The *RAND–36 Health Status Inventory* (RAND–36 HSI) is a 36-item measure of general health status that assesses the following constructs: physical functioning, role limitations due to physical health problems, pain, general health perceptions, emotional well-being, role limitations due to emotional problems, social functioning, and energy/fatigue.

The RAND–36 HSI was derived from the *Health-Related Quality of Life (HRQOL) Survey* developed at the RAND Corporation for use in the Medical Outcomes Study (Stewart et al., 1992; Tarlov et al., 1989). A 36-item short-form version of the HRQOL Survey, known as the SF–36, has been used extensively in outcomes research (Ware & Sherbourne, 1992). The RAND–36 HSI consists of the same 36 items as the SF–36 but incorporates a sophisticated scaling methodology based on item response theory (IRT), factor-based composite scores, and national norms closely stratified by age, race/ethnicity, educational level, sex, and geographic region according to the U.S. census data. Chapters 2–5 of the Manual describe the characteristics of the RAND–36 HSI standardization sample, the IRT-based scaling procedures for response options within items and items within scales, the factor analyses and derivation of composite scores, and reliability and validity studies. Chapter 6 discusses the administration, scoring, and interpretation of the RAND–36 HSI and the 12-item version, the *RAND–12 Health Status Inventory*. Chapter 7 concludes with a discussion of the clinical uses of the RAND–36 HSI. Appendixes provide procedures and tables for the computation of item scores (Appendix A); procedures and tables for the computation of scale and composite T scores (Appendix B); T scores obtained by cumulative percentages of the standardization sample (Appendix C); discrepancies between the Physical Health and Mental Health composite T scores obtained by various percentages of the standardization sample (Appendix D); a description of the RAND–12 HSI, used for predicting composite scores for aggregate-level analysis (Appendix E); and the items by scale and composite of the RAND–36 HSI (Appendix F) and the RAND–12 HSI (Appendix G). The remainder of Chapter 1 summarizes the history of the SF–36, relevant constructs, previous research, and applications.

A number of measures of functioning and well-being were developed for the Medical Outcomes Study (MOS), a 4-year observational study of the influence of characteristics of providers, patients, and health systems on outcomes of care (Stewart et al., 1992). The MOS

included both cross-sectional and longitudinal components. For the cross-sectional component, over 20,000 patients from among literate English-speaking adults visiting participating practices were sampled during a 9-day screening period (per site) in 1986. Patients from the cross-sectional study who had one or more of four chronic medical conditions (hypertension, diabetes, heart disease, depression) constituted the sampling frame for the longitudinal panel of 2,546 patients (Hays, Sherbourne, & Mazel, 1995; Stewart et al., 1992).

Because of the need to screen a large number of patients to identify those with the MOS-targeted conditions, a very brief, 20-item health-related quality-of-life survey (SF–20) was used in the cross-sectional study. The SF–20 consisted of 18 items used in a 1984 national survey fielded by Louis Harris and Associates (Ware, Sherbourne, & Davies, 1992) plus single items measuring social functioning and pain (Stewart, Hays, & Ware, 1988).

The MOS longitudinal participants were administered a much more extensive battery of items (from which the RAND–36 HSI was derived) at baseline than those administered during the cross-sectional phase. These long-form measures are described elsewhere (Hays, Sherbourne, et al., 1995; Stewart et al., 1992). A short form of the HRQOL, the SF–36, was also administered at multiple assessments during the MOS.

The items of the SF–36, which take about 8 minutes to self-administer, were selected to maximize their associations with the long-form MOS scales from which they were derived (Ware & Sherbourne, 1992). The SF–36 is an improvement over the original MOS short-form, the SF–20, because it includes an additional health concept, energy/fatigue; increases the precision of previous single-item measures (i.e., pain and social functioning) and multi-item measures (e.g., physical functioning) by the addition of items; measures the extent of physical limitations rather than the duration of the limitation; and focuses on a wider array of role limitations.

Constructs Assessed

The SF–36 and the RAND–36 HSI tap eight health constructs: physical functioning, role limitations caused by physical health problems, pain, general health perceptions, emotional well-being, role limitations caused by emotional problems, social functioning, and energy/fatigue. Both also include a single item (Item 2) that provides an indication of perceived change in health. Following are definitions of the constructs assessed by the RAND–36 HSI, as well as by the SF–36. (The RAND–36 HSI items by scale and composite are provided in Appendix F.)

Scales

Physical Functioning
The Physical Functioning Scale (PF) consists of 10 items that measure the individual's limitations in physical activities because of health. Items 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12 compose this scale.

Role Limitations due to Physical Health Problems

The Role Limitations due to Physical Health Problems Scale (RLP) consists of 4 items that measure the extent to which physical health interferes with doing work or other regular daily activities. Items 13, 14, 15, and 16 compose this scale.

Pain

The Pain Scale (PA) includes 2 items, Items 21 and 22, that measure pain frequency and the extent of role interference due to pain.

General Health Perceptions

The General Health Perceptions Scale (GHP) includes 5 items that measure the individual's perceptions of health in general, such as feeling well or ill. Items 1, 33, 34, 35, and 36 compose this scale.

Emotional Well-Being

The Emotional Well-Being Scale (EWB) includes 5 items that measure general mood or affect, including depression, anxiety, and positive well-being. Items 24, 25, 26, 28, and 30 compose this scale.

Role Limitations due to Emotional Problems

The Role Limitations due to Emotional Problems Scale (RLE) consists of 3 items—Items 17, 18, and 19—that measure the extent to which emotional problems interfere with doing work or other regular daily activities.

Social Functioning

The Social Functioning Scale (SF) includes 2 items—Items 20 and 32—that measure the extent to which health interferes with social activities with family, friends, neighbors, or groups.

Energy/Fatigue

The Energy/Fatigue Scale (EF) includes 4 items that measure feeling energetic versus tired and worn out. Items 23, 27, 29, and 31 compose this scale.

Composites

Factor analyses of the SF–36 have provided strong support for a two-factor model of health. The physical health factor is reflected primarily by measures of physical functioning, pain, and role limitations due to physical health problems; the mental health factor is reflected primarily by measures of emotional well-being and role limitations caused by emotional problems (Hays, Marshall, Wang, & Sherbourne, 1994; Hays, Sherbourne, & Mazel, 1993). General health perceptions, energy/fatigue, and social functioning have been described as reflecting both health dimensions. Physical and mental health composite scores for the SF–36 scales have been derived (Hays et al., 1993; Ware, Kosinski, Bayliss, et al., 1995).

The physical and mental health factors derived by Ware, Kosinski, Bayliss, et al. (1995) were forced to be uncorrelated (orthogonal factor rotation), whereas those derived by Hays et al. (1993) were allowed to correlate (oblique factor rotation). Correlations between physical and mental health factors at each of three intervals (baseline, 2-years post-baseline,

and 4-years post-baseline) in the MOS ranged from .32 to .41 (Hays et al., 1993). Similarly, a correlation of .53 between physical health and mental health factors was reported in a study of 1,053 older individuals (average age = 64 years) sampled from an academic general medical clinic (Dexter, Stump, Tierney, & Wolinsky, 1996). Moderate to large correlations among the SF–36 scale scores were noted to be inconsistent with two orthogonal dimensions in a study of 2,088 Australian hospital inpatients (Shadbolt, McCallum, & Singh, 1997). Oblique rotations often yield a more realistic representation of factors than do orthogonal rotations (Rummel, 1970). This conceptualization of physical health and mental health factors as correlated is consistent with the procedures presented in this Manual.

This Manual presents three composites, two representing the previously established physical health and mental health factors, and the third, global composite, which is derived from the previous two. The Physical Health Composite (PHC) is composed of the following scales: Physical Functioning, Role Limitations due to Physical Health Problems, Pain, and General Health Perceptions. The Mental Health Composite (MHC) is composed of the following scales: Emotional Well-Being, Role Limitations due to Emotional Problems, Social Functioning, and Energy/Fatigue. The Global Health Composite (GHC) is composed of all of these scales.

Reliability and Validity

Reliability estimates for all of the SF–36 scale scores were .78 or higher in the MOS (Hays et al., 1993). Reliability coefficients ranged from a low of .65 to a high of .94 for subgroups differing in age, sex, ethnicity, education, disease condition, and disease severity (McHorney, Ware, Lu, & Sherbourne, 1994). Similar reliability estimates were found for a variety of other samples including older adults (Andresen, Bowley, Rothenberg, Panzer, & Katz, 1996), hemodialysis patients (Hays, Kallich, Mapes, Coons, & Carter, 1994; Kurtin, Davies, Meyer, DeGiacomo, & Kantz, 1992), osteoarthritis patients (Kantz, Harris, Levitsky, Ware, & Davies, 1992), and epilepsy patients (Wagner et al., 1995). Reliability estimates for the physical and mental health composite scores exceed .90 (Ware, Kosinski, Bayliss, et al., 1995). Multitrait scaling analyses (Hays & Hayashi, 1990) support item convergence for hypothesized scales and item discrimination across scales (McHorney, Ware, & Raczek, 1993).

Support for convergent and discriminant validity of the SF–36 scores has been provided by a study of 1,582 residents of Sheffield, England administered the SF–36 and the *Nottingham Health Profile* (Brazier et al., 1992). The SF–36 scales were found to have a median relative efficiency of .93 in discriminating between patients differing in severity of medical and psychiatric conditions, compared to corresponding long-form scales in the MOS (McHorney, Ware, Rogers, Raczek, & Lu, 1992). Physical health measures best distinguished groups differing in severity of chronic medical illness whereas mental health measures best distinguished groups differing in the presence and severity of psychiatric disorders (McHorney et al., 1993). Epilepsy patients who were seizure-free were found to score better on the SF–36 scales than those experiencing seizures after epilepsy surgery (Vickrey et al., 1994). In a sample of osteoarthritis patients (Kantz et al., 1992), SF–36 Physical Functioning scores were significantly related (r = .65) to a two-item scale assessing walking and stair-climbing adapted from the Knee Society's physical function measure (Insall, Dorr, Scott, & Scott, 1989). Scores on the SF–36 Emotional Well-Being Scale have been shown

to correlate −.72 with the *Center for Epidemiological Studies Depression Scale* (CES–D) for a sample of 9,749 participants in the Breast Cancer Prevention Trial (Ganz, Day, Ware, Redmond, & Fisher, 1995). Patients with panic disorder were found to score worse than the general U.S. population on Role Functioning, Social Functioning, Emotional Well-Being, Energy/Fatigue, and General Health Perceptions (Sherbourne, Wells, & Judd, 1996).

Average SF–36 Physical Functioning and physician-rated Harris Hip scores (Harris, 1969) paralleled one another over time following surgery in a study of 139 patients receiving total hip replacements (Lansky, Butler, & Waller, 1992). Social Functioning scores have been shown to be the most responsive of the SF–36 scale scores to treatment of migraine with sumatriptan succinate (Solomon, 1997). Small to moderate improvements in SF–36 scores were found after varicose vein surgery, with the greatest improvements on the Physical Functioning and Pain scales (Garratt, Ruta, Abdalla, & Russell, 1996).

Research and Clinical Applications

Target Populations

Results with the SF–36 have been reported for at least 90 diseases or conditions (Shiely, Bayliss, Keller, Tsai, & Ware, 1996). Through 1996, there were five or more publications for each of at least 15 diseases or conditions: arthritis, asthma, back pain, chronic heart failure, depression, diabetes, epilepsy, hypertension, low back pain, menorrhagia, myocardial infarction, total hip replacement, older age, total knee replacement, and varicose veins; there were 10 or more publications on each of depression, diabetes, epilepsy, and hypertension.

Individual-Level Assessment

The use of the SF–36 for individual-level assessment has been relatively infrequent (Barr & Schumacher, 1995; Titler & Reiter, 1994). For example, use of the SF–36 has been reported for individuals with kidney disease (Kurtin et al., 1992), epilepsy (Wagner et al., 1995), persons hospitalized for total hip replacement (Lansky et al., 1992), and Medicaid recipients (Nelson, Hartman, Ojemann, & Wilcox, 1995). In an investigation of change over 3 months among older people referred to community continence services ($n = 18$) or mental health services ($n = 29$), Hill, Harries, and Popay (1996) concluded from qualitative interviews that there were positive impacts but found little change on SF–36 scores. On the basis of estimates of the reliability of measurement, McHorney and Tarlov (1995) suggested that the SF–36 and other HRQOL surveys included in their review have insufficient precision for individual-level applications. However, a randomized, controlled study of 163 consecutive patients with epilepsy at the New England Medical Center outpatient neurology clinic revealed that the SF–36 provided new information in a majority of the encounters (as judged by physicians) and prompted a change in therapy 13% of the time (Wagner et al., 1997).

This Manual (Chapter 7) presents clinical guidelines for use of the RAND–36 HSI scales and composites in a clinical context. Also presented are guidelines for the observation of change or stability of scores over time, which is consistent with the prevailing practice of considering confidence intervals and the relative level of *T* scores. In addition, because

responses to the RAND–36 HSI are scored and interpreted objectively, the instrument is suitable for computer administration, scoring, and interpretation such as that available with OPTAIO (The Psychological Corporation, 1998). The application of the OPTAIO program to the RAND–36 HSI is described in Chapter 6.