THE ORTHOPAEDIC FORUM

Scoring the SF-36 in Orthopaedics: A Brief Guide

Nicholas C. Laucis, BSE, Ron D. Hays, PhD, and Timothy Bhattacharyya, MD

Abstract: The Short Form-36 (SF-36) is the most widely used health-related quality-of-life measure in research to date. There are currently two sources for the SF-36 and scoring instructions: licensing them from Optum, Inc., or obtaining them from publicly available documentation from the RAND Corporation. The SF-36 yields eight scale scores and two summary scores. The physical component summary (PCS) and mental component summary (MCS) scores were derived using an orthogonal-factor analytic model that forced the PCS and MCS to be uncorrelated, and it has been shown to contribute to an inflation of the MCS in patients with substantial physical disability. Oblique scoring can reduce this inflation of the MCS in orthopaedic studies. Spreadsheets to score the SF-36, along with a copy of the questionnaire, are provided.

There has been a shift in medicine from relying on clinical biomarkers to including patient-reported outcome measures. Garratt et al. reported in 2002 that over 3900 publications existed describing the development and evaluation of patient-reported measures[1]. With so many tools available, how do researchers select the right instrument? And how do readers interpret the findings?

A very common choice is the Short Form-36 (SF-36) because of its history, brevity, reliability, and validity. According to Turner-Bowker et al., from 1988 to 2000 the SF-36 was used in more than 4000 publications[2], it has been translated into over 170 languages, and a literature search of the term "SF-36" results in more than 13,000 publications. This article will take a brief look at the history, licensing, and scoring methods of the SF-36 and its various versions, and we provide spreadsheets for researchers to use to score their own results obtained from the SF-36 and similar forms.

Origin of the SF-36

The SF-36 was a product of the Medical Outcomes Study (MOS), a four-year study that examined specific influences on outcomes of care[3,4]. From the original 149-item Functional and Well Being Profile used in the MOS, a twenty-item short survey called the SF-20 was derived, but it was limited by floor effects[5-7]. Ware and Sherbourne, working with the RAND Corporation, published the SF-36 in 1992, selecting eight health domains from the forty identified in the MOS[8,9]. This "SF-36," herein referred to as the Ware-36 (or SF-36v1), was distributed along with its scoring rules by the Medical Outcomes Trust, Inc., with strict adherence to the

item wording and scoring in order to use the SF-36 trademark[6]. The original SF-36 underwent many changes and separate derivatives were created (Fig. 1). In 1993, Hays, Sherbourne, and Mazel released a publicly available version of the SF-36, the RAND-36[6]. In contrast to the Ware-36, the RAND-36 and its scoring instructions are publicly available on the RAND Corporation web site[10]. The Ware-36 and RAND-36 have thirty-six identical items and thus both have been referred to as the "SF-36." However, the two are scored somewhat differently, as discussed in the next section. Other derivatives of the SF-36 are the SF-12, SF-8, VR-36 (Veterans RAND-36), and VR-12, discussed later.

Licensing

QualityMetric, now a part of Optum, Inc., currently holds the copyrights and licensing for the newer SF-36v2, SF-12v2, and SF-8 instruments, all designed to produce scores comparable with previous SF-36 scores[11]. The RAND Corporation maintains their publicly available version of the SF-36 (see Appendix E-1, the RAND-36 survey) on their web site, along with scoring instructions[10]. The RAND-36 Health Status Inventory, which uses scoring based on item response theory (IRT), is also available[12].

The Eight SF-36 Scale Scores

Ware and Sherbourne selected eight concepts of health, out of the forty identified in the MOS, to be scored in the SF-36. These eight
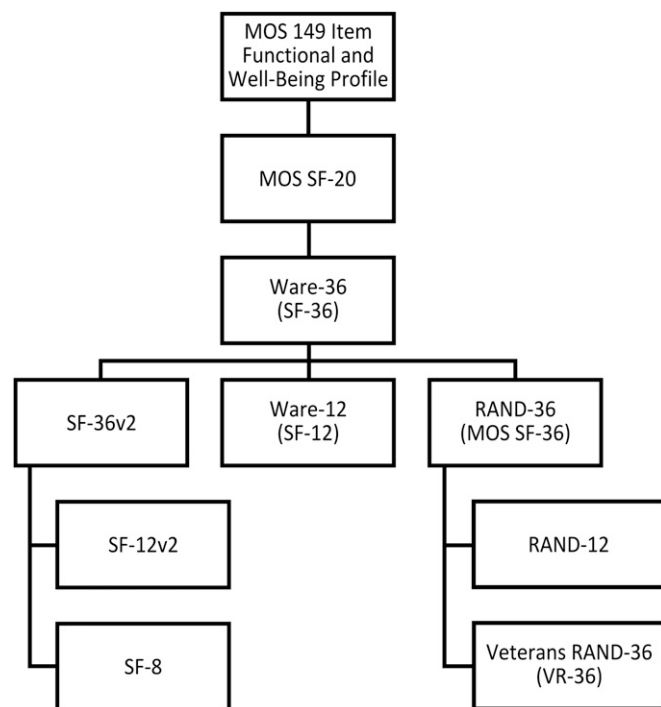

Fig. 1

Derivation hierarchy of the SF-36 (Short Form-36). The SF-36 was derived from the original MOS 149-item profile. Many derivations of the first SF-36 (Ware-36) have been released, including shortened versions consisting of twelve and eight items, updated survey layouts (SF-36v2 and VR-36), and a publicly available version, the RAND-36.

scales are physical functioning (PF), bodily pain (BP), role limitations due to physical health problems (RP), role limitations due to personal or emotional problems (RE), general mental health (MH), social functioning (SF), energy/fatigue or vitality (VIT), and general health perceptions (GH). Emotional well-being and vitality have been used interchangeably with general mental health and energy/fatigue, respectively. The items contributing to a scale are scored so that a higher score represents better health, and they are averaged together to create the scale score. Scale scoring instructions can be found on the RAND Corporation web site, and the appendix contains Microsoft Excel spreadsheets for scoring the SF-36 (see Appendix E-2) and SF-12 (see Appendix E-3) according to the RAND instructions[10].

Other Derivations of the SF-36: The SF-12, SF-8, SF-6D, VR-36, and VR-12

The SF-12 is a twelve-item subset of the Ware-36[13]. Despite the 67% reduction in the number of items, the SF-12 yields accurate estimates of the SF-36 summary scores (discussed below); however, because of the smaller number of items, the reliability and validity for the eight individual scales are less with the SF-12 than the SF-36. Therefore, the SF-36 rather than the SF-12 should be used for studies with smaller numbers of patients[14]. The SF-8 has only one question per scale and was developed for use in very large-scale and population-based health studies[11]. The SF-6D yields a preference-based score based on a subset of the SF-36 and SF-12 items that enables economic evaluation and calculation of quality-adjusted life years[15,16].

The Veterans RAND-36 or VR-36 (formerly the Veterans SF-36) was developed in 1999, with a twelve-item version (VR-12) developed later[17,18]. The VR-36 implemented 5-point response choices for the scales "role limitation due to physical function" and "role limitation due to emotional function," rather than the two choice options (yes or no) in the RAND-36. This change results in a reduction of floor and ceiling effects of these scales. The VR-36 and VR-12 also contain two additional items that assess how the patient's mental and physical health has changed over time[19].

Differences Between the Ware-36 (SF-36v1), RAND-36, and SF-36v2

The differences in scoring between the eight Ware-36 and RAND-36 scales are subtle and have been argued to not contribute to an important difference in scores in the MOS longitudinal study[6]. Both forms use the same thirty-six items and answer choices and yield identical results in six of eight scales, only differing in the BP (bodily pain) and GH (general health perceptions) subscales. For the BP subscale, the Ware-36 scores the second pain item (interference with normal work) conditionally depending on the response to the first pain item, with unequal distances between response categories, inflating the positive correlation between the two pain items. The RAND-36 scores all items assuming the distances between categories are equal, yielding a slightly higher score (meaning less pain). A complete table showing the differences in possible scores can be found in Hays et al.[6]. For the GH item "In general, would you say your health is excellent, very good,

good, fair, or poor?" the Ware-36 uses the scaled scores of 100, 85, 60, 25, and 0 for the answer options, whereas the RAND-36 scores it like other items, assuming equal intervals with scores of 100, 75, 50, 25, and 0. Despite these distinctions, Hays et al. demonstrated that the SF-36 and RAND-36 pain and general health scores achieved a correlation coefficient of 0.99 in the MOS panel sample at baseline[6]. The SF-36v2 scores have been shown to be comparable with SF-36v1 scores[9]. The SF-36v2 is a newer version of the Ware-36 created in 1996, featuring an improved item layout[20], simpler instructions, replacement of the six-level response choices for nine items in the emotional well-being and energy/fatigue scales with five-level choices, and incorporation of the increase from two to five answer choices for role evaluation items from the VR-36[9,17]. However, the wording change for the energy/fatigue item "Full of Pep" to "Full of Life" has been argued to not be equivalent[21].

The Physical Component Summary (PCS) and Mental Component Summary (MCS) Scores

The PCS and MCS are aggregates of the eight scale scores that account for >80% of the reliable variance in the scales (variance not due to measurement error). Per the original instructions by Ware and Kosinski, the summary scores should always be compared with the eight scale scores before drawing conclusions[22]. Advantages of using the physical and mental component summary measures rather than the eight scale scores include smaller confidence intervals and smaller floor and ceiling effects; a reduction of the number of analyses required from eight to two also avoids some reduction in statistical power resulting from multiple testing[23,24]. However, the method used to derive the MCS and PCS scoring coefficients has been criticized; in particular, the use of an orthogonal model yields negative coefficients for some of the scales[25].

The calculation process for the PCS and MCS has been previously described by Taft et al.[25]. After the eight scale scores are calculated, a z-score is determined for each by subtracting the scale mean of a sample of the U.S. general population from an individual's scale score and then dividing by the standard deviation from the U.S. general population. Each of the eight z-scores is then multiplied by the corresponding factor scoring coefficient for the scale[25]. There are two different sets of factor scoring coefficients, one for the PCS and another for the MCS. The products of the z-scores and factor scoring coefficients for the PCS are then summed together, and a similar calculation is performed for the MCS. Each resulting sum is multiplied by 10 and added to 50 to linearly transform the PCS or MCS to the T-score metric, which has a mean of 50 and a standard deviation of 10 for the U.S. general population.

As noted above, the factor scoring coefficients for the PCS and MCS were derived on the basis of a model that assumed two underlying orthogonal (uncorrelated) factors, physical and mental health. Because of this assumption, three SF-36 scales have negative weights in scoring the PCS: SF (social functioning), RE (role limitations due to emotional problems), and MH (mental health). Thus, having lower scores on these mental scales has the effect of increasing the PCS. Conversely, three SF-36 scales have negative weights in scoring the MCS: PF (physical functioning), RP (role limitations due to physical health problems), and BP

(bodily pain). Thus, scoring low on these physical scales raises the MCS.

The PCS and MCS scoring has relevance to orthopaedics as many orthopaedic patients score very low on the physical health scales on the SF-36 because of their musculoskeletal conditions, leading to a higher MCS than that implied by the SF-36 scale scores. This issue was documented by Nortvedt et al. in a study of individuals with a substantial physical disability, multiple sclerosis (MS). That study examined the scores of the patients with MS on the Ware-36, the SF-12, the RAND-36 (see below), and two MS-targeted instruments, the Expanded Disability Status Scale and the Incapacity Status Scale. The RAND-36 summary scores were based on an oblique (correlated) factor solution and were more strongly correlated with scores on the MS-targeted instruments than the Ware-36 PCS and MCS were. The Ware-36 and Ware-12 MCS scores underestimated the mental health issues of the MS patients by 7 and 5 points, respectively, compared with the RAND-36[26]. These differences exceed the recommended 3-point minimal important difference (MID) for the SF-36 PCS and MCS[27].

The RAND-36 Health Status Inventory (HSI) represents an alternative to the PCS and MCS scoring system. Because of the inconsistencies between the orthogonal model and observed data and the potential for anomalous results, the Psychological Corporation decided to derive an alternate scoring method[12]. In particular, only four scales (PF [physical functioning], RP [role limitations due to physical health], BP [bodily pain], and GH [general health]) are used to create the Physical Health Composite (PHC), and the other four scales (MH [mental health], RE [role limitations due to emotional problems], SF [social functioning], and VIT [vitality]) are used to create the Mental Health Composite (MHC)[12]. The RAND-36 HSI did not include overlapping scales to calculate the PHC and MHC because their factor loadings did not warrant their inclusion. In addition, a third summary score called the Global Health Composite was developed to represent the overlapping aspects of physical and mental health[12]. Another option was provided by Farivar et al. by mimicking the approach used in creating the PCS and MCS but allowing the underlying factors to be correlated (an oblique solution)[28]. Scoring coefficients based on the oblique-factor analytic model minimize or prevent the inflation of the MCS scores by poor physical health scores that commonly occur with orthopaedic patients. However, the orthogonal method of calculation is still the most commonly used. A spreadsheet in the appendix calculates summary scores for all three methods (orthogonal, oblique, and RAND HSI) from the eight SF-36 scale scores (see Appendix E-4).

There are several recent studies that illustrate the limitations of using the Ware-36 orthogonal PCS and MCS in orthopaedic populations. Gartsman et al. noted, in their study of shoulder conditions, that many of the physical functioning items focus on the lower extremities and the SF-36 was not a good indicator of physical functioning of the upper extremities, and this limitation will carry over to the PCS[29]. Agren et al. divided patients who had sustained displaced intra-articular calcaneal fractures into superior and inferior outcome groups on the basis of a visual analog scale score for pain and function[30]. The authors noted that the group with superior outcomes had a significantly

TABLE I Results from Recalculating the Data from Agren et al.[30]*										
	Orthogonal[25]		Oblique[28]		Difference to Orthogonal		RAND-36[12]		Difference to Orthogonal	
	PCS	MCS	PCS	MCS	ΔPCS	ΔMCS	PHC	MHC	ΔPCS	ΔMCS
Superior group	52.6	53.3	53.4	53.8	0.8	0.5	53.5	53.8	0.9	0.5
Inferior group	38.0	48.2	39.4	44.1	1.4	−4.1	38.6	43.8	0.6	−4.4

*Using the orthogonal, oblique, and RAND-36 HSI methods. The orthogonal values differ slightly from those reported in the original study because the normative values used were marginally different. Calculations were performed on scale means for the superior and inferior groups. The oblique and RAND-36 HSI methods decreased the inferior group's MCS scores by 4.1 and 4.4 points, respectively. The inferior group's PCS and superior group's PCS and MCS were minimally affected.

better mean SF-36 PCS score (52.2 compared with 37.7) but did not have a significantly better SF-36 MCS score. The authors speculated that this was a result of a discrepancy among the patients with respect to the ability to mentally cope with the physical handicap or could reflect different levels of disappointment or depression experienced by the patients; however, the discrepancy may simply be due to the orthogonal scoring method. Recalculating the PCS and MCS of the groups using the three scoring methods (Table I), the difference between the inferior and superior groups on the MCS (oblique estimation) and MHC scores is one standard deviation (44.1 compared with 53.8 and 43.8 compared with 53.8). Looking at the difference in the z-scores of the eight scales of the inferior and superior groups, mental health (−0.32 compared with 0.34), emotional role functioning (−0.57 compared with 0.32), vitality (−0.59 compared with 0.38), and social functioning (−0.63 compared with 0.20) all differed by 0.7 to 1 standard deviation between the two groups and this is reflected in the oblique MCS and MHC. A lower score on the SF-36 physical health scales in the inferior group elevates their MCS scores when the orthogonal method is used. Analyzing the differences in the eight scales of the SF-36 individually would help to elucidate the real difference in mental health between the two groups of patients and avoid possibly false conclusions based on the "standard" scoring of the MCS.

Another example of the orthogonal method leading to inaccurate conclusions about orthopaedic interventions involves a study by Hsu et al. investigating the X STOP interspinous device for treatment of lumbar spinous stenosis[31]. Hsu et al. concluded that there was little change in mental health between the preoperative baseline and the one-year follow-up in patients who received the X STOP device for lumbar spinal stenosis. The z-score of the mental health scale shows little difference between baseline and the one-year follow-up (0.00 compared with 0.26); however, the role emotional (−0.89 compared with −0.09), social functioning (−1.11 compared with 0.00), and vitality (−0.76 compared with −0.03) scale z-scores all showed substantial improvement. The difference in the orthogonal MCS between the baseline and one-year follow-up does not reflect the improvement seen in those subscales (51.3 compared with 54.6, a change of 3.3). However, recalculations of the MCS using the oblique method (43.4 compared with 50.4, a change of 7.0) and the RAND-36 HSI method (42.7 compared with 50.8, a change of 8.1) both more accurately reflect the improvements seen in the subscales (Table II).

Bhandari et al. investigated the effect of orthopaedic trauma on health-related quality of life[32]. They assessed orthopaedic trauma patients with the SF-36 and the Symptom Checklist-90-Revised (SCL-90-R), a measure of a patient's current psychological symptom status. The authors concluded that the mean MCS value of orthopaedic trauma patients was similar to U.S. norms (45.6 compared with 50.0); however, one in five of the patients met criteria for a psychological illness and the patients indicated higher-than-normal levels of psychological distress in all primary

TABLE II Results from Recalculating the Data from Hsu et al.[31]*										
	Orthogonal[25]		Oblique[28]		Difference to Orthogonal		RAND-36[12]		Difference to Orthogonal	
Condition	PCS	MCS	PCS	MCS	ΔPCS	ΔMCS	PHC	MHC	ΔPCS	ΔMCS
Baseline	27.6	51.3	31.5	43.4	3.9	−7.9	30.5	42.7	2.9	−8.6
1-yr follow-up	40.6	54.6	44.0	50.4	3.4	−4.2	42.9	50.8	2.3	−3.8

*Using the orthogonal, oblique, and RAND-36 HSI methods. The study involved use of the X STOP device for lumbar spinal stenosis. The orthogonal values differ slightly from those reported in the original study because the normative values used were marginally different. Calculations were performed on the scale means. Oblique and RAND-36 scoring removed the inflation of the MCS that occurs with orthogonal scoring in patients with severe physical disability while minimally affecting the relative change in PCS from baseline to one year.

SCORING THE SF-36 IN ORTHOPAEDICS: A BRIEF GUIDE

TABLE III PCS and MCS Scoring Results Reported for Various Orthopaedic Procedures and Conditions with Oblique and RAND Recalculations*

Condition	Orthogonal[25]		Oblique[28]		Difference to Orthogonal		RAND-36[12]		Difference to Orthogonal	
	PCS	MCS	PCS	MCS	ΔPCS	ΔMCS	PHC	MHC	ΔPCS	ΔMCS
Anterior glenohumeral instability[29]	38.1	48.0	38.4	44.6	0.3	−3.4	38.1	43.8	0.0	−4.2
Complete reparable rotator cuff tear[29]	34.8	47.3	35.7	42.8	0.9	−4.5	35.2	41.7	0.4	−5.6
Adhesive capsulitis[29]	37.4	50.9	39.2	46.1	1.8	−4.8	38.4	46.3	1.0	−4.6
Glenohumeral osteoarthritis[29]	36.4	52.2	39.4	47.1	3.0	−5.1	38.0	47.1	1.6	−5.1
Shoulder impingement[29]	36.6	49.2	37.9	45.0	1.3	−4.2	37.2	44.6	0.6	−4.6
Orthopaedic trauma[32]	33.5	45.7	34.3	41.1	0.8	−4.6	34.1	39.2	0.6	−6.5
Displaced acetabular fractures, 6-mo follow-up[34] (median)†	28.5	57.2	34.3	47.2	5.8	−10.0	32.7	46.6	4.2	−10.6
Displaced acetabular fractures, 24-mo follow-up[34] (median)‡	39.5	54.6	43.0	48.9	3.5	−5.7	42.0	49.0	2.5	−5.6
Total hip arthroplasty, preop.[35]†	31.6	48.8	34.0	43.3	2.4	−5.5	32.4	43.1	0.8	−5.7
Total hip arthroplasty, 6-mo follow-up[35]‡	43.4	55.5	46.6	53.0	3.2	−2.5	44.9	54.0	1.5	−1.5
Total knee arthroplasty, preop.[35]†	31.3	50.0	34.3	43.9	3.0	−6.1	32.3	44.3	1.0	−5.7
Total knee arthroplasty, 6-mo follow-up[35]‡	43.3	53.6	45.6	51.2	2.3	−2.4	44.0	52.2	0.7	−1.4
Arthroscopic partial meniscectomy, preop.[35]†	40.1	54.3	43.2	50.3	3.1	−4.0	41.8	50.9	1.7	−3.4
Arthroscopic partial meniscectomy, 3-mo follow-up[35]‡	46.5	55.0	48.9	53.5	2.4	−1.5	48.0	54.1	1.5	−0.9
Anterior cruciate ligament reconstruction, preop.[35]†	36.5	51.0	38.8	46.9	2.3	−4.1	37.5	46.5	1.0	−4.5
Anterior cruciate ligament reconstruction, 2-yr follow-up[35]‡	50.9	56.7	53.4	56.2	2.5	−0.5	52.8	56.9	1.9	0.2
Orthopaedic surgical site infections, 1-yr follow-up[36]	35.3	47.3	37.3	42.1	2.0	−5.2	35.0	41.1	−0.3	−6.2
Isolated unstable ankle fracture, 4-mo follow-up[37]†	40.9	48.4	41.8	45.8	0.9	−2.6	41.2	45.1	0.3	−3.3
Isolated unstable ankle fracture, 20-mo follow-up[37]‡	44.2	54.2	47.0	51.3	2.8	−2.9	45.5	52.0	1.3	−2.2
Primary total hip arthroplasty, preop.[38]†	31.4	52.4	35.6	45.5	4.2	−6.9	34.3	44.6	2.9	−7.8
Primary total hip arthroplasty, 2-yr follow-up[38]‡	45.2	53.2	47.4	51.1	2.2	−2.1	46.4	51.3	1.2	−1.9
Revision total hip arthroplasty, preop.[38]†	32.7	52.9	36.8	46.6	4.1	−6.3	35.8	45.4	3.1	−7.5
Revision total hip arthroplasty, 2-yr follow-up[38]‡	40.3	54.2	43.9	50.0	3.6	−4.2	42.2	50.4	1.9	−3.8
Primary total knee arthroplasty, preop.[38]†	32.8	51.0	36.1	45.2	3.3	−5.8	34.6	44.9	1.8	−6.1
Primary total hip arthroplasty, 2-yr follow-up[38]‡	42.6	51.7	44.7	49.0	2.1	−2.7	43.7	49.0	1.1	−2.7
Lumbar laminectomy, preop.[38]†	32.3	44.1	32.7	39.5	0.4	−4.6	32.5	36.9	0.2	−7.2
Lumbar laminectomy, 2-yr follow-up[38]‡	46.2	53.0	48.2	51.2	2.0	−1.8	47.6	51.2	1.4	−1.8
Scoliosis, preop.[38]†	42.2	48.9	43.2	46.1	1.0	−2.8	42.5	45.7	0.3	−3.2
Scoliosis, 2-yr follow-up[38]‡	46.4	50.6	47.1	49.4	0.7	−1.2	46.9	49.7	0.5	−0.9

*PCS and MCS scoring results reported for various orthopaedic procedures and conditions with oblique and RAND recalculations. Recalculations were performed on the basis of reported scale means. Normative values were chosen to match the original study results as closely as possible. If the original study did not calculate summary scores, 1998 U.S. population normative values were used. Values are presented as the mean unless otherwise indicated. †Studies included in baseline or preop. measurements. ‡Studies included in recovery measurements.

TABLE IV Mean Differences in the Orthogonal, Oblique, and RAND-36 HSI Summary Scores Between Baseline or Preoperative and Follow-up SF-36 Results*				
	Oblique − Orthogonal		RAND-36 − Orthogonal	
	Mean ΔPCS	Mean ΔMCS	Mean ΔPCS	Mean ΔMCS
Baseline or preop. results	2.77	−5.34	1.57	−5.91
Follow-up results	2.48	−2.50	1.41	−2.05
Difference (95% CI)†	0.29 (−0.61, 1.19)	−2.84 (−3.77, −1.90)	0.16 (−0.55, 0.88)	−3.86 (−4.84, −2.89)

*Mean differences in the orthogonal, oblique, and RAND-36 HSI summary scores between baseline or preop. and follow-up SF-36 results. The studies tabulated in Table III were averaged to demonstrate that the oblique scoring method increased the PCS by 2.77 points for preop./baseline values or 2.48 points for follow-up values. The minimal clinically important difference for the SF-36 is 3 points[27]. †Paired-sample t test. CI = confidence interval.

dimensions on the SCL-90-R. Recalculations using the oblique and RAND-36 HSI methods for the MCS yield an MCS of 41.1 and an MHC of 39.2, half a standard deviation lower than the orthogonal MCS of 45.7 recalculated using Canadian population norms[33] (Table III). Physical disabilities are inflating the orthogonally calculated MCS, resulting in an interpretation that the MCS of the cohort is no different from the U.S. norm, whereas the psychologically specific SCL-90-R indicates that the cohort is experiencing higher-than-normal psychological distress.

We also recalculated the results of several other orthopaedic studies using the oblique and RAND-36 HSI methods (Table III). We separated baseline or preoperative SF-36 results from follow-up SF-36 results and calculated the differences among the three methods. The oblique and RAND-36 scoring methods increased the difference between the baseline and follow-up MCS results by a mean of 2.84 and 3.86, respectively, while the relative differences in the PCS were not significantly affected, increasing by 0.29 and 0.16 (Table IV).

Conclusions

The SF-36 and its versions are currently the most widely used health-related quality-of-life measures in the U.S. Two versions of the SF-36 are available today: the SF-36v2, which is licensed from Optum, Inc., and the RAND-36, which is publicly available from the RAND Corporation. These two versions have been shown to create comparable scale scores. A copy of the RAND-36, along with spreadsheets for scoring the SF-36 and SF-12 according to the RAND specifications, is included in the Appendix. The SF-36 forms have been used often in examining orthopaedic patient populations, commonly with reporting of orthogonal PCS and MCS measures. However, because of the extent of physical disability in orthopaedic patients, an orthogonally calculated MCS can artificially increase the MCS value, causing mental health

distress to be underrepresented in the summary score. Using an oblique method for calculating summary scores and comparing with the results of the eight scales can mitigate this effect.

Appendix

(eA) The RAND-36 survey (Appendix E-1), which was developed by RAND as part of the Medical Outcomes Study, and spreadsheets for scoring the SF-36 (Appendix E-2) and SF-12 (Appendix E-3) and for determining the summary scores according to each method (Appendix E-4) are available with the online version of this article as a data supplement at jbjs.org. ■

Nicholas C. Laucis, BSE
Timothy Bhattacharyya, MD
Clinical and Investigative Orthopedics Surgery Unit,
National Institute of Arthritis and Musculoskeletal and Skin Diseases,
National Institutes of Health,
10 Center Drive,
Building 10-CRC, Room 4-2339,
MSC1498, Bethesda, MD.
E-mail address for T. Bhattacharyya: bhattacharyyat@mail.nih.gov

Ron D. Hays, PhD
Division of General Internal Medicine & Health Services Research,
UCLA Department of Medicine,
911 Broxton Avenue,
Los Angeles, CA 90024

References

1. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ. 2002 Jun 15;324(7351):1417.

2. Turner-Bowker DM, Bartley BJ, Ware JE. SF-36® health survey and "SF" bibliography. 3rd ed. Lincoln, RI: QualityMetric Incorporated; 2002.

3. Tarlov AR, Ware JE Jr, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study. An application of methods for monitoring the results of medical care. JAMA. 1989 Aug 18;262(7):925-30.

4. Stewart AL, Sherbourne CD, Hays RD, Wells KB, Nelson EC, Kamberg CJ, Rogers WH, Berry SH, Ware JE Jr. Summary and discussion of MOS measures. In: Stewart

SCORING THE SF-36 IN ORTHOPAEDICS: A BRIEF GUIDE

AL, Ware JE Jr, editors. Measuring functioning and well-being: the medical outcomes study approach. Durham, NC: Duke University Press; 1992. p. 345-71.

5. Bindman AB, Keane D, Lurie N. Measuring health changes among severely ill patients. The floor phenomenon. Med Care. 1990 Dec;28(12):1142-52.

6. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. Health Econ. 1993 Oct;2(3):217-27.

7. Ware JE Jr, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol. 1998 Nov;51(11):903-12.

8. Ware JE Jr, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992 Jun;30(6):473-83.

9. Ware JE Jr. SF-36 Health Survey update. Spine (Phila Pa 1976). 2000 Dec 15;25 (24):3130-9.

10. 36-Item Short Form Survey from the RAND Medical Outcomes Study. RAND Health. http://www.rand.org/health/surveys_tools/mos/mos_core_36item.html. Accessed 2014 Nov 5.

11. QualityMetric. https://www.optum.com/optum-outcomes.html. Accessed 2014 Sep 29.

12. Hays R, Prince-Embury S, Chen H. R-36 HIS: RAND-36 health status inventory. San Antonio, TX: The Psychological Corporation; 1998.

13. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996 Mar;34(3):220-33.

14. Patel AA, Donegan D, Albert T. The 36-Item Short Form. J Am Acad Orthop Surg. 2007 Feb;15(2):126-34.

15. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care. 2004 Sep;42(9):851-9.

16. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002 Mar;21(2):271-92.

17. Kazis LE, Ren XS, Lee A, Skinner K, Rogers W, Clark J, Miller DR. Health status in VA patients: results from the Veterans Health Study. Am J Med Qual. 1999 Jan-Feb;14(1):28-38.

18. Jones D, Kazis L, Lee A, Rogers W, Skinner K, Cassar L, Wilson N, Hendricks A. Health status assessments using the Veterans SF-12 and SF-36: methods for evaluating otucomes in the Veterans Health Administration. J Ambul Care Manage. 2001 Jul;24(3):68-86.

19. Kazis LE, Miller DR, Clark JA, Skinner KM, Lee A, Ren XS, Spiro A 3rd, Rogers WH, Ware JE Jr. Improving the response choices on the Veterans SF-36 Health Survey role functioning scales: results from the Veterans Health Study. J Ambul Care Manage. 2004 Jul-Sep;27(3):263-80.

20. Mullin PA, Lohr KN, Bresnahan BW, McNulty P. Applying cognitive design principles to formatting HRQOL instruments. Qual Life Res. 2000 Feb;9(1):13-27.

21. Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. Ann Med. 2001 Jul;33(5):350-7.

22. Ware JE, Kosinski M. Interpreting SF-36 summary health measures: a response. Qual Life Res. 2001;10(5):405-13; discussion 415-20.

23. Baron R, Elashaal A, Germon T, Hobart J. Measuring outcomes in cervical spine surgery: think twice before using the SF-36. Spine (Phila Pa 1976). 2006 Oct 15;31 (22):2575-84.

24. Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. Med Care. 1995 Apr;33(4)(Suppl):AS264-79.

25. Taft C, Karlsson J, Sullivan M. Do SF-36 summary component scores accurately summarize subscale scores? Qual Life Res. 2001;10(5):395-404.

26. Nortvedt MW, Riise T, Myhr KM, Nyland HI. Performance of the SF-36, SF-12, and RAND-36 summary scales in a multiple sclerosis population. Med Care. 2000 Oct;38(10):1022-8.

27. Frendl DM, Ware JE Jr. Patient-reported functional health and well-being outcomes with drug therapy: a systematic review of randomized trials using the SF-36 health survey. Med Care. 2014 May;52(5):439-45.

28. Farivar SS, Cunningham WE, Hays RD. Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.I. Health Qual Life Outcomes. 2007;5:54. Epub 2007 Sep 7.

29. Gartsman GM, Brinker MR, Khan M, Karahan M. Self-assessment of general health status in patients with five common shoulder conditions. J Shoulder Elbow Surg. 1998 May-Jun;7(3):228-37.

30. Agren PH, Mukka S, Tullberg T, Wretenberg P, Sayed-Noor AS. Factors affecting long-term treatment results of displaced intraarticular calcaneal fractures: a post hoc analysis of a prospective, randomized, controlled multicenter trial. J Orthop Trauma. 2014 Oct;28(10):564-8.

31. Hsu KY, Zucherman JF, Hartjen CA, Mehalic TF, Implicito DA, Martin MJ, Johnson DR 2nd, Skidmore GA, Vessa PP, Dwyer JW, Cauthen JC, Ozuna RM. Quality of life of lumbar stenosis-treated patients in whom the X STOP interspinous device was implanted. J Neurosurg Spine. 2006 Dec;5(6):500-7.

32. Bhandari M, Busse JW, Hanson BP, Leece P, Ayeni OR, Schemitsch EH. Psychological distress and quality of life after orthopedic trauma: an observational study. Can J Surg. 2008 Feb;51(1):15-22.

33. Hopman WM, Towheed T, Anastassiades T, Tenenhouse A, Poliquin S, Berger C, Joseph L, Brown JP, Murray TM, Adachi JD, Hanley DA, Papadimitropoulos E; Canadian Multicentre Osteoporosis Study Research Group. Canadian normative data for the SF-36 health survey. CMAJ. 2000 Aug 8;163(3):265-71.

34. Borg T, Berg P, Larsson S. Quality of life after operative fixation of displaced acetabular fractures. J Orthop Trauma. 2012 Aug;26(8):445-50.

35. Busija L, Osborne RH, Nilsdotter A, Buchbinder R, Roos EM. Magnitude and meaningfulness of change in SF-36 scores in four types of orthopedic surgery. Health Qual Life Outcomes. 2008;6:55. Epub 2008 Jul 31.

36. Whitehouse JD, Friedman ND, Kirkland KB, Richardson WJ, Sexton DJ. The impact of surgical-site infections following orthopedic surgery at a community hospital and a university hospital: adverse quality of life, excess length of stay, and extra cost. Infect Control Hosp Epidemiol. 2002 Apr;23(4):183-9.

37. Obremskey WT, Dirschl DR, Crowther JD, Craig WL 3rd, Driver RE, LeCroy CM. Change over time of SF-36 functional outcomes for operatively treated unstable ankle fractures. J Orthop Trauma. 2002 Jan;16(1):30-3.

38. Hozack WJ, Rothman RH, Albert TJ, Balderston RA, Eng K. Relationship of total hip arthroplasty outcomes to other orthopaedic procedures. Clin Orthop Relat Res. 1997 Nov;(344):88-93.