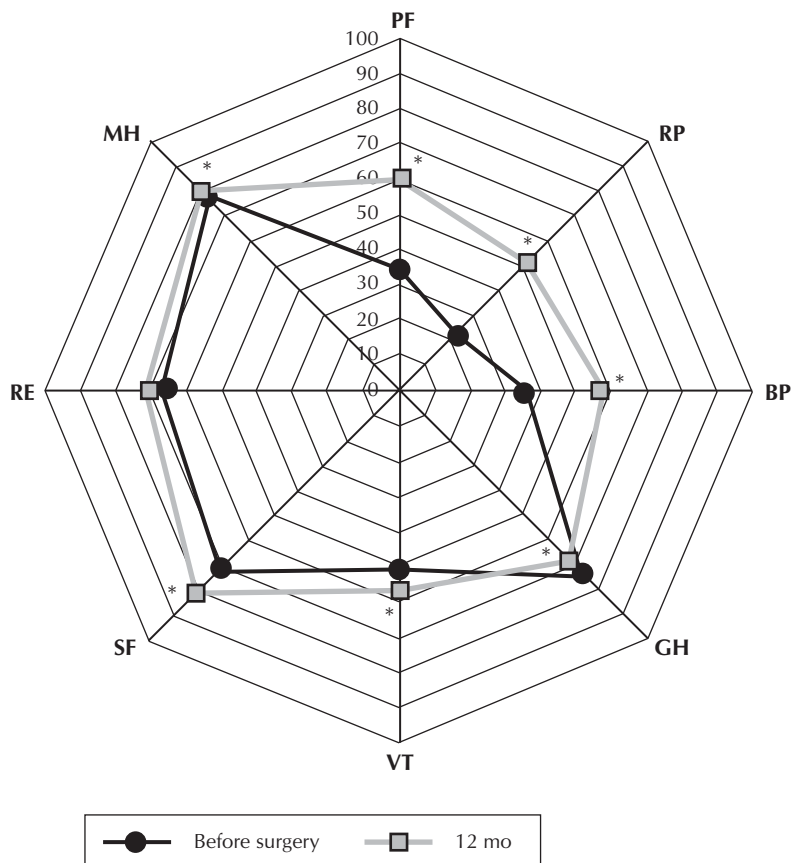


Each spoke represents a domain of the SF-36. Plots are read from the center outward along each spoke. Scores are shown on concentric circles, beginning with 0 (at the center) and increasing to 100 (outer line).

PF = physical functioning; RP = role physical; BP = bodily pain; GH = general health; VT = vitality; SF = social functioning; RE = role emotional; MH = mental health. Asterisks indicate a statistically significant difference between 12-month follow-up and baseline at  $P < .05$  using the Wilcoxon matched pairs test. Rounded 95% confidence intervals for baseline scores were as follows: PF, 28-40; RP, 16-33; BP, 33-42; GH, 69-77; VT, 45-56; SF, 63-76; RE, 56-79; and MH, 74-83. Rounded 95% confidence intervals for 12-month scores were as follows: PF, 65-77; RP, 57-78; BP, 66-77; GH, 65-74; VT, 62-70; SF, 86-94; RE, 72-89; and MH, 81-87.

ing and floor effects. A ceiling effect is when a respondent scores "100." A floor effect is when a respondent scores "0." Although operational definitions of "high and low" ceiling and floor effects are subjective, it is obvious that a floor effect was evident for role physical and that a ceiling effect was evident for role emotional. These floor and ceiling effects may affect our ability to measure the magnitude of a problem (see the previous discussion of

**Figure 8.** Radar Plot Comparing Mean 12-Month Follow-up SF-36 Scores with Baseline (Preoperative) Scores for Men Who Underwent Total Knee Arthroplasty (n = 105)



Each spoke represents a domain of the SF-36. Plots are read from the center outward along each spoke. Scores are shown on concentric circles, beginning with 0 (at the center) and increasing to 100 (outer line).

PF = physical functioning; RP = role physical; BP = bodily pain; GH = general health; VT = vitality; SF = social functioning; RE = role emotional; and MH = mental health. Asterisks indicate a statistically significant difference between 12-month follow-up and baseline at  $P < .05$  using the Wilcoxon matched pairs test. Rounded 95% confidence intervals for baseline scores were as follows: PF, 29-37; RP, 16-28; BP, 30-37; GH, 65-72; VT, 47-56; SF, 66-76; RE, 60-76; MH, 72-79. Rounded 95% confidence intervals for 12-month scores were as follows: PF, 55-65; RP, 44-60; BP, 52-62; GH, 60-69; VT, 53-62; SF, 75-86; RE, 63-79; and MH, 75-82.

the effect of age on SF-36 scores) or may impact the ability to detect change (because respondents cannot score lower than 0 or higher than 100). This may limit the usefulness of the SF-36 for evaluating change in an individual at follow-up. The SF-36 may best be used to evaluate changes in cohorts rather than in individuals because it is less likely for a group of patients to score 0 or 100. This being said,

however, there is no real evidence to suggest that the apparent floor effect for role physical and ceiling effect for role emotional (“apparent” because there is no benchmark for what constitutes “too high”) were confounding factors in our study.

In summary, our results clearly show that the SF-36 documents the burden of osteoarthritis and the benefits of total joint arthroplasty. Data from the SF-36 substantiate the fact that total joint replacement is one of the most successful modern surgical procedures and serve to temper efforts by those who would reduce availability of total joint replacement procedures by reducing reimbursement (which may heighten the surgeon’s and the patient’s threshold for electing to proceed with surgery) and decreasing quality of care by limiting length of stay (which decreases exposure to in-hospital physical therapy).<sup>47,48</sup> Many interesting secondary findings emerge, such as the lack of an age effect, differences between patients undergoing THA vs TKA, and the presence of a sex effect on HRQOL variables, and all deserve further study. But why should we use the SF-36 on a routine basis? Clinical hip and knee ratings<sup>33,34</sup> effectively document pain and function variables and are routinely used by many surgeons. Findings from physical examination and physician-patient interaction at follow-up further document the presence of pain and gait abnormalities.<sup>2</sup> Radiographic review at follow-up provides critical information about prosthesis variables and bone quality (eg, degree of periprosthetic bone

resorption, presence of osteolysis, and signs of component instability). All this information is typically recorded in the patient’s medical chart (although usually not in a systematic fashion). We knew before using the SF-36 that patients are much better after surgery.

One might assume that documenting quality of care with outcome measurements would be of bene-

**Table 3.** Floor and Ceiling Effects for Preoperative SF-36 Domains\*

	Physical Functioning	Role Physical	Bodily Pain	General Health	Vitality	Social Functioning	Role Emotional	Mental Health
<b>Men (n = 270)</b>								
Floor	3.33	57.41	4.44	0	1.48	1.85	23.70	0.37
Ceiling	0	7.78	0.37	0.74	1.11	24.44	55.19	6.67
<b>Women (n = 352)</b>								
Floor	9.66	71.31	8.52	0.28	3.98	2.56	30.68	0
Ceiling	0	3.13	0.28	2.56	0	14.20	41.76	2.56

\*Data are given as the percentage of patients who scored 0 (floor effect) or 100 (ceiling effect). Total hip arthroplasty and total knee arthroplasty data are pooled.

fit when negotiating contracts with payers. However, not one of the managed care companies that we have contracted with has ever requested HRQOL data to document the efficiency of our system and the skill of our surgeons. Use of outcome measurements to compare surgeons would lead to animosity and resistance to implementation of outcome programs. So, when should the private practice clinic use outcome assessment measures? We suggest that the SF-36 and similar survey measures for obtaining HRQOL data be used for periodic internal quality control (ie, practice surveillance when new procedures are implemented, new prosthetic designs are used, etc<sup>49</sup>) and for investigator-initiated research. There is no need to routinely track HRQOL variables after surgical procedures that have well-known favorable outcomes.<sup>50</sup> The value of tracking patient outcomes is well recognized. Many validated survey instruments are available. But before outcome assessment becomes truly entrenched in practice patterns, we must now learn how to apply the results we obtain to improve quality of care.

**Acknowledgments**

We thank Peggy Vain, BA, and Catherine Rooker, BS, for their editorial assistance and An Ly, BS, and Peggy Vain for their help with database management.

... REFERENCES ...

1. **Hozack WJ, Rothman RH, Albert TJ, Balderston RA, Eng K.** Relationship of total hip arthroplasty outcomes to other orthopaedic procedures. *Clin Orthop* 1997;344:88-93.  
 2. **McGrory BJ, Morrey BF, Rand JA, Ilstrup DM.** Correlation of patient questionnaire responses and physician history in grading clinical outcome following hip and knee arthroplasty. *J Arthroplasty* 1996;11:47-57.

3. **Dawson J, Fitzpatrick R, Murray D, Carr A.** Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg* 1998;80B:63-69.  
 4. **Dawson J, Fitzpatrick R, Carr A, Murray D.** Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg* 1996;78B:185-190.  
 5. **Naughton MJ, Shumaker SA.** Assessment of health-related quality of life in orthopaedic outcomes' studies. *Arthroscopy* 1997;13:107-113.  
 6. **Liang MH, Larson MG, Cullen KE, Schwartz JA.** Comparative measurement efficiency and sensitivity of 5 health status instruments for arthritis research. *Arthritis Rheum* 1985;28:542-547.  
 7. **Bombardier C, Melfi CA, Paul J, et al.** Comparison of a generic and a disease-specific measure of pain and physical function after knee replacement surgery. *Med Care* 1995;33:AS131-AS144.  
 8. **Hawker G, Melfi C, Paul J, Green R, Bombardier C.** Comparison of a generic (SF-36) and a disease specific (WOMAC) instrument in the measurement of outcomes after knee replacement surgery. *J Rheumatol* 1995;22:1193-1196.  
 9. **Kantz ME, Harris WJ, Levitsky K, Ware JE, Davies AR.** Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. *Med Care* 1992;30:MS240-MS252.  
 10. **Lieberman JR, Dorey F, Shekelle P, et al.** Outcome after total hip arthroplasty: Comparison of a traditional disease-specific and HRQOL measurement of outcome. *J Arthroplasty* 1997;12:639-645.  
 11. **Davies GM, Watson DJ, Bellamy N.** Comparison of the responsiveness and relative effect size of the Western Ontario and McMaster universities osteoarthritis index and the short-form medical outcomes study survey in a randomized, clinical trial of osteoarthritis patients. *Arthritis Care Res* 1999;12:172-179.  
 12. **Benroth R, Gawande S.** Patient-reported health status in total joint replacement. *J Arthroplasty* 1999;14:576-580.  
 13. **Ritter MA, Albohm MJ, Keating EM, Faris PM, Meding JB.** Comparative outcomes of total joint arthroplasty. *J Arthroplasty* 1995;10:737-741.  
 14. **Stucki G, Liang MH, Phillips C, Katz JN.** The Short Form-36 is preferable to the SIP as a generic health status measure in patients undergoing elective total hip arthroplasty. *Arthritis Care Res* 1995;8:174-181.  
 15. **Shields RK, Enloe LJ, Leo KC.** Health related HRQOL in patients with total hip or knee replacement. *Arch Phys Med Rehabil* 1999;80:572-579.  
 16. **March LM, Cross MJ, Lapsley H, et al.** Outcomes after hip or

knee replacement surgery for osteoarthritis: A prospective cohort study comparing patients' HRQOL before and after surgery with age-related population norms. *Med J Aust* 1999;171:235-238.

17. **McGuigan FX, Hozack WJ, Moriarty L, Eng K, Rothman RH.** Predicting HRQOL outcomes following total joint arthroplasty: Limitations of the SF-36 health status questionnaire. *J Arthroplasty* 1995;10:742-747.

18. **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;6:169-172.

19. **Heck DA, Robinson RL, Partridge CM, Lubitz RM, Freund DA.** Patient outcomes after knee replacement. *Clin Orthop* 1998;356:93-110.

20. **Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML.** Generic and condition-specific outcome measures for people with osteoarthritis of the knee. *Rheumatology* 1999;38:870-877.

21. **Greenfield S, Apolone G, McNeil BJ, Cleary PD.** The importance of co-existent disease in the occurrence of postoperative complications and 1-year recovery in patients undergoing total hip replacement. *Med Care* 1993;31:141-154.

22. **Katz JN, Wright EA, Guadagnoli E, et al.** Differences between men and women undergoing major orthopedic surgery for degenerative arthritis. *Arthritis Rheum* 1994;5:687-694.

23. **Young NL, Cheah D, Waddell JP, Wright JG.** Patient characteristics that affect the outcome of total hip arthroplasty: A review. *Can J Surg* 1998;41:188-195.

24. **Brinker MR, Lund PJ, Cox DD, Barrack RL.** Demographic biases found in scoring instruments of total hip arthroplasty. *J Arthroplasty* 1996;11:820-830.

25. **MacWilliam CH, Yood MU, Verner JJ, McCarthy BD, Ward RE.** Patient-related risk factors that predict poor outcome after total hip replacement. *Health Serv Res* 1996;31:623-638.

26. **Wilcock GK.** Benefits of total hip replacement to older patients and the community. *BMJ* 1978;1:37-39.

27. **Fortin PR, Clarke AE, Joseph L.** Outcomes of total hip and knee replacement: Preoperative functional status predicts outcomes at six months after surgery. *Arthritis Rheum* 1999;42:1722-1728.

28. **Anderson JG, Wixson RL, Tsai D, Stuhlberg SD, Chang RW.** Functional outcome and patient satisfaction in total knee patients over the age of 75. *J Arthroplasty* 1996;11:831-840.

29. **Wasielewski RC, Weed H, Prezioso C, Nicholson C, Puri RD.** Patient comorbidity: Relationship to outcomes of total knee arthroplasty. *Clin Orthop Rel Res* 1998;356:85-92.

30. **Cleary PD, Reilly DT, Greenfield S, et al.** Using patient reports to assess health-related HRQOL after total hip replacement. *Qual Life Res* 1993;2:3-11.

31. **Ware JE, Sherbourne CD.** The MOS 36-item short-form health survey (SF-36), I: Conceptual framework and item selection. *Med Care* 1992;30:473-483.

32. **Ware JE, Snow KK, Kosinski M, et al.** *SF-36 Health Survey: Manual and Interpretation Guide.* Boston, MA: The Health Institute, New England Medical Center; 1993.

33. **Harris WH.** Traumatic arthritis of the hip after dislocation and acetabular fractures: An end-result study using a new method of result evaluation. *J Bone Joint Surg* 1969;51A:737-755.

34. **Insall JN, Dorr LD, Scott RD, Scott WN.** Rationale of the Knee Society clinical rating system. *Clin Orthop* 1989;248:9-12.

35. **Kosinski M, Keller SD, Ware JE, Hatoum HT, Kong SX.** The SF-36 health survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis. *Med Care* 1999;37:MS23-MS29.

36. **Bayley KB, London MR, Grunkemeier GL, Lansky DJ.** Measuring the success of treatment in patient terms. *Med Care* 1995;33:AS226-AS235.

37. **Laborde JM, Powers MJ.** Life satisfaction, health control orientation, and illness-related factors in persons with osteoarthritis. *Res Nurs Health* 1985;8:183-190.

38. **Hopman-Rock M, Kraaimaat FW, Bijlisma JW.** HRQOL in elderly subjects with pain in the hip or knee. *Qual Life Res* 1997;6:67-76.

39. **Downe-Wamboldt B.** Stress, emotions, and coping: A study of elderly women with osteoarthritis. *Health Care Women Int* 1991;12:85-98.

40. **Creamer P, Hochberg MC.** Osteoarthritis. *Lancet* 1997;350:503-509.

41. **Downe-Wamboldt B.** Coping and life satisfaction in elderly women with osteoarthritis. *J Adv Nurs* 1991;16:1328-1335.

42. **Burkhardt C.** The impact of arthritis on HRQOL. *Nurs Res* 1985;34:11-16.

43. **Pincus T, Mitchell JM, Burkhauser RV.** Substantial work disability and earnings losses in individuals less than age 65 with osteoarthritis: Comparisons with rheumatoid arthritis. *J Clin Epidemiol* 1989;42:449-457.

44. **Brown GM, Dare CM, Smith PR, Meyers OL.** Important problems identified by patients with chronic arthritis. *S Afr Med J* 1987;72:126-128.

45. **Aarons H, Hall G, Hughes S, Salmon P.** Short-term recovery from hip and knee arthroplasty. *J Bone Joint Surg* 1996;78B:555-558.

46. **Mauerhan DR, Campbell M, Miller JS, Mokris JG, Gregory AM, Kiebzak GM.** Intra-articular morphine and/or bupivacaine in the management of pain after total knee arthroplasty. *J Arthroplasty* 1997;12:546-552.

47. **Mauerhan DR, Mokris JG, Ly A, Kiebzak GM.** Relationship between length of stay and manipulation rate after total knee arthroplasty. *J Arthroplasty* 1998;13:896-900.

48. **Lonergan RP, Mauerhan DR, Mokris JG.** Relationship between length of stay and dislocation rate after total hip arthroplasty. *J Arthroplasty.* In press.

49. **Callaghan JJ, Johnston RC, Pedersen DR.** Practice surveillance: A practical method to assess outcome and to perform clinical research. *Clin Orthop* 1999;369:25-38.

50. **Kravitz R.** Patient satisfaction with health care: Critical outcome or trivial pursuit? *J Gen Intern Med* 1998;13:280-282.