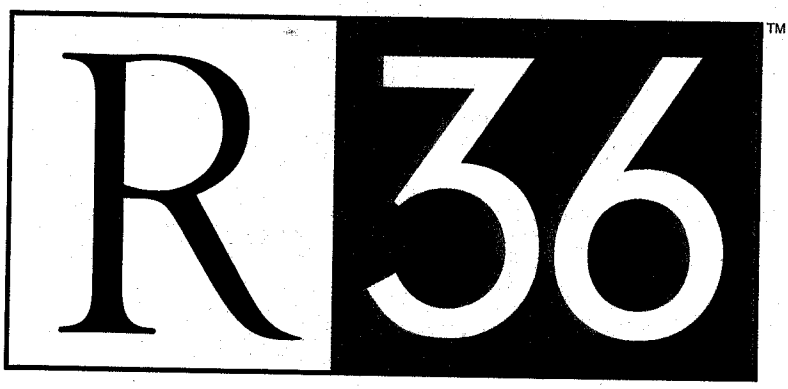


RAND–36 Health Status Inventory

Ron D. Hays

With special contributions by
Sandra Prince–Embury and Hsin–Yi Chen

THE PSYCHOLOGICAL CORPORATION®
Harcourt Brace & Company
SAN ANTONIO
Orlando • Boston • New York • Chicago • San Francisco • Atlanta • Dallas
San Diego • Philadelphia • Austin • Fort Worth • Toronto • London • Sydney

Copyright ©1998 by The Psychological Corporation

Normative data copyright ©1998 by The Psychological Corporation

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

The Psychological Corporation and the *PSI* logo are registered trademarks of The Psychological Corporation.

Beck Depression Inventory, BDI, Beck Anxiety Inventory, BAI, Beck Hopelessness Scale, and *BHS* are registered trademarks of The Psychological Corporation.

Printed in the United States of America

1234567891011 12ABCDE

Foreword

The 36 questions that comprise the *RAND–36 Health Status Inventory* are probably the most commonly asked health status questions worldwide. Also known as the SF–36, this questionnaire is being used in countless health outcome studies, health care financing studies, and clinical practice evaluations. Although there is an increasing volume of published data from various diagnostic cohorts and defined populations, including the original Medical Outcomes Study (MOS) standardization sample, there has not to date been a manual that provides U.S. census-based norms stratified by sex, age, racial/ethnic group, and educational level. These data could well become the benchmark for comparison of one's local, regional, or national results.

The 36 questions in the RAND–36 HSI were selected from the larger pool of items used in the MOS. Item selection was based on criteria that maximized item association with the longer scales. These 36 items were also selected in part to ensure coverage of the full spectrum of physical and mental health. One consequence of the commitment to cover a full range of functioning with relatively few items is the loss of precision and sensitivity to change at any given level of health, compared to disease or condition-specific assessment. However, a decided gain of this approach is the distribution of item difficulty across the continuum of health, as measured by item response theory (IRT). Therefore, although this questionnaire was not developed by IRT methodology, it is well suited to it. This Manual represents the first major effort to produce a scoring system for these 36 items that capitalizes on the strengths of IRT. The IRT method of ordering all items, and response categories within items, along a single continuum allows for the empirical weighting of responses to questions of differing difficulty according to that underlying continuum. This approach is a major advance in scoring of responses to the questionnaire and can, if developed, evolve into a major advance in health status assessment.

This Manual provides an unparalleled matrix of normative data for group and individual comparison purposes. The use of nonorthogonal factor rotations to derive separate physical and mental health composite scores is sensible, given the known relationship between these components of health, and distinguishes the RAND–36 HSI from the SF–36 scoring system. In addition, the use of a Global Health Composite score is unique and (I daresay) long awaited by the scientific community. Several health status questionnaires that tap physical and mental health domains concurrently have managed to create workable, psychometrically sound total scores. In our work with cancer and HIV patients, we have computed an IRT-based total score from the RAND–36 HSI, but this Manual marks the first time that a formal scoring system will be disseminated along with population-based normative data for adults. The Global Health Composite score has somehow previously eluded formal scoring systems for such questionnaires, but it is successfully implemented here with use of IRT methodology. This total score will have value not only at the group decision-making level but also at the individual patient-management level, where physical and mental health concerns must frequently be balanced with each other and combined into a single "bottom-line" summary.

This Manual includes over 50 tables, most of which provide *T*-score conversions for each of the individual scale and composite scores. These conversions are helpful because they transform IRT-based logistic data, which might be unfamiliar for some users, into a more familiar, standardized metric. These tables will no doubt be helpful to the benchmarking of an individual user's data, but perhaps the most useful aspects of this Manual are the sections on determining statistical significance of change and evaluating the clinical meaningfulness of change scores. Without this kind of information, interpretation of results is limited. Information about the clinical, real-world relevance of a given score or an increment of change in that score is vital to interpretation. This information, in turn, may move health care providers toward better cost-effectiveness and cost-utility modeling as we understand just what the meaning of an improvement in health afforded by a given intervention is.

In conclusion, there are thousands of people using these 36 questions in their clinical practice and clinical research. Most will benefit from having this Manual within reach when making sense of their data. Perhaps this work, carried out so meticulously with the U.S. English-speaking population, will next be expanded to include other languages and other countries where these same 36 questions are used.

David Cella, PhD
Research Professor
Institute for Health Services Research and Policy Studies
Northwestern University
Director, Center on Outcomes Research and Education
Evanston Northwestern Healthcare

Acknowledgments

I would like to acknowledge a number of people for their outstanding contributions to the development of the *RAND–36 Health Status Inventory* (RAND–36 HSI). Sandra Prince–Embury, PhD, Project Director, did an amazing job from start to finish in seeing this effort through to completion. Dr. Prince–Embury is really the major author of this work and deserves most of the credit for its completion. Hsin–Yi Chen, PhD, was the primary architect for the application of item response theory scaling in this project and played a central role in most of the statistical decisions. Larry Weiss, PhD, Director of Behavioral Healthcare and Personality Group, provided valuable consultation throughout the project, including the design of the data-collection plan and help in preparing this Manual and the normative tables. Appreciation also goes to David Tulsky, PhD, Project Director, for reviewing the Manual. Stephanie Tong, Research Assistant, dedicated many hours to the preparation of tables and revisions of the Manual. Her diligence and attention to detail in ensuring that the information presented is accurate and complete are especially valued.

Special thanks are also extended to those individuals whose meticulous and diligent efforts were essential in the preparation of the Manual for publication. Those responsible for ensuring the editorial quality of the Manual are Kathy Overstreet, Senior Editor, and Cynthia Woerner, Consulting Editor. The production of the Manual itself was in the capable hands of Debra Belknap, Senior Production Manager; Margaret Donohue, Production Manager; Michael Friedman, Production Specialist; Pat Malec, Manager of Art and Design; and Javier Flores, Associate Designer.

Joanne Lenke, PhD, President of The Psychological Corporation, and Aurelio Prifitera, PhD, Vice President and Director of the Psychological Measurement Group, provided administrative support for the project. Dr. Prifitera also personally reviewed the Manual and provided helpful feedback.

A special thanks is due to David Cella, PhD, a pioneer in using item response theory methods in health status assessment, for his diligence in reviewing the Manual and writing the foreword.

Finally, this effort would not have been possible without the contributions of the measurement experts at the RAND Corporation who worked on the RAND Medical Outcomes Study (MOS). These individuals include Anita Stewart, Cathy Sherbourne, and the principal investigator of the MOS, John Ware.

Ron D. Hays, PhD
RAND Health Sciences Program, Santa Monica, California
UCLA School of Medicine, Los Angeles, California
July 1998

Contents

Foreword iii

Acknowledgments v

Chapter 1 Introduction

Constructs Assessed 2
Scales 2
Physical Functioning 2
Role Limitations due to Physical Health Problems 3
Pain 3
General Health Perceptions 3
Emotional Well-Being 3
Role Limitations due to Emotional Problems 3
Social Functioning 3
Energy/Fatigue 3
Composites 3
Reliability and Validity 4
Research and Clinical Applications 5
Target Populations 5
Individual-Level Assessment 5

Chapter 2 Standardization

Description of the Sampling Procedure 7
Demographic Characteristics of the Age-Based Sample 8
Demographic Characteristics of the Age-Stratified Sample 11
Disability Status and Physical Conditions 12

Chapter 3 Item Scaling and Development of the Composites

Item Scaling 15
Theoretical Background 15
Previous Research 16
Application of IRT Weighting 17

Effects of IRT Scoring on Distribution 17
Development of the Composites 18

Chapter 4 Norming

Determination of Normative Groups 21
Standardization of Scale Raw Scores 24
Standardization of Composite Raw Scores 24
Cumulative Percentages for *T* Scores 24

Chapter 5 Reliability and Validity of Scores

Reliability 25
Reliability of Scale Scores 25
Reliability of Composite Scores 26
Test–Retest Reliability 26
Validity 27
Item Level 27
Scale Level 28
Composite Level 28
Previous Studies 29
Association of RAND–36 HSI With Other Measures 29
Physical Health Measures 30
Mental Health Measures 31
Global Health Composite 34
Known-Group Analyses 34
Utilization of Health-Care Resources 35

Chapter 6 Administration and Scoring

Hand-Scoring for the RAND–36 HSI 39
Hand-Scoring for the RAND–12 HSI 39
Computer-Based Administration, Scoring, and Interpretation 40

Chapter 7 Interpretation of the RAND–36 HSI Scores

Determination of Clinical Meaningfulness 46
General Parameters 46
Missing Item Responses 46
Clinical Cut Scores 46

Sequential Analysis 48
Suggested Guidelines 48
Physical Health Composite 48
Mental Health Composite 48
Physical Health and Mental Health Composite Score Discrepancy . . . 49
Global Health Composite 49
Scale Scores 50
A Final Caution 50
Longitudinal Tracking of Change 50
General Guidelines 51
Step 1: Determining Statistical Significance of Change 52
Step 2: Evaluating the Clinical Meaningfulness of Change 53
Summary 54

Appendixes

Appendix A Computation of Item Scores 55
Appendix B Computation of Scale and Composite Scores for the RAND–36 HSI 61
Appendix C *T* Scores Obtained by Cumulative Percentages of the Normative Samples 85
Appendix D Discrepancy Score Tables 105
Appendix E Description of the *RAND–12 Health Status Inventory* 107
Appendix F RAND–36 HSI Items and Response Options by Composite and Scale 115
Appendix G RAND–12 HSI Items and Response Options by Composite . . 119

References

Figures

Figure 2.1. Percentages of the RAND–36 HSI Age-Based Standardization Sample and the U.S. Population by Occupation 11

Figure 6.1. OPTAIO Session Graph of RAND–36 HSI Composite and Scale Scores 40

Figure 6.2. OPTAIO Longitudinal Graph of RAND–36 HSI Composite Scores 41

Figure 6.3A. OPTAIO Interpretive Report of RAND–36 HSI Performance (Page 1) 42

Figure 6.3B. OPTAIO Interpretive Report of RAND–36 HSI Performance (Page 2) 43

Tables

Table 2.1. Percentages of the RAND–36 HSI Age-Based Standardization Sample and the U.S. Population by Race/Ethnicity 9

Table 2.2. Percentages of the RAND–36 HSI Age-Based Standardization Sample and the U.S. Population by Education Level 9

Table 2.3. Median Ages of the RAND–36 HSI Age-Based Standardization Sample by Age Group and Sex 10

Table 2.4. Percentages of Unemployed Participants in the RAND–36 HSI Age-Based Standardization Sample and the U.S. Population by Sex 10

Table 2.5. Percentages of the RAND–36 HSI Age-Stratified Sample and the U.S. Population by Age Group and Sex 12

Table 2.6. Percentages of the RAND–36 HSI Age-Stratified Sample and the U.S. Population by Race/Ethnicity and Sex 12

Table 2.7. Percentages of the RAND–36 HSI Age-Stratified Sample and the U.S. Population by Education Level and Sex 12

Table 2.8. Percentages of RAND–36 HSI Participants Reporting Disabilities 13

Table 2.9. Percentages of Physical Conditions and Mean Level of Interference in Functioning Reported by RAND–36 HSI Standardization Participants 14

Table 3.1. Promax Factor Pattern Loadings for RAND–36 HSI Scales 20

Table 4.1. Mean Scale Raw Scores, Standard Deviations, Range Scores, and Distribution Statistics for the Seven Normative Groups 22

Table 5.1. Reliability Coefficients for RAND–36 HSI Scale and Composite Scores for the Seven Normative Groups 26

Table 5.2. Test–Retest Reliability Coefficients for RAND–36 HSI Scale and Composite Scores 27
Table 5.3. Intercorrelations of the RAND–36 HSI Scale Scores for the Age-Stratified Sample 28
Table 5.4. Promax Factor Pattern Loadings for the RAND–36 HSI Scales 29
Table 5.5. Correlations Between the RAND–36 HSI and Other Indicators of Physical Health 30
Table 5.6. Correlations Between the RAND–36 HSI and the BDI–II, BAI, and BHS for the Age-Stratified Sample 31
Table 5.7. Correlations Between the RAND–36 HSI and the *Brief Symptom Inventory* 32
Table 5.8. Correlations Between the RAND–36 HSI and the BASIS–32 33
Table 5.9. Correlations Between the RAND–36 HSI and the SAS–SR 34
Table 5.10. Correlations Between the RAND–36 HSI and Self-Reported Health-Care Resource Utilization 37
Table 7.1. Standard Errors of Prediction for the Three RAND–36 HSI Composites at 90% Level of Confidence 53
Table 7.2. *T*-Score Ranges for Evaluating the Clinical Meaningfulness of Change 54
Table A.1. Item Response-Option Weights: Physical Functioning Scale (PF) 56
Table A.2. Item Response-Option Weights: Role Limitations due to Physical Health Problems Scale (RLP) 56
Table A.3. Item Response-Option Weights: Pain Scale (PA) 57
Table A.4. Item Response-Option Weights: General Health Perceptions Scale (GHP) 57
Table A.5. Item Response-Option Weights: Emotional Well-Being Scale (EWB) 58
Table A.6. Item Response-Option Weights: Role Limitations due to Emotional Problems Scale (RLE) 58
Table A.7. Item Response-Option Weights: Social Functioning Scale (SF) 59
Table A.8. Item Response-Option Weights: Energy/Fatigue Scale (EF) 59
Table B.1. Raw-Score Ranges for the Scales 63
Table B.2. *T*-Score Equivalents of Raw Scores: Physical Functioning Scale (PF) 64

Table B.3. *T*-Score Equivalents of Raw Scores: Role Limitations due to Physical Health Problems Scale (RLP) 66
Table B.4. *T*-Score Equivalents of Raw Scores: Pain Scale (PA) 68
Table B.5. *T*-Score Equivalents of Raw Scores: General Health Perceptions Scale (GHP) 70
Table B.6. *T*-Score Equivalents of Raw Scores: Emotional Well-Being Scale (EWB) 72
Table B.7. *T*-Score Equivalents of Raw Scores: Role Limitations due to Emotional Problems Scale (RLE) 74
Table B.8. *T*-Score Equivalents of Raw Scores: Social Functioning Scale (SF) 75
Table B.9. *T*-Score Equivalents of Raw Scores: Energy/Fatigue Scale (EF) 77
Table B.10. *T*-Score Equivalents of Raw Scores: Physical Health Composite (PHC) 79
Table B.11. *T*-Score Equivalents of Raw Scores: Mental Health Composite (MHC) 81
Table B.12. *T*-Score Equivalents of Raw Scores: Global Health Composite (GHC) 83
Table C.1. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Physical Functioning Scale (PF) 85
Table C.2. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Role Limitations due to Physical Health Problems Scale (RLP) 87
Table C.3. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Pain Scale (PA) 88
Table C.4. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: General Health Perceptions Scale (GHP) 90
Table C.5. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Emotional Well-Being Scale (EWB) 92
Table C.6. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Role Limitations due to Emotional Problems Scale (RLE) 94
Table C.7. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Social Functioning Scale (SF) 95
Table C.8. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Energy/Fatigue Scale (EF) 97
Table C.9. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Physical Health Composite (PHC) 99

Table C.10. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Mental Health Composite (MHC) 101
Table C.11. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Global Health Composite (GHC) 103
Table D.1. Cumulative Percentages of the Normative Samples Obtaining PHC>MHC *T*-Score Discrepancies 105
Table D.2. Cumulative Percentages of the Normative Samples Obtaining MHC>PHC *T*-Score Discrepancies 106
Table E.1. Correlations Between RAND–12 HSI and RAND–36 HSI Item Scores for the Seven Normative Groups 112
Table E.2. Intercept Values and Parameter Estimates for Predicting Physical Health Composite *T* Scores From the RAND–12 HSI 113
Table E.3. Intercept Values and Parameter Estimates for Predicting Mental Health Composite *T* Scores From the RAND–12 HSI 113
Table E.4. Intercept Values and Parameter Estimates for Predicting Global Health Composite *T* Scores From the RAND–12 HSI 114
Table E.5. Cross-Validation Results Predicting RAND–36 HSI Composite *T* Scores From the RAND–12 HSI 114

Chapter 1
Introduction

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.

Chapter 2
Standardization

Description of the Sampling Procedure

The RAND–36 HSI normative information presented in this Manual is based on a national standardization sample of 800 respondents representative of the U.S. population of adults aged 18–89 years. A stratified sampling plan ensured that the standardization samples included representative proportions of adults according to each selected demographic variable. An analysis of the data gathered by the U.S. Bureau of the Census (1993) provided the basis for stratification by race/ethnicity and education level within age group. Due to the well-established findings that many health status constructs vary by age, two standardization samples were developed, the age-based sample, which was stratified *within* age group, and the age-stratified sample, which was stratified *across* age groups. In this way, an individual's health status may be interpreted relative both to a representative sample of adults in the individual's age cohort and to a representative U.S. population of adults in general.

First, the standardization sampling plan ensured that all age groups were sufficiently represented and, therefore, that the norms adequately represented each group. The standardization sample included equal numbers of participants by sex (n = 100 per sex) within each of the four age groups, 18–24, 25–44, 45–64, and ≥65 years. Each age group was also stratified by race/ethnicity and level of education according to U.S. population proportions. Thus, four age groups of 200 participants each composed the total age-based standardization sample of 800. Age-specific *T*-score norms and descriptive cumulative percentages are based on the data from this sample. (The norming procedure is discussed in Chapter 4; raw-score-to-*T*-score norms tables and cumulative percentage data are presented in Appendixes B and C, respectively.)

Second, for the purpose of standardization across age groups, a subsample of 500 participants was selected and stratified by age, race/ethnicity, and educational level according to U.S. population proportions. The required number of participants for each demographic variable was randomly selected from the total age-based standardization sample. The subsample included 255 female and 245 male participants. Data from this age-stratified sample were used to generate norms for the overall population and by sex.

Thus, the age-based normative sample included 200 participants in each age group, whereas the age-stratified sample included the same percentages of individuals in each age group as existed in the U.S. population. The age-based sample was stratified by race/ethnicity and educational level within each of the four age bands; the age-stratified sample was stratified by age, race/ethnicity, and educational level across that total sample. Use of the age-based norms yields a score that reflects the person's standing relative to others in the same age group. Use of the age-stratified norms yields a score that reflects the person's standing relative to the total population or to the total population of men or women. The provision of both age-based specific and general norms reflects the reality that health status may be viewed both from the perspective of what is characteristic at given points within the life cycle and from the perspective of changes that occur across the life cycle.

Regional geographic representation was controlled by the selection of 100 individuals in each of eight cities in four different regions of the country. In each of the four regions, two cities were selected, one larger and one smaller. Participants were initially selected by random telephone calling. Potential participants were screened by telephone, and those selected participated in small groups assembled by professional marketing research firms. Each participant received monetary reimbursement for 1–2 hours of participation. Each responded to a questionnaire developed by The Psychological Corporation, which consisted of several test instruments and demographic and life-history items.

McHorney, Ware, et al. (1994) have raised the issue that generalization from previous studies posed difficulties due to differences in data completeness among groups such as racial/ethnic minorities, older persons, and socioeconomically disadvantaged persons. For the standardization of the RAND–36 HSI reported in this Manual, this potential problem was addressed, first, by the initial stratification criteria for sample selection and, second, by the systematic replacement of cases with missing data with demographically matched cases.

Demographic Characteristics of the Age-Based Sample

Table 2.1 presents the racial/ethnic characteristics of the RAND–36 HSI age-based standardization sample and the U.S. population by age group. Table 2.2 presents the educational characteristics of the age-based sample and the U.S. population by age group. Table 2.3 presents the median ages of the RAND–36 HSI age-based sample by age group and sex. The data in Table 2.3 show that the median age for each age group is close to the center of each age band and is comparable for female and male participants.

Table 2.1. Percentages of the RAND–36 HSI Age-Based Standardization Sample and the U.S. Population by Race/Ethnicity

Age Group	White	African American	Hispanic	Other
18–24				
RAND–36 HSI	70.0	14.5	12.0	3.5
U.S. Population	69.9	14.5	11.7	3.9
25–44				
RAND–36 HSI	75.0	12.0	9.0	4.0
U.S. Population	74.4	12.2	9.4	4.0
45–64				
RAND–36 HSI	80.5	10.0	6.5	3.0
U.S. Population	79.8	10.2	6.7	3.3
≥65				
RAND–36 HSI	87.0	8.5	3.5	1.0
U.S. Population	85.2	8.6	4.0	2.2

Note. N = 800; for each age group, n = 200 (100 female and 100 male participants). U.S. population data are from the *Current Population Survey, March 1993, on CD–ROM* [CD–ROM] by U.S. Bureau of the Census, 1993, Washington, DC: U.S. Bureau of the Census [Producer/Distributor].

Table 2.2. Percentages of the RAND–36 HSI Age-Based Standardization Sample and the U.S. Population by Education Level

	Years of Education		
Age Group	≤12 Years	13–15 Years	≥16 Years
18–24			
RAND–36 HSI	56.0	37.0	7.0
U.S. Population	52.4	39.8	7.8
25–44			
RAND–36 HSI	46.5	28.0	25.5
U.S. Population	47.0	27.6	25.4
45–64			
RAND–36 HSI	57.5	19.5	23.0
U.S. Population	56.2	21.2	22.6
≥65			
RAND–36 HSI	67.0	18.0	15.0
U.S. Population	73.8	14.2	12.1

Note. N = 800; for each age group, n = 200 (100 female and 100 male participants). U.S. population data are from the *Current Population Survey, March 1993, on CD–ROM* [CD–ROM] by U.S. Bureau of the Census, 1993, Washington, DC: U.S. Bureau of the Census [Producer/Distributor]. Due to rounding, percentages may not sum to 100.

Table 2.3. Median Ages of the RAND–36 HSI Age-Based Standardization Sample by Age Group and Sex

Age Group	Combined	Female	Male
18–24	20.0	20.5	20.0
25–44	34.0	36.0	33.0
45–64	53.0	52.0	55.0
≥65	72.0	71.0	73.0

Note. $N = 800$; for each age group, $n = 200$ (100 female and 100 male participants).

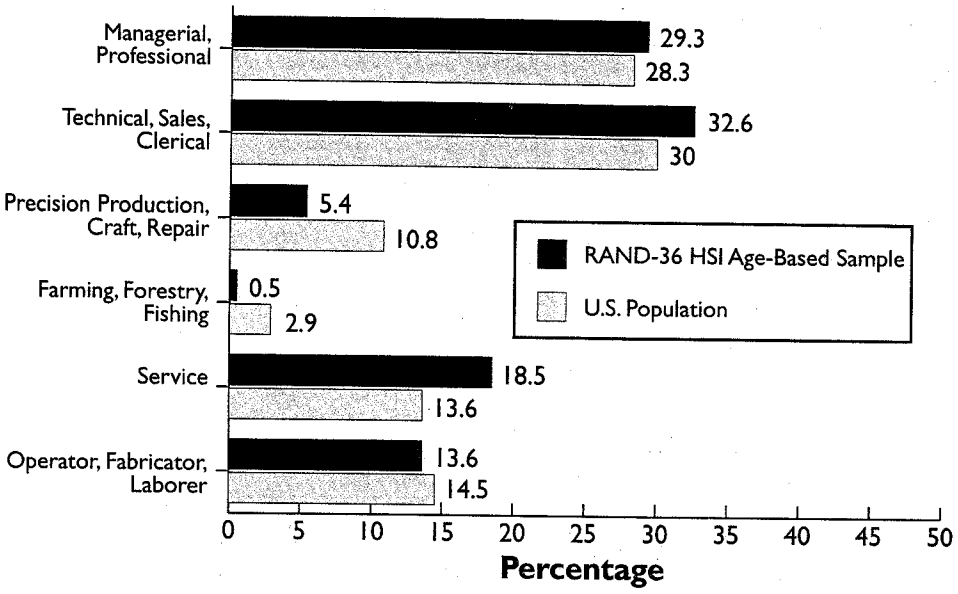
Table 2.4 presents the percentages of unemployed participants in the RAND–36 HSI age-based standardization sample and the U.S. population by total sample and by sex. Figure 2.1 graphically presents the occupational characteristics of the RAND–36 HSI age-based sample compared with the U.S. population characteristics. Although the standardization sample was not selected according to occupational characteristics, it was important to ensure that the sample was representative of the U.S. population in this respect.

Table 2.4. Percentages of Unemployed Participants in the RAND–36 HSI Age-Based Standardization Sample and the U.S. Population by Sex

Sample	Combined	Female	Male
RAND–36 HSI	5.8	5.3	6.3
U.S. Population	5.6	5.6	5.6

Note. $N = 800$; for each age group, $n = 200$ (100 female and 100 male participants). U.S. employment data for civilian noninstitutional population 16 years old and older are from the *Statistical Abstract of the United States 1995*, by U.S. Bureau of the Census, 1995, Washington, DC: Author.

Figure 2.1. Percentages of the RAND–36 HSI Age-Based Standardization Sample and the U.S. Population by Occupation



Note. N = 800 (400 female and 400 male participants). U.S. employment data for civilian noninstitutional population 16 years old and older are from the *Statistical Abstract of the United States 1995*, by U.S. Bureau of the Census, 1995, Washington, DC: Author.

Demographic Characteristics of the Age-Stratified Sample

As stated previously, the age-stratified sample was specifically designed to represent the U.S. population distribution by age for the development of age-stratified norms. Table 2.5 presents the distribution of the age-stratified sample and the U.S. population by age group and sex. Table 2.6 presents the distribution of the age-stratified sample and the U.S. population by race/ethnicity and sex. Table 2.7 presents the distribution of the age-stratified sample and the U.S. population by education level and sex. As the data in these tables show, the proportions of the RAND–36 HSI age-stratified sample closely matched those of the U.S. population on the selected demographic variables.

Table 2.5. Percentages of the RAND–36 HSI Age-Stratified Sample and the U.S. Population by Age Group and Sex

Age Group	Overall	Female	Male	U.S. Population
18–24	15.0	14.5	15.5	13.0
25–44	40.0	39.2	40.8	43.9
45–64	27.0	26.7	27.4	26.6
≥65	18.0	19.6	16.3	16.5

Note. N = 500 for the overall sample. U.S. population data are from the *Current Population Survey, March 1993, on CD–ROM* [CD–ROM] by U.S. Bureau of the Census, 1993, Washington, DC: U.S. Bureau of the Census [Producer/Distributor].

Table 2.6. Percentages of the RAND–36 HSI Age-Stratified Sample and the U.S. Population by Race/Ethnicity and Sex

Race/Ethnicity	Overall	Female	Male	U.S. Population
White	77.6	78.4	76.7	77.0
African American	11.2	11.0	11.4	11.4
Hispanic	8.0	8.2	7.8	8.1
Other	3.2	2.4	4.1	3.5

Note. N = 500 for the overall sample. U.S. population data are from the *Current Population Survey, March 1993, on CD–ROM* [CD–ROM] by U.S. Bureau of the Census, 1993, Washington, DC: U.S. Bureau of the Census [Producer/Distributor].

Table 2.7. Percentages of the RAND–36 HSI Age-Stratifed Sample and the U.S. Population by Education Level and Sex

Education Level	Overall	Female	Male	U.S. Population
≤12 Years	55.0	52.9	57.1	54.6
13–15 Years	25.4	27.1	23.7	25.3
≥16 Years	19.6	20.0	19.2	20.2

Note. N = 500 for the overall sample. U.S. population data are from the *Current Population Survey, March 1993, on CD–ROM* [CD–ROM] by U.S. Bureau of the Census, 1993, Washington, DC: U.S. Bureau of the Census [Producer/Distributor].

Disability Status and Physical Conditions

The RAND–36 HSI sampling procedure described previously also provided for limited screening for physical health. Specifically, participants had to be sufficiently ambulatory to travel or to be transported to the testing site and had to have adequate visual and hearing acuity to participate in the testing. Although the standardization sample was nonclinical, some medically diagnosed physical conditions were anticipated. Self-reported disabilities in vision, hearing, and mobility and the presence of other physically disabling conditions were

queried as part of the initial questionnaire. Table 2.8 presents percentages of those having a visual, auditory, mobility, or other type of disability by age group for those in the age-based sample and by sex for the age-stratified sample. From these data, it can be concluded that the standardization samples (both age-based and age-stratified), although nonclinical, have some representation of nondebilitating disabilities.

Table 2.8. Percentages of RAND–36 HSI Participants Reporting Disabilities

Group	Type of Disability			
	Visual	Auditory	Mobility	Other[a]
Age-Based Sample				
18–24 (n = 200)	5.0	0.5	1.0	1.0
25–44 (n = 200)	6.0	0.5	3.5	1.0
45–64 (n = 200)	5.5	2.0	4.5	5.5
≥65 (n = 200)	6.0	8.0	6.0	5.0
Age-Stratified Sample				
Overall (N = 500)	4.8	2.0	3.6	2.8
Female (n = 255)	3.9	1.6	2.7	1.6
Male (n = 245)	5.7	2.4	4.5	4.1

Note. Disability categories are not mutually exclusive.

[a]Type of disability was unspecified.

Table 2.9 presents a list of the physical conditions queried during the standardization study and the percentages of the participants reporting each symptom. Data are presented for the age-based sample by age group and for the age-stratified sample overall and by sex. Also indicated is the mean level of self-reported severity of interference in functioning experienced for each condition. The severity of interference is based on a scale from 1 (*diagnosed, no interference*) to 4 (*diagnosed, severe interference*). Mean severity of interference was calculated only for those participants who indicated the presence of the physical condition. Disability status and self-reported physical conditions reported here were used to define criterion groups for construct validation of the RAND–36 HSI measures. The relationship of disability status and physical condition to RAND–36 HSI scales is described further in Chapter 5.

Table 2.9. Percentages of Physical Conditions and Mean Level of Interference in Functioning Reported by RAND–36 HSI Standardization Participants

	Age-Based Sample								Age-Stratified Sample					
	18–24 Years		25–44 Years		45–64 Years		≥65 Years		Overall		Female		Male	
	(*n* = 200)		(*n* = 200)		(*n* = 200)		(*n* = 200)		(*n* = 500)		(*n* = 255)		(*n* = 245)	
Physical Condition	%	Mean	%	Mean	%	Mean	%	Mean	%	Mean	%	Mean	%	Mean
Back Trouble/Pain/Dislocation	10.5	2.05	15.0	2.30	15.5	2.39	17.0	2.18	14.6	2.34	14.9	2.32	14.3	2.37
High Blood Pressure	1.5	1.00	6.0	1.17	19.5	1.18	24.0	1.25	10.6	1.19	9.8	1.32	11.4	1.07
Arthritis	2.5	1.80	3.0	2.00	11.5	1.87	24.5	1.63	8.2	1.73	10.6	1.81	5.7	1.57
Asthma	9.0	1.83	6.0	1.33	4.5	1.44	4.0	1.75	6.0	1.43	7.5	1.53	4.5	1.27
Heart Problem	3.5	1.43	2.0	1.00	9.0	1.78	11.5	1.70	5.6	1.64	4.7	1.58	6.5	1.69
Diabetes	3.0	1.33	3.0	1.83	7.5	1.53	11.5	1.78	4.6	1.83	3.5	2.33	5.7	1.50
Thyroid Problem	2.0	1.25	2.5	1.20	4.5	1.22	6.5	1.00	4.2	1.19	5.9	1.13	2.4	1.33
Irritable Bowel Syndrome	3.0	1.67	5.5	1.73	4.0	1.38	3.0	1.50	4.0	1.55	5.9	1.53	2.0	1.60
Kidney Problems/Stones/Infections	2.0	2.25	3.0	2.17	5.0	1.80	2.5	1.20	3.8	1.95	2.0	2.00	5.7	1.93
Ulcer	2.0	2.00	3.0	1.83	2.5	1.20	3.5	1.29	3.4	1.47	3.5	1.33	3.3	1.63
Chronic Headache	5.5	2.27	3.5	2.43	2.5	1.80	3.5	1.86	3.2	2.25	4.7	2.42	1.6	1.75
Anemia	3.5	1.29	2.5	1.00	1.0	2.00	2.0	1.50	2.4	1.17	3.9	1.00	0.8	2.00
Other Neurological Problem	2.5	2.20	2.5	2.20	2.0	2.75	2.5	1.80	2.4	2.17	2.4	1.83	2.4	2.50
Head Injury	4.5	1.78	2.0	1.25	1.0	1.00	1.0	2.00	2.0	1.60	2.0	1.40	2.0	1.80
Chronic Fatigue Syndrome	2.0	2.00	1.5	3.33	2.0	2.50	2.5	2.00	1.8	2.44	2.0	2.20	1.6	2.75
Colitis	1.5	1.33	1.5	1.67	2.5	1.40	1.5	1.00	1.8	1.33	2.4	1.33	1.2	1.33
Other Immune System Disorder	2.0	1.75	1.5	2.00	1.5	2.67	1.5	2.67	1.6	2.25	2.0	2.60	1.2	1.67
Cancer	1.5	2.00	1.0	1.50	2.0	2.00	5.5	1.36	1.6	1.25	1.6	1.25	1.6	1.25
Other Respiratory Problems	3.5	1.86	2.0	2.25	1.5	1.33	3.0	1.50	1.6	2.00	2.4	1.83	0.8	2.50
Stroke	1.5	2.00	1.0	1.00	0.5	2.00	3.0	2.33	1.4	1.86	1.2	1.67	1.6	2.00
Epstein-Barr	1.5	1.00	1.0	2.50	1.0	4.00	1.5	1.33	1.2	2.50	1.6	2.50	0.8	2.50
Emphysema	1.5	1.33	0.5	1.00	1.0	2.50	3.0	1.83	1.2	1.67	1.6	2.00	0.8	1.00
Tachycardia	1.5	1.00	0.5	1.00	1.0	1.50	3.0	1.83	1.0	1.80	1.2	2.00	0.8	1.50
HIV	1.5	1.00	0.5	1.00	0.5	2.00	0.5	1.00	0.8	1.25	0.8	1.00	0.8	1.50
Epilepsy	2.5	1.40	0.5	1.00	1.5	2.00	1.0	1.50	0.8	1.00	1.2	1.00	0.4	1.00

Note. Mean level of interference was calculated only for those participants who indicated the presence of the condition; therefore, mean levels are based on different sample sizes, depending on the frequency of the symptom. Severity of interference in functioning is based on the following scale: 1 = *diagnosed, no interference*; 2 = *diagnosed, mild interference*; 3 = *diagnosed, moderate interference*; and 4 = *diagnosed, severe interference.*

Chapter 3
Item Scaling and Development of the Composites

Item Scaling

Theoretical Background

The original *RAND 36–Item Health Survey* (Hays et al., 1993) and 33 of the 36 items on the SF–36 utilize a traditional method of scoring, namely, a simple summation approach. However, this technique is based on certain implicit assumptions about the items. The first assumption is that each item has options that fit an interval scale. In other words, the options are assumed to be equidistant in terms of the metric of the construct underlying the scale. For example, if the three response options for an item are *not at all*, *somewhat*, and *very much*, then the assumption is that the difference in the construct being measured between *not at all* and *somewhat* is the same as that between *somewhat* and *very much*.

The second assumption in the simple summation approach is that all of the items should contribute equally to the overall scale. This assumption means that a response of *somewhat* represents the same amount of the underlying construct for every item in a scale. Further, there is the implicit assumption that the overall scores on a scale are based on an interval scale. For example, the difference between scores of 7 and 8 is the same as that between scores of 8 and 9. The difficulty presented by such an approach is that on a scale of illness with the first 8 items indicative of very minor illness and the last 2 indicative of very severe illness, a person who answered the first 7 items affirmatively is actually only slightly less ill than one who answered the first 8 items affirmatively. However, the difference in illness indicated by these two individuals and by someone who endorses the first 9 items is considerably larger.

In item response theory (IRT), the relationship between examinee item performance and the latent trait can be described by an item characteristic function that is monotonically increasing (Hambleton & Swaminathan, 1985; Hambleton, Swaminathan, & Rogers, 1991; Hays, 1998). The advantage of using latent trait estimation from IRT is that an estimate of how much each response should contribute to the overall score can be made and depends on the underlying level of the construct associated with that item response. Thus, on a measure of psychological distress, a response of "sometimes" to "I feel that I am going to die" would be

assigned a higher score (on this continuum) than a response of "often" to "I feel uncomfortable." With IRT, the assignment of differential scores to responses is possible because all of the responses, across and within items, can be placed on a latent-trait continuum.

Scoring for the RAND–36 HSI scales is based on a one-parameter IRT model (1-PL). According to this IRT model, the expected score of a respondent on a particular item is a function of both the item difficulty and the ability (latent trait) of the respondent (Rasch, 1960, 1966). Within the 1-PL family, the Rasch model was used for scales with all dichotomous-response items (e.g., 1 or 0 or yes/no) and the partial credit model was used for scales composed of items with multicategory response options (Masters, 1982). A different weight is assigned to each response option within items of the same scale by placing all response options of the same scale on the same underlying latent-trait continuum.

Previous Research

With earlier applications of the SF–36 scales, a simple summation approach for scoring was used. The assumption was that simple scoring was possible because items of the same scale had roughly equivalent relationships to the underlying health construct being measured. Researchers initially suggested that it was not necessary to standardize or weight the SF–36 items (McHorney, Ware, et al., 1994; Ware, Snow, Kosinski, & Gandek, 1993). In examining the results from 24 different patient and demographically diverse groups, McHorney, Ware, et al. (1994) also maintained that simple summation was warranted on the basis of Likert's (1932) assumptions of similar item means, score variances, and item–total correlations.

More recently, evidence contrary to Likert assumptions has been presented. Ware et al. (1993) suggested that of the SF–36 items, two were shown by empirical data not to satisfy the assumptions of a linear relationship between item scores and the underlying health construct defined by the scale. According to Keller, Ware, and Gandek (1995), observed departures from equal-interval assumptions were consistent across countries and were the greatest for the two response scales that had been recalibrated in the SF–36 scale scores (*excellent, very good, good, fair, poor* and *none, very mild, mild, moderate, severe, very severe*). One item of the Pain Scale was found to have severity ratings that did not satisfy the assumptions of equal intervals. Response-option values were consequently recalibrated for the SF–36 and values derived from the mean values of a summary criterion; the values computed were the mean value for respondents who chose each of the six levels defined in Item 7 of the same scale (Ware et al., 1993).

Recalibration had been recommended for the item measuring general health on the General Health Perceptions Scale (Davies & Ware, 1981; Stewart et al., 1988; Ware, Nelson, Sherbourne, & Stewart, 1992). These researchers found that the mean value for a criterion of general health for the respondents who chose each of the five levels defined by this item departed significantly from linearity. Intervals between adjacent response categories were unequal (Davies & Ware, 1981). Accordingly, these response options for the General Health Perceptions Scale were recalibrated for the RAND–36 HSI.

Simple summation scoring was also questioned for several of the SF–36 items by Hays et al. (1993). Haley, McHorney, and Ware (1994) employed the Rasch model to examine the hierarchical structure, unidimensionality, and reproducibility of item positions (calibrations) on

the 10-item Physical Functioning Scale. This analysis generated an empirical item hierarchy, confirmed the unidimensionality of the scale for most respondents, and established the reproducibility of item calibrations across patient populations and repeated tests. More recently, McHorney, Haley, and Ware (1997) compared simple summation scoring based on the Likert scale with the Rasch IRT scaling model for the 10-item Physical Functioning Scale. Findings favored the Rasch model in discriminating between patients who differed in disease severity. Differences were reported as most apparent in clinical groups whose scores approximated the extremes of the distribution. It was suggested that the Rasch model of scoring would be relevant to the clinical interpretation of individual scores on this scale. The development of the RAND–36 HSI stems from the significant amount of early work documenting the need for revision in scoring.

Application of IRT Weighting

IRT methodology was used for scoring the RAND–36 HSI responses in order to take into account relative item weights within each scale and item response weights within each item simultaneously.

The first step in applying the IRT scoring method was to select an appropriate sample, that is, one for which there were no missing responses to any items. Of the age-based standardization sample, 737 protocols met this criterion. Based on this sample's responses, the IRT weighting for items in each of the eight RAND–36 HSI scales was determined by the following methodology.

Calibration based on a one-parameter IRT logistic model was conducted to obtain the item characteristic curve (ICC) for each item in the scale. An ICC provides the expected item score (item response option) as a function of the individual's ability level on the construct being measured (the latent-trait continuum), given the item difficulty (or step values) of that item. For each ICC, the corresponding ability level for each response option can be obtained. Once all response options were placed on the same latent-trait continuum, they were rescaled to a 0–100 linear scale where the existing minimum ability level was set to 0 and the maximum ability level was set to 100. This new scale served as the basis for the IRT weighting of item responses.

A raw score is computed by summing the IRT weight for the selected response option for each item on the scale. Appendix A of this Manual presents the procedures and tables for these computations; Appendix B presents the procedures and tables for computing scale and composite *T* scores (the derivation of *T* scores is described in Chapter 4).

Effects of IRT Scoring on Distribution

Initial expectations were that the differences in the distribution due to IRT scoring would not be equal for all scales. Moreover, differences would be greatest for those scales with the widest range of ability levels, as determined by the number of items and the number of response options for items on that scale. Therefore, scales composed of items with only two response options, such as the 4-item Role Limitations due to Physical Health Problems and the 3-item Role Limitations due to Emotional Problems, were anticipated to show little

change in score distribution due to IRT scoring. Methods of identifying potential differences between simple summation and IRT scoring of scales included comparison of means, skewness, and kurtosis, as well as chi-square analysis.

For purposes of comparison, IRT raw scale scores were placed on the same 0–100 range as the scores derived by the simple summation method. A comparison of raw-score means based on the IRT and simple summation methods yielded significant differences for six of the eight scales. As predicted, those scales offering two response options per item did not show a significant difference (i.e., Role Limitations due to Physical Health Problems and Role Limitations due to Emotional Problems). On all six of the other scales, the IRT method yielded mean scores significantly lower than the mean scores derived with the simple summation method ($p < .0001$). Among the scales of the Physical Health Composite, Physical Functioning showed a small but significant difference between mean scores, and on Pain and General Health Perceptions, mean scores obtained by the IRT method were more than 3 points lower than mean scores obtained by simple summation. Among the Mental Health Composite scales, mean scores on Emotional Well-Being and Energy/Fatigue obtained by the IRT method were 9 and 7 points lower, respectively, than mean scores based on simple summation.

Some differences in distribution of scores were also indicated. A comparison of skewness and kurtosis demonstrated that IRT scoring yields smaller skewness (absolute value) and smaller kurtosis for seven of the eight scales. Skewness was consistently smaller with IRT scoring; the largest differences were for Pain (0.23), General Health Perceptions (0.20), and Energy/Fatigue (0.43). Kurtosis was also consistently smaller with IRT scoring. These differences were statistically significant ($p < .01$) for Physical Functioning (0.47), Pain (0.70), General Health Perceptions (0.45), Emotional Well-Being (1.58), and Energy/Fatigue (0.68). These results indicate that IRT scoring generally resulted in less skewness and smaller kurtosis, that is, in distributions that were more spread out and flatter.

The precise nature of the shift in distribution afforded by IRT scoring varied across scales. Chi-square analyses of differences by scoring method (IRT or simple summation) across scales, with scores organized into 10 ability levels, revealed shifts in the distribution of scores that were significant in five of eight scales: Pain ($\chi^2 = 46.69$, $p < .001$), General Health Perceptions ($\chi^2 = 17.08$, $p < .05$), Emotional Well-Being ($\chi^2 = 104.53$, $p < .001$), Social Functioning ($\chi^2 = 146.49$, $p < .001$), and Energy/Fatigue ($\chi^2 = 39.17$, $p < .001$).

Development of the Composites

The theoretical assumptions underlying the composite scores, as well as their psychometric development are discussed here. The steps for computing these composite scores are presented in Appendix B.

Studies of health status have consistently identified distinct physical health and mental health factors. These factors have been identified in patient groups (Hays, Marshall, et al., 1994; McHorney et al., 1993; Ware, Gandek, & the IQOLA Project Group, 1994), in the general U.S. population (Ware et al., 1993), and across different demographic and patient groups (Ware, Kosinski, Bayliss, et al., 1995).

The methodology used for deriving composite scores for the RAND–36 HSI differs from that employed with the SF–36 in several ways (Ware, Kosinski, & Keller, 1994). First, the method used for the RAND–36 HSI is based on results from principal axis factor analysis rather than principal components factor analysis, which was applied to the SF–36 (Ware, Kosinski, et al., 1994). With the principal axis factor analysis, the obtained factors (physical health and mental health) are based on common, rather than on total variance among scales (Gorsuch, 1983); they are the true underlying factors (as opposed to sample-specific components); and they explain as much of the common variance as possible.

Second, the method of composite score construction used for the RAND–36 HSI differs from those presented previously (Ware, Kosinski, et al., 1994) because the formula for the composite score includes only those scales that load highly on that factor. As a result, the Physical Health Composite score is derived from scores on the Physical Functioning, Role Limitations due to Physical Health Problems, Pain, and General Health Perceptions scales. The Mental Health Composite score is derived from the scores on the Emotional Well-Being, Role Limitations due to Emotional Problems, Social Functioning, and Energy/Fatigue scales.

An additional difference in the methodology used for the RAND–36 HSI is that it employs an oblique rotation rather than the orthogonal rotation employed previously (Ware, Kosinski, et al., 1994). The oblique rotation method is based on the assumption that the physical health and mental health factors are correlated, not independent, as assumed in orthogonal rotation methods. Previous research has found that physical and mental aspects of health are distinguishable, but also significantly correlated (Hays, Marshall, et al., 1994).

The existence of distinct physical and mental components of health status has been well documented (Hays, Marshall, et al., 1994; McHorney et al., 1993). In addition, factor patterns found in previous analyses have replicated the significant loadings on physical health and mental health factors. In previous studies that included all eight scales in the calculation of each composite, three scales—General Health Perceptions, Social Functioning, and Energy/Fatigue—loaded on both factors. This result is cited as justification for their inclusion in both the Physical Health and Mental Health composites. The composite scores for the RAND–36 HSI do not include overlapping scales because the factor loadings did not warrant their inclusion and because the Global Health Composite was developed to represent the overlapping aspects of physical health and mental health.

Thus, in addition to the Physical Health and Mental Health composites, the RAND–36 HSI yields a Global Health Composite. This composite reflects the conception of underlying global health that is composed of both physical health and mental health and potentially overlapping aspects. It can be viewed as a "thermometer" of general health. This composite is consistent with the original conception of general health as an integrative, underlying construct. In practice, the Global Health Composite score may be used in circumstances when one measure of general health status is required or when the distinction between physical health and mental health is not important.

Physical Health and Mental Health composites were derived by common factor analysis (principal axis method with two iterations and squared multiple correlations for priors) that specified two related factors with oblique rotation. For the analysis, scores on all eight scales

obtained by the age-stratified sample (N = 500) were used. The analysis was restricted to two iterations because when factors are allowed to be correlated, the model, which will continue to change through iterations, may be overfitted and lead to estimates of communalities greater than 1 (actual communalities cannot exceed 1). Such communalities are known as "Heywood" cases. Limiting the number of iterations to two with good prior estimates leads to more accurate estimates of the communalities (Gorsuch, 1983). The rotated promax factor pattern matrix presented in Table 3.1 reveals that the four scales related to physical health defined the first factor, with loadings ranging from .63 to .90. This factor was defined as the Physical Health Composite. The four scales related to mental health loaded on the second factor, with loadings ranging from .53 to .95. This factor was defined as the Mental Health Composite.

Factor scores for both the Physical Health Composite and the Mental Health Composite were derived for each member of the age-stratified sample with the UniMult program (Gorsuch, 1991). The third composite, the Global Health Composite, was then derived by factoring the two factor scores, physical health and mental health, with one common factor specified.

Table 3.1. Promax Factor Pattern Loadings for RAND–36 HSI Scales

Scale	Factor 1 Physical Health	Factor 2 Mental Health
Physical Functioning	**.90**	−.21
Role Limitations due to Physical Health Problems	**.79**	.00
Pain	**.76**	.02
General Health Perceptions	**.63**	.18
Emotional Well-Being	−.21	**.95**
Role Limitations due to Emotional Problems	.00	**.59**
Social Functioning	.30	**.53**
Energy/Fatigue	.24	**.58**

Note. N = 500. The factor analysis was based on the scores obtained by the age-stratified sample. The estimated correlation between Factors 1 and 2 was .66.

A linear equation containing beta weights was obtained for each composite by regression analysis. For the analysis, the factor score of the composite was used as the dependent variable, and the scores of the scales contributing to that composite were used as the independent variables. It should be noted that the scale weights developed from the regression equations do not match the rank ordering of the factor loadings because scale weights represent the relative contribution of what is measured by each scale to the factor, whereas the factor loadings represent the contribution of the underlying factor to what is measured by each scale.

Chapter 4
Norming

Determination of Normative Groups

A two-way (age X sex) analysis of variance (ANOVA) revealed significant age effects for all of the scales of the Physical Health Composite (Physical Functioning, $F = 60.53$, $p < .0001$; Role Limitations due to Physical Health Problems, $F = 14.48$, $p < .0001$; Pain, $F = 9.42$, $p < .0001$; General Health Perceptions, $F = 9.00$, $p < .0001$) and for the Emotional Well-Being Scale ($F = 13.30$, $p < .0001$). As the data in Table 4.1 show, mean raw scores for the Physical Health scales decrease with age whereas mean raw scores for the Emotional Well-Being Scale increase with age. The analysis also revealed significant sex effects for the Energy/Fatigue Scale ($F = 7.04$, $p < .008$), with female respondents reporting less energy. Sex effects were also found for the General Health Perceptions Scale ($F = 3.59$, $p < .06$), with male respondents obtaining slightly lower scores, and for the Role Limitations due to Emotional Problems Scale ($F = 3.73$, $p < .05$), with female respondents obtaining slightly lower scores. No interaction effects proved significant.

On the basis of these findings, the standardization sample of 800 was organized by age into four normative groups, with 200 participants in each group: 18–24, 25–44, 45–64, and ≥65. Additionally, a subgroup of 500, with 255 female and 245 male participants, was drawn from the large sample and stratified across age. Each of the four groups of the age-based sample and the age-stratified sample was stratified by race/ethnicity and education level to match the U.S. population proportions according to the 1993 census (U.S. Bureau of the Census; see Chapter 2). Thus, the RAND–36 HSI provides norms for seven groups for each scale: 18–24, 25–44, 45–64, ≥65, age-stratified sample ($n = 500$), age-stratified female sample ($n = 255$), and age-stratified male sample ($n = 245$). Table 4.1 provides the mean total raw scores; standard deviations (*SD*s); minimum, maximum, and median scores; and skewness and kurtosis values for the seven normative groups for the eight RAND–36 HSI scales.

Table 4.1. Mean Scale Raw Scores, Standard Deviations, Range Scores, and Distribution Statistics for the Seven Normative Groups

Scale	Statistic	Age-Based Sample 18–24 (*n* = 200)	25–44 (*n* = 200)	45–64 (*n* = 200)	≥65 (*n* = 200)	Age-Stratified Sample Overall (*n* = 500)	Female (*n* = 255)	Male (*n* = 245)
Physical Functioning								
	Mean	536.08	523.92	481.12	441.81	497.29	493.38	501.35
	SD	57.78	63.87	93.03	91.21	85.80	86.46	85.10
	Minimum	209.00	257.00	222.00	209.00	209.00	209.00	209.00
	Maximum	564.00	564.00	564.00	564.00	564.00	564.00	564.00
	Median	564.00	564.00	517.00	457.50	540.00	525.00	540.00
	Skewness	−3.06	−2.01	−1.18	−0.62	−1.42	−1.33	−1.53
	Kurtosis	10.60	3.81	0.33	−0.57	1.17	0.97	1.47
Role Limitations due to Physical Health Problems								
	Mean	269.30	262.04	250.38	233.44	255.53	254.87	256.21
	SD	43.54	54.24	60.55	70.33	58.59	58.66	58.62
	Minimum	112.00	112.00	112.00	112.00	112.00	112.00	112.00
	Maximum	288.00	288.00	288.00	288.00	288.00	288.00	288.00
	Median	288.00	288.00	288.00	288.00	288.00	288.00	288.00
	Skewness	−2.46	−1.98	−1.33	−0.82	−1.60	−1.57	−1.64
	Kurtosis	5.16	2.47	0.21	−1.02	1.01	0.92	1.17
Pain								
	Mean	133.71	127.15	120.89	114.97	123.51	122.58	124.48
	SD	32.06	36.16	40.82	38.38	38.59	39.98	37.15
	Minimum	6.00	6.00	6.00	29.00	6.00	6.00	6.00
	Maximum	161.00	161.00	161.00	161.00	161.00	161.00	161.00
	Median	139.00	139.00	139.00	127.00	139.00	139.00	139.00
	Skewness	−1.44	−1.36	−0.85	−0.49	−1.04	−1.03	−1.05
	Kurtosis	2.00	1.51	−0.29	−0.89	0.27	0.22	0.32
General Health Perceptions								
	Mean	333.10	329.44	307.59	297.36	319.25	323.32	315.02
	SD	79.21	81.97	86.92	77.91	82.84	80.27	85.38
	Minimum	82.00	82.00	67.00	110.00	82.00	82.00	93.00
	Maximum	434.00	434.00	434.00	434.00	434.00	434.00	434.00
	Median	339.50	344.00	316.00	304.00	330.00	332.00	319.00
	Skewness	−0.66	−0.77	−0.45	−0.11	−0.56	−0.68	−0.44
	Kurtosis	−0.05	0.25	−0.52	−0.73	−0.25	0.07	−0.50

Table 4.1. *(Continued)*

Scale	Statistic	Age-Based Sample 18–24 (*n* = 200)	25–44 (*n* = 200)	45–64 (*n* = 200)	≥65 (*n* = 200)	Age-Stratified Sample Overall (*n* = 500)	Female (*n* = 255)	Male (*n* = 245)
Emotional Well-Being								
	Mean	235.99	239.35	253.73	274.27	251.65	247.56	255.89
	SD	67.37	70.27	73.16	58.84	66.80	67.19	66.27
	Minimum	31.00	34.00	37.00	99.00	34.00	37.00	34.00
	Maximum	361.00	361.00	361.00	361.00	361.00	361.00	361.00
	Median	241.50	251.50	276.00	279.50	266.00	262.00	276.00
	Skewness	–0.47	–0.64	–0.77	–0.59	–0.74	–0.75	–0.73
	Kurtosis	–0.03	–0.07	–0.04	–0.13	0.23	0.29	0.19
Role Limitations due to Emotional Problems								
	Mean	178.60	174.91	180.45	175.12	179.83	177.80	181.95
	SD	34.39	39.42	35.33	40.89	34.85	36.21	33.32
	Minimum	72.00	72.00	72.00	72.00	72.00	72.00	72.00
	Maximum	195.00	195.00	195.00	195.00	195.00	195.00	195.00
	Median	195.00	195.00	195.00	195.00	195.00	195.00	195.00
	Skewness	–2.07	–1.76	–2.34	–1.85	–2.24	–2.04	–2.50
	Kurtosis	3.18	1.61	4.05	1.83	3.69	2.89	4.90
Social Functioning								
	Mean	165.18	160.61	162.61	166.82	164.28	163.81	164.77
	SD	33.27	38.53	39.34	32.67	36.17	35.47	36.94
	Minimum	27.00	27.00	27.00	71.00	27.00	27.00	38.00
	Maximum	188.00	188.00	188.00	188.00	188.00	188.00	188.00
	Median	188.00	188.00	188.00	188.00	188.00	188.00	188.00
	Skewness	–1.36	–1.40	–1.60	–1.45	–1.51	–1.46	–1.56
	Kurtosis	1.15	1.07	1.65	0.97	1.39	1.32	1.50
Energy/Fatigue								
	Mean	194.80	190.85	197.29	203.37	196.62	189.25	204.29
	SD	77.77	75.39	83.72	81.21	78.96	78.22	79.15
	Minimum	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	Maximum	354.00	354.00	354.00	354.00	354.00	354.00	354.00
	Median	194.50	197.50	209.00	207.00	203.50	188.00	219.00
	Skewness	–0.22	–0.27	–0.20	–0.30	–0.34	–0.25	–0.46
	Kurtosis	–0.47	–0.67	–0.80	–0.54	–0.61	–0.64	–0.50

Note. Raw scores represent IRT scoring within scale and as such are not comparable across different scales.

Standardization of Scale Raw Scores

The linear *T*-score transformation method was chosen for transforming raw scores to standardized scores, with a mean of 50 and standard deviation of 10. The computation of *T* scores for each normative group was based on the mean scores and standard deviations specific to that norm group. Table B.1 in Appendix B provides the raw-score ranges for the scales. Tables B.2–B.9 provide *T*-score conversions for the seven normative groups for each of the eight RAND–36 HSI scales. The linear method was chosen because the scale scores for the overall age-stratified sample were not distributed normally. The score distributions on five scales were skewed (i.e., greater than +1.00 or less than –1.00): Physical Functioning (–1.42), Role Limitations due to Physical Health Problems (–1.60), Pain (–1.04), Role Limitations due to Emotional Problems (–2.24), and Social Functioning (–1.51). The scores on the other three scales approximated a normal distribution: General Health Perceptions (–0.56), Emotional Well-Being (–0.74), and Energy/Fatigue (–0.34).

Standardization of Composite Raw Scores

Composite raw scores are based on the *T* scores of the four scales making up each composite. The composite raw scores were computed according to the algorithms discussed in Chapter 3. For each composite, the relevant scale *T* scores are multiplied by the appropriate beta weights (provided in Appendix A) and then summed to form the composite raw score. The composite raw score is then converted to a linear *T* score. Linear *T* scores are used because they preserve the underlying skewness of the score distributions. *T*-score equivalents of raw scores for the Physical Health, Mental Health, and Global Health composites are provided in Tables B.10, B.11, and B.12, respectively.

Cumulative Percentages for *T* Scores

T scores provide information on an individual's score relative to the total variability of that individual's normative group. In the instance of skewed distributions, however, a *T* score does not adequately reflect the relative ranking of the individual within that group. The cumulative frequency (cumulative percentage) of a particular score may be clinically significant. For instance, although a *T* score of 50 on the Physical Functioning Scale represents the mean *T* score, only 37.8% of the overall sample obtained a score equal to or less than this score, as opposed to 50% as would be the case for a sample with a normal score distribution. This point is characteristic of a skewed sample for which the mean score is lower than the median score. This finding is not surprising in a nonclinical population, for which scores on health-status instruments would be expected to cluster at the positive end of the continuum. The cumulative percentages of the seven normative groups obtaining various *T* scores are presented in Tables C.1–C.11 in Appendix C.

Chapter 5
Reliability and Validity of Scores

Reliability

The reliability of the scores on a scale reflects the extent to which the scale constitutes a homogeneous set of items (internal consistency) and the extent to which it consistently yields the same score when there is no real change in what is being measured.

Reliability of Scale Scores

Cronbach's coefficient alpha (1951) was used to estimate internal consistency for the RAND–36 HSI scales. This measure reflects the degree of item covariance in relation to the number of items in the scale. This estimate of score reliability is useful for verifying whether multiple items of the same construct yield consistent information about respondents. Table 5.1 presents the reliability coefficients for the eight scales for the seven normative groups. For the overall age-stratified sample, alpha coefficients range from a low of .71 on the Social Functioning Scale to a high of .90 on the Physical Functioning Scale. Thus, for the overall age-stratified sample, all scales meet the level of internal consistency (.70) necessary for group comparisons (Nunnally, 1978), but only the Physical Functioning Scale meets the conservative level of score reliability recommended for individual-level comparisons, .90 (Nunnally, 1978). Across the age groups, the lowest alpha coefficients occurred for young adults from ages 18 to 24, with .61 on the Pain Scale and .53 on the Social Functioning Scale. This apparently lower score reliability for this age group may be due to restriction of range in the scores; compared to middle-aged and older adults, samples of young adults are likely to exhibit less variability in pain or social functioning. This hypothesis is supported by the progression of standard deviations across age groups (see Table 4.1).

Table 5.1. Reliability Coefficients for RAND–36 HSI Scale and Composite Scores for the Seven Normative Groups

	Age-Based Sample				Age-Stratified Sample		
Scale	18–24	25–44	45–64	≥65	Overall	Female	Male
N	200	200	200	200	500	255	245
Physical Functioning	.89	.86	.92	.89	.90	.90	.91
Role Limitations due to Physical Health Problems	.82	.89	.87	.89	.88	.88	.88
Pain	.61	.76	.77	.74	.76	.73	.74
General Health Perceptions	.80	.82	.83	.77	.81	.81	.80
Emotional Well-Being	.74	.81	.82	.68	.77	.80	.76
Role Limitations due to Emotional Problems	.77	.84	.87	.89	.83	.85	.84
Social Functioning	.53	.70	.82	.70	.71	.73	.66
Energy/Fatigue	.77	.83	.89	.87	.84	.85	.83
Physical Health Composite	.91	.94	.95	.94	.94	.94	.94
Mental Health Composite	.88	.91	.94	.88	.91	.92	.89
Global Health Composite	.93	.95	.96	.94	.95	.95	.95

Note. For the eight scales, Cronbach's (1951) alpha coefficients are reported; for the composites, reliability coefficients were computed with Mosier's (1943) formula.

Reliability of Composite Scores

The reliability coefficients for the Physical Health, Mental Health, and Global Health composite scores are also provided in Table 5.1. The formula for the reliability of composite scores (Mosier, 1943) is a function of the reliability coefficients of the components of the composites, and the dispersions, intercorrelations, and respective weights of those component scores. Reliability coefficients for these composite scores range from a low of .88 to a high of .96 and approach or achieve the level of reliability recommended for individual-level comparison as well as for group comparison at one point in time and across time. For this reason, the three composite scores are recommended for clinical use, such as tracking an individual's health status over time.

Test–Retest Reliability

In addition to measures of internal consistency, reliability of scores can be estimated with respect to how consistently they measure the intended construct over time. A sample of 69 participants was administered the RAND–36 HSI items on two occasions (the median test–retest interval was 7 days). Test–retest correlation coefficients for the eight scales are presented in Table 5.2. The corrected test–retest reliability coefficients for six of the scales exceed .70. For two scales, the coefficients fall below .70: Role Limitations due to Emotional Problems (.59) and Social Functioning (.61). (Calculation of the reliability coefficients for these two scales based on simple summation scoring yielded very similar results.) An examination of the distribution of scores on the Role Limitations due to Emotional Problems Scale indicated that the relatively low coefficient was the result of the small retest sample, the bipolar nature of the response options, and the nonclinical composition of the

sample. Ninety-five percent of the participants obtained the maximum score on this scale on both test and retest occasions. These scores resulted in a restricted range of variance for this sample and yielded a reliability coefficient insufficiently reflective of the true stability of scores on the Role Limitations due to Emotional Problems Scale.

Table 5.2. Test–Retest Reliability Coefficients for RAND–36 HSI Scale and Composite Scores

Scale/Composite	First Testing Mean	First Testing *SD*	Second Testing Mean	Second Testing *SD*	Difference Score[a]	Obtained *r*	Corrected *r*[b]
Physical Functioning	50.00	9.66	49.71	10.38	–0.03	.88	.89
Role Limitations due to Physical Health Problems	49.61	10.51	50.96	9.99	0.13	.79	.77
Pain	50.07	11.21	52.09	9.87	0.20	.78	.72
General Health Perceptions	50.83	10.10	51.55	9.74	0.07	.90	.89
Emotional Well-Being	50.45	10.14	51.94	9.53	0.15	.80	.79
Role Limitations due to Emotional Problems	51.70	6.70	52.36	5.27	0.07	.11	.59
Social Functioning	50.67	10.63	51.17	9.65	0.05	.64	.61
Energy/Fatigue	49.13	10.14	51.22	10.77	0.21	.81	.81
Physical Health Composite	49.62	10.00	50.77	10.20	0.12	.89	.89
Mental Health Composite	50.42	9.62	51.97	9.09	0.16	.81	.82
Global Health Composite	49.97	9.84	51.57	9.57	0.16	.86	.87

Note: $N = 69$. Median retest interval = 7 days.

[a]The difference score is the difference between the mean scale or composite scores at first and second testings, divided by the standard deviation of that scale or composite obtained by the age-stratified sample ($N = 500$).

[b]Correlations were corrected for the variability of scores on the first testing (Guilford & Fruchter, 1978, p. 420) and calculated with the standard deviation of that scale or composite obtained by the age-stratified sample ($N = 500$).

Test–retest reliability coefficients for the three RAND–36 HSI composite scores across the seven norm groups are also provided in Table 5.2. All composite test–retest reliability coefficients consistently exceeded .80.

Validity

The validity of the RAND–36 HSI was investigated at the item, scale, and composite levels and in relation to other measures.

Item Level

The 36 items of the RAND–36 HSI were originally selected to maximize their associations with the long-form MOS scales from which they were derived (Ware & Sherbourne, 1992). Because the items composing each scale have not been altered from their previous versions, the relevance of each item to its respective scale and corresponding criteria of validity have been well documented elsewhere (Hays et al., 1993; McHorney, Ware, et al., 1994; McHorney et al., 1993; Ware & Sherbourne, 1992).

Scale Level

As discussed in Chapter 1, much previous research has been conducted to establish the validity of the scales, which retain their composition of previous versions. Item–total correlations conducted with previous versions (Ware et al., 1993) are evidence of the integrity of the item composition of each scale. Earlier research on the construct-related validity of the scales has demonstrated that the scales interrelate with each other in a way that reflects underlying theoretical constructions of health status. Results of factor analyses of the SF–36 (Hays et al., 1993; Hays, Marshall, et al., 1994) provide strong support for the two-factor model of health status reported for the RAND–36 HSI. Studies of the construct validity of the RAND–36 HSI demonstrate that the two dimensions underlying the structure of health status remain intact with IRT scoring of items. Table 5.3 presents the correlation matrix of *T* scores on the Physical Health and Mental Health scales. (See Chapter 3 for a discussion of the IRT method used for deriving these scores.)

Table 5.3. Intercorrelations of the RAND–36 HSI Scale Scores for the Age-Stratified Sample

	Physical Functioning	Role Limitations due to Physical Health Problems	Pain	General Health Perceptions	Emotional Well-Being	Role Limitations due to Emotional Problems	Social Functioning	Energy/ Fatigue
Physical Functioning		.60	.59	.59	.20	.22	.44	.41
Role Limitations due to Physical Health Problems			.63	.55	.27	.37	.58	.47
Pain				.59	.35	.30	.49	.48
General Health Perceptions					.43	.28	.51	.60
Emotional Well-Being						.49	.57	.61
Role Limitations due to Emotional Problems							.48	.39
Social Functioning								.52
Energy/Fatigue								

Note. $N = 500$.

Composite Level

Although researchers have agreed on the existence of two factors underlying the construct of health status, they have disagreed on the approach for deriving those factors and their related assumptions. Ware, Kosinski, Bayliss, et al. (1995) used a method with orthogonal rotation, assuming nonrelatedness, to derive the physical health and mental health factors, whereas Hays et al. (1993) used an oblique rotation, assuming relatedness of factors. Physical health and mental health composite scores have been derived from these factor analytic studies (Hays et al., 1993; Ware, Kosinski, Bayliss, et al., 1995). In this previous work, factor scores were based on all of the scales whether or not they loaded significantly on the factor and whether or not they were negatively loaded. As discussed previously in Chapter 3, factor analysis of the RAND–36 HSI, with scoring based on an IRT method, employed the principal axis method with oblique rotation and yielded factor pattern loadings similar to those found in the MOS studies. Table 5.4 shows the factor structure found for the overall

age-stratified sample (N = 500). A comparison of the factor structure across age groups (not reported here) confirmed the clustering of scales on the physical health and mental health factors in the same sequence across age groups. As previously discussed, only those scales that loaded significantly on each factor were used to compute composite scores based on that factor. This method treats the factor-based estimates of the underlying physical and mental dimensions of health status in a more differentiated fashion than does the use of all eight scale scores for each factor. On the other hand, the Global Health Composite, which takes into account the variance shared by the two underlying factors, is based on all eight scale scores and, as the name suggests, is a more global reflection of health status.

Table 5.4. Promax Factor Pattern Loadings for the RAND–36 HSI Scales

Scale	Factor 1 Physical Health	Factor 2 Mental Health
Physical Functioning	**.90**	–.21
Role Limitations due to Physical Health Problems	**.79**	.00
Pain	**.76**	.02
General Health Perceptions	**.63**	.18
Emotional Well-Being	–.21	**.95**
Role Limitations due to Emotional Problems	.00	**.59**
Social Functioning	.30	**.53**
Energy/Fatigue	.24	**.58**

Note. N = 500. The factor analysis was based on the scores obtained by the age-stratified sample. Estimated correlation between Factors 1 and 2 was .66.

Previous Studies

Previous research with the SF–36 scales has shown that these scales discriminate better between patients differing in the severity of medical and psychiatric conditions than corresponding long-form scales in the MOS (McHorney et al., 1992). Furthermore, the 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). Results of a study in which a sample of 1,582 residents of Sheffield, England, completed both the SF–36 and the *Nottingham Health Profile* support the convergent and discriminant validity of the SF–36 (Brazier et al., 1992). Studies of the SF–36 have been reported for at least 90 disease conditions, with results demonstrating that health status scores are consistent with chronic conditions (Shiely et al., 1996). To the extent that the RAND–36 HSI retains the scales and item composition of the SF–36, this previous research substantiates item and scale relevancy of the RAND–36 HSI.

Association of RAND–36 HSI With Other Measures

Evidence of the validity of the RAND–36 HSI scales and composites is provided by their relationships with other measures of physical health and mental health status. Various studies were conducted to demonstrate the relatedness of the Physical Health Composite and its scales to other measures of physical health, the relatedness of the Mental Health Composite

and its scales to other measures of mental health, and the correspondence of the Global Health Composite with measures of functioning that may be affected by both physical health and mental health.

Physical Health Measures

Data from the age-stratified sample ($N = 500$) was used to investigate the relationships between the RAND–36 HSI and various other indicators of physical health. Correlations between the RAND–36 HSI scale and composite scores and diagnosed disability, number of physical conditions reported, and ratings of related interference in functioning were computed for the overall sample and the female ($n = 255$) and male ($n = 245$) samples (see Table 5.5). As the data show, the RAND–36 HSI Physical Health Composite and scale scores correlated more highly with disability status, number of physical conditions reported, ratings of related interference in functioning than did the Mental Health Composite and scales. The Physical Health Composite score was a better predictor of these self-reported disability measures than were any of the individual Physical Health scale scores and was better than both the Mental Health and Global Health composite scores, although correlations between all three of the composite scores and self-reported disability measures were significant.

Table 5.5. Correlations Between the RAND–36 HSI and Other Indicators of Physical Health

	Diagnosed Disability[a]			Physical Condition[b]			Total Severity[c]		
Scale/Composite	Overall	Female	Male	Overall	Female	Male	Overall	Female	Male
Physical Functioning	–.39*	–.38*	–.41*	–.29*	–.24*	–.36*	–.42*	–.40*	–.46*
Role Limitations due to Physical Health Problems	–.36*	–.38*	–.37*	–.25*	–.16	–.39*	–.38*	–.32*	–.47*
Pain	–.36*	–.35*	–.39*	–.24*	–.18	–.31*	–.37*	–.33*	–.44*
General Health Perceptions	–.38*	–.32*	–.43*	–.32*	–.28*	–.39*	–.42*	–.38*	–.49*
Emotional Well-Being	–.11	–.10	–.14	–.16	–.08	–.26*	–.20*	–.13	–.30*
Role Limitations due to Emotional Problems	–.15	–.16	–.16	–.19*	–.13	–.29*	–.23*	–.20	–.27*
Social Functioning	–.26*	–.27*	–.26*	–.20*	–.08	–.36*	–.31*	–.21	–.44*
Energy/Fatigue	–.24*	–.23	–.28*	–.20*	–.15	–.28*	–.28*	–.23	–.35*
Physical Health Composite	–.45*	–.42*	–.48*	–.33*	–.25*	–.44*	–.48*	–.42*	–.57*
Mental Health Composite	–.23*	–.22	–.26*	–.23*	–.13	–.37*	–.31*	–.22	–.43*
Global Health Composite	–.36*	–.35*	–.39*	–.31*	–.21	–.43*	–.43*	–.35*	–.53*

Note. Data are based on the results from the age-stratified sample, which consisted of 255 female and 245 male participants.

[a]Diagnosed disability (Yes/No).

[b]Number of Physical Conditions Reported (0–25).

[c]Sum of 25 Physical Symptoms (0–100).

*$p < .0001$. All other coefficients are not significant.

Mental Health Measures

Several studies were conducted to investigate the relationship of the RAND–36 HSI and other measures and indicators of mental health. For the first analysis, participants in the standardization (N = 504) completed the RAND–36 HSI scales and the *Beck Depression Inventory—Second Edition* (BDI–II; Beck, Steer, & Brown, 1996), the *Beck Anxiety Inventory* (BAI; Beck & Steer, 1990), and the *Beck Hopelessness Scale* (BHS; Beck & Steer, 1988). The correlation coefficients between the RAND–36 HSI and these measures are provided in Table 5.6. Although most of the correlations achieved statistical significance, the correlations between the RAND–36 HSI Mental Health Composite and scales and the Beck measures were higher (from –.31 to –.57) than those between the RAND–36 HSI Physical Health Composite and scales and the Beck scales (from –.07 to –.38). The Mental Health Composite score was most highly correlated with the BDI–II (–.57) and BAI (–.54). These findings support the convergent and discriminant validity of the RAND–36 HSI scales and composites and suggest that the Mental Health Composite score is the best indicator among the RAND–36 HSI composites and scale scores of psychological symptoms of depression and anxiety.

Table 5.6. Correlations Between the RAND–36 HSI and the BDI–II, BAI, and BHS for the Age-Stratified Sample

RAND–36 HSI Scale/Composite	BDI–II	BAI	BHS
Physical Functioning	–.07	–.25	–.16
Role Limitations due to Physical Health Problems	–.15	–.32	–.18
Pain	–.20	–.33	–.16
General Health Perceptions	–.24	–.35	–.34
Emotional Well-Being	–.56	–.51	–.54
Role Limitations due to Emotional Problems	–.46	–.40	–.31
Social Functioning	–.46	–.41	–.37
Energy/Fatigue	–.31	–.37	–.34
Physical Health Composite	–.20	–.38	–.26
Mental Health Composite	–.57	–.54	–.52
Global Health Composite	–.46	–.53	–.45

Note. N = 504. Correlations ≥.25 are significant at $p < .01$.

For a second analysis, a sample of 67 respondents from the standardization sample completed the RAND–36 HSI and the *Brief Symptom Inventory* (BSI; Derogatis, 1993). The correlation coefficients between these two measures are provided in Table 5.7. All correlations of .30 or greater were significant at $p < .01$ and indicate that to some extent both physical health and mental health scales of the RAND–36 HSI are sensitive to the symptoms assessed by the BSI. Overall, the scores on the Mental Health scales were more strongly correlated with the BSI scale scores than were the Physical Health scale scores, and the Mental Health Composite score was more strongly correlated with the BSI scores than was the Physical Health Composite score. The Mental Health Composite correlations were substantially greater than the Physical Health Composite correlations with all of the BSI scales except Somatization.

Table 5.7. Correlations Between the RAND–36 HSI and the *Brief Symptom Inventory*

	Brief Symptom Inventory Scales									
RAND–36 HSI Scale/Composite	**SOM**	**O–C**	**I–S**	**DEP**	**ANX**	**HOS**	**PHOB**	**PAR**	**PSY**	**GSI**
Physical Functioning	–.54	–.33	–.28	–.45	–.35	–.24	–.37	–.30	–.27	–.41
Role Limitations due to Physical Health Problems	–.52	–.36	–.27	–.42	–.37	–.30	–.32	–.37	–.35	–.43
Pain	–.39	–.38	–.20	–.35	–.33	–.15	–.27	–.29	–.23	–.35
General Health Perceptions	–.74	–.57	–.42	–.55	–.54	–.53	–.51	–.57	–.52	–.62
Emotional Well-Being	–.60	–.55	–.56	–.69	–.64	–.59	–.56	–.61	–.64	–.66
Role Limitations due to Emotional Problems	–.47	–.65	–.55	–.71	–.56	–.53	–.46	–.52	–.63	–.65
Social Functioning	–.49	–.54	–.46	–.57	–.58	–.45	–.54	–.51	–.52	–.55
Energy/Fatigue	–.41	–.45	–.33	–.42	–.32	–.37	–.20	–.40	–.42	–.43
Physical Health Composite	–.65	–.48	–.35	–.52	–.47	–.36	–.43	–.46	–.41	–.54
Mental Health Composite	–.60	–.63	–.57	–.71	–.64	–.58	–.54	–.62	–.66	–.68
Global Health Composite	–.69	–.64	–.52	–.70	–.63	–.55	–.55	–.61	–.62	–.69

Note. $N = 67$. Correlations ≥.30 are significant at $p < .01$. SOM = Somatization, O–C = Obsessive–Compulsive, I-S = Interpersonal Sensitivity, DEP = Depression, ANX = Anxiety, HOS = Hostility, PHOB = Phobic Anxiety, PAR = Paranoid Ideation, PSY = Psychoticism, GSI = Global Severity Index.

According to the results of these two studies, the RAND–36 HSI Mental Health Composite score is the strongest indicator of psychological symptoms and is perhaps slightly more reflective of depression than of other symptoms. The Global Health Composite score is a slightly less powerful indicator of psychological symptoms except with respect to somatization. Clinical implications of these findings are that the Mental Health Composite score may be used to suggest the presence of psychological symptoms where differentiation of symptoms is not important.

The relationship between perceived health status and life functioning, another basic aspect of mental health, was also investigated. For this analysis, 500 respondents from the standardization sample completed the RAND–36 HSI and the *Behavior and Symptom Identification Scale* (BASIS–32; Eisen, Dill, & Grob, 1994), a measure of mental health

functioning commonly used with psychiatric patients. The correlation coefficients between the two measures are provided in Table 5.8. The Mental Health Composite and Emotional Well-Being Scale scores both correlated very highly with the BASIS–32 Depression and Anxiety Scale score (–.72 and –.70, respectively) and the BASIS–32 Global Overall Average score (–.69 and –.68, respectively). The Mental Health Composite score correlated more highly with the BASIS–32 scale scores than did the Physical Health Composite and scale scores. Also, the BASIS–32 Daily Living/Role Functioning Scale score correlated fairly strongly with the scores on the Mental Health Composite (–.67), the Emotional Well–Being Scale (–.64), the Role Limitations due to Emotional Problems Scale (–.50), and the Social Functioning Scale (–.50). These findings are evidence of the validity of the Mental Health Composite score as a measure of mental health status and life functioning.

Table 5.8. Correlations Between the RAND–36 HSI and the BASIS–32

	BASIS–32 Scales					
RAND–36 HSI Scale/Composite	**RE**	**DE**	**DA**	**IM**	**PS**	**GL**
Physical Functioning	–.13	–.25	–.22	–.11	–.15	–.21
Role Limitations due to Physical Health Problems	–.20	–.35	–.28	–.18	–.13	–.28
Pain	–.29	–.38	–.34	–.27	–.17	–.35
General Health Perceptions	–.30	–.39	–.38	–.29	–.21	–.38
Emotional Well-Being	–.62	–.70	–.64	–.52	–.38	–.68
Role Limitations due to Emotional Problems	–.43	–.52	–.50	–.37	–.33	–.51
Social Functioning	–.42	–.55	–.50	–.39	–.33	–.52
Energy/Fatigue	–.42	–.50	–.49	–.34	–.24	–.48
Physical Health Composite	–.28	–.41	–.36	–.26	–.20	–.36
Mental Health Composite	–.60	–.72	–.67	–.52	–.40	–.69
Global Health Composite	–.51	–.65	–.60	–.45	–.35	–.61

Note. $N = 500$. Correlations ≥.25 are significant at $p < .01$. RE = Relation to Self/Others, DE = Depression/Anxiety, DA = Daily Living/Role Functioning, IM = Impulsive/Addictive Behavior, PS = Psychosis, GL = Global/Overall Average.

The relationship between the RAND–36 HSI and the scales of the *Social Adjustment Scale–Self-Report* (SAS–SR; Weissman & Bothwell, 1976), a measure of social functioning in major life arenas, was also examined. Because both instruments are designed to reflect self-reported limitations in functioning, it was hypothesized that they would correlate highly. The correlation coefficients are reported in Table 5.9. Of the 99 intercorrelations, 77 were significant ($p < .01$). The Global SAS–SR score correlated consistently with the RAND–36 HSI scores, particularly the RAND–36 HSI Mental Health Composite score (–.70). These findings suggest that there is a significant overlap in variance between self-reported health status, particularly mental health status, and self-reported social adjustment. These findings also support the underlying construct in which health status, particularly mental health status, is an important aspect of social functioning and, conversely, social functioning is an important aspect of mental health status.

Table 5.9. Correlations Between the RAND–36 HSI and the SAS–SR

RAND–36 HSI Scale/Composite	SAS–SR Scales								
	Work Status (*n*= 345)	Social & Leisure (*n*= 487)	Family Outside Home (*n*= 480)	Primary Relationship (*n*= 306)	Parental (*n*= 183)	Family Unit (*n*= 383)	Financial (*n*= 485)	Financial & Family (*n*= 486)	Global (*n*= 488)
Physical Functioning	–.25	–.19	–.03	–.25	–.07	–.17	–.13	–.12	–.23
Role Limitations due to Physical Health Problems	–.28	–.24	–.07	–.36	–.22	–.16	–.12	–.11	–.31
Pain	–.30	–.32	–.21	–.37	–.03	–.27	–.24	–.26	–.41
General Health Perceptions	–.26	–.32	–.20	–.36	–.29	–.28	–.25	–.26	–.41
Emotional Well-Being	–.29	–.60	–.46	–.44	–.42	–.53	–.40	–.48	–.68
Role Limitations due to Emotional Problems	–.40	–.38	–.25	–.28	–.42	–.31	–.25	–.25	–.45
Social Functioning	–.31	–.47	–.30	–.37	–.31	–.35	–.29	–.30	–.50
Energy/Fatigue	–.38	–.44	–.31	–.38	–.20	–.39	–.30	–.35	–.54
Physical Health Composite	–.34	–.32	–.16	–.41	–.20	–.26	–.23	–.22	–.41
Mental Health Composite	–.41	–.61	–.43	–.48	–.41	–.51	–.40	–.45	–.70
Global Health Composite	–.43	–.54	–.35	–.49	–.35	–.44	–.36	–.40	–.64

Note. Correlations ≥.25 are significant at $p < .01$.

Global Health Composite

Examinations of the validity of the Global Health Composite must take into account its composition and purpose. Composed of all eight scales, the Global Health Composite taps the common variance of the physical health and mental health factors. Therefore, it is not designed to be as highly related to measures of physical health as the Physical Health Composite or to measures of mental health as the Mental Health Composite. Because the Global Health Composite was formulated to tap both physical and mental aspects of health, it is the second best indicator of both. The findings reported in Tales 5.5 through 5.9 support this hypothesis. The Global Health Composite score was equally or slightly less correlated than the Physical Health Composite score but more highly correlated than the Mental Health Composite score with other measures of physical health (see Table 5.5). Likewise, the Global Health Composite score was equally or slightly less correlated than the Mental Health Composite score but more highly correlated than the Physical Health Composite score with measures of psychological symptoms (See Tables 5.6, 5.7, and 5.8) and social adjustment (Table 5.9). Correlations between the Global Health Composite score and measures of health-care resource utilization also reflect this pattern (discussed later). In summary, these findings are evidence of the validity of the Global Health Composite score as a general health indicator when only one measure is requested and in cases where differentiation of mental and physical aspects of health is not required.

Known-Group Analyses

Discriminant function analyses using each composite score to discriminate selected criterion groups were conducted to determine the sensitivity of the RAND–36 HSI composite scores in differentiating groups of known health status. For the Physical Health Composite, two

criterion groups were chosen on the basis of the number of physical conditions reported by the respondent and the degree of reported interference in functioning related to these conditions. As described in Chapter 2, participants were asked to indicate any of 25 physical conditions that had been diagnosed for them and rate on a scale from 1 to 4 the degree to which each condition interfered with the individual's functioning. The total severity score was the sum of interference ratings across all physical conditions indicated. Criterion groups were composed of those representing the lowest quartile of self-reported limiting physical conditions (n = 200) and those representing the highest quartile of self-reported limiting physical conditions (n = 200). The Physical Health Composite score accurately predicted 60.5% of those reporting a high number of limiting physical conditions and 89.5% of those reporting a low number of limiting physical conditions, for a "hit" rate of 75%. Cut scores based on this classification were Physical Health Composite T scores of ≤46 for those reporting a high number of limiting physical conditions and T scores ≥47 for those reporting a low number of limiting physical conditions.

In the study of the predictive validity of the Mental Health Composite score, two sets of criterion groups were identified according to their scores on the BDI–II (Beck et al., 1996) and the BAI (Beck & Steer, 1990). For the first analysis, the *low-symptom* group was composed of respondents obtaining scores in the minimal range on either the BDI–II or the BAI (n = 372). The *high-symptom* group was composed of respondents obtaining scores in the moderate or severe range on either the BDI–II or the BAI (n = 45). With these criteria, the Mental Health Composite score accurately predicted 66.67% of those in the high-symptom group and 96.51% of those in the low-symptom group, with an overall hit rate of 93.3%. Cut scores based on this classification were Mental Health Composite T scores of ≤38 to identify high-symptom scorers and T scores of ≥39 to identify low-symptom scorers. A second discriminant analysis was performed, with one criterion group defined by scores on either one of the BDI–II or BAI in the mild, moderate, or severe range (n = 109), and the other group defined by scores in the minimal range (n = 372) on both tests. This analysis yielded a cut score of ≥41 on the Mental Health Composite, which yielded a sensitivity of 56.0% and specificity of 93.8%.

Because the Global Health Composite is composed of both physical and mental aspects of health, the criterion for the discriminant validity study could not be specific to physical or mental functioning. For this reason, the criterion for global functioning was defined by the SAS–SR (Weissman & Bothwell, 1976) Global score and the BASIS–32 (Eisen et al., 1994) Daily Living/Role Functioning score. Those scoring in the upper quartile on both scales were designated as the *low-functioning* group (n = 118), and those who scored in the lowest quartile were designated as the *high-functioning* group (n = 118). The Global Health Composite score accurately identified 80.5% of the low-functioning group and 92.4% of the high-functioning group for a hit rate of 86.4%. The cut scores based on this classification were Global Health Composite T scores ≤49 indicating low functioning and T scores ≥50 indicating high functioning. Clinical application of the cut scores derived in these analyses are discussed in Chapter 7.

Utilization of Health-Care Resources

Utilization of health-care resources is an important potential correlate of health status. Specific items about resource utilization were included in the general questionnaire

administered during standardization testing. These questions addressed self-reported frequency of office visits to a physician, sessions with a mental health counselor, admissions to an urgent care clinic or hospital emergency room, as well as hospital stays for physical or emotional problems within the previous 6 months. Reported frequencies were correlated with the RAND–36 HSI scales and composites for the age-stratified sample ($N = 500$). Coefficients are reported in Table 5.10 for the overall, female, and male samples. In summary, 76% of the correlations are significant at the $p < .01$ level or better in the predicted direction, with higher health status being associated with less reported utilization. Moreover, 61% of the correlations are significant at the $p < .001$ level or better, and 45% are significant at the $p < .0001$ level. These results support the hypothesized relationship between self-reported health status and reported recent use of health-care resources. More specifically, frequency of visits to a physician's office yielded significant correlations, ranging from –.43 to –.45, with self-reported physical health status as measured by the Physical Health Composite score. To a lesser extent, frequency of visits to an urgent care facility was correlated with the Physical Health Composite score (ranging from –.24 to –.34), and office visits to a counselor for emotional problems were correlated with the Mental Health Composite score, but primarily for women ($r = –.33$).

Table 5.10. Correlations Between the RAND–36 HSI and Self-Reported Health-Care Resource Utilization

Utilization	RAND–36 HSI Scale/Composite										
	PF	RLP	PA	GHP	EWB	RLE	SF	EF	PHC	MHC	GHC
Number of Visits to Physician in Last 6 Months											
Overall	−.31***	−.40***	−.38***	−.35***	−.16**	−.15**	−.26***	−.24***	−.44***	−.25***	−.37***
Female	−.27***	−.44***	−.35***	−.36***	−.16*	−.11	−.24***	−.21**	−.43***	−.23**	−.36***
Male	−.33***	−.36***	−.42***	−.37***	−.15	−.18*	−.28***	−.25***	−.45***	−.27***	−.38***
Number of Visits to Professional Counselor for Emotional Problems in Last 6 Months											
Overall	−.07	−.11*	−.17**	−.13*	−.26***	−.26***	−.23***	−.17***	−.14**	−.28***	−.25***
Female	−.11	−.15	−.23**	−.25***	−.29***	−.38***	−.26***	−.21**	−.22**	−.33***	−.32***
Male	−.03	−.06	−.10	−.01	−.22**	−.12	−.20*	−.13	−.06	−.22**	−.16*
Number of Visits to Urgent Care Clinic/Hospital Emergency Room in Last 6 Months											
Overall	−.15**	−.24***	−.31***	−.29***	−.14**	−.08	−.20***	−.18***	−.30***	−.20***	−.27***
Female	−.12	−.18*	−.26***	−.27***	−.17*	−.03	−.20**	−.15	−.24***	−.19*	−.24***
Male	−.18*	−.30***	−.36***	−.30***	−.13	−.13	−.21**	−.21**	−.34***	−.22**	−.30***
Number of Hospitalizations for Physical Problems in Last 6 Months											
Overall	−.24***	−.20***	−.13*	−.19***	−.06	−.15**	−.22***	−.05	−.23***	−.14*	−.20***
Female	−.19*	−.09	−.11	−.14	−.04	−.02	−.08	−.07	−.16*	−.06	−.12
Male	−.29***	−.30***	−.16*	−.22**	−.10	−.28***	−.35***	−.04	−.30***	−.21**	−.27***
Number of Hospitalizations for Emotional Problems in Last 6 Months											
Overall	−.10	−.15**	−.18***	−.14*	−.18***	−.20***	−.20***	−.09	−.17***	−.20***	−.21***
Female	−.11	−.15	−.19*	−.18*	−.15*	−.19*	−.21**	−.07	−.19*	−.19*	−.21**
Male	−.07	−.16*	−.19*	−.10	−.24***	−.23**	−.22**	−.13	−.16*	−.26***	−.24**

Note. Data are based on the age-stratified sample (N = 500), with 255 female and 245 male participants.

*p < .01. **p < .001. ***p < .0001.

Chapter 6
Administration and Scoring

The RAND–36 HSI may be administered by paper and pencil with the Question/Answer Sheet or by computer with the *OPTAIO Provider's Desktop* software (The Psychological Corporation, 1998). Question/Answer Sheets and Hand-Scoring Worksheets for the RAND–36 HSI are available from The Psychological Corporation. When the RAND–36 HSI Question/Answer Sheet is completed by a client, the responses may be hand-scored with the RAND–36 HSI Hand-Scoring Worksheet or entered into the OPTAIO program for computerized scoring and reporting. Instructions to the client are the same for both the paper-and-pencil and computerized methods of administration. With either method, points to stress to clients are to report honestly, to report for the preceding 4-week time period, and to not skip any items.

Hand-Scoring for the RAND–36 HSI

Instructions for scoring item responses and for calculating scale and composite scores for the RAND–36 HSI are provided in Appendixes A and B and on the Hand-Scoring Worksheet. RAND–36 HSI scoring is sequential. The IRT weights for the item responses are determined; the scale raw scores are based on these IRT weights and converted to *T* scores; the composite raw scores are based on the scale *T* scores and are converted to *T* scores. Appendixes A and B reflect this sequence. Although instructions for scoring are provided on the Hand-Scoring Worksheet, hand-scoring requires the use of the tables in Appendixes A and B and those in Appendix C, which provides cumulative percentages for each *T* score by scale and composite.

Hand-Scoring for the RAND–12 HSI

The RAND–12 HSI is a short form of the RAND–36 HSI and employs different scoring rules than those for the 36-item version. Detailed rules and the tables for scoring responses to the RAND–12 HSI are provided in Appendix E, along with a description of the instrument's development and intended use. Instructions for calculating estimated *T* scores for the three composites are also provided on the RAND–12 HSI Hand-Scoring Worksheet (available from The Psychological Corporation). Cumulative percentages for each *T* score are provided in Appendix C, organized by scale and composite.

Computer-Based Administration, Scoring, and Interpretation

Consistent with the increased need for expedient testing, processing, and reporting with the use of computers, the RAND–36 HSI is available on OPTAIO (The Psychological Corporation, 1998). The OPTAIO program provides on-screen administration and scoring and generates narrative interpretive reports for the RAND–36 HSI.

The RAND–36 HSI items are presented on-screen. Using a computer keyboard, the client can enter responses to the items, or the clinician can enter the client's responses from the Question/Answer Sheet. Once a client's item responses have been entered, the OPTAIO program computes the item, scale, and composite scores. The program then displays graphs of the client's *T* scores from the single administration (see Figure 6.1). In addition, longitudinal graphs can be generated for comparing a client's scores on one or more RAND–36 HSI composites over time (see Figure 6.2).

The program also generates an interpretive report that describes the client's scores in a narrative format (see Figures 6.3A and 6.3B). The availability of a computer-based analysis and interpretive report is particularly critical for initial diagnosis, determination of level of care, and priority treatment goals for the client.

Figure 6.1. OPTAIO Session Graph of RAND–36 HSI Composite and Scale Scores

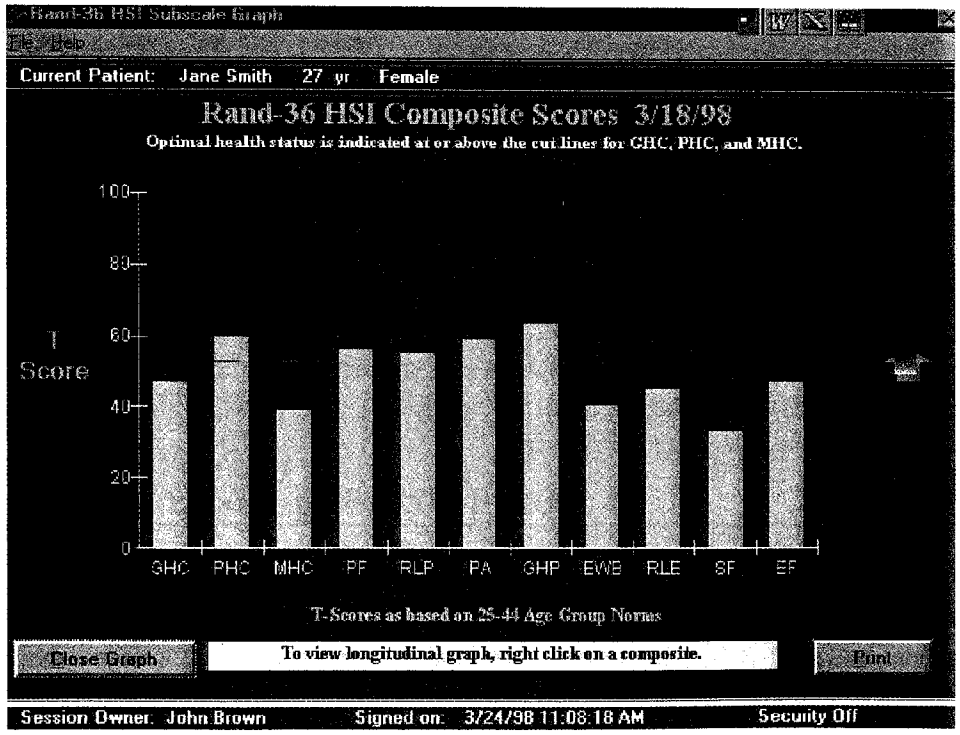


Figure 6.2. OPTAIO Longitudinal Graph of RAND–36 HSI Composite Scores

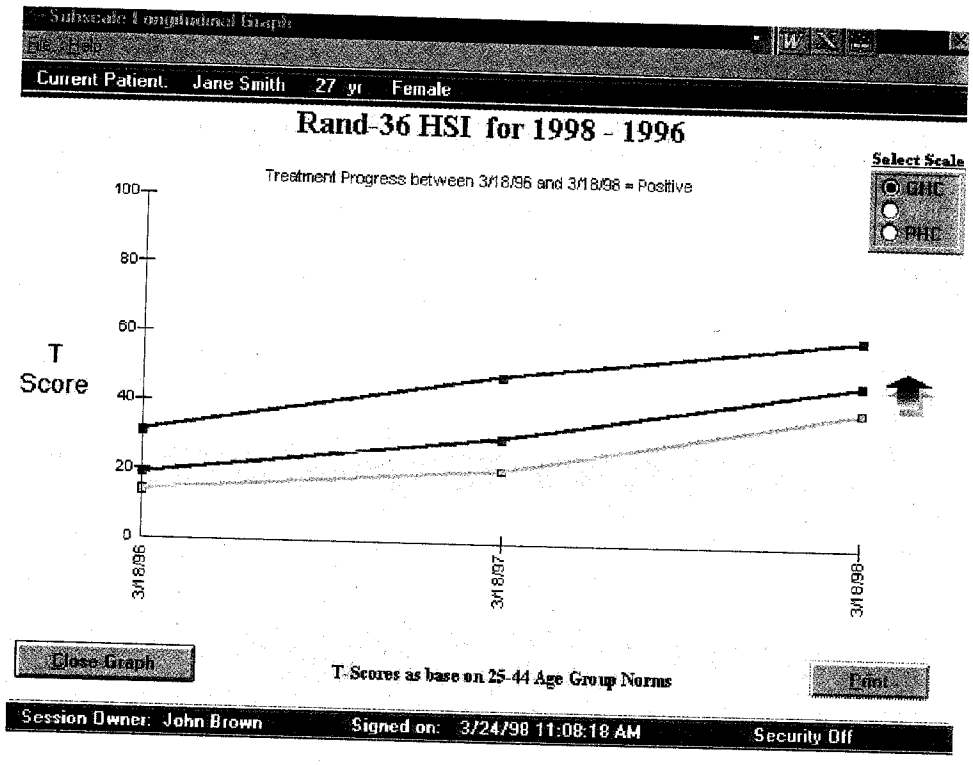


Figure 6.3A. OPTAIO Interpretive Report of RAND–36 HSI Performance (Page 1)

Brown & Associates

Created by : John Brown Patient Unit : OutPatient *RAND-36 HSI Report*

RESULTS

Scale	*T* Score	Cumulative Percentage
Global Health Composite (GHC)	47	33.0
Physical Health Composite (PHC)	60	99.9
Physical Functioning (PF)	56	99.9
Role Limitations due to Physical Health Problems (RLP)	55	99.9
Pain (PA)	59	99.9
General Health Perceptions (GHP)	63	99.9
Mental Health Composite (MHC)	39	17.0
Emotional Well-Being (EWB)	40	18.0
Role Limitations due to Emotional Problems (RLE)	45	24.0
Social Functioning (SF)	33	10.5
Energy/Fatigue (EF)	47	39.0

Missing Items: No missing items.
Invalid Scale(s): No invalid scales.

Smith, Jane T. a 27 year old Female Patient ID : 114467888

Printed on: 3/24/98 Case ID : 00000007 *2 of 4*

Figure 6.3B. OPTAIO Interpretive Report of RAND–36 HSI Performance (Page 2)

Brown & Associates

Created by : John Brown *Patient Unit : OutPatient* *RAND-36 HSI Report*

INTERPRETATION

The following report presents Ms. Smith's scores on the RAND-36 HSI with respect to other adults 25 to 44 years old.

The Global Health Composite (GHC) measures the individual's perception of her physical and mental health and the extent of any health-imposed limitations in daily functioning. The client's score on the GHC is 47, indicating that 33.0% of adults 25 to 44 years old obtained this score or lower. The physical and mental aspects of the GHC are summarized by the Physical Health Composite (PHC) and the Mental Health Composite (MHC). When interpreting results for individual clients, these three composite scores provide the most reliable source of information.

The PHC measures the individual's self-appraisal of her physical functioning, role limitations due to physical problems, bodily pain, and general health perceptions. The client's score on the PHC is 60. When compared to adults 25 to 44 years old, 99.9% obtained this score or lower. The following scores on Ms. Smith's PHC subscales illustrate her physical health status at this point in time. Her Physical Functioning (PF) score is 56, indicating that 99.9% of adults 25 to 44 years old obtained this score or lower. Her score on the Role Limitations due to Physical Health Problems (RLP) scale is 55. This score or lower was obtained by 99.9% of adults 25 to 44 years old. On the Pain (PA) scale her score is 59, indicating that 99.9% of adults 25 to 44 years old achieved this score or lower. Results on the General Health Perceptions (GHP) scale indicate that her score is 63. This score or lower was obtained by 99.9% of adults 25 to 44 years old.

The MHC assesses the individual's emotional well-being, role limitations due to emotional problems, energy and social functioning. The client's score on the MHC is 39. When compared to adults 25 to 44 years old, 17.0% obtained this score or lower. The following scores on Ms. Smith's MHC subscales illustrate her mental health status at this point in time. Her Emotional Well-Being (EWB) score is 40, indicating that 18.0% of adults 25 to 44 years old obtained this score or lower. Her score on the Role Limitations due to Emotional Problems (RLE) scale is 45. This score or lower was obtained by 24.0% of adults 25 to 44 years old. On the Social Functioning (SF) scale her score is 33, indicating that 10.5% of adults 25 to 44 years old achieved this score or lower. Results on the Energy/Fatigue (EF) scale indicate that her score is 47. This score or lower was obtained by 39.0% of adults 25 to 44 years old.

Chapter 7
Interpretation of the RAND–36 HSI Scores

Uses of measures of overall health status similar to the RAND–36 HSI have been based largely on aggregate-level analysis. These uses have included the monitoring of population health (Aaronson et al., 1992; McHorney, Kosinski, & Ware, 1994); estimating the burden of different conditions (Hays, Wells, Sherbourne, Rogers, & Spritzer, 1995; McHorney et al., 1993; Sherbourne et al., 1994); and clinical trials of treatment effects and the monitoring of outcomes in clinical practice, such as quality-assurance evaluation in hospital-based outpatient clinics. For studies involving clinical trials of treatment effects, traditional clinical parameters have been used more than have generic health status measures. However, an evaluation of general functioning when equally efficacious treatments are being compared affords an evaluation of these treatments with respect to their different trade-offs in general functioning and well-being (Fowler et al., 1988). Some researchers have recommended the combined use of general health measures, such as the RAND–36 HSI, and disease-targeted measures, for more breadth and depth in assessing outcomes (Patrick & Deyo, 1989).

As the need for the clinical application of health status instruments has increased, attention has focused on the clinical sensitivity of the instrument. For instance, research has been conducted to determine the minimum size of a group necessary for detecting clinically meaningful differences or changes and for determining the smallest mean group differences in health scale scores that would be considered clinically and socially relevant.

The increased need for instruments that can reliably detect significant change has led to a shift from individual scales, such as those on the SF–36, to composite scores because of the increased reliability and statistical sensitivity. Simply put, the change in scores on an individual scale with a few items must be much greater to be statistically significant due to lower reliability than a change in composite scores, which have more levels of difficulty and higher reliability.

Individual-level interpretations, although less frequent than aggregate-level analysis, have been based on the following approaches:

- With content-referenced interpretation, an individual's scores are referenced to the specific content of different response levels and to the percentage of individuals from a normative sample who have responded at these response levels.

- With criterion-based interpretation, an individual's scores are related to the percentage of people who report a specific criterion, such as inability to work or frequency of utilization of various services. The average amount of change reported in medical outcome studies may be cited.
- With norm-referenced interpretation, an individual's scores are compared to mean scores and designated percentile rankings for various demographic and disease-specific groups.

Determination of Clinical Meaningfulness

General Parameters

Individual-level interpretation of RAND–36 HSI performance relies primarily on the three composite scores. The Physical Health and Mental Health composites represent the physical and mental aspects of functioning and well-being consistent with the identification of predominantly physical or predominantly mental disorders. The Global Health Composite provides a global measure that encompasses overlapping aspects of physical health and mental health for use as a more integrated assessment of health-related limitation. In addition, discrepancies between the Physical Health and Mental Health composite scores can be interpreted. The tracking of client change over time based on the scale scores is not recommended because of insufficient reliability. However, the scale scores can be used to further describe an individual's composite scores at specific points in time.

Missing Item Responses

With the SF–36, a scale score could be computed as long as at least half the items were completed on each scale. This practice, which substitutes estimates for missing data, has been found to be too imprecise for individual profile analysis. For the RAND–36 HSI, only one item per scale may have a missing response, and only three items may have missing responses for the entire inventory. These restrictions maximize individual-level accuracy. Obviously, then, individuals should be encouraged to complete all items. (Computation of estimated IRT weights for missing item responses is described in Appendix A.)

Clinical Cut Scores

The clinical meaningfulness of the scores obtained on the RAND–36 HSI is determined by the integration of three sources of information—*T* scores, cumulative percentages of *T* scores, and criterion-based cut scores. As described previously, *T* scores are a way of standardizing raw scores on scales for each normative group in a way that allows comparability across normative groups. For example, although individuals of different ages may obtain different scale raw scores, their *T* scores may be the same when based on the norms for their own age cohorts. Also, placing the scale scores on the same *T*-score metric ($M = 50$, $SD = 10$) allows comparability across scales that have different raw score means and standard deviations.

On the other hand, individual-level clinical interpretation in terms of *T* scores may be difficult when distributions are strongly skewed. Although 1 *SD* below the mean score of a scale represents a significantly lower score in terms of overall variability of the scale, its clinical

meaning may vary. This variation occurs because the same linear *T* score reflects different cumulative percentages across scales and even across the seven normative groups on the same scale. Also, because of the skewed nature of these scales, the mean *T* score (50) does not necessarily represent the median point, or the *T* score obtained by 50% of the individuals in the normative sample. For these reasons, a cumulative percentage is presented for each *T* score and should be considered along with the *T* score in any clinical interpretation.

Although *T* scores and cumulative percentages provide information about an individual's relative standing in his or her normative group with respect to variance and frequency with which the score occurs, they do not provide information about independently measured health conditions. Previous research has attempted to provide anchor points by providing raw score and *T*-score means of groups of individuals manifesting specific diseases. This approach is helpful particularly if the individual has manifested that particular disease and has demographic characteristics similar to those of the reported group. It should be kept in mind, however, that general perceived-health-status measures are designed to cut across specific disease categories. For this reason, cut scores for the RAND–36 HSI are based on criteria that are not disease-specific but that pertain in a more general way to the dimensions assessed by each composite. The development of local norms by researchers working with disease-specific populations, however, is encouraged.

The guidelines presented in this chapter for evaluating an individual's scores integrate the use of *T* scores, cumulative percentages, and cut scores derived from discriminant function analyses, which were based on the scores obtained by the age-stratified normative group to ensure that results were demographically reflective of the U.S. population. It is emphasized that the cut scores are intended only as guidelines to be carefully considered in the context of local norms, disease-specific information, and circumstances unique to the individual. The following principles were used for determining these cut scores:

- *Low* scores are defined as *T* scores obtained by approximately 20% or less of the normative group.
- The *clinical cut score* is defined as the point that optimally differentiates those in an independently defined criterion group as *low* on the criterion from those defined as *high* on the criterion (see discussion in Chapter 5). The low-criterion group is actually composed of those respondents who obtained *high* scores on the criterion measures of symptom or dysfunction, and the high-criterion group is actually composed of those who obtained *low* scores on those criterion measures. Those *T* scores below the clinical cut score are defined as *below criterion;* those *T* scores above this point are defined as *above criterion.* For the Physical Health Composite, those reporting more than one physical condition, one or more of which were self-reported as limiting functioning to some degree, were compared with those who reported the presence of no limiting physical conditions. For the Mental Health Composite, those who reported moderate to severe psychological symptoms of depression and anxiety were compared with those who reported few or none of these symptoms. For the Global Health Composite, those reporting difficulty in life functioning were compared with those reporting no difficulty in life functioning.
- *High T* scores are defined as those >50 and obtained by approximately 50% or more of the normative nonclinical sample.

Sequential Analysis

The recommended logic for clinical use of the RAND–36 HSI scores is sequential, proceeding from general to specific. This logic is supported by factor analysis, which indicates that health status is an underlying integration of two predominant dimensions, physical and mental, and that these dimensions are, in turn, composed of specific aspects. This logic is also supported by the psychometric reality that the composite scores have higher reliability than the scale scores and, therefore, are more statistically appropriate for individual interpretation. The interpretive guidelines provided here recommend the consideration of individual scale scores only in the context of the statistically stronger composite scores. The following steps for the clinical interpretation of an individual's scores proceed from the general to the more specific.

Suggested Guidelines

Physical Health Composite

Low Scores

A *T* score of 42 or lower was obtained by 19.8% of the age-stratified normative sample. Such low scores, therefore, are relatively infrequent in a nonclinical sample. In addition, a subgroup of the normative sample (n = 124) who had reported diagnosed disabilities obtained a mean *T* score of 39.9. These findings support the conclusion that *T* scores ≤42 suggest that perceived physical health problems are impeding life functioning.

Clinical Cut Score

A *T* score of 47 was determined by discriminant function analyses (see Chapter 5) to be the cut score below which 60.5% of those reporting physical conditions were accurately identified. A *T* score >47 accurately identified 89.5% of those not reporting any physical conditions. A *T* score >47 is higher than the scores obtained by 29% of the age-stratified normative sample. This cut score yielded 39.5% false-negatives (high-symptom respondents who obtained *T* scores >47) and 10.5% false-positives (low-symptom respondents who obtained *T* scores ≤47).

High Scores

A *T* score of 53 or lower was obtained by 54.2% of the age-stratified normative sample. Scores >53 suggest that these individuals are less likely to have physical health problems that impede life functioning.

Mental Health Composite

Low Scores

A *T* score of 38 or lower was obtained by 14.4% or less of the age-stratified normative sample. Results of discriminant function analyses indicated that a *T* score of 38 accurately identified 66.7% of those reporting moderate or severe symptoms of depression and/or anxiety and 96.5% of those reporting minimal symptoms. A *T* score of 38 yielded 33.3% false-negatives and 3.5% false-positives. Therefore, a *T* score of 38 or lower was likely to indicate an individual in the criterion group who was reporting psychological symptoms that might impede life functioning.

Clinical Cut Score

A cut *T* score ≤41 was associated with accurate identification of 56% of those in the age-stratified normative sample who reported from mild to severe depression or anxiety. A *T* score ≥42 was obtained by 93.8% of those who reported minimal symptoms of depression or anxiety. Also, a *T* score ≤41 on the Mental Health Composite was obtained by only 18.6% of the age-stratified sample. This cut score of 41 yielded 44% false-negatives (high-symptom respondents obtaining *T* scores >41) and 6.2% false-positives (low-symptom respondents obtaining *T* scores <42) in predicting scores of depression and anxiety.

High Scores

A *T* score of 53 or lower was obtained by 53.2% of the age-stratified normative sample. Individuals obtaining scores >53 are not likely to perceive mental health problems that impede life functioning.

Physical Health and Mental Health Composite Score Discrepancy

The *T* scores obtained by an individual on the Physical Health and Mental Health composites should be compared to determine their similarity or dissimilarity. The clinical meaning of a discrepancy, although not yet explored empirically, may be reviewed for each respondent in the context of his or her unique circumstances. A discrepancy might indicate that there is a perceived difference between the person's physical health and mental health. Minimally, a significant discrepancy would raise questions about the use of the Global Health Composite score as a single indicator. The following formula is used to calculate the amount of discrepancy required for statistical significance:

$$D = z\sqrt{SE_{M_a}{}^2 + SE_{M_b}{}^2}$$

where D is the difference between the two composite scores, $z = 1.64$ at the 90% level of confidence, and $SE_M = SD\sqrt{1 - r}$, where SD is the standard deviation and r is the reliability coefficient of the respective composite score.

For the age-stratified sample, a discrepancy between the Physical Health and Mental Health composite scores greater than or equal to ±6.35 would be considered statistically significant at the 90% level of significance. However, the clinical significance of a discrepancy may be suggested more appropriately by the infrequency of the discrepancy. For this reason, a discrepancy greater than or equal to ±10 is recommended because less than 22% of the age-stratified sample obtained this discrepancy. This amount also represents 1 *SD* difference between the two composite scores. Appendix D provides tables showing the cumulative percentages of individuals in all seven normative groups obtaining various discrepancies between these composite scores. Table D.1 provides this information for those whose Physical Health Composite score exceeded the Mental Health Composite score; Table D.2 provides the information for those whose Mental Health Composite score exceeded the Physical Health Composite score.

Global Health Composite

The Global Health Composite score reflects the overall perceived health status of the individual and may be viewed as a "global thermometer" of the individual's well-being. The meaning of the Global Health Composite score is more clearly interpretable when the Physical Health and Mental Health composite scores are not significantly discrepant.

Low Scores

A *T* score of 42 or lower was obtained by 21.0% of the age-stratified sample. Such low scores may be considered relatively infrequent and as suggesting that perceived health problems are impeding life functioning.

Clinical Cut Score

Discriminant function analyses (see Chapter 5) identified a cut score of 50, which accurately identified 80.5% of *low functioners* obtaining a *T* score <50 and 92.4% of *high functioners* obtaining *T* scores ≥50. Low functioners were defined as those who were among the highest 25% of scorers on the SAS–SR (Weissman & Bothwell, 1976) Global Scale and the BASIS–32 (Eisen et al., 1994) Daily Living/Role Functioning Scale. High functioners were those who obtained among the lowest 25% of scores on these scales. A *T* score of ≤49 was reported by 37% of the RAND–36 HSI age-stratified normative sample. This cut score yielded 19.5% false-negatives (low functioners who obtained *T* scores ≥50) and 7.6% false-positives (high functioners who obtained *T* scores ≤49).

High Scores

A *T* score of 52 or lower was obtained by a least 50.8% of the age-stratified sample. Scores >52 may be considered sufficiently prevalent in the nonclinical population to indicate normal or better global health functioning.

Scale Scores

Caution should be exercised in interpreting scale scores in isolation. Scale score patterns may be productively considered as interview cues to the extent that the pattern presents clinically relevant questions. For example, "Your answers seem to suggest that while you feel a good deal of pain, it has not limited your daily activities too much. Is that accurate?"

A Final Caution

It should be noted that the recommended cut scores vary across composites. This variance occurs because the composite scores are associated with different cumulative percentage distributions and because different independent measures were used to define criterion groups. The Physical Health and Mental Health cut scores were more symptom-based, and the Global Health cut score was related to self-reported functioning. Therefore, it is possible for an individual's Physical Health and Mental Health composite scores to reflect a minimal likelihood of physical or mental problems and for the Global Health Composite score not to be in the corresponding range of global health status. Because an individual may be relatively symptom-free physically and mentally but still not enjoy good health or be functioning adequately, clinical judgment should be exercised in interpreting all scores.

Longitudinal Tracking of Change

Objective assessment to accurately determine if an individual has benefited from clinical intervention is essential. One method of assessing change that has proved useful is the *Reliable Change Index* developed by Jacobson, Follette, and Revenstorf (1984) and further described by Jacobson and Truax (1991). Briefly, with this approach, the extent of an individual's change in health status is assessed by determining if he or she has approached a normal population's responses and diverged farther from the response pattern of a dysfunctional population.

The method of assessing meaningful clinical change described for the RAND–36 HSI is related to this method but differs in important ways. The analysis recommended here is a two-step process. First, a determination is made of whether the difference in an individual's scores between one point and the next is statistically different. If the difference is statistically significant and positive, then the next step in interpretation is taken. In Step 2, the change is interpreted as *positive but insufficient, favorable, very favorable,* or *optimal.* These clinical anchors are based on the cut-score criteria provided earlier.

The three composite scores should be used for tracking clinical change because these scores are psychometrically reliable enough to allow differentiation of significant change from random fluctuation. The scale scores, on the other hand, do not consistently provide this psychometric reliability. For example, the range of reliability coefficients for the scales included in the Mental Health Composite is from .71 to .84. These values would be associated with 90% confidence intervals ranging from ±8.90 to as much as ±11.55 (on the *T*-score metric). In contrast, the confidence intervals for the Mental Health Composite range from ±5.60 to ±7.79.

The composite scores chosen for comparison are usually sequentially ordered; that is, an earlier score is compared with a more recent one in order to determine whether intervening factors, such as a clinical intervention, might have affected the individual's health status in some way. It is important for clinicians to keep in mind that a change in health status may be attributed to many things, including the course of a particular disease or the recovery process.

The clinician may choose an optimal time interval between administrations of the RAND–36 HSI. This time interval should take into account the nature of the individual's clinical condition and realistic expectations about the course of change. The clinician should also ensure that the normative group used to determine an individual's *T* scores at different times of testing is the same. The following guidelines are recommended.

- Using age-specific norms is preferred for evaluating perceived health status with the RAND–36 HSI because perceived health status is most sensitive to changes in age. For this reason, short-term comparison of an individual's health status should be based on the *T* scores for the individual's specific normative age group. A composite raw score for someone aged 18–24 will mean something quite different from the same raw score for an individual over 65 years old. On the other hand, if a longer time period has elapsed between test administrations, so that an individual shifts from one normative age group to an older one, the individual's *T* scores should be based on the specific age-group norms appropriate at each time of testing. In this way, the individual's perceived health status will be evaluated relative to what is normal for his or her age cohort at the time of testing. Moreover, an individual's perceived health status over the course of a lifetime may be evaluated relative to age. Of course, *T* scores based on age-specific norms should not be compared to *T* scores based on overall or gender-specific norms.
- On the other hand, there may be occasions when the evaluation of an individual's perceived health status should not be controlled for age so that difference due to age-related changes are discernible. In such cases, *T* scores should be based on the overall,

male, or female normative group. Moreover, the same normative group (e.g., female) should be used for all longitudinal comparisons of an individual's status.

Step 1: Determining Statistical Significance of Change

The statistical significance of a change in scores is based on the individual's estimated true score and the confidence interval of that score, which is determined by the standard error of prediction. The estimated true score (ETS) is calculated with the following formula, which corrects for regression to the mean:

$$ETS = M + r(x - M),$$

where M is the mean (50), r is the reliability coefficient of the score, and x is the observed test score. Because the composite scores used in longitudinal tracking are T scores, the mean for all samples is 50 and the standard deviation is 10. The reliability coefficients of the scores are those provided in Table 5.1.

The estimated range of anticipated fluctuation in scores from first to second testings is based on the standard error of prediction (SE_P), which is calculated with the following formula:

$$SE_P = (1.64)\ SD\ \sqrt{1 - r^2},$$

where 1.64 is the z value at the 90% level of confidence, SD is the standard deviation (10), and r is the reliability coefficient of the score.

The standard error of prediction is used to establish the confidence interval around the estimated true score from the first testing. Its value is subtracted from and added to the estimated true score. The resulting values are the lower and upper ends, respectively, of the confidence interval. The standard errors of prediction for the Physical Health, Mental Health, and Global Health composite T scores for all seven normative groups are provided in Table 7.1.

If the individual's composite score from the second testing falls within the confidence interval established for the estimated true score from the first testing, then change is rated as *equivocal.* An equivocal rating means that there has been no statistically significant change in the composite scores. If the score from the second testing is above the confidence interval, that is, greater than the highest score in the range of scores, the change is rated as *positive.* A positive rating indicates that there has been a statistically reliable increase in the composite score, reflecting some improvement in health status. If the score from the second testing is below the confidence interval, then there has been a statistically reliable decrease in the individual's health status as measured by the composite score, and progress is rated as *negative.*

Table 7.1. Standard Errors of Prediction for the Three RAND–36 HSI Composites at 90% Level of Confidence

Normative Group	Physical Health Composite	Mental Health Composite	Global Health Composite
Age-Based Sample			
18–24	±6.80	±7.79	±6.03
25–44	±5.60	±6.80	±5.12
45–64	±5.12	±5.60	±4.59
≥65	±5.60	±7.79	±5.60
Age-Stratified Sample			
Overall	±5.60	±6.80	±5.12
Female	±5.60	±6.43	±5.12
Male	±5.60	±7.48	±5.12

Briefly, the statistical significance of a change in composite scores is determined by

- calculating the individual's estimated true score for the first testing;
- establishing the confidence interval of that score with the standard errors of prediction provided in Table 7.1; and
- rating the change in scores as *equivocal, negative,* or *positive* according to the second score's position relative to the confidence interval.

If an individual's composite score comparison indicates an equivocal (or no) change, the person's condition is described as unchanged from previous assessment, and analysis is discontinued until future testing. In such cases, the clinician should keep in mind that change is sometimes so gradual that it may not be apparent if the time between testings is not sufficiently long for change to have occurred. If change has not occurred or is in a negative direction, the course of treatment should be evaluated and clinical decisions made appropriate to the specific case and appropriate treatment guidelines. If the comparison reveals a significant improvement in health status, then the clinical meaningfulness of the change is evaluated.

Step 2: Evaluating the Clinical Meaningfulness of Change

If the analysis in Step 1 revealed significant change in a positive direction, a subsequent composite *T* score of the individual is addressed in terms of relative health desirability and anticipated goals of possible interventions. Table 7.2 lists the *T*-score ranges associated with determining if a positive change is *positive but insufficient, favorable, very favorable,* or *optimal.* These *T*-score ranges are based on the empirical guidelines suggested earlier in this chapter. Thus, improvement would be rated as *optimal* if the individual's subsequent Global Health Composite *T* score is significantly greater than his or her previous *T* score and greater than or equal to 52.

Table 7.2. *T*-Score Ranges for Evaluating the Clinical Meaningfulness of Change

	Positive but Insufficient	Favorable	Very Favorable	Optimal
Physical Health Composite	≤42	43–46	47–52	≥53
Mental Health Composite	≤38	39–41	42–52	≥53
Global Health Composite	≤42	43–49	50–51	≥52

Summary

Finally, the interpretation of scores and related clinical decisions will depend on the unique clinical circumstances of the individual respondent and the prevailing practice guidelines. For instance, it should be noted that an evaluation of significant change is made independently for each composite because cut scores were based on separate score distributions and criterion groups. It is possible for an individual's improvement in mental health status to be evaluated as favorable but for his or her change in global health not to reflect a similarly favorable status. The clinician must use his or her clinical judgment and consider all known aspects of the individual's circumstances when interpreting the individual's scores.

Appendix A
Computation of Item Scores

Step 1: Scoring Responses to Completed Items

The score for an item is the IRT weight assigned to the selected response option. Item response-option weights for each of the eight RAND–36 HSI scales are provided by scale in Tables A.1–A.8. For example, the response-option weights for all of the items of the Physical Functioning Scale are provided in Table A.1.

- For each item, listed by number in the left column,
  - find the response option in column 2,
  - find the corresponding IRT weight for the response option in column 3,
  - record the IRT weight for that response on the Hand-Scoring Worksheet.
- If there is no response for *one* item on the scale, label that item "missing" on the Hand-Scoring Worksheet and continue.
- If there are *two or more* missing responses on a scale, a raw score cannot be calculated for that scale. Discontinue scoring the responses for that scale and proceed to the next scale.

Step 2: Determining the Score of a Missing Response

- Estimate the IRT weight of a missing response:
  - Add the IRT weights of the responses to the completed items on the scale.
  - Divide the sum of the IRT weights by the number of *completed* items on that scale.
- Compare that result to the range of IRT weights for the item with the missing response (found in the far right column in Tables A.1–A.8):
  - If the result is within the range for the item, use the computed average weight.
  - If the result is below the range, use the low-end weight of the range.
  - If the result is above the range, use the high-end weight of the range.
- Add the estimated IRT weight for the missing response to the sum of IRT weights for the completed items.

Table A.1. Item Response-Option Weights: Physical Functioning Scale (PF)

Item	Response Option	IRT Weight	IRT-Weight Range
3	1	50	50–100
	2	76	
	3	100	
4	1	21	21–56
	2	38	
	3	56	
5	1	17	17–50
	2	33	
	3	50	
6	1	28	28–72
	2	50	
	3	72	
7	1	13	13–44
	2	28	
	3	44	
8	1	22	22–67
	2	44	
	3	67	
9	1	31	31–65
	2	47	
	3	65	
10	1	19	19–51
	2	36	
	3	51	
11	1	8	8–35
	2	21	
	3	35	
12	1	0	0–24
	2	13	
	3	24	

Table A.2. Item Response-Option Weights: Role Limitations due to Physical Health Problems Scale (RLP)

Item	Response Option	IRT Weight	IRT-Weight Range
13	1	0	0–44
	2	44	
14	1	56	56–100
	2	100	
15	1	30	30–74
	2	74	
16	1	26	26–70
	2	70	

Table A.3. Item Response-Option Weights: Pain Scale (PA)

Item	Response Option	IRT Weight	IRT-Weight Range
21	1	100	100–6
	2	78	
	3	51	
	4	34	
	5	15	
	6	6	
22	1	61	61–0
	2	49	
	3	32	
	4	14	
	5	0	

Table A.4. Item Response-Option Weights: General Health Perceptions Scale (GHP)

Item	Response Option	IRT Weight	IRT-Weight Range
1	1	100	100–0
	2	79	
	3	46	
	4	18	
	5	0	
33	1	10	10–66
	2	20	
	3	36	
	4	52	
	5	66	
34	1	89	89–21
	2	67	
	3	46	
	4	33	
	5	21	
35	1	10	10–84
	2	25	
	3	49	
	4	70	
	5	84	
36	1	95	95–26
	2	69	
	3	48	
	4	36	
	5	26	

Table A.5. Item Response-Option Weights: Emotional Well-Being Scale (EWB)

Item	Response Option	IRT Weight	IRT-Weight Range
24	1	8	8–60
	2	15	
	3	23	
	4	31	
	5	46	
	6	60	
25	1	5	5–46
	2	13	
	3	21	
	4	27	
	5	36	
	6	46	
26	1	100	100–5
	2	73	
	3	45	
	4	32	
	5	17	
	6	5	
28	1	4	4–64
	2	13	
	3	22	
	4	31	
	5	47	
	6	64	
30	1	91	91–0
	2	64	
	3	37	
	4	26	
	5	12	
	6	0	

Table A.6. Item Response-Option Weights: Role Limitations due to Emotional Problems Scale (RLE)

Item	Response Option	IRT Weight	IRT-Weight Range
17	1	0	0–41
	2	41	
18	1	59	59–100
	2	100	
19	1	13	13–54
	2	54	

Table A.7. Item Response-Option Weights: Social Functioning Scale (SF)

Item	Response Option	IRT Weight	IRT-Weight Range
20	1	88	88–0
	2	70	
	3	44	
	4	18	
	5	0	
32	1	27	27–100
	2	38	
	3	53	
	4	79	
	5	100	

Table A.8. Item Response-Option Weights: Energy/Fatigue Scale (EF)

Item	Response Option	IRT Weight	IRT-Weight Range
23	1	100	100–4
	2	76	
	3	47	
	4	31	
	5	16	
	6	4	
27	1	90	90–1
	2	70	
	3	45	
	4	30	
	5	13	
	6	1	
29	1	0	0–73
	2	10	
	3	22	
	4	35	
	5	56	
	6	73	
31	1	0	0–91
	2	11	
	3	25	
	4	40	
	5	71	
	6	91	

Appendix B

Computation of Scale and Composite Scores for the RAND–36 HSI

Scales

Raw Scores

The scale *raw score* is the sum of the IRT weights of the responses to the items contributing to the scale. If *one* item of a scale is missing a response, the IRT weight for that item is estimated by the method described in Appendix A. If *two or more* items of a scale are missing responses, the scale raw score cannot be computed.

The items contributing to each scale and the raw-score range for each scale are provided in Table B.1.

T Scores

For each scale, the raw score is converted to a *T* score. Tables B.2–B.9 provide the *T*-score equivalents of raw scores by scale for the seven normative groups.

Convert the scale raw score to a *T* score:

- Find the specific table for the scale (e.g., Table B.2 is the conversion table for the Physical Functioning Scale).
- In the column for the selected normative group, find the range of raw scores that includes the raw score computed for that scale.
- Read across to either the far-left or far-right column for the *T* score and enter that score on the Hand-Scoring Worksheet.

Composites

Raw Scores

A composite raw score is based on the *T* scores of the scales contributing to that composite. A composite score cannot be computed if one of the contributing scale scores is missing. That is, if two or more items of any contributing scale are missing responses, that composite score cannot be calculated. In addition, the Global Health Composite score cannot be calculated if either or both scores on the Physical Health Composite and the Mental Health Composite cannot be calculated.

The calculation of each composite raw score consists of multiplying each contributing scale *T* score by its beta weight and adding the resulting products. Note that the beta weight for a scale depends on the composite score being calculated. For example, for the Physical Functioning Scale (PF), the beta weight is .2712 for calculating the Physical Health Composite raw score but .1103 for calculating the Global Health Composite raw score.

The Physical Health Composite (PHC) raw score is calculated as follows:

.2712 x Physical Functioning *T* score
+ .3487 x Role Limitations due to Physical Health Problems *T* score
+ .2755 x Pain *T* score
+ .2931 x General Health Perceptions *T* score
= Physical Health Composite Raw Score (rounded to nearest whole number)

The Mental Health Composite (MHC) raw score is calculated as follows:

.4333 x Emotional Well-Being *T* score
+ .1623 x Role Limitations due to Emotional Problems *T* score
+ .3112 x Social Functioning *T* score
+ .3136 x Energy/Fatigue *T* score
= Mental Health Composite Raw Score (rounded to nearest whole number)

The Global Health Composite (GHC) raw score is calculated as follows:

.1103 x Physical Functioning *T* score
+ .1775 x Role Limitations due to Physical Health Problems *T* score
+ .1446 x Pain *T* score
+ .1621 x General Health Perceptions *T* score
+ .2651 x Emotional Well-Being *T* score
+ .1013 x Role Limitations due to Emotional Problems *T* score
+ .2016 x Social Functioning *T* score
+ .1911 x Energy/Fatigue *T* score
= Global Health Composite Raw Score (rounded to nearest whole number)

T Scores

For each composite, the raw score is converted to a *T* score. Tables B.10–B.12 provide the *T*-score equivalents of raw scores by composite for the seven normative groups.

Convert the composite raw score to a *T* score:

- Find the appropriate table for the composite:
  - Table B.10: Physical Health Composite
  - Table B.11: Mental Health Composite
  - Table B.12: Global Health Composite
- In the column for the selected normative group, find the composite raw score.
- Read across to either the far-left or far-right column for the *T* score and enter that score on the Hand-Scoring Worksheet.

Table B.1. Raw-Score Ranges for the Scales

Scale	Items Contributing to the Scale	Raw-Score Range (Sum of IRT Weights)
Physical Functioning (PF)		
	3	209–564
	4	
	5	
	6	
	7	
	8	
	9	
	10	
	11	
	12	
Role Limitations due to Physical Health Problems (RLP)		
	13	112–288
	14	
	15	
	16	
Pain (PA)		
	21	6–161
	22	
General Health Perceptions (GHP)		
	1	67–434
	33	
	34	
	35	
	36	
Emotional Well-Being (EWB)		
	24	22–361
	25	
	26	
	28	
	30	
Role Limitations due to Emotional Problems (RLE)		
	17	72–195
	18	
	19	
Social Functioning (SF)		
	20	27–188
	32	
Energy/ Fatigue (EF)		
	23	5–354
	27	
	29	
	31	

Table B.2. *T*-Score Equivalents of Raw Scores: Physical Functioning Scale (PF)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
1	209–255	209–214						1
2	256-261	215-220						2
3	262–267	221–226						3
4	268–273	227–233						4
5	274–278	234–239						5
6	279–284	240–246						6
7	285–290	247–252						7
8	291–296	253–258						8
9	297–302	259–265						9
10	303–307	266–271						10
11	308–313	272–278						11
12	314–319	279–284						12
13	320–325	285–290						13
14	326–330	291–297						14
15	331–336	298–303						15
16	337–342	304–309			209		209–216	16
17	343–348	310–316			210–218	209–212	217–224	17
18	349–354	317–322			219–227	213–221	225–233	18
19	355–359	323–329			228–235	222–229	234–241	19
20	360–365	330–335			236–244	230–238	242–250	20
21	366–371	336–341	209–215		245–252	239–246	251–258	21
22	372–377	342–348	216–225		253–261	247–255	259–267	22
23	378–382	349–354	226–234		262–269	256–264	268–275	23
24	383–388	355–361	235–243	209	270–278	265–272	276–284	24
25	389–394	362–367	244–253	210–218	279–287	273–281	285–292	25
26	395–400	368–373	254–262	219–227	288–295	282–290	293–301	26
27	401–406	374–380	263–271	228–236	296–304	291–298	302–309	27
28	407–411	381–386	272–281	237–245	305–312	299–307	310–318	28
29	412–417	387–392	282–290	246–254	313–321	308–316	319–326	29
30	418–423	393–399	291–299	255–263	322–329	317–324	327–335	30
31	424–429	400–405	300–309	264–273	330–338	325–333	336–343	31
32	430–434	406–412	310–318	274–282	339–347	334–342	344–352	32
33	435–440	413–418	319–327	283–291	348–355	343–350	353–360	33
34	441–446	419–424	328–336	292–300	356–364	351–359	361–369	34
35	447–452	425–431	337–346	301–309	365–372	360–368	370–377	35
36	453–458	432–437	347–355	310–318	373–381	369–376	378–386	36
37	459–463	438–444	356–364	319–327	382–390	377–385	387–394	37
38	464–469	445–450	365–374	328–336	391–398	386–393	395–403	38
39	470–475	451–456	375–383	337–346	399–407	394–402	404–411	39

Table B.2. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
40	476–481	457–463	384–392	347–355	408–415	403–411	412–420	40
41	482–486	464–469	393–402	356–364	416–424	412–419	421–429	41
42	487–492	470–476	403–411	365–373	425–432	420–428	430–437	42
43	493–498	477–482	412–420	374–382	433–441	429–437	438–446	43
44	499–504	483–488	421–429	383–391	442–450	438–445	447–454	44
45	505–510	489–495	430–439	392–400	451–458	446–454	455–463	45
46	511–515	496–501	440–448	401–409	459–467	455–463	464–471	46
47	516–521	502–507	449–457	410–419	468–475	464–471	472–480	47
48	522–527	508–514	458–467	420–428	476–484	472–480	481–488	48
49	528–533	515–520	468–476	429–437	485–492	481–489	489–497	49
50	534–538	521–527	477–485	438–446	493–501	490–497	498–505	50
51	539–544	528–533	486–495	447–455	502–510	498–506	506–514	51
52	545–550	534–539	496–504	456–464	511–518	507–514	515–522	52
53	551–556	540–546	505–513	465–473	519–527	515–523	523–531	53
54	557–562	547–552	514–522	474–482	528–535	524–532	532–539	54
55	563–564	553–559	523–532	483–491	536–544	533–540	540–548	55
56		560–564	533–541	492–501	545–553	541–549	549–556	56
57			542–550	502–510	554–561	550–558	557–564	57
58			551–560	511–519	562–564	559–564		58
59			561–564	520–528				59
60				529–537				60
61				538–546				61
62				547–555				62
63				556–564				63

Table B.3. *T*-Score Equivalents of Raw Scores: Role Limitations due to Physical Health Problems Scale (RLP)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
14	112–114							14
15	115–119							15
16	120–123							16
17	124–127							17
18	128–132							18
19	133–136							19
20	137–140							20
21	141–145							21
22	146–149	112						22
23	150–153	113–118						23
24	154–158	119–123						24
25	159–162	124–129					112	25
26	163–166	130–134			112–117	112–117	113–118	26
27	167–171	135–140	112–114		118–123	118–122	119–124	27
28	172–175	141–145	115–120		124–129	123–128	125–130	28
29	176–180	146–150	121–126		130–135	129–134	131–136	29
30	181–184	151–156	127–132		136–141	135–140	137–141	30
31	185–188	157–161	133–138		142–147	141–146	142–147	31
32	189–193	162–167	139–144		148–153	147–152	148–153	32
33	194–197	168–172	145–150	112–117	154–158	153–158	154–159	33
34	198–201	173–177	151–156	118–124	159–164	159–163	160–165	34
35	202–206	178–183	157–162	125–131	165–170	164–169	166–171	35
36	207–210	184–188	163–168	132–138	171–176	170–175	172–177	36
37	211–214	189–194	169–174	139–145	177–182	176–181	178–182	37
38	215–219	195–199	175–180	146–152	183–188	182–187	183–188	38
39	220–223	200–205	181–186	153–159	189–194	188–193	189–194	39
40	224–227	206–210	187–192	160–166	195–199	194–199	195–200	40
41	228–232	211–215	193–198	167–173	200–205	200–205	201–206	41
42	233–236	216–221	199–204	174–180	206–211	206–210	207–212	42
43	237–240	222–226	205–211	181–187	212–217	211–216	213–218	43
44	241–245	227–232	212–217	188–194	218–223	217–222	219–223	44
45	246–249	233–237	218–223	195–201	224–229	223–228	224–229	45
46	250–254	238–243	224–229	202–208	230–235	229–234	230–235	46
47	255–258	244–248	230–235	209–215	236–240	235–240	236–241	47
48	259–262	249–253	236–241	216–222	241–246	241–246	242–247	48
49	263–267	254–259	242–247	223–229	247–252	247–251	248–253	49
50	268–271	260–264	248–253	230–236	253–258	252–257	254–259	50
51	272–275	265–270	254–259	237–243	259–264	258–263	260–265	51

Table B.3. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
52	276–280	271–275	260–265	244–251	265–270	264–269	266–270	52
53	281–284	276–281	266–271	252–258	271–276	270–275	271–276	53
54	285–288	282–286	272–277	259–265	277–281	276–281	277–282	54
55		287–288	278–283	266–272	282–287	282–287	283–288	55
56			284–288	273–279	288	288		56
57				280–286				57
58				287–288				58

Table B.4. *T*-Score Equivalents of Raw Scores: Pain Scale (PA)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
10	6–7							10
11	8–10							11
12	11–13							12
13	14–16							13
14	17–19							14
15	20–23							15
16	24–26	6						16
17	27–29	7–9						17
18	30–32	10–13					6–7	18
19	33–35	14–16					8–11	19
20	36–39	17–20			6–9		12–14	20
21	40–42	21–24			10–13	6–8	15–18	21
22	43–45	25–27	6–8	6–9	14–17	9–12	19–22	22
23	46–48	28–31	9–12	10–13	18–21	13–16	23–26	23
24	49–51	32–34	13–16	14–17	22–25	17–20	27–29	24
25	52–55	35–38	17–20	18–20	26–28	21–24	30–33	25
26	56–58	39–42	21–24	21–24	29–32	25–28	34–37	26
27	59–61	43–45	25–29	25–28	33–36	29–32	38–40	27
28	62–64	46–49	30–33	29–32	37–40	33–36	41–44	28
29	65–67	50–53	34–37	33–36	41–44	37–40	45–48	29
30	68–71	54–56	38–41	37–40	45–48	41–44	49–52	30
31	72–74	57–60	42–45	41–43	49–52	45–48	53–55	31
32	75–77	61–63	46–49	44–47	53–55	49–52	56–59	32
33	78–80	64–67	50–53	48–51	56–59	53–56	60–63	33
34	81–84	68–71	54–57	52–55	60–63	57–60	64–66	34
35	85–87	72–74	58–61	56–59	64–67	61–64	67–70	35
36	88–90	75–78	62–65	60–63	68–71	65–68	71–74	36
37	91–93	79–81	66–69	64–66	72–75	69–72	75–78	37
38	94–96	82–85	70–73	67–70	76–79	73–76	79–81	38
39	97–100	86–89	74–78	71–74	80–82	77–80	82–85	39
40	101–103	90–92	79–82	75–78	83–86	81–84	86–89	40
41	104–106	93–96	83–86	79–82	87–90	85–88	90–92	41
42	107–109	97–100	87–90	83–86	91–94	89–92	93–96	42
43	110–112	101–103	91–94	87–90	95–98	93–96	97–100	43
44	113–116	104–107	95–98	91–93	99–102	97–100	101–104	44
45	117–119	108–110	99–102	94–97	103–106	101–104	105–107	45
46	120–122	111–114	103–106	98–101	107–110	105–108	108–111	46
47	123–125	115–118	107–110	102–105	111–113	109–112	112–115	47
48	126–128	119–121	111–114	106–109	114–117	113–116	116–118	48
49	129–132	122–125	115–118	110–113	118–121	117–120	119–122	49

Table B.4. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
50	133–135	126–128	119–122	114–116	122–125	121–124	123–126	50
51	136–138	129–132	123–127	117–120	126–129	125–128	127–130	51
52	139–141	133–136	128–131	121–124	130–133	129–132	131–133	52
53	142–144	137–139	132–135	125–128	134–137	133–136	134–137	53
54	145–148	140–143	136–139	129–132	138–140	137–140	138–141	54
55	149–151	144–147	140–143	133–136	141–144	141–144	142–144	55
56	152–154	148–150	144–147	137–139	145–148	145–148	145–148	56
57	155–157	151–154	148–151	140–143	149–152	149–152	149–152	57
58	158–160	155–157	152–155	144–147	153–156	153–156	153–156	58
59	161	158–161	156–159	148–151	157–160	157–160	157–159	59
60			160–161	152–155	161	161	160–161	60
61				156–159				61
62				160–161				62

Table B.5. *T*-Score Equivalents of Raw Scores: General Health Perceptions Scale (GHP)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
16	67							16
17	68–75							17
18	76–83	67–71				67–70		18
19	84–91	72–79				71–78		19
20	92–99	80–87		67	67–74	79–86		20
21	100–107	88–95		68–75	75–83	87–94	67–71	21
22	108–115	96–104	67–68	76–83	84–91	95–102	72–80	22
23	116–123	105–112	69–77	84–90	92–99	103–110	81–88	23
24	124–131	113–120	78–85	91–98	100–108	111–118	89–97	24
25	132–139	121–128	86–94	99–106	109–116	119–126	98–105	25
26	140–146	129–136	95–103	107–114	117–124	127–134	106–114	26
27	147–154	137–145	104–112	115–122	125–132	135–142	115–122	27
28	155–162	146–153	113–120	123–129	133–141	143–150	123–131	28
29	163–170	154–161	121–129	130–137	142–149	151–158	132–139	29
30	171–178	162–169	130–138	138–145	150–157	159–166	140–148	30
31	179–186	170–177	139–146	146–153	158–166	167–174	149–157	31
32	187–194	178–186	147–155	154–161	167–174	175–182	158–165	32
33	195–202	187–194	156–164	162–168	175–182	183–190	166–174	33
34	203–210	195–202	165–172	169–176	183–190	191–198	175–182	34
35	211–218	203–210	173–181	177–184	191–199	199–206	183–191	35
36	219–226	211–218	182–190	185–192	200–207	207–214	192–199	36
37	227–234	219–226	191–198	193–199	208–215	215–222	200–208	37
38	235–242	227–235	199–207	200–207	216–223	223–231	209–216	38
39	243–249	236–243	208–216	208–215	224–232	232–239	217–225	39
40	250–257	244–251	217–225	216–223	233–240	240–247	226–233	40
41	258–265	252–259	226–233	224–231	241–248	248–255	234–242	41
42	266–273	260–267	234–242	232–238	249–257	256–263	243–250	42
43	274–281	268–276	243–251	239–246	258–265	264–271	251–259	43
44	282–289	277–284	252–259	247–254	266–273	272–279	260–268	44
45	290–297	285–292	260–268	255–262	274–281	280–287	269–276	45
46	298–305	293–300	269–277	263–270	282–290	288–295	277–285	46
47	306–313	301–308	278–285	271–277	291–298	296–303	286–293	47
48	314–321	309–317	286–294	278–285	299–306	304–311	294–302	48
49	322–329	318–325	295–303	286–293	307–315	312–319	303–310	49
50	330–337	326–333	304–311	294–301	316–323	320–327	311–319	50
51	338–344	334–341	312–320	302–309	324–331	328–335	320–327	51
52	345–352	342–349	321–329	310–316	332–339	336–343	328–336	52
53	353–360	350–358	330–338	317–324	340–348	344–351	337–344	53
54	361–368	359–366	339–346	325–332	349–356	352–359	345–353	54

Table B.5. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
55	369–376	367–374	347–355	333–340	357–364	360–367	354–361	55
56	377–384	375–382	356–364	341–347	365–373	368–375	362–370	56
57	385–392	383–390	365–372	348–355	374–381	376–383	371–379	57
58	393–400	391–399	373–381	356–363	382–389	384–391	380–387	58
59	401–408	400–407	382–390	364–371	390–397	392–399	388–396	59
60	409–416	408–415	391–398	372–379	398–406	400–407	397–404	60
61	417–424	416–423	399–407	380–386	407–414	408–415	405–413	61
62	425–432	424–431	408–416	387–394	415–422	416–423	414–421	62
63	433–434	432–434	417–424	395–402	423–431	424–431	422–430	63
64			425–433	403–410	432–434	432–434	431–434	64
65			434	411–418				65
66				419–425				66
67				426–433				67
68				434				68

Table B.6. *T*-Score Equivalents of Raw Scores: Emotional Well-Being Scale (EWB)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
7				22–24				7
8				25–30				8
9				31–35				9
10				36–41				10
11				42–47				11
12				48–53				12
13				54–59				13
14				60–65				14
15				66–71			22–27	15
16				72–77	22–27	22	28–33	16
17				78–83	28–34	23–29	34–40	17
18	22–23		22–23	84–88	35–41	30–35	41–47	18
19	24–30	22–25	24–30	89–94	42–47	36–42	48–53	19
20	31–37	26–32	31–37	95–100	48–54	43–49	54–60	20
21	38–43	33–39	38–45	101–106	55–61	50–56	61–67	21
22	44–50	40–46	46–52	107–112	62–67	57–62	68–73	22
23	51–57	47–53	53–59	113–118	68–74	63–69	74–80	23
24	58–64	54–60	60–67	119–124	75–81	70–76	81–86	24
25	65–70	61–67	68–74	125–130	82–87	77–82	87–93	25
26	71–77	68–74	75–81	131–135	88–94	83–89	94–100	26
27	78–84	75–81	82–89	136–141	95–101	90–96	101–106	27
28	85–91	82–88	90–96	142–147	102–108	97–103	107–113	28
29	92–97	89–95	97–103	148–153	109–114	104–109	114–120	29
30	98–104	96–102	104–111	154–159	115–121	110–116	121–126	30
31	105–111	103–109	112–118	160–165	122–128	117–123	127–133	31
32	112–118	110–116	119–125	166–171	129–134	124–129	134–139	32
33	119–124	117–123	126–133	172–177	135–141	130–136	140–146	33
34	125–131	124–130	134–140	178–183	142–148	137–143	147–153	34
35	132–138	131–137	141–147	184–188	149–154	144–150	154–159	35
36	139–145	138–144	148–154	189–194	155–161	151–156	160–166	36
37	146–151	145–151	155–162	195–200	162–168	157–163	167–173	37
38	152–158	152–158	163–169	201–206	169–174	164–170	174–179	38
39	159–165	159–165	170–176	207–212	175–181	171–177	180–186	39
40	166–171	166–172	177–184	213–218	182–188	178–183	187–192	40
41	172–178	173–179	185–191	219–224	189–194	184–190	193–199	41
42	179–185	180–186	192–198	225–230	195–201	191–197	200–206	42
43	186–192	187–193	199–206	231–236	202–208	198–203	207–212	43
44	193–198	194–200	207–213	237–241	209–214	204–210	213–219	44

Table B.6. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
45	199–205	201–207	214–220	242–247	215–221	211–217	220–226	45
46	206–212	208–214	221–228	248–253	222–228	218–224	227–232	46
47	213–219	215–221	229–235	254–259	229–234	225–230	233–239	47
48	220–225	222–228	236–242	260–265	235–241	231–237	240–245	48
49	226–232	229–235	243–250	266–271	242–248	238–244	246–252	49
50	233–239	236–242	251–257	272–277	249–254	245–250	253–259	50
51	240–246	243–249	258–264	278–283	255–261	251–257	260–265	51
52	247–252	250–256	265–272	284–288	262–268	258–264	266–272	52
53	253–259	257–263	273–279	289–294	269–275	265–271	273–279	53
54	260–266	264–270	280–286	295–300	276–281	272–277	280–285	54
55	267–273	271–277	287–293	301–306	282–288	278–284	286–292	55
56	274–279	278–285	294–301	307–312	289–295	285–291	293–298	56
57	280–286	286–292	302–308	313–318	296–301	292–297	299–305	57
58	287–293	293–299	309–315	319–324	302–308	298–304	306–312	58
59	294–299	300–306	316–323	325–330	309–315	305–311	313–318	59
60	300–306	307–313	324–330	331–336	316–321	312–318	319–325	60
61	307–313	314–320	331–337	337–341	322–328	319–324	326–332	61
62	314–320	321–327	338–345	342–347	329–335	325–331	333–338	62
63	321–326	328–334	346–352	348–353	336–341	332–338	339–345	63
64	327–333	335–341	353–359	354–359	342–348	339–344	346–351	64
65	334–340	342–348	360–361	360–361	349–355	345–351	352–358	65
66	341–347	349–355			356–361	352–358	359–361	66
67	348–353	356–361				359–361		67
68	354–360							68
69	361							69

Table B.7. *T*-Score Equivalents of Raw Scores: Role Limitations due to Emotional Problems Scale (RLE)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
17							72–73	17
18							74–76	18
19	72–73		72		72–73		77–80	19
20	74–77		73–76		74–77		81–83	20
21	78–80		77–79		78–80	72–74	84–86	21
22	81–84		80–83		81–83	75–78	87–90	22
23	85–87		84–86		84–87	79–81	91–93	23
24	88–90	72–74	87–90		88–90	82–85	94–96	24
25	91–94	75–78	91–93	72–74	91–94	86–89	97–100	25
26	95–97	79–82	94–97	75–79	95–97	90–92	101–103	26
27	98–101	83–86	98–100	80–83	98–101	93–96	104–106	27
28	102–104	87–90	101–104	84–87	102–104	97–99	107–110	28
29	105–108	91–94	105–108	88–91	105–108	100–103	111–113	29
30	109–111	95–98	109–111	92–95	109–111	104–107	114–116	30
31	112–114	99–101	112–115	96–99	112–115	108–110	117–120	31
32	115–118	102–105	116–118	100–103	116–118	111–114	121–123	32
33	119–121	106–109	119–122	104–107	119–122	115–118	124–126	33
34	122–125	110–113	123–125	108–111	123–125	119–121	127–130	34
35	126–128	114–117	126–129	112–115	126–129	122–125	131–133	35
36	129–132	118–121	130–132	116–119	130–132	126–128	134–136	36
37	133–135	122–125	133–136	120–123	133–136	129–132	137–140	37
38	136–139	126–129	137–139	124–128	137–139	133–136	141–143	38
39	140–142	130–133	140–143	129–132	140–143	137–139	144–146	39
40	143–145	134–137	144–146	133–136	144–146	140–143	147–150	40
41	146–149	138–141	147–150	137–140	147–150	144–147	151–153	41
42	150–152	142–145	151–153	141–144	151–153	148–150	154–156	42
43	153–156	146–149	154–157	145–148	154–157	151–154	157–160	43
44	157–159	150–153	158–161	149–152	158–160	155–157	161–163	44
45	160–163	154–157	162–164	153–156	161–164	158–161	164–166	45
46	164–166	158–161	165–168	157–160	165–167	162–165	167–170	46
47	167–170	162–165	169–171	161–164	168–171	166–168	171–173	47
48	171–173	166–168	172–175	165–168	172–174	169–172	174–176	48
49	174–176	169–172	176–178	169–173	175–178	173–175	177–180	49
50	177–180	173–176	179–182	174–177	179–181	176–179	181–183	50
51	181–183	177–180	183–185	178–181	182–185	180–183	184–186	51
52	184–187	181–184	186–189	182–185	186–188	184–186	187–190	52
53	188–190	185–188	190–192	186–189	189–192	187–190	191–193	53
54	191–194	189–192	193–195	190–193	193–195	191–194	194–195	54
55	195	193–195		194–195		195		55

Table B.8. *T*-Score Equivalents of Raw Scores: Social Functioning Scale (SF)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
7				27				7
8	27			28–31				8
9	28–30			32–34				9
10	31–33			35–37				10
11	34–37			38–41		27		11
12	38–40			42–44	27–28	28–30		12
13	41–43			45–47	29–32	31–34	27–29	13
14	44–47			48–50	33–35	35–37	30–33	14
15	48–50	27		51–54	36–39	38–41	34–37	15
16	51–53	28–31	27–30	55–57	40–43	42–44	38–41	16
17	54–57	32–35	31–34	58–60	44–46	45–48	42–44	17
18	58–60	36–39	35–38	61–63	47–50	49–52	45–48	18
19	61–63	40–43	39–42	64–67	51–53	53–55	49–52	19
20	64–67	44–46	43–46	68–70	54–57	56–59	53–55	20
21	68–70	47–50	47–50	71–73	58–61	60–62	56–59	21
22	71–73	51–54	51–54	74–76	62–64	63–66	60–63	22
23	74–76	55–58	55–58	77–80	65–68	67–69	64–66	23
24	77–80	59–62	59–62	81–83	69–72	70–73	67–70	24
25	81–83	63–66	63–66	84–86	73–75	74–76	71–74	25
26	84–86	67–70	67–70	87–90	76–79	77–80	75–77	26
27	87–90	71–73	71–74	91–93	80–82	81–84	78–81	27
28	91–93	74–77	75–78	94–96	83–86	85–87	82–85	28
29	94–96	78–81	79–81	97–99	87–90	88–91	86–89	29
30	97–100	82–85	82–85	100–103	91–93	92–94	90–92	30
31	101–103	86–89	86–89	104–106	94–97	95–98	93–96	31
32	104–106	90–93	90–93	107–109	98–100	99–101	97–100	32
33	107–110	94–97	94–97	110–112	101–104	102–105	101–103	33
34	111–113	98–100	98–101	113–116	105–108	106–108	104–107	34
35	114–116	101–104	102–105	117–119	109–111	109–112	108–111	35
36	117–120	105–108	106–109	120–122	112–115	113–115	112–114	36
37	121–123	109–112	110–113	123–125	116–119	116–119	115–118	37
38	124–126	113–116	114–117	126–129	120–122	120–123	119–122	38
39	127–130	117–120	118–121	130–132	123–126	124–126	123–125	39
40	131–133	121–124	122–125	133–135	127–129	127–130	126–129	40
41	134–136	125–127	126–129	136–139	130–133	131–133	130–133	41
42	137–140	128–131	130–133	140–142	134–137	134–137	134–137	42
43	141–143	132–135	134–137	143–145	138–140	138–140	138–140	43
44	144–146	136–139	138–140	146–148	141–144	141–144	141–144	44
45	147–150	140–143	141–144	149–152	145–148	145–147	145–148	45
46	151–153	144–147	145–148	153–155	149–151	148–151	149–151	46

Continued on next page.

Table B.8. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
47	154–156	148–150	149–152	156–158	152–155	152–154	152–155	47
48	157–160	151–154	153–156	159–161	156–158	155–158	156–159	48
49	161–163	155–158	157–160	162–165	159–162	159–162	160–162	49
50	164–166	159–162	161–164	166–168	163–166	163–165	163–166	50
51	167–170	163–166	165–168	169–171	167–169	166–169	167–170	51
52	171–173	167–170	169–172	172–174	170–173	170–172	171–174	52
53	174–176	171–174	173–176	175–178	174–176	173–176	175–177	53
54	177–180	175–177	177–180	179–181	177–180	177–179	178–181	54
55	181–183	178–181	181–184	182–184	181–184	180–183	182–185	55
56	184–186	182–185	185–188	185–188	185–187	184–186	186–188	56
57	187–188	186–188			188	187–188		57

Table B.9. *T*-Score Equivalents of Raw Scores: Energy/Fatigue Scale (EF)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
25		5–6					5–10	25
26	5–12	7–13		5–12	5–11	5	11–18	26
27	13–19	14–21	5–8	13–20	12–18	6–13	19–26	27
28	20–27	22–28	9–17	21–28	19–26	14–21	27–34	28
29	28–35	29–36	18–25	29–36	27–34	22–28	35–42	29
30	36–43	37–43	26–34	37–45	35–42	29–36	43–49	30
31	44–50	44–51	35–42	46–53	43–50	37–44	50–57	31
32	51–58	52–58	43–50	54–61	51–58	45–52	58–65	32
33	59–66	59–66	51–59	62–69	59–66	53–60	66–73	33
34	67–74	67–73	60–67	70–77	67–74	61–68	74–81	34
35	75–82	74–81	68–75	78–85	75–82	69–75	82–89	35
36	83–89	82–89	76–84	86–93	83–90	76–83	90–97	36
37	90–97	90–96	85–92	94–101	91–97	84–91	98–105	37
38	98–105	97–104	93–101	102–109	98–105	92–99	106–113	38
39	106–113	105–111	102–109	110–118	106–113	100–107	114–121	39
40	114–120	112–119	110–117	119–126	114–121	108–114	122–129	40
41	121–128	120–126	118–126	127–134	122–129	115–122	130–137	41
42	129–136	127–134	127–134	135–142	130–137	123–130	138–144	42
43	137–144	135–141	135–142	143–150	138–145	131–138	145–152	43
44	145–152	142–149	143–151	151–158	146–153	139–146	153–160	44
45	153–159	150–156	152–159	159–166	154–161	147–154	161–168	45
46	160–167	157–164	160–167	167–174	162–168	155–161	169–176	46
47	168–175	165–172	168–176	175–183	169–176	162–169	177–184	47
48	176–183	173–179	177–184	184–191	177–184	170–177	185–192	48
49	184–190	180–187	185–193	192–199	185–192	178–185	193–200	49
50	191–198	188–194	194–201	200–207	193–200	186–193	201–208	50
51	199–206	195–202	202–209	208–215	201–208	194–200	209–216	51
52	207–214	203–209	210–218	216–223	209–216	201–208	217–224	52
53	215–222	210–217	219–226	224–231	217–224	209–216	225–231	53
54	223–229	218–224	227–234	232–239	225–232	217–224	232–239	54
55	230–237	225–232	235–243	240–248	233–240	225–232	240–247	55
56	238–245	233–239	244–251	249–256	241–247	233–240	248–255	56
57	246–253	240–247	252–260	257–264	248–255	241–247	256–263	57
58	254–260	248–254	261–268	265–272	256–263	248–255	264–271	58
59	261–268	255–262	269–276	273–280	264–271	256–263	272–279	59
60	269–276	263–270	277–285	281–288	272–279	264–271	280–287	60
61	277–284	271–277	286–293	289–296	280–287	272–279	288–295	61
62	285–292	278–285	294–301	297–304	288–295	280–287	296–303	62
63	293–299	286–292	302–310	305–312	296–303	288–294	304–311	63
64	300–307	293–300	311–318	313–321	304–311	295–302	312–319	64
65	308–315	301–307	319–327	322–329	312–319	303–310	320–326	65

Continued on next page.

Table B.9. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
66	316–323	308–315	328–335	330–337	320–326	311–318	327–334	66
67	324–330	316–322	336–343	338–345	327–334	319–326	335–342	67
68	331–338	323–330	344–352	346–353	335–342	327–333	343–350	68
69	339–346	331–337	353–354	354	343–350	334–341	351–354	69
70	347–354	338–345			351–354	342–349		70
71		346–352				350–354		71
72		353–354						72

Table B.10. *T*-Score Equivalents of Raw Scores: Physical Health Composite (PHC)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
1	13–15							1
2	16							2
3	17							3
4	18							4
5	19							5
6	20							6
7	21							7
8	22	18						8
9	—	19						9
10	23	20						10
11	24	21						11
12	25	22						12
13	26	23						13
14	27	24					24	14
15	28	25			25	25	25	15
16	29	26			26	26	26	16
17	30	27			27	27	27	17
18	31	28			28	28	28	18
19	—	29	28		29	29	29	19
20	32	30	29	30	30	30	30	20
21	33	31	30	31	31	31	31	21
22	34	32	31	32	32	32	32	22
23	35	33	32	33	33	33	33	23
24	36	34	33	34	34	34	34	24
25	37	35	34	35	35	35	35	25
26	38	36	35	36	36	36	36	26
27	39	37	36	37	37	37	37	27
28	40	38	37	38	38	38	38	28
29	—	39	38	39	39	39	39	29
30	41	40	39	40	40	40	40	30
31	42	41	40	41	41	41	41	31
32	43	42	41	42	42	42	42	32
33	44	43	42	43	43	43	43	33
34	45	44	43	44	44	44	—	34
35	46	45	44	45	45	45	44	35
36	47	46	45	46	46	46	45	36
37	48	47	46	47	47	47	46	37
38	49	48	47	48	48	48	47	38
39	—	49	48	49	49	49	48	39

Continued on next page.

Table B.10. ***(Continued)***

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
40	50	50	49	50	50	50	49	40
41	51	51	50	51	51	51	50	41
42	52	52	51	52	52	52	51	42
43	53	53	52	53	53	53	52	43
44	54	54	53	54	54	54	53	44
45	55	55	54	55	55	55	54	45
46	56	—	55	56	56	56	55	46
47	57	56	56	57	57	57	56	47
48	58	57	57	58	58	58	57	48
49	59	58	58	—	59	59	58	49
50	—	59	59	59	60	60	59	50
51	60	60	60	60	61	61	60	51
52	61	61	61	61	62	—	61	52
53	62	62	62	62	63	62	62	53
54	63	63	63	63	64	63	63	54
55	64	64	64–65	64	65	64	64	55
56	65	65	66	65	66	65	65	56
57	66	66	67	66	67	66	66	57
58	67	67	68	67	68	67	67	58
59	68	68	69	68	69	68	68	59
60		69	70	69	70	69	69	60
61			71	70	71	70	70	61
62				71		71		62
63				72				63
64				73				64
65				74				65

Table B.11. *T*-Score Equivalents of Raw Scores: Mental Health Composite (MHC)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
4				17				4
5				18				5
6				19				6
7				20				7
8				21				8
9				22			21	9
10				23			22	10
11	22			24	22	22	23	11
12	23			25	23	23	24	12
13	24			26	24	24	25	13
14	25	25	24	27	25	25	26	14
15	26	26	25	28	26	26	27	15
16	27	27	26	29	27	27	28	16
17	28	28	27	30	28	28	29	17
18	29	29	28	31	29	29	30	18
19	30	30	29	32	30	30	31	19
20	31	31	30	33	31	31	32	20
21	32	32	31	—	32	32	33	21
22	33	33	32	34	33	33	34	22
23	34	34	33	35	34	34	35	23
24	35	35	34	36	35	35	36	24
25	36	36	35	37	36	36	37	25
26	37	37	36	38	37	37	38	26
27	38	38	37–38	39	38	38	39	27
28	39	39	39	40	39	39	40	28
29	40	40	40	41	40	40	41	29
30	41	41	41	42	41	41	42	30
31	42	42	42	43	42	42	43	31
32	43	43	43	44	43	43	44	32
33	44	44	44	45	44	44	45	33
34	45	45	45	46	45	45	—	34
35	46	46	46	47	46	46	46	35
36	47	47	47	48	47	47	47	36
37	48	48	48	49	48	48	48	37
38	49	49	49	50	49	49	49	38
39	50	50	50	51	50	50	50	39
40	51	51	51	52	51	51	51	40
41	52	52	52	—	52	52	52	41
42	53	53	53	53	53	53	53	42

Continued on next page.

Table B.11. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
43	54	54	54	54	54	54	54	43
44	55	55	55	55	55	55	55	44
45	56	56	56	56	56	56	56	45
46	57	57	57	57	57	57	57	46
47	58	58	58	58	58	58	58	47
48	59	59	59	59	59	59	59	48
49	60	60	60	60	60	60	60	49
50	61	61	61	61	61	61	61	50
51	62	62	62	62	62	62	62	51
52	63	63	63	63	63	63	63	52
53	64	64	64	64	64	64	64	53
54	65	65	65	65	65	65	65	54
55	66	66	66	66	66	66	66	55
56	67	67	67	67	67	67	67	56
57	68	68	68	68	68	68	68	57
58	69	69	69	69	69	69	69	58
59	70	70	70	70	70	70	70	59
60	71	71	71	—	71	71	71	60
61	72	72	72	71	72	72	72	61
62	73	73	73	72	73	73	73	62
63	74	74	74	73	74	74	74	63
64	75	75	75	74	75	75	75	64
65	76	76	76	75	76	76	76	65
66	77	77		76	77	77		66
67	78	78				78		67
68	79							68

Table B.12. *T*-Score Equivalents of Raw Scores: Global Health Composite (GHC)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
1	20							1
2	21							2
3	22							3
4	23							4
5	24							5
6	25							6
7	26							7
8	27	25			26	26	25	8
9	28	26		26	27	27	26	9
10	29	27		27	28	28	27	10
11	30	28		28	29	29	28	11
12	31	29		29	30	30	29	12
13	32	30	29	30	31	31	30	13
14	33	31	30	31	32	32	31	14
15	34	32	31	32	33	33	32	15
16	35	33	32	33	34	34	33	16
17	—	34	33	34	35	35	34	17
18	36	35	34	35	36	36	35	18
19	37	36	35	36	37	37	36	19
20	38	37	36	37	38	38	37	20
21	39	38	37	38	39	39	38	21
22	40	39	38	39	40	40	39	22
23	41	40	39	40	41	41	40	23
24	42	41	40	41	42	42	41	24
25	43	42	41	42	43	43	42	25
26	44	43	42	43	44	44	43	26
27	45	44	43	44–45	45	45	44	27
28	46	45	44	46	46	46	45	28
29	47	46–47	45–46	47	47	47	46	29
30	48	48	47	48	48	48	47	30
31	49	49	48	49	49	49	48	31
32	50	50	49	50	50	50	49–50	32
33	51	51	50	51	51	51	51	33
34	52	52	51	52	52	52	52	34
35	53	53	52	53	53	53	53	35
36	54	54	53	54	54	54	54	36
37	55	55	54	55	55	55	55	37
38	56	56	55	56	56	56	56	38
39	57	57	56	57	57	57	57	39

Continued on next page.

Table B.12. (*Continued*)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
40	58	58	57	58	58	58	58	40
41	59	59	58	59	59	59	59	41
42	60	60	59	60	60	60	60	42
43	61	61	60	61	61	61	61	43
44	62	62	61	62	62	62	62	44
45	63	63	62	63	63	63	63	45
46	64	64	63	64	64	64	64	46
47	65	65	64	65	65	65	65	47
48	66	66	65–66	66	66	66	66	48
49	67	67	67	67	67	67	67	49
50	68	68	68	68	68	68	68	50
51	69	69	69	69	69	69	69	51
52	70	70	70	70	70	70	70	52
53	71	71	71	71	71	71	71	53
54	72	72	72	72	72	72	72	54
55	73	73	73	73	73	73	73	55
56	74	74	74	74	74	74	74	56
57	75	75	75	75	75	75	75	57
58	76	76	76	76	76	76	76	58
59	—	77	77	77	77	77	77	59
60	77	78	78	78	78	78	78	60
61	78	79	79	79	79	79	79	61
62	79	80	80	80	80	80	80	62
63	80	81	81	81	81	81	81	63
64	81	82	82	82	82	82	82	64
65	82	83	83	83	83	83	83	65
66	83			84		84		66
67				85				67

Appendix C
T Scores Obtained by Cumulative Percentages of the Normative Samples

Table C.1. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Physical Functioning Scale (PF)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
1	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	1
2	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	2
3	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	3
4	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	4
5	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	5
6	1.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	6
7	2.0	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	7
8	2.0	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	8
9	2.0	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	9
10	2.0	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	10
11	2.0	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	11
12	2.0	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	12
13	2.0	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	13
14	2.0	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	14
15	2.0	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	15
16	2.0	1.0	<0.1	<0.1	0.6	<0.1	0.4	16
17	2.0	2.0	<0.1	<0.1	0.6	0.8	0.8	17
18	2.0	2.5	<0.1	<0.1	0.8	0.8	1.2	18
19	2.0	2.5	<0.1	<0.1	1.0	0.8	1.2	19
20	2.0	2.5	<0.1	<0.1	1.0	0.8	1.2	20
21	3.0	3.0	<0.1	<0.1	1.4	0.8	1.6	21
22	3.0	3.0	0.5	<0.1	2.0	2.0	2.0	22
23	3.0	3.5	1.0	<0.1	2.4	2.4	2.0	23
24	3.5	3.5	2.0	1.0	2.8	3.5	2.9	24
25	3.5	5.5	2.5	1.0	3.6	3.9	4.5	25
26	4.5	5.5	3.0	1.0	4.2	3.9	4.9	26

Continued on next page.

Table C.1. (*Continued*)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
27	4.5	5.5	4.0	1.5	5.2	4.7	5.7	27
28	5.0	6.5	5.0	2.0	6.0	5.5	6.9	28
29	6.0	7.0	6.5	2.5	7.6	7.5	7.8	29
30	6.5	7.0	8.0	3.0	7.8	7.5	8.2	30
31	7.0	7.0	9.0	5.5	8.4	7.8	9.0	31
32	8.0	7.0	11.5	6.5	9.0	9.0	9.4	32
33	8.5	8.5	12.0	8.5	9.6	9.4	10.2	33
34	8.5	10.0	12.0	11.0	10.8	9.8	11.8	34
35	9.0	10.5	14.5	13.0	12.0	11.8	11.8	35
36	9.0	11.0	14.5	15.5	12.2	12.2	13.1	36
37	9.5	11.5	15.0	16.5	13.2	12.2	13.5	37
38	9.5	12.0	15.5	17.0	14.0	13.3	13.5	38
39	10.5	13.0	15.5	17.5	14.8	14.9	14.7	39
40	10.5	14.0	17.0	19.0	16.0	16.9	15.9	40
41	12.0	14.5	18.0	21.0	18.4	18.4	17.6	41
42	12.0	16.5	20.5	23.0	19.8	20.8	19.2	42
43	13.5	18.0	21.5	24.5	21.6	23.1	20.0	43
44	14.0	18.5	22.5	27.0	23.2	25.1	21.6	44
45	15.5	21.0	24.5	29.0	25.0	25.9	23.7	45
46	16.5	24.0	27.0	31.5	26.6	28.6	24.9	46
47	19.5	25.0	31.0	35.5	28.2	29.0	26.5	47
48	20.5	26.5	34.5	38.5	30.6	32.2	27.3	48
49	20.5	32.0	35.0	42.5	33.2	35.7	32.7	49
50	20.5	34.5	37.0	45.0	37.8	38.0	35.9	50
51	36.5	34.5	39.0	49.5	38.6	40.8	38.0	51
52	38.0	34.5	45.0	53.0	44.2	42.4	44.1	52
53	38.0	46.0	46.0	55.0	47.4	49.4	44.5	53
54	38.0	47.0	56.0	59.0	47.4	50.2	44.5	54
55	99.9	47.0	58.0	61.5	60.4	60.8	58.8	55
56	99.9	99.9	71.5	70.0	61.6	64.3	58.8	56
57	99.9	99.9	72.5	71.5	61.6	64.3	99.9	57
58	99.9	99.9	72.5	78.0	99.9	99.9	99.9	58
59	99.9	99.9	99.9	80.5	99.9	99.9	99.9	59
60	99.9	99.9	99.9	80.5	99.9	99.9	99.9	60
61	99.9	99.9	99.9	91.5	99.9	99.9	99.9	61
62	99.9	99.9	99.9	92.0	99.9	99.9	99.9	62
63	99.9	99.9	99.9	99.9	99.9	99.9	99.9	63

Table C.2. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Role Limitations due to Physical Health Problems Scale (RLP)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
≤21	3.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	≤21
22	3.5	8.0	<0.1	<0.1	<0.1	<0.1	<0.1	22
23	3.5	8.0	<0.1	<0.1	<0.1	<0.1	<0.1	23
24	7.0	8.0	<0.1	<0.1	<0.1	<0.1	<0.1	24
25	7.0	8.0	<0.1	<0.1	<0.1	<0.1	9.8	25
26	7.0	8.0	<0.1	<0.1	9.4	9.0	9.8	26
27	7.0	8.0	9.0	<0.1	9.4	9.0	9.8	27
28	7.0	8.0	9.0	<0.1	9.4	9.0	9.8	28
29	7.0	8.0	9.0	<0.1	9.4	9.0	9.8	29
30	7.0	11.5	9.0	<0.1	9.4	9.0	9.8	30
31	7.0	11.5	9.0	<0.1	9.4	9.0	9.8	31
32	7.0	11.5	9.0	<0.1	9.4	9.0	9.8	32
33	7.0	11.5	9.0	18.5	15.2	16.1	14.3	33
34	12.0	11.5	18.5	18.5	15.2	16.1	14.3	34
35	12.0	11.5	18.5	18.5	15.2	16.1	14.3	35
36	12.0	11.5	18.5	18.5	15.2	16.1	14.3	36
37	12.0	11.5	18.5	18.5	15.2	16.1	14.3	37
38	12.0	11.5	18.5	18.5	15.2	16.1	14.3	38
39	12.0	16.5	18.5	26.0	15.2	16.1	14.3	39
40	12.0	16.5	18.5	26.0	15.2	16.1	20.4	40
41	12.0	16.5	18.5	26.0	20.2	20.0	20.4	41
42	12.0	16.5	24.0	26.0	20.2	20.0	20.4	42
43	12.0	16.5	24.0	26.0	20.2	20.0	20.4	43
44	20.0	16.5	24.0	26.0	20.2	20.0	20.4	44
45	20.0	16.5	24.0	34.5	20.2	20.0	20.4	45
46	20.0	16.5	24.0	34.5	20.2	20.0	20.4	46
47	20.0	23.0	24.0	34.5	20.2	20.0	20.4	47
48	20.0	23.0	24.0	34.5	29.0	30.2	27.8	48
49	20.0	23.0	34.0	34.5	29.0	30.2	27.8	49
50	20.0	23.0	34.0	34.5	29.0	30.2	27.8	50
51	20.0	23.0	34.0	34.5	29.0	30.2	27.8	51
52	20.0	23.0	34.0	45.0	29.0	30.2	27.8	52
53	20.0	23.0	34.0	45.0	29.0	30.2	27.8	53
54	99.9	23.0	34.0	45.0	29.0	30.2	27.8	54
55	99.9	99.9	34.0	45.0	29.0	30.2	99.9	55
56	99.9	99.9	99.9	45.0	99.9	99.9	99.9	56
57	99.9	99.9	99.9	45.0	99.9	99.9	99.9	57
58	99.9	99.9	99.9	99.9	99.9	99.9	99.9	58

Table C.3. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Pain Scale (PA)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
≤12	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	≤12
13	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	13
14	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	14
15	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	15
16	1.0	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	16
17	1.5	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	17
18	1.5	1.0	<0.1	<0.1	<0.1	<0.1	0.4	18
19	1.5	1.0	<0.1	<0.1	<0.1	<0.1	0.4	19
20	1.5	2.0	<0.1	<0.1	0.6	<0.1	0.4	20
21	1.5	2.0	<0.1	<0.1	0.6	0.8	0.4	21
22	1.5	2.0	0.5	<0.1	0.8	0.8	1.2	22
23	3.0	4.0	0.5	<0.1	1.2	1.2	1.2	23
24	3.0	4.0	0.5	<0.1	1.2	1.2	2.9	24
25	3.0	4.0	0.5	<0.1	1.2	1.2	2.9	25
26	3.0	4.0	0.5	<0.1	4.6	1.2	2.9	26
27	3.0	4.0	5.5	<0.1	4.6	6.3	3.3	27
28	3.5	7.0	5.5	3.0	5.0	6.3	3.3	28
29	6.0	7.5	5.5	3.0	5.0	6.7	7.3	29
30	6.0	7.5	6.5	3.0	7.8	6.7	7.8	30
31	6.0	7.5	6.5	3.0	8.0	8.2	7.8	31
32	6.5	7.5	9.5	4.0	8.0	8.2	7.8	32
33	6.5	8.5	9.5	8.0	8.0	8.2	7.8	33
34	12.0	8.5	9.5	8.0	8.0	8.2	12.2	34
35	12.0	8.5	9.5	8.0	13.4	9.0	12.2	35
36	12.0	8.5	9.5	8.0	13.4	14.5	12.2	36
37	12.0	8.5	16.0	20.0	13.4	14.5	12.2	37
38	14.0	13.5	16.0	20.0	13.4	14.5	12.2	38
39	17.0	13.5	16.0	20.0	13.4	14.5	18.8	39
40	17.0	13.5	16.0	20.0	19.0	19.2	18.8	40
41	17.0	14.5	22.0	20.0	19.0	19.2	18.8	41
42	17.0	25.5	22.0	28.5	19.0	19.2	19.6	42
43	24.5	25.5	22.0	28.5	21.0	22.4	28.6	43
44	24.5	25.5	24.5	28.5	29.8	31.0	28.6	44
45	24.5	25.5	34.5	32.0	29.8	31.0	28.6	45
46	24.5	29.0	34.5	41.0	29.8	31.0	28.6	46
47	24.5	29.0	34.5	41.0	33.6	34.9	32.2	47
48	34.0	29.0	39.0	41.0	33.6	34.9	32.2	48
49	34.0	29.0	39.0	45.0	33.6	34.9	32.2	49
50	34.0	42.0	39.0	45.0	33.6	34.9	32.2	50

Table C.3. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
51	34.0	42.0	48.5	45.0	45.6	46.7	44.5	51
52	60.0	42.0	48.5	45.0	45.6	46.7	44.5	52
53	60.0	69.0	48.5	56.0	45.6	46.7	44.5	53
54	60.0	69.0	67.0	56.0	69.2	68.2	70.2	54
55	60.5	69.0	67.0	56.0	69.2	68.2	70.2	55
56	60.5	69.0	67.0	77.5	69.2	68.2	70.2	56
57	60.5	69.0	67.0	77.5	69.2	68.2	70.2	57
58	60.5	69.0	67.0	77.5	69.2	68.2	70.2	58
59	99.9	99.9	67.0	78.5	69.2	68.2	70.2	59
60	99.9	99.9	99.9	78.5	99.9	99.9	99.9	60
61	99.9	99.9	99.9	78.5	99.9	99.9	99.9	61
62	99.9	99.9	99.9	99.9	99.9	99.9	99.9	62

Table C.4. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: General Health Perceptions Scale (GHP)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
≤17	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	≤17
18	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	18
19	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	19
20	0.5	0.5	<0.1	<0.1	<0.1	0.4	<0.1	20
21	0.5	1.0	<0.1	<0.1	0.2	0.4	<0.1	21
22	0.5	1.0	0.5	<0.1	0.2	0.4	<0.1	22
23	1.0	2.5	0.5	<0.1	0.4	2.0	<0.1	23
24	1.5	2.5	0.5	<0.1	0.8	2.0	0.4	24
25	1.5	2.5	0.5	<0.1	1.6	2.4	0.4	25
26	3.0	3.0	0.5	0.5	1.8	3.1	1.2	26
27	3.0	4.0	1.5	0.5	2.6	3.5	1.2	27
28	3.5	4.0	1.5	1.0	3.8	3.5	1.6	28
29	4.0	4.0	2.5	1.5	4.0	3.5	4.5	29
30	4.0	5.5	4.5	2.5	4.4	4.3	4.5	30
31	6.0	5.5	5.0	2.5	6.0	5.1	5.3	31
32	6.5	6.0	7.0	3.5	6.8	6.7	6.5	32
33	7.0	7.0	8.0	4.5	7.8	8.2	8.6	33
34	8.0	7.5	9.5	6.5	9.2	9.0	9.0	34
35	8.0	8.5	12.0	6.5	10.4	10.2	10.2	35
36	9.5	10.0	13.5	9.5	11.4	11.0	11.8	36
37	11.5	11.5	13.5	13.5	12.6	12.2	12.7	37
38	13.5	12.5	15.5	16.0	13.8	13.7	13.9	38
39	16.0	15.0	17.0	19.5	15.4	14.9	15.5	39
40	17.0	17.0	18.0	22.0	17.6	16.5	18.0	40
41	20.5	18.5	20.5	24.0	18.4	18.0	19.6	41
42	23.0	22.0	22.5	24.5	20.2	19.6	20.0	42
43	24.0	24.5	24.0	26.0	23.2	22.4	22.9	43
44	27.0	28.5	26.0	29.0	25.6	25.9	27.3	44
45	30.0	29.0	27.5	31.0	29.2	30.6	31.0	45
46	36.0	34.5	33.5	36.0	33.6	31.8	36.3	46
47	37.0	35.5	40.0	40.0	36.8	35.7	37.6	47
48	41.0	38.0	41.5	45.5	41.0	38.0	42.0	48
49	41.5	40.0	45.0	46.5	43.0	42.7	46.5	49
50	49.0	45.5	48.5	49.5	47.2	44.3	50.2	50
51	52.0	48.5	53.5	54.5	51.8	51.4	51.0	51
52	56.0	51.0	54.0	58.0	55.2	53.7	57.1	52
53	56.5	62.0	60.0	61.5	57.8	60.0	58.8	53
54	64.5	66.5	62.0	67.5	66.2	65.5	65.7	54
55	66.5	67.5	70.5	69.5	67.2	71.4	68.2	55
56	68.5	70.5	71.5	72.0	72.4	72.2	70.6	56

Table C.4. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
57	74.5	74.0	73.5	76.5	74.4	73.3	75.5	57
58	77.0	78.5	75.0	77.5	78.4	80.4	77.6	58
59	79.0	78.5	81.0	82.0	79.6	82.7	78.8	59
60	81.5	82.5	81.0	83.5	82.0	83.1	80.8	60
61	84.0	85.5	84.5	84.5	85.6	86.3	84.9	61
62	84.0	85.5	89.0	87.0	87.8	88.2	87.3	62
63	99.9	99.9	91.5	89.0	87.8	88.2	87.3	63
64	99.9	99.9	91.5	90.0	99.9	99.9	99.9	64
65	99.9	99.9	99.9	93.0	99.9	99.9	99.9	65
66	99.9	99.9	99.9	94.5	99.9	99.9	99.9	66
67	99.9	99.9	99.9	94.5	99.9	99.9	99.9	67
68	99.9	99.9	99.9	99.9	99.9	99.9	99.9	68

Table C.5. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Emotional Well-Being Scale (EWB)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
≤16	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	≤16
17	<0.1	<0.1	<0.1	<0.1	0.2	<0.1	0.4	17
18	<0.1	<0.1	<0.1	<0.1	0.4	<0.1	0.8	18
19	<0.1	<0.1	<0.1	<0.1	0.6	0.4	0.8	19
20	0.5	<0.1	0.5	0.5	0.6	0.4	0.8	20
21	0.5	0.5	0.5	0.5	0.8	0.4	0.8	21
22	0.5	1.0	0.5	0.5	1.0	0.8	0.8	22
23	0.5	1.0	0.5	1.0	1.6	1.2	0.8	23
24	1.5	1.0	1.0	1.0	1.8	2.7	1.2	24
25	1.5	1.5	2.0	1.0	2.2	3.1	2.0	25
26	2.0	2.5	2.0	2.0	2.8	3.1	2.4	26
27	2.0	3.0	2.5	2.5	3.0	3.5	2.4	27
28	2.5	4.0	3.5	3.0	3.4	3.9	2.4	28
29	3.5	4.5	4.5	3.5	3.6	4.3	3.7	29
30	4.0	5.0	6.0	5.0	4.4	4.7	3.7	30
31	5.0	5.0	7.5	5.0	5.0	5.9	5.7	31
32	5.5	6.0	8.5	6.5	6.6	6.7	6.1	32
33	7.0	6.5	9.5	8.0	7.6	7.5	6.5	33
34	7.5	7.5	10.5	8.5	8.4	9.4	7.3	34
35	8.5	9.0	10.5	11.0	8.6	9.4	10.2	35
36	10.0	11.5	10.5	12.5	10.8	10.2	11.8	36
37	11.0	12.5	13.0	14.0	12.2	11.0	14.3	37
38	14.0	15.0	14.0	14.0	14.6	11.4	15.1	38
39	15.5	16.5	16.5	16.0	15.6	15.3	18.0	39
40	18.5	18.0	19.5	18.5	19.4	16.1	20.0	40
41	20.0	19.5	21.0	20.0	20.4	20.0	20.4	41
42	23.0	22.0	23.5	22.5	22.4	22.0	21.6	42
43	25.5	25.0	24.0	23.5	23.8	23.5	23.7	43
44	28.0	27.5	28.0	26.5	27.0	26.7	25.7	44
45	30.5	28.5	28.5	27.5	28.4	29.8	28.6	45
46	34.0	33.5	31.0	32.0	32.0	31.8	29.0	46
47	35.5	34.5	32.0	33.0	32.6	34.9	35.1	47
48	41.0	38.5	35.0	38.5	37.8	38.4	36.3	48
49	42.0	44.5	41.0	42.5	40.8	40.4	40.8	49
50	49.5	47.0	43.0	48.0	44.8	44.3	43.3	50
51	51.0	49.5	44.5	51.0	46.2	47.8	45.3	51
52	56.5	53.5	47.5	51.0	51.6	50.6	48.6	52
53	57.5	55.0	55.0	60.5	52.8	55.7	57.6	53
54	62.0	60.0	58.0	60.5	62.4	60.8	60.8	54

Table C.5. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
55	63.0	63.5	70.0	62.0	62.8	64.7	67.3	55
56	74.0	68.0	70.0	77.0	75.2	71.4	72.2	56
57	76.5	74.0	83.0	78.5	75.2	78.0	73.9	57
58	85.0	80.5	83.5	80.0	86.8	78.4	84.5	58
59	85.0	81.5	85.5	80.0	87.0	89.4	85.3	59
60	86.0	90.0	85.5	88.5	89.0	89.4	87.3	60
61	90.0	91.0	92.5	88.5	89.0	90.6	87.8	61
62	92.5	91.0	93.0	89.5	94.6	91.0	93.9	62
63	92.5	96.5	93.5	89.5	94.6	95.3	94.3	63
64	92.5	96.5	93.5	89.5	95.6	96.1	94.7	64
65	96.0	97.5	99.9	99.9	95.6	96.5	94.7	65
66	96.0	97.5	99.9	99.9	99.9	96.5	99.9	66
67	96.0	99.9	99.9	99.9	99.9	99.9	99.9	67
68	96.0	99.9	99.9	99.9	99.9	99.9	99.9	68
69	99.9	99.9	99.9	99.9	99.9	99.9	99.9	69

Table C.6. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Role Limitations due to Emotional Problems Scale (RLE)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
≤18	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	5.7	≤18
19	5.5	<0.1	7.0	<0.1	6.4	<0.1	5.7	19
20	5.5	<0.1	7.0	<0.1	6.4	<0.1	5.7	20
21	5.5	<0.1	7.0	<0.1	6.4	7.1	5.7	21
22	5.5	<0.1	7.0	<0.1	6.4	7.1	5.7	22
23	5.5	<0.1	7.0	<0.1	6.4	7.1	5.7	23
24	5.5	8.5	7.0	<0.1	6.4	7.1	5.7	24
25	5.5	8.5	7.0	11.0	6.4	7.1	5.7	25
26	5.5	8.5	7.0	11.0	6.4	7.1	5.7	26
27	5.5	8.5	7.0	11.0	6.4	7.1	5.7	27
28	5.5	8.5	7.0	11.0	6.4	7.1	5.7	28
29	5.5	8.5	7.0	11.0	6.4	7.1	10.6	29
30	5.5	8.5	7.0	11.0	6.4	7.1	10.6	30
31	12.0	8.5	11.5	11.0	11.6	7.1	10.6	31
32	12.0	8.5	11.5	11.0	11.6	12.5	10.6	32
33	12.0	8.5	11.5	11.0	11.6	12.5	10.6	33
34	12.0	16.5	11.5	11.0	11.6	12.5	10.6	34
35	12.0	16.5	11.5	15.0	11.6	12.5	10.6	35
36	12.0	16.5	11.5	15.0	11.6	12.5	10.6	36
37	12.0	16.5	11.5	15.0	11.6	12.5	10.6	37
38	12.0	16.5	11.5	15.0	11.6	12.5	10.6	38
39	12.0	16.5	11.5	15.0	11.6	12.5	10.6	39
40	12.0	16.5	11.5	15.0	11.6	12.5	10.6	40
41	12.0	16.5	11.5	15.0	11.6	12.5	10.6	41
42	12.0	16.5	11.5	15.0	11.6	12.5	15.5	42
43	22.5	16.5	17.0	15.0	19.0	22.4	15.5	43
44	22.5	16.5	17.0	15.0	19.0	22.4	15.5	44
45	22.5	24.0	17.0	22.5	19.0	22.4	15.5	45
46	22.5	24.0	17.0	22.5	19.0	22.4	15.5	46
47	22.5	24.0	17.0	22.5	19.0	22.4	15.5	47
48	22.5	24.0	17.0	22.5	19.0	22.4	15.5	48
49	22.5	24.0	17.0	22.5	19.0	22.4	15.5	49
50	22.5	24.0	17.0	22.5	19.0	22.4	15.5	50
51	22.5	24.0	17.0	22.5	19.0	22.4	15.5	51
52	22.5	24.0	17.0	22.5	19.0	22.4	15.5	52
53	22.5	24.0	17.0	22.5	19.0	22.4	15.5	53
54	22.5	24.0	99.9	22.5	99.9	22.4	99.9	54
55	99.9	99.9	99.9	99.9	99.9	99.9	99.9	55

Table C.7. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Social Functioning Scale (SF)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
7	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	7
8	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	8
9	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	9
10	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	10
11	0.5	<0.1	<0.1	<0.1	<0.1	0.4	<0.1	11
12	0.5	<0.1	<0.1	<0.1	0.2	0.4	<0.1	12
13	0.5	<0.1	<0.1	<0.1	0.2	0.4	<0.1	13
14	0.5	<0.1	<0.1	<0.1	0.2	0.4	<0.1	14
15	0.5	0.5	<0.1	<0.1	0.4	0.4	<0.1	15
16	0.5	0.5	0.5	<0.1	0.4	0.4	0.4	16
17	1.0	0.5	0.5	<0.1	0.8	0.8	0.4	17
18	1.0	0.5	1.0	<0.1	0.8	0.8	0.8	18
19	1.0	0.5	1.0	<0.1	1.0	0.8	0.8	19
20	1.0	0.5	2.0	<0.1	2.0	1.6	1.2	20
21	1.0	0.5	2.0	2.0	2.0	1.6	2.4	21
22	1.0	1.0	2.5	2.0	2.0	1.6	2.4	22
23	1.0	3.5	4.0	2.0	2.0	1.6	2.4	23
24	1.5	3.5	4.0	4.0	4.2	3.1	2.4	24
25	2.0	3.5	4.0	4.0	4.2	3.1	5.3	25
26	2.0	3.5	4.0	4.0	4.2	3.1	5.3	26
27	2.0	6.5	6.5	4.0	5.4	4.3	5.3	27
28	2.0	6.5	6.5	4.0	5.4	4.3	6.5	28
29	2.0	6.5	6.5	9.0	5.4	4.3	6.5	29
30	7.0	7.5	7.0	9.0	5.4	4.3	6.5	30
31	7.0	7.5	7.0	9.0	9.6	10.2	6.5	31
32	7.0	7.5	7.0	9.5	9.6	10.2	9.0	32
33	7.5	10.5	12.0	9.5	9.6	10.2	9.0	33
34	7.5	10.5	12.0	10.0	10.4	11.4	9.0	34
35	12.5	10.5	12.0	11.0	10.4	11.4	9.4	35
36	13.5	11.0	12.5	11.0	13.6	14.1	9.4	36
37	21.5	11.0	12.5	16.0	13.8	14.1	13.5	37
38	21.5	16.0	13.5	17.0	13.8	18.8	13.5	38
39	21.5	16.0	13.5	17.0	19.8	20.0	19.2	39
40	21.5	21.5	19.5	17.0	19.8	20.0	19.6	40
41	21.5	22.5	19.5	17.0	19.8	20.0	19.6	41
42	21.5	22.5	19.5	20.0	19.8	20.0	19.6	42
43	23.0	22.5	19.5	22.0	19.8	20.0	19.6	43
44	23.5	22.5	19.5	22.0	22.8	22.7	22.9	44
45	30.5	24.0	24.5	28.0	22.8	22.7	22.9	45
46	30.5	24.0	24.5	28.0	29.8	29.8	29.8	46

Continued on next page.

Table C.7. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
47	30.5	33.5	30.0	28.0	29.8	29.8	29.8	47
48	30.5	33.5	30.0	28.0	29.8	29.8	29.8	48
49	30.5	33.5	30.0	28.0	29.8	29.8	29.8	49
50	30.5	33.5	30.0	35.5	29.8	29.8	29.8	50
51	41.5	33.5	37.0	39.0	36.0	38.0	37.6	51
52	41.5	47.0	41.5	39.0	40.8	43.9	37.6	52
53	41.5	47.0	41.5	39.0	40.8	43.9	37.6	53
54	41.5	47.0	41.5	39.0	40.8	43.9	37.6	54
55	41.5	47.0	41.5	39.0	40.8	43.9	37.6	55
56	41.5	47.0	99.9	99.9	40.8	43.9	99.9	56
57	99.9	99.9	99.9	99.9	99.9	99.9	99.9	57

Table C.8. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Energy/Fatigue Scale (EF)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
25	<0.1	0.5	<0.1	<0.1	<0.1	<0.1	1.2	25
26	1.5	0.5	<0.1	1.5	1.4	1.6	1.2	26
27	1.5	0.5	1.0	1.5	1.4	1.6	1.6	27
28	2.0	1.0	1.0	1.5	1.6	1.6	2.4	28
29	2.5	1.5	1.0	1.5	2.2	1.6	3.7	29
30	3.5	3.5	2.0	2.0	3.6	2.0	3.7	30
31	4.0	5.0	3.0	5.5	5.0	3.5	4.9	31
32	4.0	5.0	5.0	5.5	5.4	5.9	6.1	32
33	6.5	6.0	6.5	8.5	7.6	6.3	6.1	33
34	7.5	6.5	8.5	9.0	8.0	9.4	7.3	34
35	8.5	9.0	8.5	10.0	10.0	10.6	9.0	35
36	9.5	11.0	10.0	10.5	10.6	11.8	12.2	36
37	11.5	15.0	12.0	11.0	13.2	12.2	13.1	37
38	13.5	16.5	14.0	15.0	14.4	14.5	16.3	38
39	16.0	17.5	18.0	15.0	16.8	16.1	16.7	39
40	17.0	18.0	19.5	18.5	18.6	17.3	18.4	40
41	20.0	21.5	22.5	18.5	20.4	20.8	22.9	41
42	22.5	22.5	23.5	25.0	24.8	22.4	24.9	42
43	27.0	27.5	30.0	25.0	27.4	27.1	26.1	43
44	30.5	29.5	32.0	30.5	29.6	29.8	29.0	44
45	33.0	32.0	36.0	31.0	32.0	33.3	32.2	45
46	38.0	34.5	39.5	38.0	36.6	34.9	34.7	46
47	41.5	39.0	42.5	39.0	40.0	41.2	35.5	47
48	43.5	41.5	43.0	43.0	40.4	45.1	42.4	48
49	48.0	43.5	49.0	44.5	47.4	45.1	43.7	49
50	50.5	48.5	49.0	50.5	48.8	52.5	46.9	50
51	53.5	51.5	50.0	51.5	52.6	53.7	48.6	51
52	56.0	55.5	53.5	56.0	53.8	58.0	56.3	52
53	64.5	57.0	59.0	58.0	60.6	58.8	56.7	53
54	66.0	65.5	59.5	61.0	61.2	64.7	59.2	54
55	68.0	66.5	62.0	69.0	64.0	65.5	65.7	55
56	72.5	70.0	69.0	70.5	69.8	68.6	69.0	56
57	74.0	75.5	71.5	72.0	72.2	73.7	71.8	57
58	75.0	77.0	73.5	73.0	74.4	75.3	72.2	58
59	77.5	79.0	84.0	82.5	75.2	76.9	82.9	59
60	87.0	79.5	84.0	83.5	85.2	78.0	83.3	60
61	87.5	89.0	92.0	90.0	85.6	87.5	91.8	61
62	92.0	89.0	92.0	90.0	93.6	87.8	91.8	62
63	92.5	94.5	94.5	94.5	94.0	95.3	96.3	63

Continued on next page.

Table C.8. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
64	93.0	96.0	94.5	94.5	97.4	95.7	96.7	64
65	95.5	96.0	94.5	94.5	97.6	98.4	96.7	65
66	96.5	98.5	96.5	96.5	97.6	98.4	98.0	66
67	96.5	98.5	96.5	96.5	98.2	98.4	98.4	67
68	98.0	98.5	96.5	96.5	98.4	98.4	98.4	68
69	98.0	99.0	99.9	99.9	98.4	98.4	99.9	69
70	99.9	99.0	99.9	99.9	99.9	98.4	99.9	70
71	99.9	99.0	99.9	99.9	99.9	99.9	99.9	71
72	99.9	99.9	99.9	99.9	99.9	99.9	99.9	72

Table C.9. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Physical Health Composite (PHC)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
≤7	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	≤7
8	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	8
9	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	9
10	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	10
11	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	11
12	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	12
13	0.5	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	13
14	1.5	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	14
15	1.5	1.0	<0.1	<0.1	<0.1	<0.1	<0.1	15
16	1.5	1.5	<0.1	<0.1	<0.1	<0.1	<0.1	16
17	1.5	2.0	<0.1	<0.1	<0.1	<0.1	<0.1	17
18	1.5	2.5	<0.1	<0.1	0.2	0.4	<0.1	18
19	1.5	3.0	<0.1	<0.1	0.6	1.2	0.4	19
20	1.5	3.5	<0.1	<0.1	1.0	1.2	0.8	20
21	1.5	4.0	<0.1	<0.1	1.0	1.2	0.8	21
22	1.5	4.5	<0.1	<0.1	1.4	1.6	1.2	22
23	2.5	4.5	1.5	0.5	1.8	2.4	1.6	23
24	3.0	4.5	1.5	0.5	3.6	4.3	2.9	24
25	3.5	5.0	2.0	0.5	4.4	5.1	4.1	25
26	4.0	5.5	2.5	0.5	5.0	5.1	6.1	26
27	5.0	6.5	4.0	0.5	5.6	5.1	6.1	27
28	5.5	6.5	6.0	1.5	6.0	5.1	6.9	28
29	5.5	6.5	6.0	2.5	6.6	5.9	7.3	29
30	6.0	6.5	7.0	5.0	7.0	6.7	8.2	30
31	7.0	6.5	7.0	6.0	8.4	7.5	9.4	31
32	8.0	6.5	8.5	8.5	9.4	9.0	9.8	32
33	8.5	7.0	10.0	9.5	10.0	9.8	10.6	33
34	10.5	7.0	10.5	10.5	11.2	11.0	10.6	34
35	11.0	8.5	11.5	13.5	11.8	11.8	11.4	35
36	12.0	9.5	13.0	15.5	12.2	12.2	11.8	36
37	13.0	10.5	15.5	18.0	13.2	12.9	12.2	37
38	14.0	11.0	15.5	19.0	14.6	15.3	14.3	38
39	14.0	12.0	17.5	20.0	16.8	17.3	15.1	39
40	15.0	12.5	20.0	21.0	17.6	18.8	16.3	40
41	16.0	13.5	20.5	23.5	18.6	20.4	16.7	41
42	17.0	14.5	24.0	25.0	19.8	20.8	18.4	42
43	18.0	17.5	24.5	27.0	21.2	22.0	19.2	43
44	20.0	20.0	26.0	29.0	22.4	22.4	20.8	44
45	21.5	23.5	28.0	32.5	23.6	23.5	22.4	45
46	24.5	23.5	28.0	33.5	26.4	26.3	24.9	46

Continued on next page.

Table C.9. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
47	29.0	25.5	29.0	35.5	29.0	28.6	28.6	47
48	31.0	28.0	31.0	36.5	31.8	32.2	30.2	48
49	35.5	32.5	34.0	36.5	36.2	36.5	33.1	49
50	35.5	36.5	36.5	39.0	39.8	40.0	37.1	50
51	39.0	41.5	42.0	45.0	42.6	44.3	40.4	51
52	44.0	44.0	45.5	47.5	49.2	44.3	43.7	52
53	53.0	51.0	47.0	50.5	54.2	49.4	51.0	53
54	58.0	56.5	52.5	55.5	59.8	54.5	57.1	54
55	64.0	66.5	63.5	63.0	67.0	58.4	63.7	55
56	72.5	70.0	69.5	69.5	72.2	66.7	70.2	56
57	79.5	79.0	78.0	72.5	78.0	72.2	74.7	57
58	89.5	88.5	82.0	77.0	85.6	81.2	79.2	58
59	99.9	93.5	88.5	81.5	91.4	88.2	89.4	59
60	99.9	99.9	95.0	84.5	95.2	92.9	95.5	60
61	99.9	99.9	99.9	87.0	99.9	94.9	99.9	61
62	99.9	99.9	99.9	91.5	99.9	99.9	99.9	62
63	99.9	99.9	99.9	95.5	99.9	99.9	99.9	63
64	99.9	99.9	99.9	98.0	99.9	99.9	99.9	64
≥65	99.9	99.9	99.9	99.9	99.9	99.9	99.9	≥65

Table C.10. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Mental Health Composite (MHC)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
≤12	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	≤12
13	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	0.4	13
14	<0.1	<0.1	<0.1	<0.1	0.2	<0.1	0.4	14
15	<0.1	<0.1	<0.1	<0.1	0.4	<0.1	0.4	15
16	<0.1	<0.1	<0.1	<0.1	0.6	0.8	0.4	16
17	0.5	<0.1	<0.1	<0.1	0.6	0.8	0.4	17
18	0.5	0.5	0.5	<0.1	0.8	0.8	0.8	18
19	0.5	1.0	0.5	<0.1	1.2	1.2	1.2	19
20	1.0	1.0	0.5	<0.1	1.2	1.6	1.2	20
21	1.5	1.5	1.0	<0.1	1.4	1.6	1.2	21
22	1.5	2.5	1.0	0.5	1.4	1.6	2.0	22
23	2.5	2.5	1.5	1.5	2.0	2.0	2.0	23
24	3.0	3.0	1.5	1.5	2.6	2.7	2.9	24
25	3.0	3.0	2.0	2.5	2.8	3.1	3.3	25
26	4.0	3.0	3.0	3.0	3.0	3.1	3.3	26
27	4.0	3.5	5.0	3.0	3.4	3.1	4.1	27
28	4.5	3.5	5.0	4.0	3.6	3.5	4.5	28
29	6.0	3.5	5.0	4.0	4.8	4.3	4.9	29
30	6.0	4.0	5.5	4.0	5.4	5.5	5.3	30
31	6.5	6.0	6.5	6.5	6.2	6.3	6.1	31
32	7.0	6.0	8.0	7.5	6.6	7.5	6.1	32
33	8.0	7.5	8.5	8.5	7.0	7.5	7.8	33
34	8.5	8.5	9.5	10.0	8.0	7.8	7.8	34
35	9.5	9.0	10.5	12.0	10.4	9.4	9.4	35
36	11.0	10.5	12.5	12.5	12.0	12.2	11.8	36
37	12.5	11.5	13.5	13.5	13.6	12.9	13.9	37
38	14.5	14.5	16.0	14.5	14.4	14.5	14.3	38
39	15.0	17.0	17.0	18.0	15.8	15.7	15.9	39
40	16.5	17.5	17.5	19.0	17.0	17.3	17.6	40
41	17.0	19.5	20.0	19.0	18.6	18.4	18.8	41
42	17.0	20.5	21.5	21.0	19.8	18.8	20.0	42
43	19.5	23.5	22.5	23.0	21.2	20.8	21.6	43
44	22.5	24.0	24.5	24.5	23.8	22.4	23.3	44
45	27.0	27.0	26.0	26.0	25.8	25.9	24.1	45
46	30.0	30.0	28.0	29.0	28.6	29.8	26.5	46
47	33.5	34.0	31.0	32.5	32.4	34.5	31.0	47
48	35.5	37.0	36.5	36.0	36.6	37.6	34.3	48
49	42.0	40.5	38.5	38.0	39.4	41.2	37.1	49
50	45.0	45.0	43.0	41.0	41.6	43.5	39.6	50

Continued on next page.

Table C.10. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
51	46.5	46.5	46.5	47.0	44.0	46.7	42.9	51
52	51.5	48.5	50.5	49.5	48.6	48.6	48.2	52
53	57.5	53.0	54.5	56.5	53.2	53.7	55.1	53
54	61.0	58.5	60.0	64.5	58.8	57.3	62.4	54
55	66.0	65.5	63.5	70.5	65.6	64.7	67.3	55
56	70.5	72.0	65.5	74.5	71.2	69.8	71.4	56
57	76.5	76.0	74.5	77.5	75.0	72.9	75.9	57
58	79.0	81.0	82.0	81.5	79.0	78.0	83.3	58
59	84.5	84.5	89.0	86.5	87.2	84.3	89.4	59
60	90.5	88.5	90.5	86.5	92.2	89.4	91.8	60
61	93.5	94.5	93.0	91.0	93.8	93.7	95.1	61
62	95.0	95.5	96.0	93.5	96.2	94.9	95.9	62
63	95.5	98.0	97.0	94.0	97.2	96.9	98.0	63
64	96.5	99.0	98.0	97.0	99.0	98.4	99.2	64
65	98.0	99.5	99.9	98.0	99.2	99.2	99.9	65
66	98.5	99.5	99.9	99.9	99.9	99.2	99.9	66
67	98.5	99.9	99.9	99.9	99.9	99.9	99.9	67
≥68	99.9	99.9	99.9	99.9	99.9	99.9	99.9	≥68

Table C.11. *T* Scores Obtained by Cumulative Percentages of the Normative Samples: Global Health Composite (GHC)

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
≤10	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	≤10
11	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	11
12	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	12
13	0.5	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	13
14	1.5	1.0	<0.1	<0.1	0.4	<0.1	<0.1	14
15	1.5	1.0	<0.1	<0.1	0.6	0.8	<0.1	15
16	1.5	1.0	<0.1	<0.1	0.8	1.2	0.4	16
17	1.5	2.0	<0.1	<0.1	1.0	1.2	0.8	17
18	1.5	2.0	<0.1	<0.1	1.0	1.2	0.8	18
19	1.5	2.0	<0.1	<0.1	1.2	1.2	0.8	19
20	1.5	2.5	1.0	<0.1	1.6	1.6	1.6	20
21	1.5	2.5	1.0	<0.1	2.0	1.6	2.0	21
22	2.0	3.0	1.5	<0.1	2.0	2.0	2.0	22
23	2.0	3.0	1.5	<0.1	2.2	2.0	2.4	23
24	2.0	3.0	2.0	<0.1	2.8	2.0	2.9	24
25	2.5	3.0	2.0	0.5	3.2	2.7	3.3	25
26	2.5	3.5	2.5	1.0	3.4	2.7	4.1	26
27	3.0	3.5	3.0	2.5	4.2	3.1	4.5	27
28	3.0	4.5	3.5	3.5	4.8	4.3	4.5	28
29	3.0	4.5	4.5	4.5	5.2	4.3	5.7	29
30	4.0	5.0	5.5	5.5	5.8	5.1	6.1	30
31	6.0	6.5	6.5	7.5	6.6	5.9	6.5	31
32	6.0	7.5	8.0	8.5	7.8	7.8	7.8	32
33	7.0	7.5	9.0	10.5	8.2	8.6	8.2	33
34	9.5	8.0	10.5	10.5	8.8	9.0	9.0	34
35	11.5	8.5	11.5	12.0	9.4	9.4	9.4	35
36	12.5	9.5	11.5	14.0	10.8	11.0	11.0	36
37	13.5	10.5	13.0	15.5	12.4	12.5	11.8	37
38	15.0	13.0	14.5	16.5	14.0	13.7	14.3	38
39	16.0	13.5	16.0	17.5	16.0	15.3	16.3	39
40	16.0	16.0	17.0	19.5	17.6	17.6	17.6	40
41	17.5	16.5	20.5	23.0	18.6	18.4	18.4	41
42	18.5	18.5	22.0	24.5	21.0	20.0	21.6	42
43	21.0	21.0	23.5	26.5	22.8	22.7	22.9	43
44	24.0	22.5	27.5	28.0	24.2	24.3	24.5	44
45	25.5	24.0	29.0	29.0	26.0	26.3	26.1	45
46	27.0	27.0	32.0	30.0	28.8	29.4	28.6	46
47	30.5	33.0	33.0	32.0	32.2	34.5	31.0	47
48	34.0	36.5	35.5	37.5	34.4	36.1	32.7	48

Continued on next page.

Table C.11. *(Continued)*

T Score	Age-Based Sample				Age-Stratified Sample			*T* Score
	18–24	25–44	45–64	≥65	Overall	Female	Male	
49	37.5	40.5	38.5	41.0	37.0	40.4	36.7	49
50	41.0	44.0	40.5	43.0	41.0	43.1	40.0	50
51	47.0	49.5	45.0	47.0	46.2	47.8	46.5	51
52	52.0	55.0	49.5	52.5	50.8	52.5	52.2	52
53	56.5	59.5	54.0	57.0	56.6	56.5	55.1	53
54	60.5	61.5	58.5	62.0	59.6	61.2	60.4	54
55	70.0	65.5	63.0	68.5	65.2	65.5	66.9	55
56	75.5	72.0	71.0	72.0	71.2	72.2	71.8	56
57	79.5	78.0	73.0	76.0	75.8	76.5	78.0	57
58	85.0	82.5	79.0	79.0	80.8	80.8	85.7	58
59	85.0	87.0	86.5	82.5	87.4	85.9	91.0	59
60	90.0	92.0	91.0	85.0	92.2	91.4	93.5	60
61	94.0	96.0	94.5	90.5	95.6	94.9	95.9	61
62	95.5	98.0	95.5	94.5	97.2	98.0	97.6	62
63	97.0	99.0	98.0	97.0	98.6	98.0	99.6	63
64	98.0	99.9	99.0	98.5	99.8	99.6	99.6	64
65	98.5	99.9	99.9	99.5	99.9	99.9	99.9	65
≥66	99.9	99.9	99.9	99.9	99.9	99.9	99.9	≥66

Appendix D
Discrepancy Score Tables

Table D.1. Cumulative Percentages of the Normative Samples Obtaining PHC>MHC *T*-Score Discrepancies

Amount of Discrepancy	Age-Based Sample				Age-Stratified Sample		
	18–24	25–44	45–64	≥65	Overall	Female	Male
≥28	0.5	0.0	1.0	0.5	0.4	0.8	0.0
27	0.5	0.0	1.0	0.5	0.6	1.2	0.0
26	0.5	0.5	1.0	0.5	1.2	2.4	0.0
25	0.5	0.5	1.0	0.5	1.4	2.7	0.8
24	1.0	1.5	1.0	0.5	1.4	2.7	0.8
23	1.0	1.5	1.5	0.5	1.8	2.7	0.8
22	1.0	3.0	2.0	1.0	2.4	3.9	0.8
21	1.0	3.0	2.0	1.0	2.4	4.3	0.8
20	1.0	3.0	2.0	1.5	2.6	4.3	1.2
19	1.0	4.0	3.0	1.5	3.0	4.7	1.6
18	1.0	4.5	3.5	1.5	3.6	5.5	2.4
17	2.5	4.5	4.5	1.5	4.0	5.5	3.7
16	4.5	5.0	4.5	3.0	4.6	5.5	4.1
15	6.0	5.5	5.0	4.5	5.0	5.5	4.5
14	6.5	6.0	5.5	6.0	5.2	6.7	4.9
13	7.0	6.5	6.5	7.0	6.8	8.2	5.7
12	8.5	7.0	8.0	8.0	8.2	10.2	6.9
11	8.5	9.5	8.5	8.0	10.2	13.7	8.6
10	12.0	12.0	10.0	8.5	12.0	14.9	11.0
9	13.5	13.0	11.5	10.5	14.4	17.6	13.5
8	17.5	15.5	14.5	12.5	16.6	19.2	16.7
7	20.0	17.5	16.5	15.5	19.2	22.7	20.8
6	24.5	21.0	20.0	20.0	23.0	25.1	24.9
5	26.0	24.0	22.0	24.5	26.4	29.0	26.9
4	31.0	29.0	27.0	27.5	29.4	32.5	30.6
3	37.5	35.5	31.5	35.0	32.4	36.1	35.9
2	41.5	42.5	35.5	44.0	36.8	38.8	41.2
1	44.0	47.5	42.5	47.0	42.2	44.3	46.9
0[a]	7.0	7.5	6.0	6.0	5.8	5.5	6.2
Mean	−0.2	0.2	0.0	0.1	−0.4	0.0	0.2
SD	8.9	8.5	8.1	7.6	8.9	9.9	8.1
Median	0.0	0.0	−1.0	0.0	−1.0	−1.0	0.0

Note. PHC–MHC discrepancies >10 were obtained by 15% or less of individuals within each normative group. PHC–MHC discrepancies >6 are statistically significant at $p < .10$. This table does not include negative PHC–MHC discrepancies. Mean difference scores, standard deviations, and medians are based on the entire distribution.

[a]This row provides the percentage of individuals in each normative group for whom PHC–MHC composite score discrepancies were equal to zero or <1.

Table D.2. Cumulative Percentages of the Normative Samples Obtaining MHC>PHC *T*-Score Discrepancies

Amount of Discrepancy	Age-Based Sample				Age-Stratified Sample		
	18–24	25–44	45–64	≥65	Overall	Female	Male
≥26	0.5	0.0	0.0	0.0	0.2	0.0	0.4
25	1.0	0.0	0.0	0.0	0.6	0.4	0.4
24	1.5	0.0	0.0	0.0	0.6	0.8	0.8
23	2.0	0.0	0.5	0.0	0.8	0.8	0.8
22	2.0	0.0	0.5	0.0	1.2	0.8	0.8
21	2.0	1.0	0.5	1.0	1.2	1.2	1.6
20	2.0	1.0	1.0	1.0	1.6	2.0	1.6
19	2.5	1.5	1.5	1.0	1.8	2.0	1.6
18	3.0	2.5	2.0	1.5	2.2	3.1	2.0
17	3.5	3.0	2.0	2.0	3.0	4.3	2.4
16	6.5	3.5	3.5	2.0	3.6	5.1	2.4
15	6.5	4.5	4.0	2.0	4.4	6.3	2.9
14	7.5	5.0	4.5	2.5	5.2	7.8	4.9
13	8.5	8.0	5.0	3.0	7.6	8.6	6.1
12	8.5	8.5	6.0	5.0	8.6	10.6	6.9
11	9.0	10.0	7.0	5.5	11.6	13.7	8.6
10	9.5	12.0	8.5	8.0	12.4	14.5	10.6
9	13.5	13.5	11.0	10.0	15.4	17.6	11.4
8	14.5	15.5	13.5	13.5	18.4	20.8	14.3
7	17.5	17.5	15.0	19.5	22.0	25.5	17.6
6	23.0	21.0	19.5	22.5	26.6	27.8	20.4
5	24.5	25.5	23.5	26.0	31.2	30.2	24.5
4	32.5	31.5	31.5	31.5	35.8	35.3	30.6
3	38.5	36.5	35.5	36.0	42.8	41.6	35.1
2	44.5	40.0	46.5	40.0	47.0	45.9	43.3
1	49.0	45.0	51.5	47.0	52.0	50.2	46.9
0[a]	7.0	7.5	6.0	6.0	5.8	5.5	6.2
Mean	0.2	–0.2	0.0	–0.1	0.4	0.0	–0.2
SD	8.9	8.5	8.1	7.6	8.9	9.9	8.1
Median	0.0	0.0	1.0	0.0	1.0	1.0	0.0

Note. MHC–PHC discrepancies >10 were obtained by 15% or less of individuals within each normative group. MHC–PHC discrepancies >6 are statistically significant at $p < .10$. This table does not include negative MHC–PHC discrepancies. Mean difference scores, standard deviations, and medians are based on the entire distribution.

[a] This row provides the percentage of individuals in each normative group for whom MHC–PHC composite score discrepancies were equal to zero or <1.

Appendix E
Description of the *RAND–12 Health Status Inventory*

The RAND–12 HSI, a 12-item version of the RAND–36 HSI, is reported in this appendix, including a discussion of the selection of the items, the methodology of the scoring, and an example. The RAND–12 HSI items by scale and composite are provided in Appendix G. The purpose of the RAND–12 HSI is to provide estimated scores on the Physical Health, Mental Health, and Global Health composites of the 36-item instrument. For this reason, separate norms tables and validity analyses are not provided. The intended use of the 12-item instrument is aggregate-level analysis, when summary reports are desired; its scores are not considered precise enough for individual-level analysis. Instructions for scoring responses are also provided on the RAND–12 HSI Hand-Scoring Worksheet. (RAND–12 HSI Question/Answer Sheets and Hand-Scoring Worksheets are available from The Psychological Corporation.)

Item Selection

The RAND–12 HSI contains at least one item from each of the eight scales of the RAND–36 HSI so that the abbreviated form adequately represents the wide range of relevant aspects of health status (Ware, Kosinski, & Keller, 1995, 1996). Table E.1 summarizes the correlations between the 12 items and the three RAND–36 HSI composite scores that they were selected to estimate.

Item Scoring Methodology

The scores derived from the RAND–12 HSI represent composite estimates of the corresponding RAND–36 HSI Physical Health, Mental Health, and Global Health composite scores. The items included in each composite are drawn from the items on the scales included in the corresponding RAND–36 HSI composites. Six items contribute to each of the Physical Health Composite and Mental Health Composite estimates, with all 12 items contributing to the Global Health Composite estimate. The three composite estimates are based on three regression equations for each of the seven normative groups (described in Chapter 2), for a total of 21 regression formulas. For each regression, the RAND–36 HSI

composite *T* score was the dependent variable, and the IRT-weighted item scores were the independent variables. Scoring methodology was developed to be as pragmatic as possible while maintaining the IRT weighting at the level of response option for each item. For this reason, the value for each item response in the prediction formula was the IRT weighting for the response option in the RAND–36 HSI for that item (see Appendix A, Tables A.1–A.8). Although these response-option IRT weights are based on scale composition that is not retained in the RAND–12 HSI, these weights retain the item-specific response-option weighting. The relative weights of item responses in the new formula are determined by the regression analyses.

Each of the 21 prediction formulas was developed on the data from 60% of each original normative sample and cross-validated on the data from the remaining 40% of each sample. The general formula for calculating an estimated composite *T* score is

Estimated Composite *T* Score = Intercept Value + Sum of Derived Item Scores

The derived score for each of the contributing items is the IRT weight for the response option multiplied by the item's parameter estimate. The specific formula for each of the three composites follows. Note that the item numbers are RAND–36 HSI item numbers. The IRT weights for the formulas are found in Appendix A, Table A.1–A.8. The norm-specific intercept values and the parameter estimates are provided in Tables E.2–E.4.

Estimated **Physical Health Composite** *T* Score =

Intercept Value (Table E.2)
+ (Item 1 IRT Weight) (Parameter Estimate)
+ (Item 4 IRT Weight) (Parameter Estimate)
+ (Item 6 IRT Weight) (Parameter Estimate)
+ (Item 14 IRT Weight) (Parameter Estimate)
+ (Item 15 IRT Weight) (Parameter Estimate)
+ (Item 22 IRT Weight) (Parameter Estimate)

Estimated **Mental Health Composite** *T* Score =

Intercept Value (Table E.3)
+ (Item 18 IRT Weight) (Parameter Estimate)
+ (Item 19 IRT Weight) (Parameter Estimate)
+ (Item 26 IRT Weight) (Parameter Estimate)
+ (Item 27 IRT Weight) (Parameter Estimate)
+ (Item 28 IRT Weight) (Parameter Estimate)
+ (Item 32 IRT Weight) (Parameter Estimate)

Estimated **Global Health Composite** *T* Score =

Intercept Value (Table E.4)
+ (Item 1 IRT Weight) (Parameter Estimate)
+ (Item 4 IRT Weight) (Parameter Estimate)
+ (Item 6 IRT Weight) (Parameter Estimate)
+ (Item 14 IRT Weight) (Parameter Estimate)
+ (Item 15 IRT Weight) (Parameter Estimate)
+ (Item 18 IRT Weight) (Parameter Estimate)
+ (Item 19 IRT Weight) (Parameter Estimate)
+ (Item 22 IRT Weight) (Parameter Estimate)
+ (Item 26 IRT Weight) (Parameter Estimate)
+ (Item 27 IRT Weight) (Parameter Estimate)
+ (Item 28 IRT Weight) (Parameter Estimate)
+ (Item 32 IRT Weight) (Parameter Estimate)

Scoring Example

Following is an example demonstrating the calculation of the three RAND–12 HSI estimated composite *T* scores. The example is based on the responses of a 21-year-old female respondent and the intercept value and parameter estimates for the 18–24 Age Group.

Estimated Physical Health Composite *T* Score

Intercept Value = –19.0938

RAND–12 HSI Item	RAND–36 HSI Item	Response	IRT Weight		Parameter Estimate		Derived Item Score
1	1	3	46	x	.1108	=	5.0968
2	4	3	56	x	.1566	=	8.7696
3	6	3	72	x	.1605	=	11.5560
4	14	2	100	x	.1076	=	10.7600
5	15	2	74	x	.1877	=	13.8898
8	22	2	49	x	.3283	=	16.0867

Estimated *T* Score = –19.0938 + 66.1589
= 47.0651
= 47 (rounded to nearest whole number)

Actual *T* Score = 46

Estimated Mental Health Composite *T* Score

Intercept Value = 10.4916

RAND–12 HSI Item	RAND–36 HSI Item	Response	IRT Weight	x	Parameter Estimate	=	Derived Item Score
6	18	1	59	x	.0521	=	3.0739
7	19	1	13	x	.1044	=	1.3572
9	26	4	32	x	.1054	=	3.3728
10	27	5	13	x	.1268	=	1.6484
11	28	3	22	x	.1642	=	3.6124
12	32	2	38	x	.1170	=	4.4460

Estimated *T* Score = 10.4916 + 17.5107
= 28.0023
= 28 (rounded to nearest whole number)

Actual *T* Score = 29

Estimated Global Health Composite *T* Score

Intercept Value = –8.4989

RAND–12 HSI Item	RAND–36 HSI Item	Response	IRT Weight	x	Parameter Estimate	=	Derived Item Score
1	1	3	46	x	.0517	=	2.3782
2	4	3	56	x	.0428	=	2.3968
3	6	3	72	x	.1118	=	8.0496
4	14	2	100	x	.0300	=	3.0000
5	15	2	74	x	.0901	=	6.6674
6	18	1	59	x	.0303	=	1.7877
7	19	1	13	x	.0491	=	.6383
8	22	2	49	x	.1285	=	6.2965
9	26	4	32	x	.0723	=	2.3136
10	27	5	13	x	.0962	=	1.2506
11	28	3	22	x	.1296	=	2.8512
12	32	2	38	x	.0989	=	3.7582

Estimated *T* Score = –8.4989 + 41.3881
= 32.8892
= 33 (rounded to nearest whole number)

Actual *T* Score = 34

Note on Negative Regression Weightings

In multiple linear regression, the set of regression weights is chosen to maximize the variance of the dependent variables that can be explained by the set of independent variables (R^2). Unless all of the independent variables are uncorrelated, a simple relationship does not exist between regression weights and the correlation between the independent variable and dependent variable. Therefore, a negative regression weight in multiple regressions does not indicate anything about the relationship between the item score and the composite score. The relationship between any given item score and composite score is demonstrated by the correlation between the two scores (see Table E.1).

Cross-Validation

Cross-validation results correlating predicted composite *T* scores from the RAND–12 HSI with actual *T* scores on the RAND–36 HSI for the cross-validation sample of each normative group are presented in Table E.5. The cross-validated R^2 values are all adequate, with 19 of the 21 R^2 values $\geq .90$.

Table E.1. Correlations Between RAND–12 HSI and RAND–36 HSI Item Scores for the Seven Normative Groups

Item Number		RAND-36 HSI Composites																					
		Physical Health Composite							Mental Health Composite							Global Health Composite						
R–12	R–36[a]	18–24	25–44	45–64	≥65	Overall	Female	Male	18–24	25–44	45–64	≥65	Overall	Female	Male	18–24	25–44	45–64	≥65	Overall	Female	Male
1	1	.69	.73	.77	.72	.74	.73	.74	.51	.55	.57	.53	.52	.48	.57	.65	.70	.73	.67	.69	.68	.70
2	4	.52	.68	.72	.73	.70	.69	.71	.16**	.32	.49	.51	.33	.30	.36	.33	.52	.64	.66	.55	.53	.56
3	6	.63	.54	.71	.65	.66	.65	.68	.32	.35	.46	.41	.32	.25	.40	.49	.47	.63	.55	.52	.47	.56
4	14	.69	.69	.74	.75	.73	.71	.74	.40	.43	.57	.58	.47	.41	.52	.57	.59	.70	.71	.65	.62	.68
5	15	.58	.74	.84	.79	.77	.76	.78	.22****	.43	.54	.58	.42	.30	.55	.40	.62	.73	.73	.64	.57	.71
8	22	.75	.79	.84	.82	.81	.83	.80	.47	.54	.55	.61	.53	.49	.57	.65	.71	.74	.76	.73	.73	.73
6	18	.40	.36	.31	.40	.30	.18***	.46	.61	.64	.56	.53	.59	.60	.60	.58	.57	.49	.51	.52	.47	.58
7	19	.41	.28	.37	.39	.32	.25	.41	.63	.56	.61	.50	.54	.53	.55	.60	.49	.55	.48	.49	.46	.53
9	26	.25******	.41	.41	.37	.33	.30	.36	.63	.72	.77	.66	.73	.75	.70	.53	.65	.66	.57	.61	.63	.61
10	27	.44	.54	.56	.60	.53	.48	.57	.72	.68	.77	.73	.70	.72	.68	.67	.69	.74	.72	.70	.70	.69
11	28	.42	.33	.44	.31	.29	.30	.29	.71	.67	.80	.59	.66	.67	.66	.66	.57	.70	.50	.55	.58	.54
12	32	.36	.43	.68	.49	.50	.48	.52	.55	.67	.77	.71	.71	.71	.69	.53	.63	.80	.66	.69	.70	.67

Note. 18–24, n = 200; 25–44, n = 200; 45–64, n = 200; ≥65, n = 200; Overall, n = 500; Female, n = 255; Male, n = 245. Items above the bold line are included in the Physical Health Composite estimate. Items below the bold line are included in the Mental Health Composite estimate.

[a]The RAND–36 HSI item numbers are listed so that IRT-weight ranges for each item can be found in Appendix A, Tables A.1–A.8.

*All correlations are significant at $p < .0001$, unless otherwise indicated.

**$p < .0224$. ***$p < .0033$. ****$p < .0021$. *****$p < .0003$.

Table E.2. Intercept Values and Parameter Estimates for Predicting Physical Health Composite *T* Scores From the RAND–12 HSI

		Aged-Based Sample				Age-Stratified Sample		
		18–24	25–44	45–64	≥65	Overall	Female	Male
Intercept		-19.0938	-11.0710	10.5244	9.1548	2.8864	–.4386	2.7032
Item Number								
R–12	**R–36[a]**							
1	1	.1108	.0896	.0943	.0923	.0967	.1071	.0928
2	4	.1566	.2342	.0489	.1143	.0799	.1215	.0748
3	6	.1605	.1602	.1146	.0876	.1306	.1144	.1231
4	14	.1076	.1107	.0763	.0822	.0911	.1425	.0713
5	15	.1877	.1482	.1172	.1010	.1331	.0919	.1773
8	22	.3283	.2160	.2005	.2597	.2151	.2121	.2206

Note. Normative groups represent 60% of the original standardization samples. The IRT weight is multiplied by the parameter estimate for the selected normative group.

[a]The RAND–36 HSI item numbers are listed so that IRT-weight ranges for each item can be found in Appendix A, Tables A.1–A.8.

Table E.3. Intercept Values and Parameter Estimates for Predicting Mental Health Composite *T* Scores From the RAND–12 HSI

		Aged-Based Sample				Age-Stratified Sample		
		18–24	25–44	45–64	≥65	Overall	Female	Male
Intercept		10.4916	7.5204	7.8641	4.5878	5.0185	5.6996	1.9757
Item Number								
R–12	**R–36[a]**							
6	18	.0521	.0741	.0978	.0308	.0939	.1105	.1241
7	19	.1044	.1407	.0454	.1050	.1112	.0898	.0862
9	26	.1054	.1097	.0951	.1133	.1026	.1158	.1040
10	27	.1268	.1104	.0992	.1529	.1125	.1102	.1440
11	28	.1642	.1633	.1507	.1672	.1621	.1600	.1529
12	32	.1170	.1206	.1478	.1547	.1326	.1214	.1233

Note. Normative groups represent 60% of the original standardization samples. The IRT weight is multiplied by the parameter estimate for the selected normative group.

[a]The RAND–36 HSI item numbers are listed so that IRT-weight ranges for each item can be found in Appendix A, Tables A.1–A.8.

Table E.4. Intercept Values and Parameter Estimates for Predicting Global Health Composite *T* Scores From the RAND–12 HSI

		Aged-Based Sample				Age-Stratified Sample		
		18–24	25–44	45–64	≥65	Overall	Female	Male
Intercept		–8.4989	–7.6688	5.2822	2.1652	–3.1424	–3.5430	–1.7177
Item Number								
R–12	R–36[a]							
1	1	.0517	.0512	.0650	.0547	.0581	.0541	.0423
2	4	.0428	.1178	–.0031	.0357	.0057	.0725	–.0063
3	6	.1118	.0709	.0628	.0480	.0714	.0518	.0742
4	14	.0300	.0642	.0518	.0558	.0598	.0821	.0501
5	15	.0901	.0700	.0581	.0580	.0812	.0227	.0882
6	18	.0303	.0476	.0675	–.0015	.0524	.0623	.0468
7	19	.0491	.0869	–.0159	.0621	.0701	.0505	.0688
8	22	.1285	.1264	.1027	.1393	.1169	.1263	.1132
9	26	.0723	.0737	.0745	.0751	.0701	.0792	.0687
10	27	.0962	.0619	.0445	.0764	.0626	.0688	.0860
11	28	.1296	.0860	.0855	.0978	.0928	.1037	.0959
12	32	.0989	.0709	.0837	.0951	.0869	.0738	.0778

Note. Normative groups represent 60% of the original standardization samples. The IRT weight is multiplied by the parameter estimate for the selected normative group.

[a]The RAND–36 HSI item numbers are listed so that IRT-weight ranges for each item can be found in Appendix A, Tables A.1–A.8.

Table E.5. Cross-Validation Results Predicting RAND–36 HSI Composite *T* Scores From the RAND–12 HSI

		Aged-Based Sample				Age-Stratified Sample		
Composite		18–24 (*n* = 80)	25–44 (*n* = 80)	45–64 (*n* = 80)	≥65 (*n* = 80)	Overall (*n* = 200)	Female (*n* = 102)	Male (*n* = 98)
Physical Health	R^2	.91	.88	.93	.92	.94	.92	.94
	$S_{y\cdot x}$	3.00	3.46	2.65	2.83	2.45	2.83	2.45
Mental Health	R^2	.92	.92	.95	.87	.92	.93	.93
	$S_{y\cdot x}$	2.83	2.83	2.24	3.61	2.83	2.65	2.65
Global Health	R^2	.94	.93	.95	.94	.95	.94	.95
	$S_{y\cdot x}$	2.45	2.65	2.24	2.45	2.24	2.45	2.24

Note. Cross-validation samples are based on 40% of the original standardization sample. R^2 = the squared correlation between actual score and predicted score. $S_{y\cdot x}$ = the Standard Error of Multiple Estimate: $S_{y\cdot x} = S_y\sqrt{1 - R^2}$.

Appendix F

RAND–36 HSI Items and Response Options by Composite and Scale

Physical Health Composite

Physical Functioning Scale

The following items are about activities you might do during a typical day. Does **your health now limit you** in these activities? If so, how much?

Response Options: 1 = Yes, limited a lot
2 = Yes, limited a little
3 = No, not limited at all

3. **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports
4. **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf
5. Lifting or carrying groceries
6. Climbing **several** flights of stairs
7. Climbing **one** flight of stairs
8. Bending, kneeling, or stooping
9. Walking **more than a mile**
10. Walking **several blocks**
11. Walking **one block**
12. Bathing or dressing yourself

Role Limitations due to Physical Health Problems Scale

During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities **as a result of your physical health**?

Response Options: 1 = Yes
2 = No

13. Cut down the **amount of time** you spent on work or other activities
14. **Accomplished less** than you would like
15. Were limited in the **kind** of work or other activities
16. Had **difficulty** performing the work or other activities (for example, it took extra effort)

Pain Scale

21. How much **bodily** pain have you had during the **past 4 weeks?**

Response Options: 1 = None
2 = Very mild
3 = Mild
4 = Moderate
5 = Severe
6 = Very severe

22. During the **past 4 weeks**, how much did **pain** interfere with your normal work (including both work outside the home and housework)?

Response Options: 1 = Not at all
2 = A little bit
3 = Moderately
4 = Quite a bit
5 = Extremely

General Health Perceptions Scale

1. In general, would you say your health is:

Response Options: 1 = Excellent
2 = Very good
3 = Good
4 = Fair
5 = Poor

How true or false is **each** of the following statements for you?

Response Options: 1 = Definitely true
2 = Mostly true
3 = Don't know
4 = Mostly false
5 = Definitely false

33. I seem to get sick a little easier than other people.
34. I am as healthy as anybody I know.
35. I expect my health to get worse.
36. My health is excellent.

Mental Health Composite

Emotional Well-Being Scale

The following questions are about how you feel and how things have been with you **during the past 4 weeks**. For each question, please give the one answer that comes closest to the way you have been feeling.

How much time during the **past 4 weeks**:

Response Options: 1 = All of the time
2 = Most of the time
3 = A good bit of the time
4 = Some of the time
5 = A little of the time
6 = None of the time

24. Have you been a very nervous person?
25. Have you felt so down in the dumps that nothing could cheer you up?
26. Have you felt calm and peaceful?
28. Have you felt downhearted or blue?
30. Have you been a happy person?

Role Limitations due to Emotional Problems Scale

During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities **as a result of any emotional problems** (such as feeling depressed or anxious)?

Response Options: 1= Yes
2 = No

17. Cut down the **amount of time** you spent on work or other activities
18. **Accomplished less** than you would like
19. Didn't do work or other activities as **carefully** as usual

Social Functioning Scale

20. During the **past 4 weeks**, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbors, or groups?

    Response Options: 1 = Not at all
    2 = Slightly
    3 = Moderately
    4 = Quite a bit
    5 = Extremely

32. During the **past 4 weeks**, how much of the time has your **physical health or emotional problems** interfered with your social activities (like visiting with friends, relatives, etc.)?

    Response Options: 1 = All of the time
    2 = Most of the time
    3 = Some of the time
    4 = A little of the time
    5 = None of the time

Energy/Fatigue Scale

The following questions are about how you feel and how things have been with you **during the past 4 weeks**. For each question, please give the one answer that comes closest to the way you have been feeling.

How much time during the **past 4 weeks**:

Response Options: 1 = All of the time
2 = Most of the time
3 = A good bit of the time
4 = Some of the time
5 = A little of the time
6 = None of the time

23. Did you feel full of pep?
27. Did you have a lot of energy?
29. Did you feel worn out?
31. Did you feel tired?

Appendix G
RAND–12 HSI Items and Response Options by Composite

Physical Health Composite

1. In general, would you say your health is:

Response Options: 1 = Excellent
2 = Very good
3 = Good
4 = Fair
5 = Poor

The following items are about activities you might do during a typical day. Does **your health now limit you** in these activities? If so, how much?

Response Options: 1 = Yes, limited a lot
2 = Yes, limited a little
3 = No, not limited at all

2. **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf
3. Climbing **several** flights of stairs

During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities **as a result of your physical health**?

Response Options: 1 = Yes
2 = No

4. **Accomplished less** than you would like
5. Were limited in the **kind** of work or other activities

8. During the **past 4 weeks**, how much did **pain** interfere with your normal work (including both work outside the home and housework)?

Response Options: 1 = Not at all
2 = A little bit
3 = Moderately
4 = Quite a bit
5 = Extremely

Mental Health Composite

During the **past 4 weeks**, have you had any of the following problems with your work or other regular daily activities **as a result of any emotional problems** (such as feeling depressed or anxious)?

Response Options: 1 = Yes
2 = No

6. **Accomplished less** than you would like
7. Didn't do work or other activities as **carefully** as usual

The following questions are about how you feel and how things have been with you **during the past 4 weeks**. For each question, please give the one answer that comes closest to the way you have been feeling.

How much time during the **past 4 weeks**:

Response Options: 1 = All of the time
2 = Most of the time
3 = A good bit of the time
4 = Some of the time
5 = A little of the time
6 = None of the time

9. Have you felt calm and peaceful?
10. Did you have a lot of energy?
11. Have you felt downhearted or blue?

12. During the **past 4 weeks**, how much of the time has your **physical health or emotional problems** interfered with your social activities (like visiting with friends, relatives, etc.)?

Response Options: 1 = All of the time
2 = Most of the time
3 = Some of the time
4 = A little of the time
5 = None of the time

References

Aaronson, N. K., Acquadro, C., Alonso, J., Apolone, G., Bucquet, D., Bullinger, M., Bungay, K., Fukuhara, S., Gandek, B., Keller, S., Razavi, R., Sanson–Fisher, M., Sullivan, S., Wood–Dauphinee, S., Wagner, A., & Ware, J. E., Jr. (1992). International quality of life assessment (IQOLA) project. *Quality of Life Research, 1*(5), 349–351.

Andresen, E. M., Bowley, N., Rothenberg, B. M., Panzer, R., & Katz, P. (1996). Test–retest performance of a mailed version of the Medical Outcomes Study 36-Item Short-Form Health Survey among older adults. *Medical Care, 34*(12), 1165–1170.

Barr, J. T., & Schumacher, G. E. (1995). A teaching strategy to incorporate patient-centered outcomes assessment in health professional curricula: The class-generated patient profile. *Journal of Allied Health, 24*(1), 65–78.

Beck, A. T., & Steer, R. A. (1988). *Beck Hopelessness Scale: Manual.* San Antonio, TX: The Psychological Corporation.

Beck, A. T., & Steer, R. A. (1990). *Beck Anxiety Inventory: Manual.* San Antonio, TX: The Psychological Corporation.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory: Manual* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Brazier, J. E., Harper, R., Jones, N. M., O'Cathain, A., Thomas, K. J., Usherwood, T., & Westlake, L. (1992). Validating the SF–36 health survey questionnaire: New outcome measure for primary care. *British Medical Journal, 305,* 160–164.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Davies, A. R., & Ware, J. E., Jr. (1981). *Measuring health perceptions in the health insurance experiment* (Publication No. R–2711–HHS). Santa Monica, CA: The RAND Corporation.

Derogatis, L. R. (1993). *Brief Symptom Inventory: Administration, scoring, and procedures manual.* Minneapolis, MN: National Computer Systems.

Dexter, P. R., Stump, T. E., Tierney, W. M., & Wolinsky, F. D. (1996). The psychometric properties of the SF–36 health survey among older adults in a clinical setting. *Journal of Clinical Geropsychology, 2*(3), 225–237.

Eisen, S. V., Dill, D. L., & Grob, M. C. (1994). Reliability and validity of a brief patient-report instrument for psychiatric outcome evaluation. *Hospital and Community Psychiatry, 45*(3), 242–247.

Fowler, F. J., Jr., Wennberg, J. E., Timothy, R. P., Barry, M. J., Mulley, A. G., Jr., & Hanley, D. (1988). Symptom status and quality of life following prostatectomy. *Journal of the American Medical Association, 259*(20), 3018–3022.

Ganz, P. A., Day, R., Ware, J. E., Jr., Redmond, C., & Fisher, B. (1995). Base-line quality-of-life assessment in the National Surgical Adjuvant Breast and Bowel Project Breast Cancer Prevention Trial. *Journal of the National Cancer Institute, 87*(18), 1372–1382.

Garratt, A. M., Ruta, D. A., Abdalla, M. I., & Russell, I. T. (1996). Responsiveness of the SF–36 and a condition-specific measure of health for patients with varicose veins. *Quality of Life Research, 5*(2), 223–234.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gorsuch, R. L. (1991). *UniMult guide for univariate and multivariate data analysis.* Altadena, CA: UniMult.

Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education* (6th ed.). New York: McGraw–Hill.

Haley, S. M., McHorney, C. A., & Ware, J. E., Jr. (1994). Evaluation of the MOS SF–36 Physical Functioning Scale (PF–10): I. Unidimensionality and reproducibility of the Rasch Item Scale. *Journal of Clinical Epidemiology, 47*(6), 671–684.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston: Kluwer.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Harris, W. H. (1969). Traumatic arthritis of the hip after dislocation and acetabular fractures: Treatment by mold arthroplasty: An end-result study using a new method of result evaluation. *Journal of Bone and Joint Surgery, 51*(4), 737–755.

Hays, R. D. (1998). Item response theory models. In M. J. Staquet, R. D. Hays, & P. M. Fayers (Eds.), *Quality of life assessment in clinical trials: Methods and practice* (pp. 183–190). Oxford: Oxford University Press.

Hays, R. D., & Hayashi, T. (1990). Beyond internal consistency reliability: Rationale and user's guide for Multitrait Analysis Program on the microcomputer. *Behavior Research Methods, Instruments, & Computers, 22*(2), 167–175.

Hays, R. D., Kallich, J. D., Mapes, D. L., Coons, S. J., & Carter, W. B. (1994). Development of the Kidney Disease Quality of Life (KDQOL) Instrument. *Quality of Life Research, 3*(5), 329–338.

Hays, R. D., Marshall, G. N., Wang, E. Y. I., & Sherbourne, C. D. (1994). Four-year cross-lagged associations between physical and mental health in the medical outcomes study. *Journal of Consulting and Clinical Psychology, 62*(3), 441–449.

Hays, R. D., Sherbourne, C. D., & Mazel, R. M. (1993). The RAND 36–Item Health Survey 1.0. *Health Economics, 2,* 217–227.

Hays, R. D., Sherbourne, C. D., & Mazel, R. M. (1995). *User's manual for the Medical Outcomes Study (MOS) core measures of health-related quality of life.* Santa Monica, CA: The RAND Corporation.

Hays, R. D., Wells, K. B., Sherbourne, C. D., Rogers, W., & Spritzer, K. (1995). Functioning and well-being outcomes of patients with depression compared with chronic general medical illnesses. *Archives of General Psychiatry, 52*(1), 11–19.

Hill, S., Harries, U., & Popay, J. (1996). Is the short form 36 (SF–36) suitable for routine health outcomes assessment in health care for older people? Evidence from preliminary work in community based health services in England. *Journal of Epidemiology and Community Health, 50*(1), 94–98.

Insall, J. N., Dorr, L. D., Scott, R. D., & Scott, W. N. (1989). Rationale of the Knee Society clinical rating system. *Clinical Orthopaedics and Related Research, 248,* 13–14.

Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15,* 336–352.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19.

Kantz, M. E., Harris, W. J., Levitsky, K., Ware, J. E., Jr., & Davies, A. R. (1992). Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. *Medical Care, 30*(5, Suppl.), MS240–MS252.

Keller, S. D., Ware, J. E., Jr., & Gandek, B. (1995). Equivalence of translations of response choices widely used in health questionnaire: Comparison of results from nine countries in the IQOLA project. *Quality of Life Research, 4,* 445.

Kurtin, P. S., Davies, A. R., Meyer, K. B., DeGiacomo, J. M., & Kantz, M. E. (1992). Patient-based health status measures in outpatient dialysis. Early experiences in developing an outcomes assessment program. *Medical Care, 30*(5, Suppl.), MS136–MS149.

Lansky, D., Butler, J. B., & Waller, F. T. (1992). Using health status measures in the hospital setting: From acute care to "outcomes management." *Medical Care, 30* (5, Suppl.), MS57–MS73.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140,* 5–55.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149–174.

McHorney, C. A., Haley, S. M., & Ware, J. E., Jr. (1997). Evaluation of the MOS SF–36 Physical Functioning Scale (PF–10): II. Comparison of relative precision using Likert and Rasch scoring methods. *Journal of Clinical Epidemiology, 50*(4), 451–461.

McHorney, C. A., Kosinski, M., & Ware, J. E., Jr. (1994). Comparisons of the costs and quality of norms for the SF–36 health survey collected by mail versus telephone interview: Results from a national survey. *Medical Care, 32*(6), 551–567.

McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research, 4*(4), 293–307.

McHorney, C. A., Ware, J. E., Jr., Lu, J. F. R., & Sherbourne, C. D. (1994). The MOS 36-Item Short-Form Health Survey (SF–36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care, 32*(1), 40–66.

McHorney, C. A., Ware, J. E., Jr., & Raczek, A. E. (1993). The MOS 36-Item Short-Form Health Survey (SF–36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care, 31(3),* 247–263.

McHorney, C. A., Ware, J. E., Jr., Rogers, W., Raczek, A. E., & Lu, J. F. R. (1992). The validity and relative precision of MOS Short- and Long-Form Health Status Scales and Dartmouth COOP Charts: Results from the Medical Outcomes Study. *Medical Care, 30*(5, Suppl.), MS253–MS265.

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika, 8,* 161–168.

Nelson, D. C., Hartman, E., Ojemann, P. G., & Wilcox, M. (1995). Outcomes measurement and management with a large Medicaid population: A public–private collaboration. *Behavioral Healthcare Tomorrow, 4*(3), 31–37.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw–Hill.

Patrick, D. L., & Deyo, R. A. (1989). Generic and disease-specific measures in assessing health status and quality of life. *Medical Care, 27*(3, Suppl.), S217–S232.

The Psychological Corporation. (1998). *OPTAIO Provider's Desktop* (Version 1.3) [Computer software]. San Antonio, TX: Author.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1966). An item analysis which takes individual differences into account. *British Journal of Mathematical and Statistical Psychology, 19,* 49–57.

Rummel, R. J. (1970). *Applied factor analysis.* Evanston, IL: Northwestern University Press.

Shadbolt, B., McCallum, J., & Singh, M. (1997). Health outcomes by self-report: Validity of the SF–36 among Australian hospital patients. *Quality of Life Research, 6*(4), 343–352.

Sherbourne, C. D., Wells, K. B., Hays, R. D., Rogers, W., Burnam, M. A., & Judd, L. L. (1994). Subthreshold depression and depressive disorder: Clinical characteristics of general medical and mental health specialty outpatients. *American Journal of Psychiatry, 151*(12), 1777–1784.

Sherbourne, C. D., Wells, K. B., & Judd, L. L. (1996). Functioning and well-being of patients with panic disorder. *American Journal of Psychiatry, 153*(2), 213–218.

Shiely, J-C., Bayliss, M. S., Keller, S. D., Tsai, C., & Ware, J. E., Jr. (1996). *SF–36 Health Survey Annotated Bibliography: First Edition (1988–1995).* Boston: The Health Institute, New England Medical Center.

Solomon, G. D. (1997). Evolution of the measurement of quality of life in migraine. *Neurology, 48*(3, Suppl. 3), S10–S15.

Stewart, A. L., Hays, R. D., & Ware, J. E., Jr. (1988). The MOS Short-Form General Health Survey: Reliability and validity in a patient population. *Medical Care, 26*(7), 724–735.

Stewart, A. L., Sherbourne, C. D., Hays, R. D., Wells, K. B., Nelson, E. C., Kamberg, C. J., Rogers, W. H., Berry, S. H., & Ware, J. E., Jr. (1992). Summary and discussion of MOS measures. In A. L. Stewart & J. E. Ware, Jr. (Eds.), *Measuring functioning and well-being: The Medical Outcomes Study approach* (pp. 345–371). Durham, NC: Duke University Press.

Tarlov, A. R., Ware, J. E., Jr., Greenfield, S., Nelson, E. C., Perrin, E., & Zubkoff, M. (1989). The Medical Outcomes Study: An application of methods for monitoring the results of medical care. *Journal of the American Medical Association, 262*(7), 925–930.

Titler, M. G., & Reiter, R. C. (1994). Outcomes measurement in clinical practice. *MEDSURG Nursing, 3*(5), 395–398.

U.S. Bureau of the Census. (1993). *Current Population Survey, March 1993, on CD–ROM* [CD–ROM]. Washington, DC: U.S. Bureau of the Census (Producer/Distributor).

U.S. Bureau of the Census. (1995). *Statistical Abstract of the United States 1995.* Washington, DC: Author.

Vickrey, B. G., Hays, R. D., Rausch, R., Sutherling, W. W., Engel, J., Jr., & Brook, R. H. (1994). Quality of life of epilepsy surgery patients as compared with outpatients with hypertension, diabetes, heart disease, and/or depressive symptoms. *Epilepsia, 35*(3), 597–607.

Wagner, A. K., Cynn, D. J., Ehrenberg, B. L., Tran, T. A., Bungay, K. M., & Rogers, W. H. (1995). The impact of routine health status measurement on aspects of epilepsy patients' care. *Epilepsia, 36*(Suppl. 4), 96.

Wagner, A. K., Ehrenberg, B. L., Tran, T. A., Bungay, K. M., Cynn, D. J., & Rogers, W. H. (1997). Patient-based health status measurement in clinical practice: A study of its impact on epilepsy patients' care. *Quality of Life Research, 6*(4), 329–341.

Ware, J. E., Jr., Gandek, B., & the IQOLA Project Group. (1994). The SF–36 Health Survey: Development and use in mental health research and the IQOLA Project. *International Journal of Mental Health, 23*(2), 49–73.

Ware, J. E., Jr., Kosinski, M., Bayliss, M. S., McHorney, C. A., Rogers, W. H., & Raczek, A. (1995). 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. *Medical Care, 33*(4), AS264–AS279.

Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1994). *SF–36 physical and mental health summary scales: A user's manual.* Boston: The Health Institute.

Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1995). *SF–12: How to score the SF–12 physical and mental health summary scales* (2nd ed.). Boston: The Health Institute.

Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*(3), 220–233.

Ware, J. E., Jr., Nelson, E. C., Sherbourne, C. D., & Stewart, A. L. (1992). Preliminary tests of a 6-item general health survey: A patient application. In A. L. Stewart & J. E. Ware, Jr. (Eds.), *Measuring functioning and well-being: The Medical Outcomes Study approach* (pp. 291–308). Durham, NC: Duke University Press.

Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-Item Short-Form Health Survey (SF–36): I. Conceptual framework and item selection. *Medical Care, 30*(6), 473–483.

Ware, J. E., Jr., Sherbourne, C. D., & Davies, A. R. (1992). Developing and testing the MOS 20-item short-form health survey: A general population application. In A. L. Stewart & J. E. Ware, Jr. (Eds.), *Measuring functioning and well-being: The Medical Outcomes Study approach* (pp. 277–290). Durham, NC: Duke University Press.

Ware, J. E., Jr., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF–36 Health Survey: Manual and interpretation guide.* Boston: The Health Institute.

Weissman, M. M., & Bothwell, S. (1976). Assessment of social adjustment by patient self-report. *Archives of General Psychiatry, 33*(9), 1111–1115.