

# Health-Related Quality of Life After Elective Surgery

## Measurement of Longitudinal Changes

Carol M. Mangione, MD, MSPH, Lee Goldman, MD, MPH, E. John Orav, PhD, Edward R. Marcantonio, MD, Alex Pedan, MS, Lynn E. Ludwig, RNC, ANP, Magruder C. Donaldson, MD, David J. Sugarbaker, MD, Robert Poss, MD, Thomas H. Lee, MD, MSc

**OBJECTIVE:** To examine the responsiveness of the 36-Item Short Form Health Survey (SF-36) to clinical changes in three surgical groups and to study how health-related quality of life (HRQL) changes with time among patients who undergo total hip arthroplasty, thoracic surgery for treatment of non-small-cell lung cancer, or abdominal aortic aneurysm (AAA) repair.

**DESIGN:** Prospective cohort study with serial evaluations of HRQL preoperatively and at 1, 6, and 12 months after surgery.

**SETTING:** University tertiary care hospital.

**PATIENTS:** Of 528 patients, more than 50 years of age, who were admitted for these elective procedures, 454 (86%) provided preoperative health status data and are members of the study cohort. At 12 months after surgery, 439 (93%) of the cohort was successfully contacted and 390 (90%) provided follow-up interviews.

**MEASUREMENTS AND MAIN RESULTS:** The Medical Outcomes Study SF-36, the Specific Activity Scale, five validated health transition questions, and a 0 to 100 scale measure of global health were used to assess changes in health status at 1, 6, and 12 months after surgery. Change in health status as measured by the SF-36 demonstrated that physical function and role limitations due to physical health problems were worse 1 month after these three surgeries. However, by 6 months after surgery, most patients experienced significant gains in the majority of the dimensions of health, and these gains were sustained at 12 months after surgery. Longitudinal changes in the SF-36 were positively associated with responses to the five health transition questions, to changes on the Specific Activity Scale and global health rating question, and to clinical parameters for persons who had AAA repair. These findings indicate that the SF-36 has evidence of validity and is responsive to expected changes in HRQL after elective surgery for these procedures.

**CONCLUSIONS:** For the total hip arthroplasty patients, responsiveness was greatest for the SF-36 scales that measure physical constructs. However, for the two other procedures and at various points of recovery, significant changes were observed for all eight subscales, suggesting that responsiveness was dependent on the type of surgery and the timing of follow-up, and that multidimensional measures are needed to fully capture changes in HRQL after surgery.

**KEY WORDS:** health-related quality of life; elective surgery.  
J GEN INTERN MED 1997;12:686-697.

broaden the focus beyond traditional clinical outcomes to include the patient's perception of health-related quality of life (HRQL).<sup>5-9</sup> This investigation assesses the responsiveness of the Medical Outcomes Study (MOS) 36-Item Short Form Health Survey (SF-36 or RAND 36-Item Health Survey 1.0) when it is used to monitor changes in HRQL after major elective surgery and also examines how HRQL changes with time after three specific surgical procedures.<sup>10-13</sup> The elective surgical population is of particular interest because of the high costs and uncertain benefits associated with many procedures.

The SF-36 is an attractive measure because of its brevity, patient acceptability, and rigorous psychometric development.<sup>9-18</sup> In addition, published population-based data greatly enhance the interpretation of SF-36 scores.<sup>17</sup> The SF-36 has well-established reliability and validity in cross-sectional studies of medical and psychiatric patients,<sup>11,18-21</sup> and in longitudinal studies of selected medical conditions such as diabetes mellitus,<sup>15,22</sup> end-stage renal disease,<sup>16</sup> HIV infection,<sup>23,24</sup> asthma,<sup>25</sup> coronary disease,<sup>26</sup> and psychiatric diseases.<sup>27-29</sup>

Except for arthroplasty surgery,<sup>30-35</sup> however, only a few surgery-specific studies of highly select patients have provided data to support the SF-36's responsiveness to the effects of elective surgery over time.<sup>36-39</sup> This is the first investigation to examine simultaneously the responsiveness of the SF-36 across multiple procedures at various postoperative time intervals. These analyses describe

---

Received from the Departments of Medicine, Surgery, and Orthopedic Surgery, Brigham and Women's Hospital and Harvard Medical School, and Department of Biostatistics, Harvard School of Public Health, Boston, Mass.

Funded in part by a grant from the Agency for Health Care Policy and Research (1R01-HS06573). Dr. Mangione is the recipient of a Clinical Investigator Award (1K08-AG00605) from the National Institute on Aging, and is an awardee of the Robert Wood Johnson Foundation Generalist Physicians Faculty Scholars Program (029250).

Dr. Mangione is currently a member of the Division of General Internal Medicine and Health Services Research, University of California-Los Angeles School of Medicine, and Dr. Goldman is Chairman of the Department of Medicine, University of California-San Francisco School of Medicine.

Address correspondence and reprint requests to Dr. Mangione: UCLA Dept. of Medicine, Division of GIM and HSR, 10833 Le Conte Ave., Los Angeles, CA 90095-1736.

In recent years, pressures to prioritize health care spending have amplified the need to assess and compare the outcomes of medical interventions.<sup>1-4</sup> To this end, investigators have developed questionnaires that

686

preoperative HRQL and observed changes in health status after three major elective surgical procedures: total hip arthroplasty (THA); thoracic surgery for the treatment of non-small-cell lung cancer; and resection of abdominal aortic aneurysm (AAA).

**METHODS**

**Subjects**

Eligible subjects were patients older than 50 years of age who were admitted to Brigham and Women's Hospital, a 720-bed urban teaching hospital, for unilateral THA for the treatment of hip dysfunction caused by osteoarthritis, thoracic surgery (lobectomy or pneumonectomy) for the treatment of non-small-cell lung cancer, or repair of AAA, between November 1, 1990, and May 30, 1993. Eligibility criteria included the ability to speak English and adequate hearing and cognitive function to complete a preoperative self-administered health status questionnaire and postoperative telephone interviews. During the study period, 528 (95%) of the eligible patients who were approached by research staff gave informed consent and were enrolled in the study. Of these, 454 patients (86%), completed the preoperative health status questionnaire and constituted the cohort for this report. The distributions of gender, age, and race among the persons who did not return the preoperative questionnaire were similar to those among persons who did. However, a greater proportion of those who did not return the survey had three or more coexistent chronic medical conditions ( $p < .05$  for linear trend).

Of the 454 participating patients, 5 died before the 1-month follow-up interview, another died before the 6-month follow-up, and 17 more deaths occurred before the 12-month interview (Fig. 1). Of these 23 deaths, 14 were among thoracic surgery patients, 6 among those who had AAA repair, and 3 among the THA patients. Because changes in HRQL are relevant only for persons who survive, subjects were deleted from the analyses from the time of their death forward.

**Data Collection**

All subjects were evaluated a median of 1 day before surgery. Data were prospectively collected on medical history, including 26 underlying chronic medical conditions, physical examination, and mental status. The MOS SF-36 was included preoperatively and at 1, 6, and 12 months after surgery. The SF-36 includes questions that evaluate eight health dimensions: health perception, physical function, role limitations due to physical health problems and role limitations due to emotional problems, mental health, social function, vitality, and bodily pain.<sup>9</sup> Subjects also completed the Specific Activity Scale (SAS), an ordinal scaled, four-class measure of cardiovascular physical functioning (1 = best to 4 = worst) based on the metabolic expenditures of various personal care, housework,

occupational, and recreational activities before and at 6 months after surgery.<sup>40,41</sup>

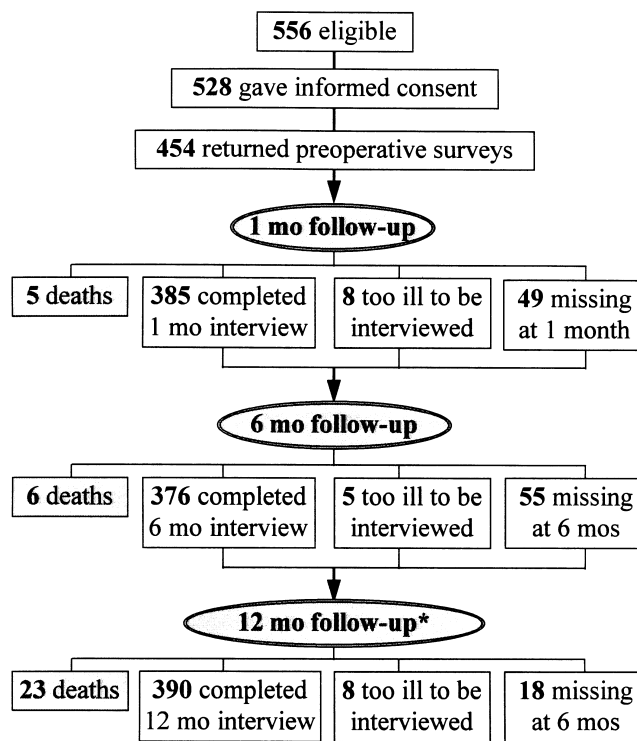
To determine whether participants considered their health to be unchanged, better, or worse when compared with their preoperative health, the 6-month follow-up interview included five dimension-specific health transition questions (Appendix A).<sup>42</sup> Finally, during all four interviews patients were asked the following question as a measure of global health:

*If you were to rate your current health on a scale from 0 to 100 with 100 being perfect health and 0 being death, what number would you rate yourself today?*

Responses from the non-SF-36 questions were used to evaluate the validity of the observed changes in the SF-36 after each of the elective procedures. Patients were interviewed at 1, 6, and 12 months after surgery to determine whether the SF-36 would be sensitive to short-term disability from elective surgery, and to the long-term benefits of surgery or declines in health due to progression of the underlying condition.

**Rationale for the Selection of Surgical Procedures and Expected Changes in Quality of Life**

We used the framework provided by the physical, mental, and combined constructs represented in the SF-36,<sup>11</sup> to classify the expected surgery-specific changes in SF-36



\* 93% follow-up rate at 12 months after surgery

FIGURE 1. Enrollment and follow-up rates.

scores over time (Table 1). A range of anticipated changes in HRQL were derived from published condition-specific descriptions of outcomes after THA, thoracic surgery for lung cancer, and repair of AAA. We have used these to evaluate the responsiveness of the SF-36 relative to expected longitudinal changes in clinical status over time.

**Total Hip Arthroplasty.** Previous studies of outcomes after THA have demonstrated dramatic improvements in HRQL, particularly for the dimensions of pain and physical functioning.<sup>36-39</sup> From published findings, we anticipated that after transient procedure-related declines in physical health, THA patients would report large improvements on the SF-36 subscales that share variance with the physical construct and moderate improvements on the combined construct of the SF-36.<sup>11</sup> Because chronic pain can influence the severity of symptoms related to emotional well-being, we also expected subscales that share variance with the mental construct to improve moderately after surgery (Table 1).

**Thoracic Surgery for Lung Cancer.** As described by Dales et al.,<sup>43</sup> we expected patients to report substantial declines in physical functioning during the immediate postoperative period followed by improvements with possible return to preoperative levels by 12 months after surgery for those who remained disease-free. If the SF-36 is a useful tool for the longitudinal assessment of persons with lung cancer, it should be sensitive to the immediate physical and emotional consequences of lost respiratory capacity due to surgery and postoperative pain, as well as the long-term effect of uncertainty about prognosis (Table 1).

**Abdominal Aortic Aneurysm Repair.** Repair of AAA is a procedure with known moderate to severe short-term morbidity, therefore we expected improvement among survivors on the subscales that represent the mental health construct. Because vascular disease of the aorta rarely occurs in isolation, we expected these patients to have more coexistent vascular disease at the time of surgery. Previous investigators have demonstrated that the SF-36 is sensitive to the effect of coexistent medical conditions.<sup>44-46</sup> For these reasons we expected the AAA patients to have pronounced short-term declines in SF-36 subscales that represent the physical and combined constructs. For those who survived the procedure, we ex-

pected a return to preoperative levels of HRQL, rather than improvements related to the procedure (Table 1).

## Statistical Analysis

**Health Status Scores.** The scoring algorithm for the SF-36 was identical to that described by Ware et al.<sup>17</sup> Mean scores were calculated by surgical procedure. The percentages of persons with the highest and lowest possible preoperative SF-36 scores also were recorded.

**Changes in Health Status over Time.** Mean changes in health status scores at 1, 6, and 12 months after surgery relative to preoperative scores were calculated for the three surgical groups. Student's paired *t* tests were used to assess the significance of within-subject changes in the SF-36 subscales and the 0 to 100 global rating scale. The Wilcoxon Sign-Rank Test was used to test for significant change in the SAS at 6 months after surgery.<sup>47</sup>

**Responsiveness of the SF-36.** To assess the responsiveness of each SF-36 subscale, we compared the square of the absolute value of the *T* statistic (12-month after-surgery score – preoperative score/SD of the difference) for each of the subscales relative to the subscale with the largest *T* statistic. This comparison simultaneously accounts for the magnitude of change over time and the variability in each subscale. The relative efficiency statistic  $(T_1/T_2)^2$ , first described by Liang et al.,<sup>48</sup> identifies which SF-36 subscale at each point in time is most responsive for the three procedures under investigation.

**Validity of the SF-36.** To evaluate the validity of the longitudinal changes observed with the SF-36 for the three surgical groups, we conducted three statistical comparisons.<sup>49</sup> The first test compared the magnitude of the correlations between changes in the SF-36 and changes in the overall-health-rating question (0–100) and SAS class. We interpreted the clinical significance of these observed correlations by adopting a framework described by McHorney et al.,<sup>11</sup> in which correlations that are greater than or equal to 0.70 are classified as “high,” those between 0.7 and 0.3 are considered to be “moderate,” and those less than 0.3 are considered to be “low.” Because response to the 0 to 100 rating question was based on a person's synthesis of overall health, our expectation was that changes

**Table 1. Hypothesized Changes in the 36-Item Short Form Health Survey's Physical, Mental, and Combined Health Constructs Among Survivors\***

Constructs	THA <sup>†</sup>			Thoracic Surgery for Lung Cancer (Without Recurrence)			AAA Repair <sup>†</sup>		
	1 Mo	6 Mo	12 Mo	1 Mo	6 Mo	12 Mo	1 Mo	6 Mo	12 Mo
Physical	↓	↑↑	↑↑	↓↓	↓	↔	↓↓	↔	↔
Mental	↑	↑	↑	↓	↑	↑	↑	↑	↔
Combined	↓	↑	↑	↓	↔	↔	↓	↔	↔

\*Symbols: ↑ indicates improved relative to preoperative levels; ↓, declined relative to preoperative levels; ↔, unchanged from preoperative levels.

<sup>†</sup>THA indicates total hip arthroplasty; AAA, abdominal aortic aneurysm.

in the response to this question would be moderately correlated with the SF-36 subscales that contribute to the combined construct (health perception and vitality) and that the remainder of the subscales would have low correlations ( $<0.3$ ) with the 0 to 100 rating question.

The SAS covers some physical activities that are similar to those found in the SF-36 physical functioning scale; therefore we expected changes in this measure to correlate moderately (0.3–0.7) with changes in the SF-36 subscales that measure the physical construct (physical functioning, role limitations due to physical problems, and bodily pain). Because poorer cardiovascular functioning can lead to fatigue, we expected moderate correlations with changes in SAS and the SF-36 vitality scale. We expected low correlations between the SAS and the remaining SF-36 scales.

A person's impression of transition in HRQL is the closest proxy to a "gold standard" for change in HRQL (Appendix A).<sup>42</sup> Our second statistical comparison uses data from transition questions to compare the magnitude of the dimension-specific changes in SF-36 scores for participants who reported that they were better, the same, or worse at 6 months after surgery. To provide a reference point for the clinical interpretation of the SF-36 scores, the third test statistic uses Student's *t* tests for pairwise comparisons of preoperative and postoperative scores to published age- and gender-adjusted population-based scores.<sup>17</sup> To adjust the published scores, they were weighted to match the surgery-specific distribution of age and gender in our sample.

**Tests of Clinical Validity: Procedure-Specific Analyses of SF-36 Scores.** Within the AAA subgroup, validity of the observed SF-36 subscale changes was also assessed using clinical parameters by comparing mean change at 1 year after surgery among those with more than three coexistent medical conditions relative to those with fewer than three comorbidities. From the work of Greenfield et al.,<sup>44–46</sup> we expected that those with more medical comorbidity would have greater declines in HRQL over time.

**Reliability of the SF-36.** To determine whether the SF-36 was equally reliable across different surgical categories, we calculated Cronbach's coefficient  $\alpha$  as a measure of internal consistency for each of the multi-item subscales of the SF-36 in each of the groups.<sup>50</sup> Significance was defined as a two-tailed *p* value  $\leq .05$  for all statistical tests.

## RESULTS

### Population

The 454 patients (54% female; 96% white) had a mean ( $\pm$  SD) age of  $67 \pm 9$  years (Table 2). Common comorbid conditions were hypertension (42%), arthritis (22%), ischemic heart disease (22%), and diabetes mellitus (8%); 238 (52%) of the patients had only one or no comorbid conditions (Table 2). Patients in the three surgical groups varied by demographic and medical characteristics.

The preoperative SF-36 scores that contribute to the physical construct (physical function, role limitations due to physical limitations, and bodily pain) were lowest for patients scheduled for THA, whereas scores on the SF-36 health perception subscale were lower for patients with potentially life-threatening conditions (lung cancer or AAA) (Table 3). More than one third of the preoperative scores were at the ceiling or floor of the role-physical and role-emotional subscales and the social-functioning subscale, indicating that for these three subscales, scores after surgery could only stay the same or change in one direction. Reliability estimates based on Cronbach's coefficient  $\alpha$  ranged from 0.92 to 0.86 for the SF-36 subscales.

### Responsiveness of the SF-36 to Within-Procedure Change in Health Status over Time

In this section we first compare the observed changes in SF-36 scores to the expected changes described in Table 1. We then report which SF-36 scale is most responsive for each procedure, and finally we compare the preoperative and postoperative SF-36 scores to a population-based sample. As was hypothesized, for patients undergoing any one of the elective surgeries under consideration, SF-36 scores obtained 1 month after surgery demonstrated significant declines in physical functioning and role limitations due to physical problems (Table 4).

**Total Hip Arthroplasty.** An unexpected early benefit from surgery was large and significant improvement in bodily pain. Improvements in pain became even more pronounced by 6 months and were sustained at 12 months after surgery. The dramatic improvements observed across the mental and combined health constructs were of greater magnitude than anticipated. By 12 months after surgery, the THA patients had substantial and significant ( $p < .05$ ) improvements across all eight SF-36 dimensions of HRQL.

For patients undergoing THA, bodily pain was the most responsive subscale at all three points in time. However, at 6 and 12 months after surgery, the scales for physical functioning and role limitations due to physical problems had relative efficiency statistics above 0.47, indicating that they were also moderately responsive.

The clinical relevance of the observed changes in SF-36 scores can be appreciated when the THA patients' scores are compared with those from an age- and gender-adjusted population-based sample (Fig. 2). For seven of the eight subscales, the preoperative scores were significantly lower than scores from the population sample. However, by 6 months after surgery, all eight subscales were the same or better for the THA patients relative to the population-based sample. This improvement was sustained at 12 months after surgery.

**Thoracic Surgery for Lung Cancer.** As expected, patients who had thoracotomy reported increased pain and significant declines in all dimensions of health that contribute

Table 2. Description of the Study Population (n = 454)

Characteristic	Surgery Type			p Value <sup>†</sup>
	THA* n = 236 (52%)	AAA Repair* n = 95 (21%)	Thoracic Surgery for Lung Cancer n = 123 (27%)	
Mean age + SD	67 ± 9	72 ± 8	64 ± 8	<.001
Female gender, n (%)	137 (58)	17 (18)	56 (46)	<.001
White race, n (%)	225 (95)	93 (98)	119 (97)	NS
Specific Activity Scale, <sup>‡</sup> n (%)				
1	46 (20)	31 (33)	53 (43)	<.001
2	59 (25)	35 (37)	39 (32)	
3	111 (47)	29 (31)	30 (24)	
4	19 (8)	0 (0)	1 (7)	
0–100 Scale rating, mean ± SD	75 ± 16	70 ± 21	75 ± 20	NS
Insurance				
Private (fee for service)	180 (76)	83 (87)	87 (71)	NS
HMO	38 (16)	6 (6)	26 (21)	
Medicare only	15 (6)	5 (5)	6 (5)	
Medicaid only	1 (0.4)	0 (0)	3 (2)	
Self-pay	1 (0.4)	0 (0)	1 (1)	
Number of comorbidities, <sup>§</sup> n (%)				
0	65 (28)	6 (6)	21 (17)	<.001
1	88 (37)	20 (21)	38 (31)	
2	52 (22)	23 (24)	32 (26)	
3	23 (10)	17 (18)	17 (14)	
4 or more	8 (3)	29 (31)	15 (12)	
ASA class, <sup>  </sup> n (%)				
I	16 (7)	0 (0)	3 (2)	<.001
II	164 (69)	27 (29)	53 (43)	
III	56 (24)	60 (64)	65 (53)	
IV	0 (0)	7 (7)	2 (2)	

\*THA indicates total hip arthroplasty; AAA, abdominal aortic aneurysm.

<sup>†</sup> p Value for pairwise comparison of age and 0–100 rating, and comparison of proportions for all categorical variables. NS indicates not significant.

<sup>‡</sup> An ordinal scaled four-class physical functioning instrument, 1 (best) to 4 (worst), based on the metabolic expenditures of various personal care, housework, occupational, and recreational activities (data missing, n = 6).

<sup>§</sup> History of comorbidities by patient self-report and from medical record review.

<sup>||</sup> American Society of Anesthesiologists Physical Status Class, a 5-level classification scheme on the physician's judgment of severity of systemic disease; higher value represents greater severity.

to the physical construct. Unexpectedly, at 1 month after surgery, mental health, role limitations due to mental health, and health perception did not decline. However, by 6 months after surgery, health perception became significantly poorer than preoperative levels and did not improve by 12 months after surgery. Patients continued to report significant declines in physical function, bodily pain, and vitality at 6 and 12 months after surgery. As expected, subscales that contribute to the mental construct improved by 12 months after surgery.

For persons undergoing resection of lung cancer, the bodily pain scale was the most responsive SF-36 scale at 1 month after surgery. However, at later follow-up times, 6 of the 8 subscales were moderately to highly responsive (relative efficiency statistic 0.4–0.9), suggesting that the patients treated with thoracotomy for lung cancer experienced positive and negative changes across many dimensions of health, rather than along a particular construct.

When compared with a population-based sample, preoperative patients with non-small-cell lung cancer had significantly poorer HRQL principally among the scales that share variance with the mental construct (social function, role limitations due to mental health problems, and mental health) and the combined construct (health perception and vitality) (Fig. 2). As only patients with localized disease are considered candidates for surgery, it is not surprising that preoperative HRQL along the physical construct was not significantly worse than that of an age- and gender-adjusted population-based sample. At 1 month after surgery, all SF-36 subscales except for health perception were significantly lower than the scores reported by a comparable population-based sample. Although lower than preoperative levels, by 12 months after surgery, the SF-36 physical functioning, bodily pain, and health perception scales were no longer different from those reported in the population-based sample (Fig. 2).

Table 3. Preoperative 36-Item Short Form Health Survey Scores by Surgical Procedure (n = 454)

SF-36 Subscales*	Surgery Type		
	THA† n = 236 (52%)	Thoracic Surgery for Lung Cancer n = 123 (27%)	AAA Repair† n = 95 (21%)
Physical function, mean (SD)	31 (23)	69 (27)	64 (26)
Ceiling, %	<1	20	3
Floor, %	9	<1	0
Role-physical, mean (SD)	21 (34)	51 (46)	52 (44)
Ceiling, %	11	39	38
Floor, %	63	39	32
Bodily pain, mean (SD)	31 (17)	76 (28)	72 (26)
Ceiling, %	0	47	33
Floor, %	8	2	0
Mental health, mean (SD)	72 (18)	70 (20)	71 (21)
Ceiling, %	2	3	5
Floor, %	0	0	0
Role-emotional, mean (SD)	61 (45)	66 (43)	67 (42)
Ceiling, %	54	57	55
Floor, %	28	24	22
Social function, mean (SD)	58 (29)	73 (27)	80 (24)
Ceiling, %	17	33	44
Floor, %	3	<1	0
Vitality, mean (SD)	46 (21)	56 (25)	54 (23)
Ceiling, %	<1	5	2
Floor, %	<1	<1	0
Health perception, mean (SD)	72 (19)	68 (22)	66 (19)
Ceiling, %	6	5	1
Floor, %	0	0	0

\*All health status scores are calibrated such that 0 points is the worst possible response on the scale and 100 points is the best possible response.

†THA indicates total hip arthroplasty; AAA, abdominal aortic aneurysm.

**Abdominal Aortic Aneurysm Repair.** As was expected, patients undergoing AAA repair reported significant declines in most of the scales that contribute to the physical and combined constructs at 1 month after surgery, but these declines returned to preoperative levels or better by 6 months after surgery (Table 4). The return to preoperative levels was sustained at 12 months after surgery. Unexpectedly, by 6 months after surgery, patients reported significant improvements in scales that contribute to the mental construct relative to preoperative levels.

Because patients who had AAA repair were still recovering from surgery at 1 month, the most responsive subscales changed in the negative direction; these included role limitations due to physical problems, vitality, physical functioning, and social functioning (relative efficiency statistic > 0.5). By 6 months after surgery, the responsiveness of the AAA patients was greatest for the subscales that contribute to the mental construct of the SF-36. As was expected, the 12-month change scores for six of the eight subscales were negative (-7 to -12 points) for AAA patients who had three or more medical comorbidities before surgery. These change scores were significantly lower than those for AAA patients who had fewer comorbidities.

When compared with a population-based sample, preoperative AAA patients had significantly greater role limita-

tions due to physical or emotional problems, and significantly poorer mental health scores. At 1 month after surgery, except for pain and role limitations due to mental health problems, all of the SF-36 scores were significantly lower for AAA patients. However, by 6 months after surgery, all eight subscales were the same or better than those of a comparable population-based sample, and these improvements were sustained at 12 months after surgery (Fig. 2).

The 0 to 100 health-rating scores improved significantly for THA patients at 6 and 12 months after surgery and declined significantly for the patients with lung cancer. The SAS improved only for the THA patients.

### Validity of the SF-36

There was evidence of correlational validity when change scores on the SF-36 were compared with the observed changes using other questions designed to capture similar information. The strongest evidence was moderate correlation between changes at 6 months for the SF-36 subscales that contribute to the combined construct (vitality and health perception) and 6-month changes observed with the second measure of global health, the 0 to 100 rating question ( $r = .4$ ;  $p = .0001$ ). Of note, changes in the 0 to 100 rating question were also moderately

Table 4. Responsiveness of the 36-Item Short Form Health Survey by Surgical Procedure

SF-36 Subscale	THA*			Thoracic Surgery for Lung Cancer			AAA Repair*		
	1 Mo n = 212	6 Mo n = 210	12 Mo n = 224	1 Mo n = 104	6 Mo n = 95	12 Mo n = 88	1 Mo n = 69	6 Mo n = 71	12 Mo n = 78
<b>Physical construct</b>									
Physical function									
Mean change <sup>†</sup>	-5.1	29.4	32.6	-19.8	-6.8	-7.3	-15.0	1.7	13.8
(SD)	(26.5) <sup>‡</sup>	(30.5) <sup>‡</sup>	(29.4) <sup>‡</sup>	(29.2) <sup>‡</sup>	(27.1) <sup>‡</sup>	(25.8)	(29.5) <sup>‡</sup>	(26.8)	(27.6)
T statistic	-2.8	13.9	16.6	-6.9	-2.4	-2.6	-4.2	0.5	0.6
Relative efficiency <sup>‡</sup>	<0.1	0.6	0.5	0.6	0.4	0.4	0.5	<0.1	<0.1
Role-physical									
Mean change	-7.6	47.3	45.4	-36.9	1.1	-0.5	-37.9	15.3	16.1
(SD)	(42.9) <sup>‡</sup>	(48.8) <sup>‡</sup>	(46.3) <sup>‡</sup>	(49.1) <sup>‡</sup>	(52.9) <sup>‡</sup>	(57) <sup>‡</sup>	(80.5) <sup>‡</sup>	(56.7) <sup>‡</sup>	(80.9) <sup>‡</sup>
T statistic	-2.6	14.0	14.6	-7.5	0.2	-0.1	-6.2	2.3	2.8
Relative efficiency	<0.1	0.6	0.4	0.8	<0.1	<0.1	1	0.8	0.7
Bodily pain									
Mean change	33.9	36.1	39.4	-27.1	-7.4	-11.5	-4.3	1.1	-2.1
(SD)	(27.7) <sup>‡</sup>	(28.1) <sup>‡</sup>	(24.7) <sup>‡</sup>	(32.0) <sup>‡</sup>	(28.8) <sup>‡</sup>	(30.8) <sup>‡</sup>	(32.2)	(27.8)	(29.3)
T statistic	17.8	18.6	23.9	-8.6	-2.5	-3.5	-1.1	0.3	-0.6
Relative efficiency	1	1	1	1	0.4	0.7	<0.1	<0.1	<0.1
<b>Mental construct</b>									
Mental health									
Mean change	6.4	9.3	8.7	0.8	7.4	6.7	1.5	4.6	6.1
(SD)	(20.4) <sup>‡</sup>	(19.9) <sup>‡</sup>	(19.5) <sup>‡</sup>	(19.4)	(18.8) <sup>‡</sup>	(17.3) <sup>‡</sup>	(21.2)	(18.90) <sup>‡</sup>	(19.3) <sup>‡</sup>
T statistic	4.5	6.8	6.6	0.4	3.9	3.6	0.6	2.0	2.8
Relative efficiency	<0.1	0.1	0.1	<0.1	1	0.8	<0.1	0.6	0.7
Role-mental									
Mean change	22.9	32.8	24.8	2.9	18.9	12.7	5.5	15.4	16.7
(SD)	(52.1) <sup>‡</sup>	(49.7) <sup>‡</sup>	(46.8) <sup>‡</sup>	(54.5)	(49) <sup>‡</sup>	(52.1) <sup>‡</sup>	(48.4)	(49.3) <sup>‡</sup>	(43.8) <sup>‡</sup>
T statistic	6.4	9.5	7.9	0.5	3.7	2.3	0.9	2.6	3.3
Relative efficiency	0.1	0.3	0.1	<0.1	0.9	0.3	<0.1	1	1
Social function									
Mean change	-0.1	28.3	28.3	-18.3	7.6	-0.8	-19.0	5.3	3.8
(SD)	(41.9)	(32.2) <sup>‡</sup>	(30.1) <sup>‡</sup>	(38) <sup>‡</sup>	(28) <sup>‡</sup>	(32)	(37.1) <sup>‡</sup>	(29.0)	(22.4)
T statistic	-0.04	12.8	14.1	-4.9	2.6	-0.2	-4.22	1.53	1.5
Relative efficiency	0	0.5	0.4	0.3	0.4	<0.1	0.5	0.3	0.2
<b>Combined construct</b>									
Vitality									
Mean change	7.1	16.6	15.7	-19.9	-6.0	-7.0	-15.6	1.0	-0.2
(SD)	(25.9) <sup>‡</sup>	(22.9) <sup>‡</sup>	(25.0) <sup>‡</sup>	(23.8) <sup>‡</sup>	(24.3) <sup>‡</sup>	(25.0) <sup>‡</sup>	(21.3) <sup>‡</sup>	(19.8)	(21.7)
T statistic	4.0	10.5	9.4	-8.5	-2.4	-2.6	-6.0	0.4	-0.1
Relative efficiency	<0.1	0.3	0.2	1.0	0.4	0.4	0.9	<0.1	<0.1
Health perception									
Mean change	7.3	3.9	2.7	-2.8	-7.2	-11.1	5.8	-1.0	0.1
(SD)	(16.3) <sup>‡</sup>	(18.2) <sup>‡</sup>	(19.3) <sup>‡</sup>	(18.1)	(22.4) <sup>‡</sup>	(25.2) <sup>‡</sup>	(19.1) <sup>‡</sup>	(18.5)	(20.4)
T statistic	6.5	3.1	2.1	-1.6	-3.1	-4.1	2.5	-0.4	0.05
Relative efficiency	0.1	<0.1	<0.1	0.3	0.6	1	0.2	<0.1	<0.1
0-100 Rating									
Mean change	2.5	3.9	2.9	-8.0	-6.7	-10.7	3.4	2.5	1.6
(SD)	(18.8)	(16.9) <sup>‡</sup>	(17.0) <sup>‡</sup>	(19.1) <sup>‡</sup>	(18.6) <sup>‡</sup>	(23.7) <sup>‡</sup>	(24.1)	(19.0)	(21.8)
T statistic	1.9	3.2	2.5	-4.1	-3.3	-4.2	1.1	1.1	0.6
SAS <sup>  </sup>									
Mean change	NA	0.5	NA	NA	0.11	NA	NA	0.2	NA
(SD)		(1.1) <sup>‡</sup>			(0.9)			(0.9)	
T statistic		6.7			1.13			1.9	

\*THA indicates total hip arthroplasty; AAA, abdominal aortic aneurysm.

<sup>†</sup>All changes are relative to the observed scores before surgery. NA indicates not applicable.

<sup>‡</sup> p < .05 for the paired t test

<sup>§</sup>The relative efficiency statistic (RES) is a ratio of the absolute value of the T statistic for the change in each subscale over time relative to the subscale with the largest T statistic, squared:  $(T_1/T_2)^2$ . The closer the RES is to 1, the more responsive a subscale is relative to the most responsive subscale in the SF-36.

<sup>||</sup>Participants were only asked to complete the Specific Activity Scale (SAS) before and at 6 months after surgery.

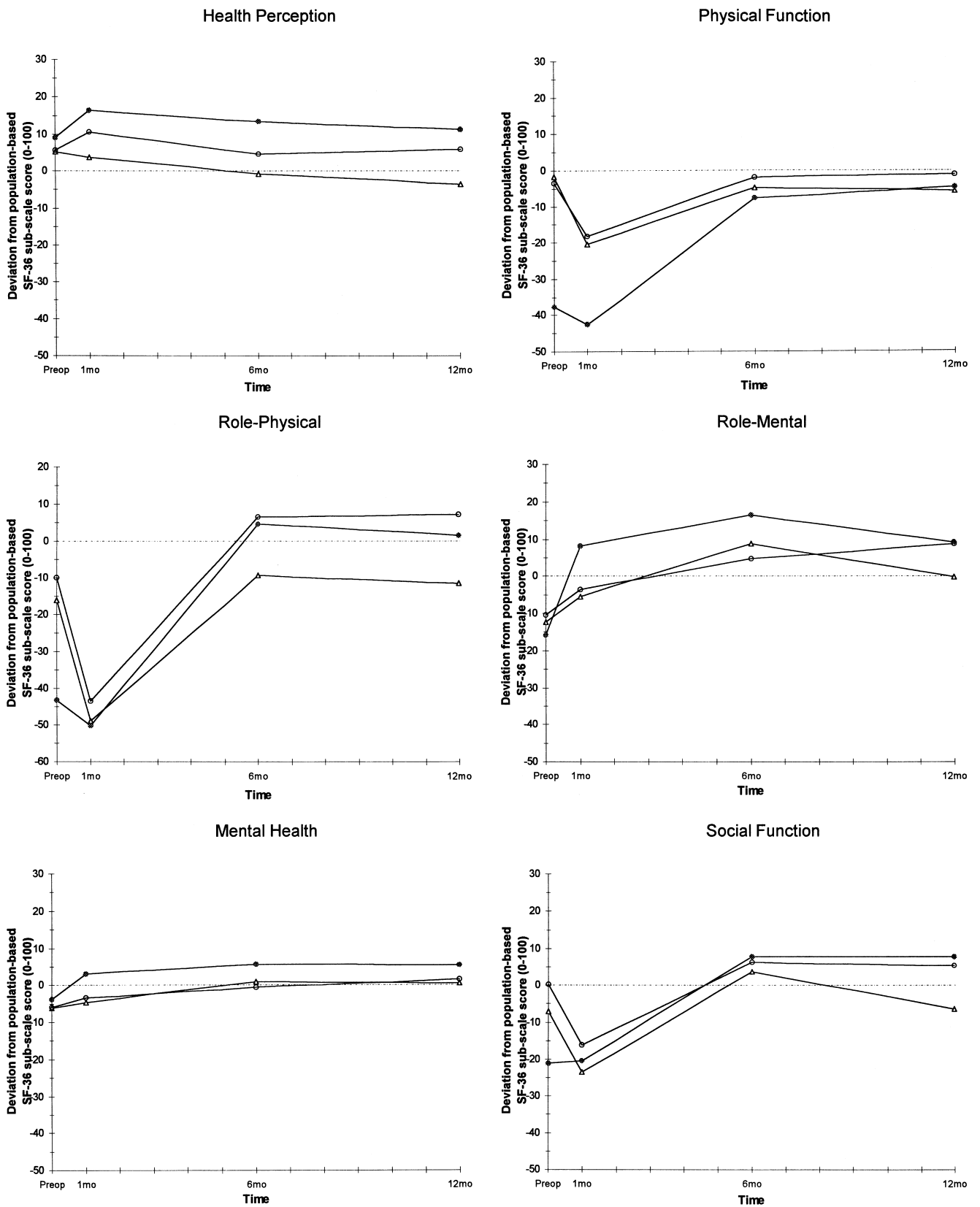
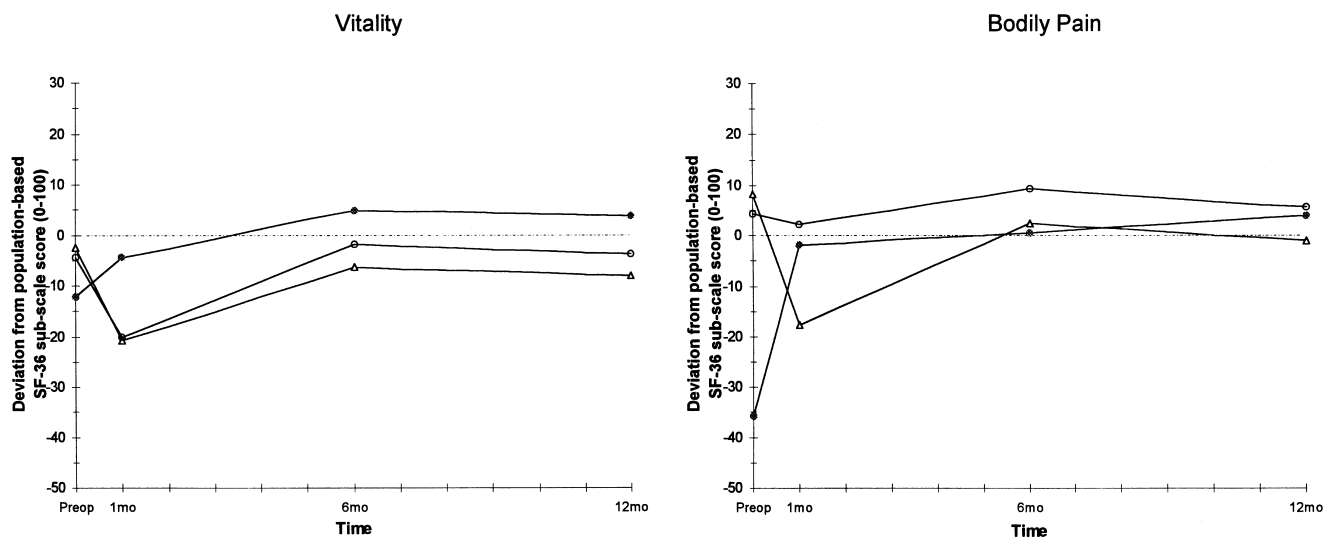


FIGURE 2. Continued.





**FIGURE 2.** (continued) Deviation from age- and gender-adjusted population-based SF-36 scores by surgical procedure. Triangles indicate thoracic surgery for lung cancer; filled circles, total hip arthroplasty; open circles, abdominal aortic aneurysm; dotted line, age and gender adjusted population based value.

correlated with changes in the SF-36 mental health score ( $r = .4$ ;  $p = .0001$ ). Statistically significant but lower correlations were observed for the remainder of the SF-36 subscales ( $r = .1$  to  $.3$ ). Similarly, changes in the two measures of physical functioning (the SF-36 physical function subscale and the SAS) were also moderately correlated ( $r = .3$ ;  $p = .0001$ ). As would be expected, change on the SAS did not correlate with SF-36 subscales designed to measure mental health.

A second test of validity compared the mean change in SF-36 scores for persons who categorized themselves as improved, the same, or worse after surgery. These analyses demonstrated that the mean improvements in SF-36 subscale scores were greatest for those subjects who reported improvement in their health at 6 months after surgery (Table 5). Declines in dimension-specific health were observed for all five subscales for those who reported that these dimensions were worse at 6 months after surgery; however, for four of the five subscales, the declines were not statistically significant (Table 5). For those subjects who reported no change in HRQL at 6 months after surgery, on average, SF-36 scores improved significantly for two of the five dimensions tested, but the magnitude of these SF-36 improvements was significantly lower when compared with subjects who reported HRQL improvements.

## DISCUSSION

In this report, we demonstrate that the SF-36 was responsive in the expected direction over time, as judged by what is known clinically about recovery after the three surgeries studied; that no one subscale or construct was responsive for all three procedures at all of the points in time after surgery; and that the observed changes in the

SF-36 tended to have a biphasic pattern after these elective surgeries.

Our strongest evidence for the validity of the SF-36 as an outcome tool after major elective surgery lies in the relation between the expected clinical response for each of these surgeries and the observed longitudinal changes in SF-36 scores. These data are consistent with and extend the findings of other investigators who have found the SF-36 to be responsive after knee and hip arthroplastic surgeries,<sup>31-35</sup> and surgery for primary hyperparathyroidism,<sup>38</sup> and run contrary to the minimal change in SF-36 scores observed after transurethral prostatectomy,<sup>37</sup> and cataract surgery.<sup>51</sup>

Our findings with regard to the relative responsiveness of each of the SF-36 subscales demonstrate that conclusions about the most sensitive subscale are greatly influenced by the type of procedure and the timing of postoperative follow-up. These data support the importance of multidimensional assessments of health after elective surgery and indicate that the specific timing of the postoperative evaluation with the SF-36 may influence the interpretation of benefit derived from a procedure. It is noteworthy that preoperative ceiling effects, such as those observed for the two role-functioning and social-functioning subscales in this cohort, greatly influence responsiveness.

A biphasic pattern of decline in HRQL followed by significant long-term gains was observed for the two procedures for which it would be expected, after THA and after thoracic surgery for patients with lung cancer. The observed changes were most dramatic for patients who underwent THA, a procedure specifically designed to improve physical capacity and to decrease pain. The average decline in social function at 1 month after surgery for the

**Table 5. Relation of Changes in Dimension-Specific 36-Item Short Form Health Survey Scores and Dimension-Specific Health Transition Questions at 6 Months After Surgery**

SF-36 Subscale	Mean Change (SE) in SF-36 Stratified by Response to Dimension-Specific Transition Questions*		
	Better†	No Change	Worse
Health perception (n = 345)	1.7‡ (1.4)	1.3 (1.8)	-15.2 (2.9)
Bodily pain (n = 260)	17.9 (2.7)	4.9 (3.8)	-1.4 (3.8)
Physical function (n = 337)	23.1 (2.7)	16.1 (3.3)	-0.6 (3.2)
Social function (n = 124)	21.1 (9.0)	5.5 (4.7)	-3.7 (4.8)
Mental health (n = 353)	10.9 (2.2)	7.7 (1.2)	-0.8 (3.0)

\*SF-36 scores range 0–100, 0 = worst health, 100 = best health. See Appendix A for the dimension-specific questions. Participants with preoperative scores at the ceiling (score = 100) and floor (score = 0) for each subscale were omitted prior to this analysis.

†Between-group differences in SF-36 scores for those who reported better or unchanged versus worse dimension-specific health were significantly higher ( $p < .05$ ) for health perception, bodily pain, physical function, social function, and mental health.

‡All within-group changes in SF-36 scores were significant ( $p < .05$ ) except for those subjects who reported the same or better health perception, worse or the same bodily pain or social functioning, and worse physical functioning or mental health.

cohort overall was equivalent to a change from being limited in one's ability to visit with friends some of the time to being limited most of the time. These declines in health status encountered 1 month after surgery are unlikely to be captured reliably in routinely collected administrative or medical record data and have implications for the provision of in-home services as economic pressures lead to shorter hospital lengths of stay.

Some dimension-specific changes in health did not conform to a biphasic pattern. For example, patients who had AAA repairs reported significant improvement in health perception despite worsening across many dimensions of health at 1 month after surgery. This positive change early during the postoperative recovery may reflect their relief at surviving a major life-threatening condition. Conversely, for those patients who had thoracic surgery, persistent declines in health perception, physical function, and vitality were observed at 12 months after surgery. The declines in HRQL for this group may have been influenced by the poor or uncertain long-term prognosis for many of these patients.

The validity of the SF-36 as a measure of outcome after elective surgery is supported by many of our results. As mentioned above, the strongest evidence for validity is the direction and magnitude of dimension-specific changes observed in each surgical group relative to what would be expected clinically. Moderate correlational validity is sup-

ported by the associations between the SF-36 subscales and the other indicators of physical function and global health perception that were measured in the cohort. Further evidence for construct validity of the SF-36 is provided by the positive association between the responses to the health transition questions, an independent patient impression of change in HRQL, and the observed change in SF-36 scores over time. However, it is important to note that participants who reported no change in dimension-specific health status over time actually had positive changes in their SF-36 scores. It is possible that at 6 months after surgery subjects answering the health transition questions were biased to recall preoperative functioning as better than it actually was at the time of the preoperative interview.

A strength of this study is that we have comparative data for three common surgical procedures collected simultaneously. Data are available that describe changes in health status after elective surgery.<sup>30–40,44,47–50</sup> But, for the most part, these investigations have been limited to a small sample size and one procedure. In the aggregate, changes in SF-36 scores may prove to be useful indicators of quality of care. However, before comparisons can be made across different health care settings, reliable and valid methods for case-mix adjustment are needed.<sup>52</sup>

Our study is limited by a problem that is generalizable to all research evaluating health status outcomes—the lack of a single reference standard for HRQL against which the SF-36 and other instruments can be measured. The sensitivity of instruments for detecting small changes in health status therefore remains uncertain, but the patterns of change and the consistency of results with the different instruments suggest that the differences detected with the SF-36 are valid. Other limitations are the uncertain generalizability of data from a study conducted in a university-based tertiary care population and the 1-year follow-up time. We assessed preoperative health status after the patient had decided to undergo surgery, but before the surgery itself. Therefore the timing of data collection may have biased the preoperative scores and the subsequent differences that were observed.

Future research should examine the outcomes of specific operations in different populations, define the duration of impaired functional status after elective surgery, and when possible, compare patients undergoing surgery with those randomized to alternative treatments. Policy makers who set priorities for allocating health care resources should use standard approaches to compare the outcomes of different medical interventions.

## REFERENCES

1. Klevit HD, Bates AC, Castanares T, Kirk EP, Sipes-Metzler PR, Wopat R. Prioritization of health care services: a progress report by the Oregon Health Services Commission. *Arch Intern Med.* 1991;151:912–6.
2. Hadorn DC. Setting health care priorities in Oregon: cost-effective-

- ness meets the rule of rescue. *JAMA*. 1991;265:2218-25.
3. Daniels N. Is the Oregon plan fair? *JAMA*. 1991;265:2232-5.
  4. Lansky D, Butler JBV, Waller FT. Using health status measures in the hospital setting: from acute care to 'outcomes management.' *Med Care*. 1992;30:MS57-73.
  5. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the Index of Well-Being. *Health Serv Res*. 1976;11:478-507.
  6. Bergner M, Bobbit RA, Carter WB, et al. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care*. 1981;19:787-805.
  7. Jette AM, Davies AR, Cleary PD, et al. The functional status questionnaire: reliability and validity when used in primary care. *J Gen Intern Med*. 1986;1:143-9.
  8. Hunt SM, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. *J R Coll Gen Pract*. 1985;35:857-9.
  9. Stewart AL, Hays RD, Ware JE. The MOS Short-form General Health Survey: reliability and validity in a patient population. *Med Care*. 1988;26:724-35.
  10. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36), I: conceptual framework and item selection. *Med Care*. 1992;30:473-83.
  11. 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;31:247-63.
  12. McHorney CA, Ware JE Jr, Lu JFR, et al. The MOS 36-Item Short-Form Health Survey (SF-36), III: tests of data quality, scaling assumptions, and validity among diverse patient groups. *Med Care*. 1994;32(1):40-66.
  13. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Econ*. 1993;2:217-27.
  14. Ware JE, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raezle A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcome Study. *Med Care*. 1995;33:AS264-79.
  15. Nerenz DR, Repansky DP, Whitehouse MD, et al. Ongoing assessment of health status in patients with diabetes mellitus using the SF-36 and diabetes TYPE Scale. *Med Care*. 1992;30:MS112-24.
  16. Kurtin PS, Davies AR, Meyer KB, et al. Patient-based health status measurements in outpatient dialysis: early experiences in developing an outcomes assessment program. *Med Care*. 1992;30:MS136-49.
  17. Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey: Manual and Interpretation Guide. Boston, Mass: Nimrod Press; 1993.
  18. Stewart AL, Greenfield S, Hays RD, et al. Functional status and well-being of patients with chronic medical conditions, results from the Medical Outcomes Study. *JAMA*. 1989;262:907-13.
  19. Brazier JE, Harper R, Jones NMB, et al. Validating the SF-36 Health Survey questionnaire: new outcome measure for primary care. *BMJ*. 1992;305:160-4.
  20. Fryback DG, Dasbach EJ, Klein R, et al. The Beaver Dam Health Outcomes Study: initial catalog of health state quality factors. *Med Decis Making*. 1993;13(2):89-102.
  21. Johnson PA, Goldman L, Orav EJ, et al. Comparison of the Medical Outcomes Study Short-Form 36-Item Health Survey in black patients and white patients with acute chest pain. *Med Care*. 1995;33:145-60.
  22. Weinberger M, Kirkman MS, Samsa GP, et al. The relationship between glycemic control and health-related quality of life in patients with non-insulin-dependent diabetes mellitus. *Med Care*. 1994;32:1173-81.
  23. Wu AW, Rubin HR, Mathews WC, et al. A health status questionnaire using 30 items from the Medical Outcomes Study: preliminary validation in persons with early HIV infection. *Med Care*. 1991;29:786-98.
  24. Hays RD, Shapiro MF. An overview of generic health-related quality of life measures for HIV research. *Qual Life Res*. 1992;1:91-7.
  25. Bousquet J, Knani J, Dhivert H, et al. Quality of life in asthma, I: internal consistency and validity of the SF-36 questionnaire. *Am J Respir Crit Care Med*. 1994;149:371-5.
  26. Krumholz HM, McHorney CA, Clark L, et al. Changes in health after elective percutaneous coronary revascularization. *Med Care*. 1996;34:754-9.
  27. Fifer S, Mathias SD, Patrick DL, et al. Untreated anxiety among adult primary care patients in a health maintenance organization. *Arch Gen Psychiatry*. 1994;51:740-50.
  28. Hays RD, Wells KB, Sherbourne CB, et al. Functioning and well-being outcomes of patients with depression compared to chronic medical illness. *Arch Gen Psychiatry*. 1995;52:11-9.
  29. Wells KB, Burnam MA, Rogers W, et al. The course of depression in adult outpatients: results from the Medical Outcomes Study. *Arch Gen Psychiatry*. 1992;49:788-94.
  30. Kantz ME, Harris WJ, Levitsky K, et al. Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. *Med Care*. 1992;30:MS240-52.
  31. Ritter MA, Albohm MJ, Keating EM, et al. Comparative outcomes of total joint arthroplasty. *J Arthroplasty*. 1995;10:737-41.
  32. Dawson J, Fitzpatrick R, Murray D, Carr A. Comparison of measures to assess outcomes in total hip replacement surgery. *Qual Health Care*. 1996;5:81-8.
  33. Anderson JG, Wixson RL, Tsai D, et al. Functional outcomes and patient satisfaction in total knee patients over the age of 75. *J Arthroplasty*. 1996;11:831-40.
  34. O'Boyle CA, McGess H, Hickey A, O'Malley K, Joyce CRB. Individual quality of life in patients undergoing hip replacement. *Lancet*. 1992;339:1088-91.
  35. Cleary PD, Reilly DT, Greenfield S, et al. Using patient reports to assess health-related quality of life after total hip replacement. *Qual Life Res*. 1993;2:3-11.
  36. Phillips RC, Lansky DJ. Outcomes management in heart valve replacement surgery: early experience. *J Heart Valve Dis*. 1992;1:42-50.
  37. Keoghane SR, Lawrence KC, Jenkinson CP, et al. The Oxford laser prostate trial: sensitivity to change of three measures of outcome. *Urology*. 1996;47:43-7.
  38. Burney RE, Jones KR, Coon JW, et al. Assessment of patient outcomes after operation for primary hyperparathyroidism. *Surgery*. 1996;120:1013-9.
  39. Currie IC, Wilson YG, Baird RN, Larmont PM. Treatment of intermittent claudication: the impact on quality of life. *Eur J Vasc Endovasc Surg*. 1995;10:356-61.
  40. Goldman L, Hashimoto B, Cook EF, Loscalzo A. Comparative reproducibility and validity of systems for assessing cardiovascular functional class: advantages of a new Specific Activity Scale. *Circulation*. 1981;64:1227-34.
  41. Lee TH, Shammash JB, Ribeiro JP, Hartley LH, Sherwood J, Goldman L. Estimation of maximum oxygen uptake from clinical data: performance of the Specific Activity Scale. *Am Heart J*. 1988;115:203-4.
  42. Bindman AB, Keane D, Lurie N. Measuring health changes among severely ill patients: the floor phenomenon. *Med Care*. 1990;28:1142-52.
  43. Dales RE, Belanger R, Shamji FM, et al. Quality-of-life following thoracotomy for lung cancer. *J Clin Epidemiol*. 1994;47:1443-9.
  44. Greenfield S, Apolone G, McNeil BJ, Cleary PD. The importance of co-existent disease in the occurrence of postoperative complications and one-year recovery in patients undergoing total hip replacement. Comorbidity and outcomes after hip replacement. *Med Care*. 1993;31:141-54.
  45. Cleary PD, Greenfield S, Mulley AG, et al. Variations in length of stay and outcomes for six medical and surgical conditions in Massachusetts and California. *JAMA*. 1991;266:73-9.
  46. Greenfield S, Sullivan L, Dukes KA, Silliman R, D'Agostino R, Kaplan SH. Development and testing of a new measure for case mix for use in office practice. *Med Care*. 1995;33(suppl):AS47-55.

47. Rosner B. *Fundamentals of Biostatistics*. Boston, Mass: Duxbury Press; 1986:286-7.
48. Liang MH, Larson MG, Cullen KE, Schwartz JA. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum*. 1985;28:542-7.
49. McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires*. New York, NY: Oxford University Press; 1987:14.
50. Nunally JC. *Psychometric Theory*. New York, NY: McGraw Hill; 1978:190-255.
51. Mangione CM, Phillips RS, Lawrence MG, et al. Improved visual function and attenuation of age-related declines in health-related quality of life after cataract extraction. *Arch Ophthalmol*. 1994; 112:1419-25.
52. Cleary PD, Greenfield S, McNeil BJ. Assessing quality of life after surgery. *Control Clin Trials*. 1991;12:189-203S.

## APPENDIX A

### *Transition Questions*

---

#### **Health Perception**

In the past 6 months, has there been a change in your overall health and sense of well-being? (check one)

- No
- Yes, I am a lot worse
- Yes, I am a little worse
- Yes, I am a little better
- Yes, I am a lot better

#### **Mental Health**

In the past 6 months, has there been a change in your emotional or mental health? (check one)

- No
- Yes, I am a lot worse
- Yes, I am a little worse
- Yes, I am a little better
- Yes, I am a lot better

#### **Social Functioning**

In the past 6 months, has there been a change in the amount or level of your social activity? (check one)

- No
- Yes, I am a lot less active
- Yes, I am a little less active
- Yes, I am a little more active
- Yes, I am a lot more active

#### **Physical Functioning**

In the past 6 months, has there been a change in the type of physical activity you can do or the amount you can do them? (check one)

- No
- Yes, I am a lot more limited
- Yes, I am a little more limited
- Yes, I am a little less limited
- Yes, I am a lot less limited

#### **Pain**

In the past 6 months, has there been a change in how much bodily pain you have? (check one)

- No
  - Yes, a lot more pain
  - Yes, a little more pain
  - Yes, a little less pain
  - Yes, a lot less pain
-