A GUIDE TO THE DEVELOPMENT OF CERTIFIED MODES OF SHORT FORM SURVEY ADMINISTRATION

Mark E. Maruish, PhD
Diane M. Turner-Bowker, PhD
# Table of Contents

Copyright .......................................................................................................................................................... ii  
Table of Contents ............................................................................................................................................ iii  
Preface ............................................................................................................................................................. vii  
How to Use This Guide ................................................................................................................................... ix  

## Chapter 1 Introduction

What is Standardization? ............................................................................................................................... 1  
Advantages of Standardization .................................................................................................................. 2  
Objectivity ................................................................................................................................................... 2  
Communication .......................................................................................................................................... 2  
Quantification .............................................................................................................................................. 2  
Scientific Generalization ........................................................................................................................... 2  
Economy ..................................................................................................................................................... 3  
QualityMetric’s Vendor Certification Program ........................................................................................ 3  
Purpose of this Guide .................................................................................................................................... 3  

## Chapter 2 The Short Form Family of Health Surveys

QualityMetric’s Short Form Surveys ........................................................................................................... 5  
The SF-36® Health Survey ........................................................................................................................... 5  
The SF-36v2® Health Survey ....................................................................................................................... 6  
The SF-12® Health Survey ........................................................................................................................... 6  
The SF-12v2® Health Survey ....................................................................................................................... 7  
The SF-8™ Health Survey ........................................................................................................................... 7  
The DYNHA® SF-36® Health Survey ........................................................................................................ 7  
The SF-10™ Health Survey for Children .................................................................................................... 7  
Enhanced Versions of the Short Form Surveys .......................................................................................... 8  
The SF-36v2®/SF-12v2® Health Survey–PIQ-6 Enhanced ........................................................................ 8  
The SF-12v2® Health Survey–SET ............................................................................................................ 8  
The SF-12v2® Health Survey–MH Enhanced .......................................................................................... 9  
The SF-12v2® Health Survey–PF Enhanced .............................................................................................. 9  
The SF-12v2® Health Survey with HRA ................................................................................................. 9  
Short Form Standard and Acute Recall Forms .......................................................................................... 10  
Translated Versions of the Short Form Surveys ...................................................................................... 10  
Short Form Concepts and Measures ........................................................................................................ 11  
Health Domain Scales ............................................................................................................................... 11  
Physical Functioning (PF) ......................................................................................................................... 11  
Role-Physical (RP) .................................................................................................................................... 11  
Bodily Pain (BP) .......................................................................................................................................... 13  
General Health (GH) .................................................................................................................................. 13
# Table of Contents

- Vitality (VT) .......................................................................................................... 13
- Social Functioning (SF) ........................................................................................ 13
- Role-Emotional (RE) ............................................................................................ 13
- Mental Health (MH) .............................................................................................. 13
- Self-Evaluated Transition (SET) .......................................................................... 13
- Physical and Mental Component Summary Measures (PCS and MCS) .......... 13
- Short Form Survey Profile of Scores ................................................................. 16
- SF-6D Health Utility Index ................................................................................... 17
- Medical Expenditure Prediction .......................................................................... 17
- Short Form Survey Applications ......................................................................... 17
- Evaluating and Monitoring Individual Patients in Clinical Practice .......... 17
- Monitoring Populations ..................................................................................... 18
- Estimating the Burden of Disease .................................................................... 18
- Evaluating Treatment Effects in Clinical Trials .............................................. 18
- Disease Management ......................................................................................... 19
- Risk Prediction and Cost-Effectiveness ............................................................ 19
- Patient-Provider Relations ................................................................................ 19
- Direct-to-Consumer Information ...................................................................... 19

## Chapter 3  Survey Administration: General Considerations

- Determining Respondent Eligibility ................................................................. 21
- Age .................................................................................................................... 21
- Reading Ability ............................................................................................... 22
- Non-English Speaking Respondents ............................................................... 22
- Guidelines for Administration .......................................................................... 23
- When to Administer the Survey ..................................................................... 23
- Administration of a Full Survey versus Selected Scales or Items .................. 24
- Administration of Short Form Surveys with Other Generic or Disease-Specific Surveys .......................................................... 24
- Introducing the Short Form Surveys to the Respondent ............................... 24
- Addressing Problems and Questions .............................................................. 24
- Closing ............................................................................................................. 24
- Additional Considerations ............................................................................... 25
- When to Readminister a Short Form Survey ................................................ 25
- Order effects .................................................................................................. 26

## Chapter 4  Considerations for Modes of Short Form Survey Administration

- General Considerations for All Modes of Short Form Survey Administration 27
- Administration by Paper Form .......................................................................... 28
- Mode-Specific Considerations ........................................................................ 28
- Certification Standards .................................................................................... 28
- Administration by Mail ..................................................................................... 29
- Mode-Specific Considerations ........................................................................ 29
- Certification Standards .................................................................................... 29
- Administration by Interview .......................................................................... 32
- Mode-Specific Considerations ........................................................................ 32
- Certification Standards .................................................................................... 32
- Administration by Internet, Desktop, or Handheld Device .......................... 33
- Mode-Specific Considerations ........................................................................ 33
- Certification Standards .................................................................................... 33
Administration by IVRS ................................................................. 35
Mode-Specific Considerations ......................................................... 35
Certification Standards .................................................................. 35
Administration by Translated Version of a Short Form Survey .......... 36
Mode-Specific Considerations ......................................................... 36
Certification Standards .................................................................. 36
Effects of Mode of Administration on Short Form Survey Results .... 37

References .................................................................................. 47

Appendix  Checklists for Development of Certified Modes of Administration
A.1 Checklist for the Development of a Certified Mode of Administration:
   Paper Form ................................................................................. 55
A.2 Checklist for the Development of a Certified Mode of Administration:
   Mail-Out Form ........................................................................... 57
A.3 Checklist for the Development of a Certified Mode of Administration:
   Internet, Desktop, or Handheld Device ..................................... 59
A.4 Checklist for the Development of a Certified Mode of Administration:
   Interactive Voice Response Systems ...................................... 61

Figures
Figure 2.1 Short Form Health Status Measurement Model .................. 15
Figure 2.2 Sample Short Form Health Survey Profile of Scores .......... 16
Figure 3.1 Recommended Steps for Administering the Short Form Surveys 23
Figure 4.1 Item 5 from the SF-36v2 Paper Form ................................ 31
Figure 4.2 Item 6 from the SF-36v2 Paper Form ................................ 31
Figure 4.3 Items 5 and 6 from an Online SF-36v2 Administration ........ 34

Tables
Table 2.1 Summary of Short Form Health Survey Similarities and Differences........ 11
Table 2.2 Abbreviated Item Content for the Short Form Health Domain Scales .......... 12
Table 3.1 Short Form Survey Administration Dos and Don’ts .................. 25
Table 4.1 Recommended Design Characteristics for HRQOL Self-Report Instruments 30
Table 4.2 Effects of Method of Data Collection on Short Form Results: Findings from
   Selected Studies ....................................................................... 38
Some of the “Short Form family” of health status instruments developed by QualityMetric Incorporated—the SF-36\textsuperscript{v2} \textsuperscript{®} Health Survey (SF-36\textsuperscript{v2}), DYNHA\textsuperscript{®} SF-36\textsuperscript{®} Health Survey (DYNHASF-36), SF-36\textsuperscript{®} Health Survey (SF-36), SF-12\textsuperscript{v2} \textsuperscript{®} Health Survey (SF-12\textsuperscript{v2}), SF-12\textsuperscript{®} Health Survey (SF-12), SF-8\textsuperscript{™} Health Survey (SF-8), and SF-10\textsuperscript{™} Health Survey for Children (SF-10)—is the most widely used set of health status assessment tools in clinical trials (Garratt, Schmidt, Mackintosh, & Fitzpatrick, 2002). The availability of more than 110 language translations and English-language adaptations of these instruments facilitates the adoption of the Short Form surveys in randomized clinical trials (RCTs); the more than 1,800 published articles documenting RCTs utilizing the Short Form instruments attest to their utility and acceptance in the scientific community.

Widespread use of the Short Form instruments for research and clinical purposes has been accompanied by the growing need for varied modes for administering, scoring, and reporting the results of administration, modes employing the latest technology. QualityMetric Incorporated has responded to this demand by making available a wide range of assessment capabilities for the Short Form surveys. In addition to the standard self-report paper-form format, the Short Form surveys can now be administered, scored, and/or have their results reported using interview scripts, approved translations of the surveys, and applications developed for desktop computers, the Internet, handheld devices, interactive voice response systems (IVRS), and faxback systems.

Because of the rapidly expanding worldwide demand for its commercial products and services, QualityMetric™ is expanding its licensing program for vendors and those involved in academic and commercially sponsored research who intend to develop their own means for administering, scoring, and reporting the results of the administration of Short Form instruments using QualityMetric’s trademarks, copyrighted forms, scoring algorithms, norms, and interpretive guidelines. As a condition for licensing, QualityMetric now requires that any Short Form survey products developed by outside parties be certified as meeting standards that ensure end-users that these products meet standards of high quality, in terms of maintaining standardization and accuracy in survey administration, scoring, and reporting and interpretation of survey results. To assist licensees, QualityMetric provides a series of three guides that describe important aspects of administration, scoring, and reporting and interpretation of survey results and the guidelines and criteria that must be met in order for Short Form assessment products to be certified by QualityMetric. The underlying theme of each of these guides is the importance of maintaining standardization in any assessment process utilizing any of the Short Form surveys.

The primary purpose of this first guide in the three-part series is to provide those developing Short Form survey forms and software with the guidelines and criteria that QualityMetric uses to license and certify modes of administration developed by parties other than itself. This guide also serves to provide the background information necessary for understanding the requirements of QualityMetric’s certification program. In this regard, the first chapter emphasizes the importance of standardization in the
development and use of health status surveys and other psychometric instruments. Chapter 2 provides a description of Short Form concepts, measures, and forms as an introduction to the administration procedures and development criteria presented in this guide, as well as to the scoring and interpretation criteria discussed in the second and third guides, respectively, in the series. Chapter 3 presents general guidelines for administering the Short Form surveys, including determining eligibility for completing the surveys, when and how to administer them, and other considerations. In Chapter 4, guidelines and criteria are provided for developing common, approved modes of Short Form survey administration, along with information on the comparability of scores across modes (including results from several published studies).

It is important to note that with the publication of this certification guide series, QualityMetric introduces three changes in terminology that it has been using in its commercial and peer-reviewed publications for over a decade. First, what was previously called “norm-based scores” is now referred to as “T scores.” Also, the set of procedures used to maximize the amount of useable Short Form data, previously referred to as “Missing Data Estimation (MDE),” is now called “Missing Score Estimation (MSE).” Finally, the “Reported Health Transition (HT)” item that is part of the SF-36 and SF-36v2 surveys is now referred to as the “Self-Evaluated Transition (SET)” item. The reason for these changes is to more precisely describe what each term represents and thus minimize misconceptions about what the term means among users of the Short Form surveys and other consumers of information derived from the Short Form surveys. Although the terminology has changed, what each term represents and how it’s used remains unchanged.

Information about each Short Form survey is available in its user’s manual, referenced in Chapter 1. In particular, the *User’s Manual for the SF-36v2 Health Survey, Second Edition* (Ware et al., 2007) provides a comprehensive history of the Short Form surveys, including a synopsis of the investigations leading to the development of the original SF-36 and a discussion of the improvements to the original version of the survey that have been incorporated into the SF-36v2 and SF-12v2. The SF-36v2 manual also provides form-to-form comparisons that can be useful for selecting the most appropriate form for a given application (i.e., for various clinical or research needs).

Additional information about the history, development, psychometric properties, and use of the Short Form surveys is available from several sources. A comprehensive overview of the SF-36 survey is found in “SF-36 Health Survey Update,” a chapter by J.E. Ware, Jr., in *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment, Volume 3, Third Edition* (Maruish [Ed.], 2004). More recently, Ware (2008) published an invited commentary in the *Journal of Clinical Epidemiology* summarizing the improvements to the SF-36 that led to the development of the SF-36v2. This commentary was the first article in a special series, called “Improvements in Short Form (SF) Measures of Health Status,” appearing in this journal. Other sources of information about the development and empirical testing of the Short Forms include several articles published in the November 1998 issue of the *Journal of Clinical Epidemiology* (Gandek & Ware [Eds.], 1998b) documenting domestic and international studies of improvements in item wording and response categories that were evaluated pursuant to the International Quality of Life Assessment (IQOLA) Project, as well as more current studies on translations (e.g., Bjorner & Turner-Bowker, in press; Han, Lee, Iwaya, Kataoka, & Kohzuki, 2004; Turner-Bowker, DeRosa, & Ware, 2007).

Numerous other independent articles have been published and can be found at [http://www.qualitymetric.com/](http://www.qualitymetric.com/), QualityMetric Incorporated’s marketplace Web site, and at [http://www.sf-36.org/](http://www.sf-36.org/), the Web site for the users of the Short Form family of instruments. Information about the Short Form surveys and the opportunity to discuss related issues with its developers are also available on the Internet at [http://www.sf-36.org/](http://www.sf-36.org/).
Suggestions on how to use this guide and quickly find particular information are offered below. The reader is also referred to the Table of Contents to locate more specific information.

**Introduction**

*Chapter 1* details the purpose of this guide and includes a discussion of the importance of maintaining standardization in the development of surveys and modes of survey administration.

**Concepts, measures, and forms**

*Chapter 2* presents an overview of QualityMetric’s Short Form survey concepts, measures, and forms, as well as a discussion of the surveys’ applications.

**Administration**

*Chapter 3* provides important, general considerations and guidelines for maintaining standardization in the administration of the Short Form surveys.

**Modes of Short Form survey administration**

*Chapter 4* reviews acceptable modes for administering the Short Form surveys and QualityMetric’s criteria for certifying Short Form forms and electronic products developed by its licensees.

**Research on modes of administration**

*Chapter 4* also presents a summary of key findings from selected studies investigating different modes of administration using the Short Form surveys.
Introduction

Like many other psychometric measurement tools, the keystone of the “Short Form family” of health status instruments developed by QualityMetric Incorporated—the SF-36v2® Health Survey (SF-36v2; Ware et al., 2007), DYNHA® SF-36® Health Survey (DYNHA SF-36), SF-36® Health Survey (SF-36; Ware, Snow, Kosinski, & Gandek, 1993), SF-12v2® Health Survey (SF-12v2; Ware, Kosinski, Turner-Bowker, & Gandek, 2002), SF-12® Health Survey (SF-12; Ware, Kosinski, & Keller, 1995), SF-8™ Health Survey (SF-8; Ware, Kosinski, Dewey, & Gandek, 2001), and SF-10™ Health Survey for Children (SF-10; Saris-Baglama, DeRosa, et al., 2007)—is standardization. The purpose of this chapter is to briefly discuss what standardization means and the advantages of using standardized metrics and methods in health status assessment. Maintaining standardization in the administration, scoring, and reporting capabilities of Short Form and other QualityMetric™ products developed by licensed vendors is the primary concern of QualityMetric. It serves as the basis (and criterion) for QualityMetric’s QM Certified Scoring program, which is also discussed in this chapter.

Finally, this chapter will also describe the purpose of this and the other two guides in this Short Form certification series and identify resources for additional information on each of QualityMetric’s Short Form instruments.

What Is Standardization?

In the context of psychometric tests and surveys, standardization refers to “uniformity of procedure in administering and scoring a test [or survey]” (Anastasi, 1988, p. 25). Standardization is necessary to help ensure comparability of results from a test administered by different survey users. Anastasi notes that “the formulation of directions is a major part of the standardization of a new test. Such standardization extends to the exact materials employed, time limits, oral instructions, preliminary demonstrations, ways of handling queries, and every other detail of the testing situation” (p. 25). As Urbina (2004) states, “The purpose of standardizing test procedures is to make all of the variables that are under the control of the examiner as uniform as possible, so that everyone who takes the test will be taking it the same way” (p. 2). In addition, both Anastasi and Urbina point to the development and use of norms as being a major component of standardization.

The American Education Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) specify the same criteria for standardization—consistency in test materials, administration procedures, and application of scoring procedures—in Standards for Educational and Psychological Testing (1999). This 1999 edition is the sixth in a series of such publications (the first of which was published in 1954) that is designed to guide the development and use of tests. For 50 years, these standards have guided developers and users of tests, surveys, and other psychometric measures by providing “criteria for the evaluation of tests, testing practices, and the effects of test use” and acting as a “frame of reference to assure that relevant issues are addressed”
This comprehensive, detailed set of standards guided the development of all Short Form instruments and their accompanying user’s manuals, not only to conform to the expectations of the scientific and clinical communities but also to ensure that these instruments provide a meaningful contribution to the field of health status and quality of life measurement.

Standard 11.19 of the Standards document states that:

When a test user contemplates an approved change in test format, mode of administration, instructions, or language used in administering the test, the user should have a sound rationale for concluding that validity, reliability, and appropriateness of norms will not be compromised. (p. 117)

This standard provides the rationale for QualityMetric’s program of certification for all Short Form surveys, administration, and scoring materials and of software developed and used by certified vendors for resale and by researchers for their own purposes. AERA, APA, and NCME recognize that there are times when “flexibility” in administration or scoring may be justified by the demands of the situation. However, they are also quick to note that “each step toward greater flexibility almost inevitably enlarges the scope and magnitude of measurement error” (p. 26). In this regard, it is important to understand that any changes in the scoring of the Short Form survey items, scales, or component summary measure algorithms may compromise their validity and reliability, invalidate normative comparisons, and/or complicate or prevent meaningful comparisons across studies.

Advantages of Standardization

Maintaining standardization in the administration, scoring, and interpretation of psychological, health status, and other types of psychometric measurement instruments has several advantages for the users of the instruments. In their discussion of standardization, Nunnally and Bernstein (1994) identified five advantages of standardized measurement over personal judgments. Although nonstandardized measures, or standardized measures being administered in ways other than the standardized manner, are not the same as personal judgments or subjective assessments, these same five advantages, discussed in the following sections, nonetheless apply: objectivity, communication, quantification, scientific generalization, and economy.

Objectivity

As Nunnally and Bernstein (1994) point out, one of the key principles of science is that any factual statement made by one observer should be able to be verified by other observers acting independently. In order to do this, the observers must be able to agree on what is being measured and how to measure it. Only through the use of measurement that is standardized can this be accomplished.

Communication

Use of standardized methods of measurement greatly facilitates accurate and efficient communication of results between observers of the same phenomenon (Nunnally & Bernstein, 1994). Use of a standardized instrument of measurement, administered and scored in a standardized manner, ensures that what is communicated by one observer will be accurately understood by other observers because the findings will have the same meaning to all who are familiar with the instrument.

Quantification

Standardized measures allow findings to be quantified; that is, represented numerically (Nunnally & Bernstein, 1994). This certainly is more advantageous than subjective assessments from the standpoint of allowing finer differentiations and more exact estimates of changes in that which is being measured. Moreover, it allows the observer to make use of mathematical and statistical methods of analyses with greater confidence in the reliability and validity of the findings.

Scientific Generalization

Quantification of findings is necessary for the formulation and testing of hypotheses,
which in turn is necessary for generalizing the findings beyond one particular situation or set of circumstances—an important consideration for scientific research (Nunnally & Bernstein, 1994). Having findings with the known reliability and validity that comes from using a data gathering instrument in its standardized manner strengthens one’s confidence in results of hypothesis testing and, consequently, one’s ability to generalize the findings to other circumstances. This is a particularly important consideration when using patient-reported outcomes (PRO) data from a Short Form survey to help guide clinical decision making and treatment outcome prediction or to support the psychometric integrity of the measurement when making claims of treatment efficacy, such as those presented for approval to the Federal Drug Administration for the purpose of making labeling claims for a new drug.

**Economy**

Nunnally and Bernstein (1994) note that, when compared to assessments based on personal judgments, standardized assessments are more economical in terms of both time and money. Stated in a somewhat different way, maintaining standardization in the administration and scoring of measures versus using measures in nonstandardized ways can result in savings of time and/or money. Such savings come from not having to provide evidence of the validity or reliability of findings obtained when using an instrument in a manner that differs from the way it was administered or scored when its reported psychometric properties were established. Particularly when used for research purposes, studies must be conducted to demonstrate that changes to the “standard” way of administering or scoring an instrument have not affected its validity or reliability.

**QualityMetric’s Vendor Certification Program**

QualityMetric’s vendor certification program provides licensed vendors of QualityMetric surveys a means of demonstrating or verifying to their customers or other parties that the licensed products they develop and use themselves or offer to others have met the high quality standards that QualityMetric sets for its assessment products. These standards ensure that standardization of Short Form survey administration, scoring, and interpretation and reporting products is maintained and that the results yielded by their use are consistent with those from products developed by QualityMetric.

The vendor certification standards adopted by QualityMetric for administration of its surveys are specified in this guide; certification standards for scoring and interpreting/reporting the results of QualityMetric surveys are specified in the two accompanying guides in this series. Vendors desiring QualityMetric certification for any type of product they develop under a QualityMetric license (i.e., administration, scoring, and/or interpretation/reporting product) must submit documentation demonstrating that the product meets the specified standards. In the case of administration products, this would include a copy of the printed form containing QualityMetric survey items or other intellectual property. For electronic modes of administration, this might include eForms, printed screen shots, or a recording or script designed for oral administration (e.g., live interviewer, IVRS) of a survey, depending on the mode being developed.

**Purpose of This Guide**

This guide was developed to meet the needs of organizations that are developing paper forms, audio files, and/or software for administering, scoring, and reporting results from any of the Short Form surveys, under license from QualityMetric Incorporated. The purpose of this guide is to provide these organizations with information and guidance that will enable them to develop Short Form paper forms and software that meet their particular survey administration needs while maintaining standardization. Specifically, this guide provides the licensee with: a description of the Short Form surveys and possible applications of these instruments; general instructions for Short Form survey administration; and specific
guidelines and certification criteria for different modes of administration, including a summary of relevant findings from the professional literature.

Additional information regarding the history, administration, scoring, interpretation, and psychometric properties of the Short Form instruments are encouraged to obtain copies of the *User’s Manual for the SF-36v2® Health Survey, Second Edition* (Ware et al., 2007), *How to Score Version 2 of the SF-12 Health Survey (with a Supplement Documenting Version 1)* (Ware, Kosinski, Turner-Bowker, & Gandek, 2002), *How to Score and Interpret Single-Item Health Status Measures: A Manual for Users of the SF-8 Health Survey* (Ware, Kosinski, Dewey, & Gandek, 2001), *The SF-10™ for Children Health Survey: A User’s Guide* (Saris-Baglama, DeRosa, et al., 2007), and/or *QualityMetric Health Outcomes™ Scoring Software 2.0 User’s Guide* (Saris-Baglama, Dewey, et al., 2007). All are available from QualityMetric.

Because of their similarity in content and presentation, and because they are now the most commonly licensed Short Form instruments for certified development, the focus of this and the other guides in this series will be on the SF-36v2, SF-12v2, and SF-8. However, the general guidelines and recommendations for certification of modes of administration, scoring, data quality evaluation, interpretation, and reporting also apply to the SF-36, SF-12, DYNHA SF-36, and SF-10.
The Short Form Family of Health Surveys

Each health survey in the Short Form family of instruments was developed to be a brief, broad, generic measure of eight domains, or aspects, of health status that are considered important in describing and monitoring individuals suffering from a disease or illness. Although the surveys measure these domains in terms of functioning, behaviorally more subjective states, and/or personal evaluations, they are not intended to be comprehensive surveys of health status. The adult Short Form surveys (SF-36, SF-36v2, SF-12, SF-12v2, SF-8, and DYNHA SF-36) maintain comparability with each other by measuring the same health domains, deriving the same component summary measures, and employing the same or similar items. Because of its intended purpose, the SF-10 offers only similar, but more limited, measures of health status for children.

The purpose of this chapter is to provide a broad overview of each of the Short Form surveys. This overview is followed by sections describing the Short Form health domain scales, the items they comprise, the two component summary measures, and the time-relevant forms (standard, acute, 24-hour) in which the surveys are available. In addition, the applications for which the Short Form surveys have been found to be useful are briefly discussed. QualityMetric considers this information important for developers of Short Form survey products to know as it provides a basic understanding of the surveys, the aspects of health status they measure, and how they measure them. This in turn should facilitate product development in general and may help resolve problems and questions that may arise during the course of development.

QualityMetric’s Short Form Surveys

The “developmental” version of the SF-36, published in 1988 (Ware), represented a significant advancement in short-form instrumentation available to measure the self-reported health status of patient and nonpatient populations. Since that time, one revised version of the SF-36 and three abbreviated Short Form surveys have been developed and published. The SF-36v2 represents the most current fixed-form version of the SF-36 available, incorporating the use of more comprehensive normative data with the knowledge and advancements gained from over a decade of applications in research and clinical settings. The SF-36v2 is recommended for all new studies requiring one of the two 36-item measures. However, all members of the Short Form family of instruments for adults—the SF-8, SF-12, SF-12v2, SF-36, SF-36v2, and DYNHA SF-36—are cross-calibrated and scored on the same norm-based metric to maximize their comparability, and all have demonstrated their usefulness in assessing health status. The SF-10™ Health Survey for Children is also a member of the Short Form family but is not calibrated with the adult surveys.

There are many commonalities among the Short Form surveys. The following sections offer a brief description of each of the seven forms and versions.

The SF-36® Health Survey

The SF-36® Health Survey (SF-36; McHorney, Ware, Lu, & Sherbourne, 1994;
McHorney, Ware, & Raczek, 1993; Ware, 2004; Ware, Kosinski, & Gandek, 2000; Ware, Kosinski, & Keller, 1994; Ware & Sherbourne, 1992) contains 36 items used to measure eight domains of health-related quality of life (HRQOL): Physical Functioning, Role-Physical (role limitations due to physical health), Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional (role limitations due to mental/emotional health), and Mental Health. The information obtained from these eight health domains can be further aggregated into two summary measures: the Physical Component Summary (PCS) measure and the Mental Component Summary (MCS) measure. Data from the survey have been successfully used to compare health outcomes among groups of patients; to screen individuals for health concerns; to monitor health outcomes of individual patients, groups of patients, and populations over time; to predict the costs of healthcare; and to predict job loss, work productivity, mortality, and other important consequences of differences in health status.

The SF-36 is widely used throughout the world because it is brief, comprehensive, psychometrically sound, and available in over 80 translations, several of which are accompanied by translation-specific norms. Translations of the instrument for which norms have been developed include Italian (Apolone, Mosconi, & Ware, 1997); German (Bullinger & Kirchberger, 1998), Dutch (Aaronson et al., 1998), Danish (Bjorner et al., 1997); Japanese (Fukuhara, Suzukamo, Bito, & Kurokawa, 2001); French (Leplege, Ecosse, Pouchot, Coste, & Perneger, 2001); and Swedish (Sullivan, Karlsson, & Ware, 1994), as well as adaptations for Australia (Australian Bureau of Statistics, 1997), Canada (Hopman et al., 2000) and United Kingdom (Jenkinson, Layte, Wright, & Coulter, 1996). Moreover, it has proven its usefulness in measuring health status and outcomes in both general and specific populations. Up-to-date information about the SF-36, including citations for the most recently published studies and the developers’ responses to frequently asked questions, are available online at http://www.qualitymetric.com.

The SF-36v2® Health Survey

Based on the SF-36, the SF-36v2® Health Survey (SF-36v2; Ware, 2000, 2004; Ware et al., 2007) offers significant improvements in the measurement of HRQOL. These include:

- Improved instructions and questionnaire items, designed to simplify the wording and make the language more familiar.
- Improved layout for questions and response choices that makes items easier to read and complete, thus reducing the number of missing responses.
- Greater comparability with the translations and cultural adaptations widely used in the United States and in other countries.
- Five-level response choices, replacing yes/no response choices, for items in the Role-Physical (RP) and Role-Emotional (RE) health domain scales, which extend the range of functioning measured and increase score precision.
- Five-level response choices, replacing six-level response categories, designed to eliminate an ambiguous response choice (A good bit of the time) in the Mental Health (MH) and Vitality (VT) health domain scales.
- Adoption of T scores (formerly referred to as “norm-based scores”) for the health domain scales. (The component summary measures were always scored using T scores.)
- Use of the more comprehensive 1998 T-score norms for both the standard (4-week) and acute (1-week) forms.

The SF-12® Health Survey

The SF-12® Health Survey (SF-12; Ware et al., 2002; Ware, Kosinski, & Keller, 1995, 1996) contains a subset of 12 items taken directly from the SF-36. Because of its brevity, it has become a popular choice in population health surveys and in studies that combine the tool with additional surveys (such as disease-specific surveys). At the time the SF-12 was developed (1994), the dichotomous format for the role scale items precluded the calculation of reliable health domain scale scores. Thus, it yields scores on only the PCS and MCS measures.
The SF-12v2® Health Survey

Based on the SF-12, the SF-12v2® Health Survey (SF-12v2; Ware et al., 2002) offers significant improvements in the measurement of health status. Its 12 items were taken directly from the SF-36v2; as a result, the improvements found in the SF-12v2 are similar to the improvements made to the SF-36v2. In addition to the substantial gains in the range and precision of measurement achieved, the eight health domain scales can be scored on the SF-12v2. Thus, for those seeking a brief but comprehensive measure of health status, the SF-12v2 is proving to be an even more viable alternative to the SF-12 was for the SF-36.

The SF-8™ Health Survey

The SF-8™ Health Survey (SF-8; Ware, Kosinski, Dewey, & Gandek, 2001) contains 8 items, only one of which is identical to any of the items in the SF-36v2. Although the SF-8 items are not a direct subset of SF-36v2 items, both the SF-8 and the SF-36v2 measure the same eight health domains. Whereas the SF-36v2 uses between 2 and 10 items to measure each health domain, the SF-8 uses just one item for each health domain, making it less burdensome to complete and a good alternative to the SF-36v2 and the SF-12v2 for large-scale population survey efforts. Similar to the SF-36v2 and the SF-12v2, the PCS and MCS measures can be calculated from SF-8 results. The one disadvantage is that its scores generally cover a narrower range of the measured constructs, are coarser (i.e., define fewer levels) for some scales, and are less precise. Therefore, the SF-8 is not the Short Form survey of choice when one is interested in individual-level interpretations of scores, in conducting studies with smaller sample sizes in which enhanced precision is especially important, or in conducting studies requiring more statistical power.

The DYNHA® SF-36® Health Survey

For the most demanding applications of health status surveys, fixed-form, Short Form tools are no longer the most efficient, practical, or precise measures. Ongoing research is demonstrating that software based on computerized adaptive testing (CAT) logic delivers the best of both worlds: more practical and more precise measures that cover the very wide range of levels of health and well-being required to monitor and compare generic health outcomes across diverse populations, all while being administered with only the minimum of necessary items. By matching questions to each respondent’s health level, CAT is also able to estimate scores much more efficiently than fixed-form surveys.

The core general health measures in the DYNHA software are based on the Short Form family of instruments. The software uses item response theory (IRT) models to calibrate item pools (using items from the SF-36v2 and other widely used questionnaires) and to select the best items for each individual, items that are then scored using the same metric (T scores) as the SF-36v2. The resulting survey scores are quite accurate over a very wide range and can be quickly estimated at greatly reduced costs. This approach to survey administration offers efficiency, comparability of results using T scores, and availability of interpretation guidelines based on the Short Form surveys.

A prototype of computerized dynamic health assessments is available on the Internet at http://www.amlhealthy.com/.

The SF-10™ Health Survey for Children

The SF-10™ for Children Health Survey (SF-10; Saris-Baglama, DeRosa, et al., 2007) is a 10-item, parent-completed Short Form survey designed to measure children’s health status through the assessment of the physical and psychosocial functioning of children aged 5 through 17 years. It is an alternative to the short-form Child Health Questionnaire™ (CHQ™; Landgraf, Abetz, & Ware, 1999), which was developed in the early 1990s from findings of the Child Health Assessment Project at New England Medical Center’s Health Institute in response to a need for a comprehensive, generic measure of functional health and well-being in children and adolescents.

Much like the SF-12v2 and SF-8, the SF-10 was developed to be an assessment instrument that is brief, reliable, and valid yet comprehensive in its coverage of content areas relevant to
children’s physical and psychosocial functioning and well-being. Specifically, the objective was to reproduce the Physical Summary (PhS) and Psychosocial Summary (PsS) scores from the CHQ using only one item from each of the 10 domains represented. The SF-10 was developed as an alternate form to the CHQ that would enable the reproduction of the PhS and PsS scores on the 50-item CHQ while using significantly fewer items. The SF-10 is intended for use in population-based studies, for studies involving large sample sizes, and for group-level comparisons in which precision is less of a concern due to large sample sizes. When used in population studies, it yields results that are comparable to those that can be obtained with the longer form CHQ.

**Enhanced Versions of the Short Form Surveys**

Recent additions to the Short Form family of health surveys include “enhanced versions” of existing surveys. These enhanced versions are existing Short Form surveys plus additional items that either add considerable precision to a particular scale of the Short Form survey or allow the simultaneous assessment of a generic or disease-specific health domain that is frequently measured along with the standard Short Form domains, using a separate, stand-alone instrument. Generally, the items that are added are placed at the end of the Short Form survey, to be administered after the standardized survey has been completed; the SF-12v2–SET and SF-12v2–PIQ-6™ Enhanced versions are constructed this way. However, there are instances when integrating the additional items with the items from the standardized form is more beneficial or advantageous. For example, on the SF-12v2-MH Enhanced form, the three additional SF-36v2 MH items are integrated into the SF-12v2 Item 6 question-response grid, which includes the VT item and the other two MH items.

**The SF-36v2®/SF-12v2® Health Survey–PIQ-6™ Enhanced.** The SF-36v2 and SF-12v2 Health Surveys with PIQ-6 are versions of the SF-36v2 and SF-12v2 that include all items necessary to administer and score the Pain Impact Questionnaire (PIQ-6™; Becker, Saris-Baglama, Kosinski, Williams, & Bjorner, 2005). The PIQ-6 is a brief, six-item, patient-based assessment designed to measure pain severity and the impact of pain on work and leisure activities, as well as on emotional well-being. The PIQ-6 consists of one pain severity item with response options scaled from 1 (none) to 6 (very severe) and five pain impact items with varied response options that are scaled from 1 to 5. These six items yield two pain-related rating scores. The Pain Impact Score represents the sum of the weights assigned to all of the PIQ-6 item response options selected by the patient, with higher scores indicating a greater degree of pain severity and pain impact on a person’s life. The Pain Severity Rating is simply the response to Item 1 of the PIQ-6 and does not require additional scoring. Higher ratings of this item indicate a higher severity of pain.

The PIQ-6 is intended for use with adults suffering from infrequent and mild pain symptoms (e.g., back, neck, abdominal, knee, hip, and/or facial pain symptoms; headaches) to severe pain syndromes. It is suitable for several purposes in a wide variety of settings, ranging from clinical practice to population assessments, disease management, and clinical trials. The survey was developed using data from the U.S. general population and chronic pain patients aged from 18 to 97 years. Results obtained from the administration of the PIQ-6 to younger patients may not be reliable or valid.

**The SF-12v2® Health Survey–SET.** Both the SF-36v2 and SF-36 include a general-health-rating item (in addition to the five items in the surveys’ GH scale) that asks respondents to estimate the amount of change in their health in general, over a 1-year (standard form) or a 1-week (acute form) period. The five-point rating response for this item ranges from much better to much worse. This item, the Self-Evaluated Transition (SET) item (referred to as the Reported Health Transition [HT] item on the SF-36v2 and SF-36), is not used when scoring any of the eight Short Form health domain scales. Although easily overlooked, it does provide useful information about perceived changes in health status that occurred during the year (on the standard form) or week (on the acute form) prior to the administration of
the survey. Its inclusion with the SF-12v2 items at the end of the survey form constitutes the SF-12v2 Health Survey–SET.

The SET item can be analyzed as a categorical variable, or as an ordinal-level or interval-level scale for research or clinical purposes. It can also serve as a straightforward indicator of self-perceived change in an individual’s health status. When treated as categorical data, SET item responses can be used to analyze the percentage of respondents who select each of this item’s response choices, to estimate measured change (observed changes in health domain scale scores) reported for each response category, or to determine the likelihood of future health events based on the respondent’s rating of his or her own health.

**The SF-12v2® Health Survey–MH Enhanced.** The SF-12v2® Health Survey–MH Enhanced is the standard SF-12v2 survey plus three additional items from the Mental Health domain of the SF-36v2. This tool contains 15 items, including all five items from the SF-36v2 MH scale (also referred to as the Mental Health Inventory [MHI-5; Veit & Ware, 1983]).

The MH scale enhancement adds considerable precision to the SF-12v2 in terms of measuring the mental health domain. The five-item SF-36v2 MH scale, which is what ultimately is scored from SF-12v2–MH Enhanced form responses, has been shown to be an effective screener for major affective disorders, such as depression (Berwick et al., 1991; Ware & Kosinski, 2001; Weinstein, Berwick, Goldman, Murphy, & Barsky, 1989). In addition, this scale has been cross-calibrated with the Beck Depression Inventory® (BDI®; Katz, Katz, & Shaw, 1994) in order to enhance the understanding and interpretation of MH scores from the SF-36v2 and SF-12v2. Using the BDI score estimated from the SF-36v2 MH scale, one can evaluate the mental health burden observed at baseline in BDI score units and analyze health benefits of treatment in those same units.

**The SF-12v2® Health Survey–PF Enhanced.** Similar to the SF-12v2–MH Enhanced, the SF-12v2–PF Enhanced is the SF-12v2 survey with the addition of the eight PF items from the SF-36v2. This instrument allows for more precise measurement of the physical function-
behavioral change activities. Finally, a brief assessment of member satisfaction of care can be conducted with items that have been developed by members of QualityMetric staff and adopted by the NCQA and other accredited organizations for assessment of quality of care.

Short Form Standard and Acute Recall Forms

Both versions of the SF-36 and the SF-12 are available in two forms, each covering a specific recall period. The standard, or 4-week recall, form asks the respondent to answer the questions as they pertain to the way he or she felt or acted during the past 4 weeks. The acute, or 1-week recall, form asks the respondent to answer the questions as they pertain to the way he or she felt or acted during the past week. In addition to these two recall forms, the SF-8 is also available in a 24-hour recall form.

Use of the standard (4-week recall) form of the SF-36v2 is appropriate for cases in which the instrument will be administered only once to the respondent or when at least 4 weeks will pass before re-administration. In most cases, the standard version will meet the needs of the clinician for patient monitoring and the needs of the researcher for many types of investigations, particularly those of a longitudinal nature.

The acute (1-week recall) form provides a better description than the standard form of health status during the most recent week. When more frequent readministration is required, the acute form is most appropriate. For example, the acute form is recommended when a clinician or researcher wants to closely monitor the effects of a physical (e.g., pharmacological) or behavioral (e.g., psychotherapeutic) intervention on a patient or group of patients when such effects are likely to occur rapidly (e.g., asthma therapy). However, 1 or more weeks must pass between administrations of the acute version in order to obtain valid information. Generally, the results from administrations of the standard and acute forms substantially agree. However, at times users may find that results from the acute form will differ from those obtained from the standard form (see Keller et al., 1997).

The SF-8’s 24-hour recall form allows the user to assess patient health status during the 24 hours prior to administration. This form is particularly useful when a more “point-in-time” assessment is required or when one wishes to measure changes in patients whose health status can change quickly owing to the nature of their condition or to some fast-acting intervention that has been introduced.

Table 2.1 summarizes some of the general similarities and differences among the Short Form surveys.

Translated Versions of the Short Form Surveys

More than 110 translations and English-language adaptations of the Short Form surveys have been completed since the beginning of the International Quality of Life Assessment (IQOLA) Project in 1991. The IQOLA Project began with the goal of translating the SF-36 for international use in 14 countries. To this end, the IQOLA Project team adopted a multi-stage translation procedure designed to assure that translations of the SF-36 were not only conceptually equivalent to the U.S. source form but also linguistically and culturally relevant (Aaronson et al., 1992; Bullinger et al., 1998). In brief, the IQOLA translation process included development of an initial forward translation based on multiple independent translations; backward translation of the forward translation into English; review of the backward translation for conceptual equivalence with the source form; and pilot testing of the translation among native speakers. In many countries, independent judges rated the translation on clarity, use of common language, conceptual equivalence, and overall acceptability. A Thurstone-like scaling exercise was also used to inform the selection of response choices in many countries (Keller et al., 1998). In addition, a harmonization meeting was held among investigators from the first dozen countries to join the project (Wagner et al., 1998). Overall, the psychometric properties of the IQOLA translations have been thoroughly evaluated, as documented elsewhere (see Gandek & Ware, 1998b).
### Table 2.1

*Summary of Short Form Health Survey Similarities and Differences*

<table>
<thead>
<tr>
<th></th>
<th>SF-36</th>
<th>SF-36v2</th>
<th>SF-12</th>
<th>SF-12v2</th>
<th>SF-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved item wording</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Increased range</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Improved format</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Standard form (4-week recall)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Acute form (1-week recall)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Acute form (24-hour recall)</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Eight-scale profile</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Component summary measures</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Translated versions</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Use for individual patient assessment</td>
<td>X</td>
<td>X</td>
<td>X&lt;sup&gt;a&lt;/sup&gt;</td>
<td>X&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>X&lt;sup&gt;ab&lt;/sup&gt;</td>
</tr>
<tr>
<td>Use for detection of small differences in group data</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Use for large samples</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Use with population surveys</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

<sup>a</sup> Most appropriate when using PCS and MCS scores.

<sup>b</sup> Health domain scales are appropriate for use with individuals only when very large score differences are expected.

---

### Short Form Concepts and Measures

#### Health Domain Scales

The SF-36v2 includes one favorably scored scale measuring each of eight health domains: physical functioning, role participation with physical health problems (role-physical), bodily pain, general health, vitality, social functioning, role participation with emotional health problems (role-emotional), and mental health. These scales are the same as those developed for the SF-36, and the items that constitute them address the same general content found in the items of the original Short Form survey. A summary of item content for the SF-36v2, SF-12v2, and SF-8 is presented in Table 2.2.

The following scale descriptions pertain to the SF-36v2. The reader is referred to Table 2.2 to determine which aspects of the SF-36v2 survey content are not included in the shorter SF-12v2 and SF-8.

**Physical Functioning (PF).** The content of the 10-item PF scale reflects the importance of distinct aspects of physical functioning and the necessity of sampling a range of severe and minor physical limitations. Items represent levels and kinds of limitations between the extremes of physical activities, including lifting and carrying groceries; climbing stairs; bending, kneeling, or stooping; and walking moderate distances. One self-care item is included to represent limitations in self-care activities. The PF items capture both the presence and extent of physical limitations using a three-level response continuum. Low scores indicate significant limitations in performing physical activities while high scores reflect little or no such limitations.

**Role-Physical (RP).** The four-item RP scale covers an array of physical health-related role limitations, including (a) limitations in the kind of work or other usual activities, (b) reductions in the amount of time spent on work or other usual activities, (c) difficulty performing work
<table>
<thead>
<tr>
<th>Short Form Scale</th>
<th>Item Number</th>
<th>SF-36v2¹</th>
<th>SF-12v2²</th>
<th>SF-8</th>
<th>Abbreviated Item Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Functioning (PF)</td>
<td>3a</td>
<td></td>
<td></td>
<td></td>
<td>Vigorous activities, such as running, lifting heavy objects, or participating in strenuous sports</td>
</tr>
<tr>
<td></td>
<td>3b 2a</td>
<td></td>
<td></td>
<td></td>
<td>Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf</td>
</tr>
<tr>
<td></td>
<td>3c</td>
<td></td>
<td></td>
<td></td>
<td>Lifting or carrying groceries</td>
</tr>
<tr>
<td></td>
<td>3d 2b</td>
<td></td>
<td></td>
<td></td>
<td>Climbing several flights of stairs</td>
</tr>
<tr>
<td></td>
<td>3e</td>
<td></td>
<td></td>
<td></td>
<td>Climbing one flight of stairs</td>
</tr>
<tr>
<td></td>
<td>3f</td>
<td></td>
<td></td>
<td></td>
<td>Bending, kneeling, or stooping</td>
</tr>
<tr>
<td></td>
<td>3g</td>
<td></td>
<td></td>
<td></td>
<td>Walking more than a mile</td>
</tr>
<tr>
<td></td>
<td>3h</td>
<td></td>
<td></td>
<td></td>
<td>Walking several hundred yards</td>
</tr>
<tr>
<td></td>
<td>3i</td>
<td></td>
<td></td>
<td></td>
<td>Walking 100 yards</td>
</tr>
<tr>
<td></td>
<td>3j</td>
<td></td>
<td></td>
<td></td>
<td>Bathing or dressing oneself</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>Physical health problems limit usual physical activities</td>
</tr>
<tr>
<td>Role-Physical (RP)</td>
<td>4a</td>
<td></td>
<td></td>
<td></td>
<td>Cut down the amount of time spent on work or other activities</td>
</tr>
<tr>
<td></td>
<td>4b 3a</td>
<td></td>
<td></td>
<td></td>
<td>Accomplished less than you would like</td>
</tr>
<tr>
<td></td>
<td>4c 3b</td>
<td></td>
<td></td>
<td></td>
<td>Limited in kind of work or other activities</td>
</tr>
<tr>
<td></td>
<td>4d</td>
<td></td>
<td></td>
<td></td>
<td>Had difficulty performing work or other activities (e.g., it took extra effort)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>Difficulty doing daily work because of physical health</td>
</tr>
<tr>
<td>Bodily Pain (BP)</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>Intensity of bodily pain</td>
</tr>
<tr>
<td></td>
<td>8 5</td>
<td></td>
<td></td>
<td></td>
<td>Extent pain interfered with normal work</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>Severity of bodily pain</td>
</tr>
<tr>
<td>General Health (GH)</td>
<td>1 1</td>
<td></td>
<td></td>
<td></td>
<td>Is your health: excellent, very good, good, fair, or poor</td>
</tr>
<tr>
<td></td>
<td>11a</td>
<td></td>
<td></td>
<td></td>
<td>Seem to get sick a little easier than other people</td>
</tr>
<tr>
<td></td>
<td>11b</td>
<td></td>
<td></td>
<td></td>
<td>As healthy as anybody I know</td>
</tr>
<tr>
<td></td>
<td>11c</td>
<td></td>
<td></td>
<td></td>
<td>Expect my health to get worse</td>
</tr>
<tr>
<td></td>
<td>11d</td>
<td></td>
<td></td>
<td></td>
<td>Health is excellent</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>Is your health: excellent, very good, good, fair, poor, or very poor</td>
</tr>
<tr>
<td>Vitality (VT)</td>
<td>9a</td>
<td></td>
<td></td>
<td></td>
<td>Feel full of life</td>
</tr>
<tr>
<td></td>
<td>9e 6b</td>
<td></td>
<td></td>
<td></td>
<td>Have a lot of energy</td>
</tr>
<tr>
<td></td>
<td>9g</td>
<td></td>
<td></td>
<td></td>
<td>Feel worn out</td>
</tr>
<tr>
<td></td>
<td>9i</td>
<td></td>
<td></td>
<td></td>
<td>Feel tired</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>Amount of energy</td>
</tr>
</tbody>
</table>

*(continued on next page)*
Table 2.2
Abbreviated Item Content for the Short Form Health Domain Scales (Continued)

<table>
<thead>
<tr>
<th>Short Form Scale</th>
<th>Item Number</th>
<th>SF-36v2</th>
<th>SF-12v2</th>
<th>SF-8</th>
<th>Abbreviated Item Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Functioning (SF)</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>Extent health problems interfered with normal social activities</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>7</td>
<td></td>
<td></td>
<td>Frequency health problems interfered with social activities</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>How much physical health or emotional problems limit usual social activities</td>
</tr>
<tr>
<td>Role-Emotional (RE)</td>
<td>5a</td>
<td></td>
<td></td>
<td></td>
<td>Cut down the amount of time spent on work or other activities</td>
</tr>
<tr>
<td></td>
<td>5b</td>
<td>4a</td>
<td></td>
<td></td>
<td>Accomplished less than you would like</td>
</tr>
<tr>
<td></td>
<td>5c</td>
<td>4b</td>
<td></td>
<td></td>
<td>Did work or other activities less carefully than usual</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>How much personal or emotional problems keeps from doing usual work or daily activities</td>
</tr>
<tr>
<td>Mental Health (MH)</td>
<td>9b</td>
<td></td>
<td></td>
<td></td>
<td>Been very nervous</td>
</tr>
<tr>
<td></td>
<td>9c</td>
<td></td>
<td></td>
<td></td>
<td>Felt so down in the dumps that nothing could cheer you up</td>
</tr>
<tr>
<td></td>
<td>9d</td>
<td>6a</td>
<td></td>
<td></td>
<td>Felt calm and peaceful</td>
</tr>
<tr>
<td></td>
<td>9f</td>
<td>6a</td>
<td></td>
<td></td>
<td>Felt downhearted and depressed</td>
</tr>
<tr>
<td></td>
<td>9h</td>
<td></td>
<td></td>
<td></td>
<td>Been happy</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>Bothered by emotional problems</td>
</tr>
<tr>
<td>Self-Evaluated Transition (SET)</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>How health is now compared to 1 year ago</td>
</tr>
</tbody>
</table>

1 Also applies to the SF-36.
2 Also applies to the SF-12.
3 Formerly referred to as Reported Health Transition (HT).

or other usual activities, and (d) accomplishing less. Low scores on the RP scale reflect problems with work or other activities as a result of physical problems. High scores indicate little or no problems with work or other daily activities stemming from physical problems.

**Bodily Pain (BP).** The BP scale comprises two items: one pertaining to the intensity of bodily pain and one measuring the extent of interference with normal work activities due to pain. Low scores indicate high levels of pain that impact normal activities, whereas high scores indicate no pain and no related impact on normal activities.

**General Health (GH).** The GH scale consists of five items, including a rating of health (excellent to poor) and four items addressing the respondent’s views and expectations of his or her health. Low scores indicate evaluation of general health as poor and likely to get worse. High scores indicate that the respondent evaluates his or her health most favorably.

**Vitality (VT).** This four-item measure of vitality (i.e., energy level and fatigue) was developed to capture differences in subjective well-being. Low scores indicate feelings of tiredness and being worn out. High scores indicate feeling full of energy all or most of the time.
**Social Functioning (SF).** This two-item scale assesses health-related effects on quantity and quality of social activities, asking specifically about the impact of either physical or emotional problems on social activities. The degree to which physical and emotional problems interfere with normal social activities increases with decreasing SF scores. The lowest score equates to extreme or frequent interference with normal social activities due to physical and emotional problems; the highest score indicates that the individual performs normal social activities without interference from physical or emotional problems.

**Role-Emotional (RE).** The three-item RE scale assesses mental health-related role limitations in terms of (a) time spent doing work or other usual activities, (b) amount of work or activities accomplished, and (c) the care with which work or other activities were performed. Low scores on this scale reflect problems with work or other activities as a result of emotional problems. High scores reflect no such limitations due to emotional problems.

**Mental Health (MH).** The five-item MH scale includes one or more items from each of four major mental health dimensions (anxiety, depression, loss of behavioral/emotional control, and psychological well-being). Low scores on MH are indicative of frequent feelings of nervousness and depression, whereas high scores indicate feelings of peace, happiness, and calm all or most of the time.

**Self-Evaluated Transition (SET).** In addition to health domain scale items, the SF-36 includes a general health item asking respondents to rate the amount of change they experienced in their health in general over a 1-year or 1-week period on the standard form or acute form, respectively. Formerly referred to as Reported Health Transition (HT), this item is not used to score any of the eight health domain scales or component summary measures; however, it does provide useful information about perceived changes in health status that occurred during the year (on the standard form) or week (on the acute form) prior to the administration of the survey. If clinical or research needs require the measurement of reported health transition over a period other than 1 year or 1 week (e.g., during the past 3 months), the user may use this item as a template for developing a more time-relevant item that would be administered in addition to the standard HT item.

**Physical and Mental Component Summary Measures (PCS and MCS)**

Figure 2.1 illustrates the measurement model, or conceptual framework, underlying the construction of the SF-36v2 and SF-12v2 multi-item health domain scales and component summary measures. This model has three levels: (a) items, (b) health domain scales that aggregate items, and (c) component summary measures that aggregate the health domain scales. The aggregates of the health domain scales are referred to as component summary measures because they were derived and scored using a factor analytic method called principal components analysis (Harman, 1976). Although they reflect the two broad components or aspects of health—physical and mental—all of the eight health domain scales are used to score both component summary measures. All but 1 of the 36 items (i.e., the SET item) are used to score the eight health domain scales.

Factor analyses of correlations among the eight health domain scales of each version of the survey have consistently identified two factors (Ware, Kosinski, Bayliss, et al., 1995; Ware, Kosinski, Gandek, et al., 1998; Ware, Kosinski, & Keller, 1994). On the strength of the pattern of their correlations with the eight scales, the two factors have been interpreted as physical and mental components of health status. Three scales (PF, RP, and BP) correlate most highly with the physical component and contribute most to the scoring of the Physical Component Summary (PCS) measure. The mental component correlates most highly with the MH, RE, and SF scales, which contribute most to the scoring of the Mental Component Summary (MCS) measure. Three of the scales have noteworthy correlations with both components: VT correlates substantially with both but higher with the mental component, GH correlates with both but higher with the physical component, and SF correlates much higher with the mental component.
Figure 2.1  Short Form Health Status Measurement Model

<table>
<thead>
<tr>
<th>Items</th>
<th>Scales</th>
<th>Component Summary Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a. Vigorous Activities</td>
<td>Physical Functioning (PF)</td>
<td></td>
</tr>
<tr>
<td>3b. Moderate Activities</td>
<td>Role-Physical (RP)</td>
<td></td>
</tr>
<tr>
<td>3c. Lift, Carry Groceries</td>
<td>Bodily Pain (BP)</td>
<td></td>
</tr>
<tr>
<td>3d. Climb Several Flights</td>
<td>General Health (GH)</td>
<td></td>
</tr>
<tr>
<td>3e. Climb One Flight</td>
<td></td>
<td>Mental Health</td>
</tr>
<tr>
<td>3f. Bend, Kneel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3g. Walk Mile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3h. Walk Several Hundred Yards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3i. Walk One Hundred Yards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3j. Bathe, Dress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a. Cut Down Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4b. Accomplished Less</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4c. Limited in Kind</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4d. Had Difficulty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Pain - Magnitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Pain - Interference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a. EVGFP Rating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11a. Sick Easier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11b. As Healthy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11c. Health To Get Worse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11d. Health Excellent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9a. Full of Life</td>
<td></td>
<td>Vitality (VT)</td>
</tr>
<tr>
<td>9b. Nervous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9c. Down in Dumps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9d. Peaceful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9e. Energy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9f. Depressed/Downhearted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9g. Worn Out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9h. Happy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9i. Tired</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Social-Extent</td>
<td></td>
<td>Social Functioning (SF)</td>
</tr>
<tr>
<td>10. Social-Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5a. Cut Down Time</td>
<td></td>
<td>Role-Emotional (RE)</td>
</tr>
<tr>
<td>5b. Accomplished Less</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5c. Less Careful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5d. Had Difficulty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5e. Limited in Kind</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5f. Accomplished Less</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5g. Lift, Carry Groceries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5h. Walk Mile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5i. Walk One Hundred Yards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5j. Bathe, Dress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a. Cut Down Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4b. Accomplished Less</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4c. Limited in Kind</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4d. Had Difficulty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Pain - Magnitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Pain - Interference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a. EVGFP Rating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11a. Sick Easier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11b. As Healthy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11c. Health To Get Worse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11d. Health Excellent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9a. Full of Life</td>
<td></td>
<td>Vitality (VT)</td>
</tr>
<tr>
<td>9b. Nervous</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9c. Down in Dumps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9d. Peaceful</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9e. Energy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9f. Depressed/Downhearted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9g. Worn Out</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9h. Happy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9i. Tired</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Item numbers correspond to SF-36v2 item numbers. SF-12v2 items are indicated in **highlighted bold**. All health domain scales contribute to the scoring of both the Physical and Mental Component Summary measures. Scales contributing most to the scoring of the summary measures are indicated by a connecting solid line (—). Scales contributing to the scoring of the summary measures to a lesser degree are indicated by a dotted line (······).
The PCS and MCS measures were constructed and scored to achieve a number of advantages in addition to reducing the eight-scale profile to two component summary measures without substantial loss of information. The advantages include a very large increase in the number of levels defined, smaller confidence intervals (CIs) relative to each of the eight health domain scales, and the elimination of both floor and ceiling effects. Practical advantages include the reduction of the number of statistical comparisons required in an outcome study or clinical trial and improved validity in distinguishing between physical and mental health outcomes.

Low scores on the PCS measure indicate limitations in physical functioning, role participation due to physical problems, a high degree of bodily pain, and poor general health. A high score on the PCS measure indicates little or no physical limitations, disabilities, or decrements in well-being; a high energy level; and good general health. For the MCS measure, a low score is indicative of frequent psychological distress, social and role disability due to emotional problems, and poor general health. A high score on the MCS measure indicates frequent positive affect, little or no psychological distress or limitations in usual social/role activities due to emotional problems, and good general health. A strength of the PCS and MCS measures is their value in distinguishing a physical health outcome from a mental health outcome (Ware & Kosinski, 2001; Ware, Kosinski, Bayliss, et al., 1995).

**Short Form Survey Profile of Scores**

Figure 2.2 illustrates the Short Form surveys’ profile of T scores. The incorporation of the component summary measures at the beginning of the profile emphasizes the importance of considering the findings from these more general measures of health status in the interpretation of results from any of the surveys in the SF family of instruments. It also facilitates interpretation by immediately establishing what the general burden of illness or effects of treatment are (i.e., physical or mental) before examining the more specific health domain scales. The PCS and MCS scores provide, as their labels suggest, a summary of the respondent’s health status from both a broad physical health perspective and a broad mental health perspective, respectively. Results on the PCS and MCS measures should serve as a starting place for determining whether functional limitations exist
in either of the two major components of health; if so, the health domains contributing greatest to the affected dimension(s) and the items they comprise require further examination to ascertain their potential contribution to the respondent’s impaired functioning.

The eight health domain scales are ordered, left to right, from the best physical health scale (PF) to the best mental health scale (MH). This ordering further facilitates interpretation of the profile, with domain scales on the left side of the health domain scale profile reflecting physical health status and the domain scales on the right side reflecting mental health status.

**SF-6D Health Utility Index**

Although not originally designed for use in economic evaluations, research has shown that a meaningful health state classification can be created by applying a scoring method that focuses on seven of the health domains covered by both versions of the SF-36 and SF-12. The resulting SF-6D Health Utility Index (Brazier, Roberts, & Deverill, 2002; Brazier, Usherwood, Harper, & Thomas, 1998) is the first preference-based index constructed from a psychometric measure of health status. It is scored from 0.0 (worst measured health state) to 1.0 (best measured health state). Using a six-domain classification of health states (totaling 18,000 states in all), it can be used in the determination of the cost-effectiveness of various healthcare interventions and quality adjusted life years (QALYs). Currently, the SF-6D is the only Short Form measure that provides a description of health suitable for both an evaluation of a psychometric profile and an economic evaluation.

**Medical Expenditure Prediction**

QualityMetric’s SF Expenditure Prediction™ Solution enables new interpretation of Short Form surveys based on scientifically valid algorithms, allowing prediction of average monthly medical expenditures for patients and populations over the next 6 months based on current health status and selected demographic variables. In a recent study (Fleishman, Cohen, Manning, & Kosinski, 2006), the summary scores of the SF-12 were shown to improve the prediction of expenditures over and above some of the variables traditionally used to predict expenditures, such as historical claims data, diagnoses, and demographics. The SF Expenditure Prediction Solution can be derived from the administration of either version of the SF-36 or SF-12.

**Short Form Survey Applications**

Until the early 1990s, most clinical trials, disease management programs, population monitoring efforts, and health research studies defined the results or outcomes of interest relatively narrowly; that is, in terms of clinical variables. When patient-reported outcomes were considered, definitions tended to focus on disease-specific indicators. Increasingly, the variety of uses and users of patient-reported health assessments have expanded the definition of outcomes to include measures of both generic and disease-specific concepts. Used together, generic and disease-specific health assessments provide a comprehensive definition of health in its multiple dimensions as experienced by the individual.

Each of the Short Form instruments can be used alone or in combination with disease-specific, patient-reported outcomes (PRO) measures in several ways and for several purposes. For example, Short Form surveys have been accepted by the Food and Drug Administration (FDA) as valid measures of health outcomes that can be used in clinical studies. It is this utility, along with brevity, normative data, and demonstrated psychometric grounding, which makes them valuable tools in both clinical and research settings for several purposes, described in the following sections.

**Evaluating and Monitoring Individual Patients in Clinical Practice**

The SF-36v2 and SF-12v2 have proven valuable to physicians and other healthcare providers as a means of evaluating and monitoring individuals seeking treatment for physical or
mental health problems. Unlike standard means of assessing health status (e.g., physician examination, lab tests, mental status examinations), the SF-36v2 and SF-12v2 provide a broad overview of a patient’s health status and its effect on his or her functioning. Incorporation of either Short Form instrument into a standard office procedure is facilitated by the fact that it is a brief, patient self-report measure.

When administered at the beginning of an episode of care, the Short Form surveys can be used to help identify aspects of the patient’s health (e.g., functional impairment or distress) that might not otherwise be detected. The results of the initial administration can also serve as a baseline measure of health status that can be compared to results obtained from one or more readministrations of the survey during the course of treatment, thus providing objective means of documenting the outcomes of the treatment. The results from one episode of care can also be used as comparison data for subsequent episodes of care. In addition, scores on the PCS and MCS measures may be used to roughly stratify patients according to who is more likely to utilize healthcare services (Ware & Kosinski, 2001) or consume more healthcare dollars (Fleischman, Cohen, Manning, & Kosinski, 2006). Wetzler, Lum, and Bush (2000) provide a detailed discussion of the use of the SF-36 in primary care settings for various decision-making purposes related to patients presenting with possible behavioral healthcare problems. Moreover, some studies have shown that the SF-12v2 can also be used to determine whether one treatment option is likely to have a more significant impact on a patient’s health status or quality of life (Fransen, Nairn, Winstanley, Lam, & Edmonds, 2007; Secor, Markow, MacKenzie, & Thrall, 2004).

**Monitoring Populations**

The brevity of the Short Form surveys lends to comprehensive population monitoring. The effectiveness of the Short Form surveys in monitoring functioning and well-being, assessing disease burden, and comparing the health of different populations and patient groups has been reported in a total of more than 12,000 publications as of April 2009. A prime example of how a member of the Short Form family of health surveys can be used in population monitoring is the Medicare Health Outcomes Survey (HOS; Gandek, Sinclair, Kosinski & Ware, 2004; Ware, Gandek, Sinclair, & Kosinski, 2004). From 1998 to 2004, the HOS consisted of the SF-36 items along with questions about activities of daily living (ADLs) and case-mix and risk-adjustment questions for Medicare beneficiaries enrolled in managed care programs.

**Estimating the Burden of Disease**

The value of general and disease-specific norms has been demonstrated for the Short Form scales. They appear to be increasingly accepted, on the strength of validation studies to date, as valid health measures for the purposes of documenting disease burden. The disease-specific normative data provided in the SF-36v2 and SF-12v2 manuals provide estimates of the burden of disease for 18 disease or physically impaired subsamples of the 1998 SF-36v2 and SF-12v2 normative groups. Estimates of the unique effects of chronic conditions on each survey’s scales and measures are also presented in the manuals. The usefulness of the Short Form surveys in describing the burden of disease is found in those publications describing more than 150 diseases and conditions, with 16 conditions each being addressed in more than 100 publications. Included among the more recent articles employing the SF-12v2 are those reporting the burden of disease or its treatment for prostate cancer (Kouba, Hubbard, Moore, Wallen, & Pruthi, 2007), osteoarthritis (Fransen et al., 2007), diabetic peripheral neuropathy (Gore et al., 2005), Parkinson’s disease and psychogenic movement disorders (Anderson et al., 2007), premenstrual dysphoric disorder (Yang et al., 2008), and disordered gambling among disability services recipients (Morasco & Petry, 2006).

**Evaluating Treatment Effects in Clinical Trials**

Medical researchers conducting clinical trials now recognize the need to define benefits more broadly than traditional clinical endpoints by including PROs in clinical trials. Additional
clinical evidence based on PROs also commands increasing attention from the FDA, making it critical to the drug review and approval process. The Short Form surveys are becoming widely recognized as leading PRO measures in clinical trials. They have been cited in a total of more than 1,800 published articles (as of January 2009) reporting randomized controlled trial results. With more than 110 translations or adaptations available, the Short Form surveys represent an international benchmark for health outcomes measurement and have been accepted by the FDA to measure health outcomes in clinical studies.

When included in a clinical trial protocol, the Short Form surveys can quantify the consumer’s experience of improved HRQOL, deliver proof of efficacy that goes beyond traditional clinical endpoints, and provide a scientifically valid body of evidence to facilitate timely regulatory approval. For example, as indicated earlier, Secor et al. (2004) investigated the effects of acupuncture, chiropractic, and naturopathy approaches in decreasing pain and improving quality of life, and Fransen et al. (2007) sought to determine the clinical benefits of hydrotherapy and tai chi for patients with osteoarthritis of the hips or knees to a waiting-list control group in a randomized controlled trial. Furthermore, in a study of the short-term effects of radical retropubic prostatectomy (RRP) on HRQOL, Kouba et al. (2007) administered the acute (1-week) form of the SF-12v2 to 51 prostate cancer patients undergoing the procedure.

The ability to choose from among several administration options (e.g., paper form, handheld, and interactive voice response [IVRS] technology) is another feature that enables the survey to meet the needs of clinical trials. Comparability of Short Form survey scores across administration modalities is addressed in Chapter 4.

**Disease Management**

Increasingly, disease management providers are incorporating PRO surveys into their measurement systems. Data from such surveys add significant value because they improve risk prediction, service planning, and outcomes monitoring efforts, and they ensure that program planning and evaluation efforts incorporate the patient’s perspective. The Short Form surveys can provide practical solutions to disease management’s most pressing measurement challenges. Moreover, the Short Form surveys’ reliability and validity in assessing the burden of disease has been demonstrated for several patient populations. Many studies document their ability to predict hospitalization, total medical expenditures, job loss and work productivity, future health, risk of depression, use of mental healthcare, and mortality. For example, Isetts et al. (2006) investigated the effects of collaborative drug therapy management (CDTM) on patients’ HRQOL (as measured by the SF-12v2) and perceptions of care (as measured by the Consumer Assessment of Health Plans [CAHPS]). In addition, disease-specific surveys can be paired with the Short Form surveys to screen patients with common chronic conditions (e.g., asthma) and capture a more comprehensive picture of HRQOL benefits.

**Risk Prediction and Cost-Effectiveness**

Including the Short Form scales and measures in predictive models can improve forecasts of future expenditures, resource utilization, health outcomes, likelihood of hospitalization, risk of depression, use of mental health specialty care, job loss, return to work and work productivity, future health, and mortality. For example, based on the work of Fleishman et al. (2006) in collaboration with the Agency for Healthcare Research and Quality (AHRQ), QualityMetric Incorporated developed the SF Expenditure Prediction Solution (see previous discussion). Use of this feature allows for the (a) prediction of average monthly medical expenditures for patients and populations over the subsequent 6 months based on their current health status, (b) improvement in the identification of patients at risk for incurring high future medical costs; and (c) provision of health status information to organizations, thus enabling informed economic decisions to be made on preventive care and/or behavior modification.
Patient-Provider Relations

One result of high healthcare costs can be seen in the amount of time that clinicians spend each day with patients. Limited contact between patients and providers has only increased the need and demand for member-focused services that promote information flow and foster improved care delivery through consumer involvement. When incorporated as part of the standard care process, the Short Form surveys can improve and enhance the communication process by providing information that enables healthcare providers to make the best use of the limited time they have to see patients. As previously discussed, Short Form survey results can be used to establish an objective baseline measure of health status against which health problems can be identified, the effects of treatment monitored, and the outcomes of that treatment quantitatively assessed. Employing electronic means or other capabilities for administering, scoring, reporting, and capturing the results of one of the Short Form surveys at the time of an office visit can afford the busy provider the additional benefits of providing immediate feedback for members and aggregated survey results for groups of patients. The Short Form surveys also can be used to measure the effects of other attempts at improving communication between patients and their healthcare providers (e.g., see Llewellyn, McGurk, & Weinman, 2006).

Direct-to-Consumer Information

Pharmaceutical companies increasingly engage in providing information directly to potential consumers. Critical to the success of direct-to-consumer (DTC) information campaigns is consumer recognition that the information provided has immediate relevance to them. Increasingly, DTC materials include short, self-report health assessments, results from which link directly to guidelines regarding likelihood of diagnosis and/or recommended self-care, physician consultation, and treatment options. To be most effective, such assessments should meet scientific standards of reliability and validity and have demonstrated acceptance and relevance among consumers and clinicians.

When health assessments meet measurement standards and are selected or developed with their planned use in mind, benefiting populations can be identified, key data can be collected, and recommendations can be provided, all with a solid return on investment. Those employing a Short Form survey as part of the DTC assessment have the added benefit of being able to administer the survey in fixed-form format either in print, by telephone with speech recognition, by handheld devices, or online via the Internet. Also, one or more disease-specific measures can be administered along with the Short Form survey to provide consumers and their clinicians with information required to screen and monitor common chronic conditions such as asthma, congestive heart failure, or depression. Overall, using a Short Form survey as part of a DTC marketing effort can help garner consumer acceptance by providing a first-stage screener for conditions having substantial impact on generic domains and, when used longitudinally, gathering proof of improved outcomes.
Survey Administration: General Considerations

This chapter presents considerations for the administration of any of the Short Form surveys, regardless of the mode or format in which they are administered. Person-specific considerations—age, reading level, language, and level of cooperation and understanding—for determining how appropriate it is for the respondent to complete the instrument are discussed first. Specific guidelines for administration are also provided, including suggested scripts for introducing and concluding administrations to individuals and groups. In addition, a tabular summary of the most important Dos and Don’ts of Short Form survey administration is presented. Finally, matters pertaining to the administration environment are discussed, as is the inclusion of a Short Form survey as part of a longer interview, survey, or other data collection effort. Understanding how an in-person, paper-form administration of a Short Form survey should be conducted will provide helpful considerations and guidelines for developing other modes of administration in a manner that will help maintain standardization.

Proper administration of any of the Short Form surveys assumes that a trained person oversees the administration and that the respondent meets the eligibility requirements for completing the survey. Following the administration instructions and recommendations provided in this guide is particularly important when the survey administrator or proctor administers the paper-and-pencil version of a Short Form survey in person, to one or more individuals. For in-person administrations, it is very important for the administrator to establish rapport with the respondent, encourage completion of the survey, emphasize the importance of respondent’s answers, address questions or concerns about the Short Form survey, and ensure the survey is correctly and completely filled out. It is beyond the scope of this guide to address common questions and concerns that may be raised by administrators (e.g., What should I do if the respondent does not answer all the items?) and respondents (e.g., What do my answers mean?) during the course of a survey administration. However, these types of issues are identified and addressed in the SF-36v2 User’s Guide (Ware et al., 2007).

Determining Respondent Eligibility

Age

The Short Form surveys were normed for use with adults; thus, use of the published norms should be limited to individuals aged 18 years and older. Items like those in the SF-36 have been successfully administered to persons as young as 14 years using self-administration and interviewer administration over the telephone and in-person (Brook et al., 1979; Ware et al., 1980), and SF-36 translations have been successfully administered to those as young as age 15 (Gandek & Ware, 1998a). One can expect similar success with the other Short Form surveys when norms for adolescents become available.

Reading Ability

In situations where participation requires completion of a self-administered survey, potential respondents should be excluded if they are unable to read the survey due to limited reading
ability. Before giving a respondent a survey form, the administrator should determine if any information is available regarding the respondent’s ability to read. Using the Microsoft® Word readability determination feature, the SF-36v2, SF-12v2, and SF-8 standard forms were found to have Flesch-Kincaid Grade Levels of 6.9, 6.8, and 8.2, respectively, and Flesch Reading Ease scores of 68, 68, and 62, respectively, on a 100-point scale. Putting these scores in perspective, the closer a Flesch Reading Ease score is to 100, the easier the text is to read. In most cases, a Flesch Reading Ease score of 60 to 70 is desirable (Millhollan & Murray, 2001).

If a respondent is unable to read the Short Form survey items for any reason, he or she should not be offered a survey form; rather, the administrator should consider other modes of administration. Administration via an interactive voice response system (IVRS), if available, is a good solution to this situation. Otherwise, the administrator can conduct the assessment using the appropriate (standard form or acute form) interview script (available from QualityMetric Incorporated at www.qualitymetric.com), recording that the survey was not self-administered due to reading ability. The interviewer-administered script can also be used if the survey is administered to a large group of respondents who are unable to read. In this case, printed survey forms and pencils would be provided to the respondents, the items would be read aloud, the numbers corresponding to the response options for each item would be read along with the responses, and the respondent would be asked to record his or her response using the item response numbers on the survey form as a guide.

It is important to note that the administration order of items 7 and 8 from the Bodily Pain health domain scale is reversed on the SF-36v2 standard and acute form interview scripts. Thus, the scores for these items obtained using an interview script must be reversed (i.e., the response to item 7 from the interview script should be entered in the item 8 response area on the paper form, and vice versa) before data entry.

If a study is expected to have a large number of respondents who have visual impairments, a large-type version of the survey should be prepared under the permission and supervision of QualityMetric Incorporated. It should be noted that the printing of special forms does add to the cost and complexity of data collection and administration; however, when necessary, this is a good investment. Also note that any large-type version must maintain the instrument’s standardized content and format.

**Non-English Speaking Respondents**

If a respondent cannot read English but can understand and speak English, the Short Form surveys can be administered using one of the standardized interview scripts or, if available, IVRS mode (see Chapter 4). If a respondent does not speak English but is believed to be able to read English at least at the required level, proceed with the administration of the survey. If he or she is unable to read English at this level or prefers to complete a translated version of the survey, the administrator is encouraged to provide the respondent with a version of the survey that is translated into his or her native language. Bilingual respondents should be given the choice of completing either the English form or a translated form, if one is available. A list of translated versions of the Short Form surveys can be found at http://www.iqola.org/ or http://www.qualitymetric.com/. In lieu of the availability of either option, record that the survey was not completed due to a language barrier.

**Guidelines for Administration**

The Short Form surveys should be administered in a standardized manner using the standardized administration formats approved by QualityMetric Incorporated. Any change to the physical format of the survey form or, in the case of interview administration, the interview script may affect the way patients respond to the items, thus compromising the validity of the results. This includes removing specific questions from the printed form or interview script.
Maintaining standardization in administration helps to ensure the accuracy and correct interpretation of results (see Chapter 1).

Each Short Form survey standardized paper form includes specific instructions, questions, and response choices presented in a standardized format. Using the standardized standard (4-week) and acute (1-week) forms that are available from or approved by QualityMetric Incorporated helps ensure standardization of administration and accuracy in the interpretation of survey results.

The flow chart in Figure 3.1 summarizes recommended steps for in-person administration using either the standard or acute version of the paper form.

**When to Administer the Survey**

In a clinical setting, the Short Form surveys should be administered before the respondent sees a healthcare provider so that the interaction between the respondent and the provider does not influence the respondent’s answers to the survey. Ideally, the survey should also be administered before the respondent is asked other health questions or about concurrent illnesses, again so that any discussion of health problems does not influence the respondent’s answers to the survey questions.

**Administration of a Full Survey versus Selected Scales or Items**

The eight Short Form health domain scales cover content areas that can be scored and meaningfully interpreted separately. Administration of all health domain scales, however, allows one to compute the PCS and MCS measures, which yield even more information. Even so, there may be circumstances in which administration of only a subset of Short Form scales is desired. It is not uncommon for QualityMetric Incorporated to grant permission to use one or more individual Short Form health domain scales apart from the others. An example is the use of a single health domain scale in a randomized clinical trial. The validity of

---

**Figure 3.1** Recommended Steps for Administering the Short Form Surveys

- Greet and evaluate the respondent
- If the respondent does not read English or is bilingual, determine which approved language version to use or use interviewer administration
- Determine if visual problems exist. If so, administer a large-font form or use interviewer administration
- Introduce the survey
- Give the respondent the survey form
- Instruct the respondent on how to fill out the form
- Answer any respondent questions before, during, or after the administration
- Retrieve the form upon completion
- Check the form for completeness before the respondent leaves
- Thank the respondent for completing the form
an extracted scale can be maintained, depending on the context in which it is administered. Nevertheless, in some instances, the comparability and/or interpretation of a single scale administered apart from its source could be compromised. If one does administer a single scale from a Short Form survey, it is recommended that it be administered before any disease-specific instrument that may also be administered to the respondent. An exception, however, should be made with regard to the MH scale. MH scale items, which may be upsetting to some respondents experiencing emotional problems, are rarely administered first for that reason.

In other circumstances, one may wish to extract and use only specific Short Form survey items. It is important to be aware that administration of single items from a health domain scale may yield data with limited interpretability. If one wishes to use a brief instrument, either of the two briefest Short Form surveys—the SF-8 or SF-12v2—should be considered. In either case, be aware that single items usually provide coarser measures than multi-item scales or measures.

Administration of Short Form Surveys with Other Generic or Disease-Specific Surveys

The Short Form surveys can be used with a disease-specific survey or another generic survey. However, the Short Form survey items should remain in the same order and format and should not be mixed with items from other instruments. When used with a disease-specific survey, the Short Form survey should be administered before the other measure to avoid sensitizing the respondent to disease-specific health status issues that may then influence his or her responses to the Short Form survey questions about general health status.

Introducing Short Form Surveys to the Respondent

QualityMetric Incorporated recommends the following script (or a variation appropriately reworded to sound more like the administrator’s style of speech) for introducing each of the Short Form surveys:

*We would like to better understand how well you are able to do your usual activities and how you rate your own health. To help us better understand these things about you, please complete this questionnaire about your general health.*

*The questionnaire is simple to fill out. Be sure to read the instructions on the top of the first page [point to them]. Remember, this is not a test and there are no right or wrong answers. Choose the response that best represents the way you feel. I will quickly review the questionnaire when you are done to make sure that all the items have been completed.*

*Please fill out the questionnaire now. I will be nearby in case you want to ask me any questions. Return the questionnaire to me when you have completed it.*

*[As appropriate, add:] You should answer these questions by yourself. Spouses or other family members, or visitors, should not assist you in completing the questionnaire.*

Addressing Problems and Questions

It is not unusual for respondents to ask questions or display certain types of behaviors before, during, or after the administration of the survey. Several common questions or behaviors that experienced Short Form survey administrators have encountered over the years, as well suggestions as to how to respond to them, are addressed in the SF-36v2 User’s Guide (Ware et al., 2007).

Closing

When the respondent returns the survey form, QualityMetric recommends that the administrator check it for completeness. Note whether all of the survey questions have been answered. If the survey is not complete, ask the respondent whether he or she had any difficulty completing it, and record the reasons for non-completion. Finally, thank the respondent using the following exit script (or a variation appropriately reworded to sound more like the administrator’s style of speech):

*Thank you for taking the time to complete this survey. It is possible you will be asked to complete the questionnaire again at a later date.*
In some instances, the respondent may be providing other information in addition to completing the survey. In such cases, a specific thank you for completing the survey may not be required or appropriate. Finally, the completed survey form should be stored in a safe and secure place to ensure confidentiality.

Specific Dos and Don’ts for those administering a Short Form survey are summarized in Table 3.1. They represent efforts that are recommended to maximize the probability of obtaining complete response sets from survey respondents while maintaining the standardization necessary for obtaining valid results.

### Additional Considerations

In addition to the guidelines already provided, the following considerations should be taken into account when administering a Short Form survey as part of a clinical routine or a research protocol.

**When to readminister a Short Form survey.** When a Short Form survey should be readministered is dependent upon which form (standard, acute, 24-hour) has been selected to gather a respondent’s health status information. A Short Form survey should not be readministered any sooner than the time interval (4 weeks, 1 week, 24 hours) that the respondent is asked to consider when answering the survey questions. For example, when using a standard form, responses are based on the 4 weeks just prior to survey administration; therefore, because of the standard form’s 4-week time interval, it is recommended that a minimum of 4 weeks elapse between administrations of the

<table>
<thead>
<tr>
<th>DOs</th>
<th>DON’Ts</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO introduce the Short Form survey and explain the reasons for completing it and the importance and advantages for the respondent of doing so</td>
<td>DO NOT minimize the importance of the Short Form survey</td>
</tr>
<tr>
<td>DO have respondents complete the survey before they fill out any other health data forms and before they see their healthcare provider</td>
<td>DO NOT discuss respondents’ health, health data, or emotions with them before they complete the survey</td>
</tr>
<tr>
<td>DO be warm, friendly, and helpful</td>
<td>DO NOT force or command respondents to complete the survey</td>
</tr>
<tr>
<td>DO request and encourage respondents to complete the entire survey</td>
<td>DO NOT accept incomplete survey forms without first encouraging respondents to respond to any unanswered items</td>
</tr>
<tr>
<td>DO read and repeat a question and its response choices verbatim for respondents if they ask for clarification</td>
<td>DO NOT change the wording of questions or response choices</td>
</tr>
<tr>
<td>DO tell respondents to answer items based on what they think each item means</td>
<td>DO NOT interpret or explain items</td>
</tr>
<tr>
<td>DO have respondents complete the survey by themselves</td>
<td>DO NOT allow spouses, family members, or friends to help respondents complete the survey. Ideally, caregivers should not be present during this assessment</td>
</tr>
<tr>
<td>DO inform respondents if they will be asked to fill out the same survey again</td>
<td>DO thank respondents for completing the survey</td>
</tr>
</tbody>
</table>

Table 3.1

*Short Form Survey Administration Dos and Don’ts*
Short Form surveys’ standard forms. Similarly, when an acute form is selected, only a 1-week interval between administrations is required. Thus, in both clinical and research settings, consider the expected interval between administrations of a Short Form survey when determining which of the available forms is most appropriate for the situation. Note that if the expected interval is less than 1 week, the SF-8’s 24-hour recall form is the only Short Form survey that can meet this particular requirement.

Order effects. In some cases, the Short Form survey will be a component of an assessment battery that the respondent will undergo more than once. In these cases, its place in the order of the initial administration in a battery of assessment instruments and/or procedures should be maintained during follow-up assessments. A clear and concise instruction set should precede the administration of the Short Form survey, regardless of its placement in a battery of instruments.
The increasing recognition of the importance and benefits of and, in some cases, the requirements for assessing health status in clinical and research settings has led to the increased use of the Short Form surveys and other health status measures. Accompanying this increase has been a demand by clinicians and researchers to be able to administer and score these instruments in a mode that best meets their particular needs and circumstances. Consequently, QualityMetric Incorporated has developed and has licensed and certified others to develop materials and software for administering the Short Form surveys in a variety of modes. In addition to the standard paper form, completed in the presence of a clinician or researcher, the Short Form surveys can also be administered via mail-out/mail-back or faxback paper form; face-to-face or telephone interview; Internet, desktop, or handheld device (e.g., personal digital assistant [PDA]) software; or interactive voice response system (IVRS) technology. All administration and scoring products developed by QualityMetric and its certified licensees are reviewed by QualityMetric to ensure that they do not violate the minimum criteria for maintaining standardization as intended by the Short Form survey developers.

The purpose of this chapter is to present both general and specific development and administration considerations and guidelines for various Short Form administration modes. In addition, results from selected published research studies of mode effects are summarized and the results of selected studies are discussed.

General Considerations for All Modes of Short Form Survey Administration

Regardless of the Short Form survey selected or the administration mode being considered, several considerations must be addressed in order to maintain standardization, thus ensuring results interpretability and generalizability. These considerations include doing the following:

1. **Provide a means for identifying the survey respondent.** There should be a means of entering one or more unique identifiers (e.g., name, assigned identification number) for each set of responses obtained from the administration of any of the Short Form surveys. Moreover, entry of an identifier should be required for administration and/or scoring of the Short Form survey.

2. **Maintain the exact wording of survey instructions and items.** Instructions and items should be exactly the same as those on the forms published by QualityMetric. If licensees wish to modify a Short Form to be more relevant to their particular needs or interests, they should instead create a new item that would be administered in addition to the standardized items. Further, the revised and/or new item(s) should be administered after the original Short Form has been completed intact/in its entirety. For example, the SF-36v2 Self-Evaluated Transition (SET) item asks respondents to rate the amount of change they
experienced in their health in general over a 1-year period on the standard (4-week) form or over a 1-week period on the acute (1-week) form. If clinic or research needs require the measurement of reported health transition over a 3-month period, the licensee can create a revised SET item with the 3-month recall period, and administer it after all of the other original SF-36v2 items have been completed.

3. **Maintain the order of Short Form item presentation.** Short Form survey items should be presented in the same order as shown on the forms published by QualityMetric. Reordering of Short Form items may have an impact on the survey taker’s response choice selection, thus affecting the survey results.

4. **Administer Short Form survey items together.** If presented on the same form as items from another survey (e.g., Short Form items included as a part of a case report form), administer the Short Form items together. Do not intersperse Short Form items with items from other measures, or vice versa.

5. **Administer generic Short Form survey content first.** If integrated with a larger survey (e.g., with a disease-specific module), administer the generic Short Form survey content first, then remaining survey modules.

6. **Exercise caution when administering only select items from the Short Form surveys.** There may be circumstances in which administration of only selected Short Form survey items is desired (e.g., administering/scoring only the SF-36v2 PF scale) in a randomized clinical trial. In doing so, the validity of an extracted scale may be compromised. If one does administer items to score a single scale from one of the Short Form surveys, they should be administered together. As noted previously, generic content should be administered before remaining survey modules (e.g., disease-specific survey). An exception, however, may be made for items in the MH scale. Because some MH items are sensitive in nature, administering them early in an inventory may impact participant responses to subsequent survey items. In other circumstances, a user may wish to extract and use only specific Short Form items. It is important to be aware that administration of single items from a health domain scale may yield data with limited interpretability. If one wishes to use a brief instrument, full administration of the SF-8 or SF-12v2 should be considered. However, licensees should be aware that single item measures provide coarser measurement than multi-item scales or measures and may not provide the level of precision required to achieve the assessment/study goals (e.g., monitoring person-level changes in health status over time).

7. **Display the appropriate copyright notice provided by QualityMetric as part of any Short Form administration.** Specific considerations for each of the Short Form administration modes are presented in the sections that follow.

### Administration by Paper Form

#### Mode-Specific Considerations

The instructions, recommendations, and considerations presented in Chapter 3 apply to paper-form administration when the survey administrator or proctor administers the SF-36v2 to one or more individuals in person.

#### Certification Standards

QualityMetric has developed paper forms for all forms (4-week, 1-week, and/or 24-hour recalls) of the Short Form surveys and makes them available at no charge to licensees in Microsoft Word or PDF format. Each of the forms was developed to be generally consistent with the formatting methods recommended by Mullin, Lohr, Bresnahan, and McNulty (2000). Mullin et al.’s recommendations are based on
cognitive design principles and formatting tech-
iques that were identified from an extensive
review of the literature. From these findings,
the authors developed a set of formatting guide-
lines for HRQOL instruments, applied them to
several existing HRQOL instruments used in-
ternationally (including the SF-36), and then
presented the modified forms to the instru-
ments’ developers, survey research methodolo-
gists, and research experts for approval.

Based on their findings, Mullin et al. (2000)
identified six characteristics of a “soundly de-
dsigned” survey form that help to eliminate navi-
gation errors, increase ease of responding, and
increase respondents’ motivation to complete
self-administered survey forms. These charac-
teristics include:

1. Simplicity in design, such that the form
does not contain any unnecessary or
redundant information or graphical
complexities.

2. Consistency in design, whereby the re-
sponse tasks of similar-looking ques-
tions are consistent and different types
of questions are visually distinct.

3. Organization in design, in which the
placement of groups of items or the in-
structions for their completion is based
on their commonality or relevancy to
each other.

4. Naturalness in design, so that the format
of the form is consistent with or takes
advantage of what comes naturally to
the respondent.

5. Clarity in design, in which important
features of the form are presented in a
clear, distinct manner.

6. Attractiveness in design, achieved
through ensuring the inclusion of the
other five characteristics along with a bit
of “artistry” that can facilitate comple-
tion of the form.

Specific ways in which each of these six
characteristics can be achieved are summarized
in Table 4.1, which also directs the reader to
examples of these characteristics that can be
found on the SF-36v2 form.

A checklist to help guide the development
of certified Short Form paper form modes of
administrations is presented in Appendix A.1.

Administration by Mail

Mode-Specific Considerations

Administration of a Short Form survey
using a mail-out and mail-back (MO/MB) sys-
tem is a common and efficient means of con-
ducting research that involves large numbers
of subjects sampled from across a large geo-
ographical area and/or multiple administrations
of the instrument over long periods of time. This
method can also be useful for clinical purposes.
For example, it can provide a means of moni-
toring patients with chronic conditions during
long intervals (e.g., 6 months) between sched-
uled visits. It can also be used to assess the en-
during effects of treatment long after treatment
has been terminated.

There are many issues to consider when
deciding whether to use an MO/MB system. In
addition to concerns about maintaining patient
privacy and confidentiality and standardization
of administration, there are other practical con-
siderations, such as identifying the most effec-
tive MO/MB methodology for the population
being assessed, the cost of implementing such
a system, and the expected return on that in-
vestment. It is beyond the scope of this guide
to adequately address these issues. Those plan-
ing a MO/MB study can find additional infor-
mation regarding the benefits and limitations
of this design in Dillman, Smyth, and Chris-
tian (2009) or other resources that specifically
address these and other important issues to con-
sider when conducting mail surveys.

Certification Standards

The same general standards for paper ad-
ministration apply to paper forms designed
specifically for MO/MB administration; how-
ever, some additional considerations are nec-
essary. For example, respondents must be
provided with contact information for the
study sponsor. Instructions and/or a telephone
number to call if questions arise should be in-
cluded in a cover letter accompanying the
mailed form. The mail-back address should
also be included in the event that the mail-back
envelope is lost. Alternatively, this information
can be printed on the form in an unobtrusive
Table 4.1  
Recommended Design Characteristics for HRQOL Self-Report Instruments

<table>
<thead>
<tr>
<th>Design Characteristic</th>
<th>Means of Achievement</th>
</tr>
</thead>
</table>
| Simplicity            | • Eliminate grid lines from items presented in a tabular format (see Figure 4.1)  
                       | • De-emphasize or eliminate information that is irrelevant to the respondent  
                       | • Employ graphical information, such as check boxes, to identify the area in which responses are to be indicated (see Figure 4.1)  
                       | • Be selective in the use of emphasis of terms or phrases (e.g., use underlining to emphasize important criteria in selecting a response or to differentiate highly similar adjacent text)  
                       | • Do not repeat instructions unnecessarily |
| Consistency           | • Maintain the graphical structure of items throughout the form (see Figures 4.1 and 4.2); avoid mixing formatting aspects such as page orientation (landscape vs. portrait) and one- and two-column formatting of response choices |
| Organization          | • Place item completion instructions in proximity to the item  
                       | • Group items according to topics which are covered by the items; identify new groups of items by short headings and lead-in statements |
| Naturalness           | • Format questions and response choices to create a natural reading flow (e.g., left to right, up and down)  
                       | • Left-justify text to be consistent with the natural flow of reading |
| Clarity               | • Use graphical guides, such as the three-side box over response options and upside-down triangles under the response choices on the Short Form survey forms (see Figure 4.1)  
                       | • Use at least a 60% contrast between survey text and background  
                       | • Make judicious use of space to separate questions from responses, items from other items, groups of items from other groups of items, and columns of text  
                       | • Based on research, use the following recommended print and font specifications:  
                         o Times or Times Roman font; also consider other serif fonts such as Garamond, Century, Roman, Georgia, and Calisto (see Dillman, Smyth, & Christian, 2009)  
                         o A minimum of an 8-point font, with 12-point font generally recommended and a 16- to 18-point font for elderly and vision-impaired respondents; Dillman et al. recommend 10 to 12-point fonts.  
                         o Avoidance of italicized text and words in ALL UPPERCASE letters  
                         • On paper forms, number each page, with the word “Page” before the number  

(continued on next page)
Table 4.1
Recommended Design Characteristics for HRQOL Self-Report Instruments (continued)

<table>
<thead>
<tr>
<th>Attractiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Make item and page numbers highly visible</td>
</tr>
<tr>
<td>• Incorporate suggestions for the other five</td>
</tr>
<tr>
<td>characteristics</td>
</tr>
<tr>
<td>• Incorporate asymmetric composition of text (e.g.,</td>
</tr>
<tr>
<td>make question text bold or larger than response</td>
</tr>
<tr>
<td>choice text; see Figure 4.1)</td>
</tr>
<tr>
<td>• Include a cover page (or introductory screen) that</td>
</tr>
<tr>
<td>will help increase the respondent’s interest in</td>
</tr>
<tr>
<td>completing the survey and thank him/her for</td>
</tr>
<tr>
<td>completing the survey</td>
</tr>
</tbody>
</table>

Note: From Mullin et al. (2000).

Figure 4.1 Item 5 from the SF-36v2 Paper Form

5. During the past 4 weeks, how much of the time 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)?

<table>
<thead>
<tr>
<th>All of the time</th>
<th>Most of the time</th>
<th>Some of the time</th>
<th>A little of the time</th>
<th>None of the time</th>
</tr>
</thead>
<tbody>
<tr>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
</tr>
</tbody>
</table>

a. Cut down on the amount of time you spent on work or other activities.............................................. □ 1 □ 2 □ 3 □ 4 □ 5
b. Accomplished less than you would like .............................................. □ 1 □ 2 □ 3 □ 4 □ 5
c. Did work or other activities less carefully than usual .............................................. □ 1 □ 2 □ 3 □ 4 □ 5

Figure 4.2 Item 6 from the SF-36v2 Paper Form

6. 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?

<table>
<thead>
<tr>
<th>Not at all</th>
<th>Slightly</th>
<th>Moderately</th>
<th>Quite a bit</th>
<th>Extremely</th>
</tr>
</thead>
<tbody>
<tr>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
<td>▼</td>
</tr>
<tr>
<td>□ 1</td>
<td>□ 2</td>
<td>□ 3</td>
<td>□ 4</td>
<td>□ 5</td>
</tr>
</tbody>
</table>
location (e.g., at the bottom of the last page of
the form).

A checklist to help guide the development of
Certified Short Form mail-out paper form
modes of administrations is presented in Ap-
pendix A.2.

Administration by Interview

Mode-Specific Considerations

The Short Form surveys can be adminis-
tered by interview, either face-to-face or over
the telephone. In either case, they should be
administered using the approved script avail-
able from QualityMetric. The interviewer
should request that the patient’s caregiver (if
present) leave the room during administration
of the survey, unless circumstances indicate
that it would be better for the caregiver to be present.

As with any health survey, interviewers
should familiarize themselves with Short Form
administration guidelines prior to conducting
the interview administration and should ensure
that the assessment environment is conducive
to its purpose. An introduction to the adminis-
tration, such as the following, should be given
prior to reading the first question:

We would like to better understand how
you feel, how well you are able to do your usual
activities, and how you rate your own health.
To help us better understand these things about
you, please answer some questions about your
general health.

This is not a test and there are no right or
wrong answers. Choose the response that best
represents the way you feel. Please answer ev-
ery question. As we proceed, please feel free
to ask me any questions.

If the respondent is to indicate his or her
answers on a paper form rather than by giving
an oral response, the interviewer should pro-
vide the respondent with a firm writing surface,
such as a clipboard or table top, and a pen or a
#2 pencil (if using a scannable answer sheet).

The interviewer should not attempt to in-
terpret or explain any of the items; however, he
or she may repeat an item verbatim if asked.
The interviewer should request and encourage
respondents to provide an answer for each ques-
tion but should not force them to do so. Addi-
tional instructions specific to each section of
the assessment are presented in QualityMetric’s
Short Form interview scripts.

When administered to respondents with
mild cognitive impairment or early dementia,
it is recommended that the interviewer be suit-
ably prepared and trained to properly adminin-
ister the survey. Respondents with mild cognitive
impairment may demonstrate some behaviors
unlike other groups of respondents, and pa-
tience and redirection may be necessary to en-
courage survey completion.

If possible, the same administrator should
interview each respondent if subsequent
readministration of the Short Form survey is
required.

Note that when using the interview script
to administer the SF-36v2, items 7 and 8 from the Bodily Pain health domain
scale are administered in reverse order from
the way they appear on the printed SF-36v2
paper form. Reversing the order of the pre-
sentation of these two items facilitates the flow
of the administration. When conducting an
interview-administered survey, the interviewer
must inform those persons responding on the
standardized SF-36v2 paper form of the item
order discrepancy to ensure that the intended
responses to each of these items are marked in
the appropriate response areas. That is, the re-

dose to item 7 from the interview script should
be entered in the item 8 response area on the
paper form, and vice versa. If the administrator
is writing down the respondent’s oral answers,
he or she must be mindful of the reverse or-
dering of the items when entering and scoring
the responses.

Certification Standards

QualityMetric has developed interview
scripts for the Short Form surveys that should
be used when circumstances require face-to-
face or live telephone administration of a sur-
vey. These scripts include standardized probes
and transition text, which the interviewer should
use during the course of administration. As with
the paper forms, there should be no deviation
from the item, transition, or probe text indicated
Administration by Internet, Desktop, or Handheld Device

Mode-Specific Considerations

Administration of the Short Form surveys by the Internet, desktop computer, or a handheld device (e.g., PDA) allows licensed vendors and researchers to take advantage of current technology to facilitate rapid and efficient data gathering. Each of these electronic modes of administration has some of the same advantages and involves some of the same issues as the MO/MB methodology. In addition, online screen presentation should maintain as much as possible the standardized paper-form format of the items until other ways of item presentation have been empirically investigated. At the same time, Internet administration design requires many considerations that are distinct from the paper form considerations. The works of Dillman and his colleagues (Christian, Dillman, & Smyth, 2007; Dillman, 2000, 2006; Dillman & Smyth, 2007; Dillman et al., 2009; Smyth, Dillman, Christian, & Stern, 2006) and others (e.g., Couper, Traugott, Lamias, 2001; Peytchev, Couper, McCabe, & Crawford, 2006) are excellent resources to consult for those developing Internet administration modes for the Short Form surveys. A demonstration of an online administration of the SF-36v2 developed by QualityMetric is available at http://www.amihealth.com/.

Certification Standards

As with other modes of administration developed for its Short Form surveys, QualityMetric’s criteria for certifying electronic survey administration—Internet, desktop, handheld device, and IVRS—are focused on ensuring that the administration of a Short Form survey maintains the standardization for that survey. Because the survey instructions, questions, and response choices are presented via auditory recording, the IVRS mode is addressed separately in the next section of this chapter.

In general, the online presentation of the Short Form survey items should be consistent with the presentation described earlier for the paper form mode of administration, across the entire Short Form survey administration. For example, in Figure 4.3, note the similarity of the electronic visual presentation of Items 5 and 6 from the SF-36v2 and their recommended paper form presentations (shown earlier in Figures 4.1 and 4.2). However, unlike the paper form, sans-serif fonts (Verdana, Arial, etc.) are preferred over the serif fonts for documents displayed electronically. Also, administering items typically presented together in a grid (see Item 5 in Figure 4.3) one at a time as single items may be considered. If this approach is taken, the licensee should contact QualityMetric for the correct text that should be used in presenting these items singly.

Please note that the initial instructions to the respondent for completing an electronic version of any Short Form survey should indicate how to enter a response. For example, the instructions for completing the electronic form represented in Figure 4.3 indicate, “Please take time to read and answer each question carefully, and click the circle that best represents your answer” [italics added]. Instructions for other means of response entry (e.g., touch screen) should be specific to that mode.

Aside from item presentation, Internet, desktop, and handheld device administration must allow the respondent to engage in the same type of survey completion behaviors as would be possible during a paper form administration. At the minimum, the software must instruct how and allow the respondent to perform the following tasks:

1. Enter the respondent’s name or identification number
2. Enter an item response
3. Change an item response
4. Have instructions and items repeated
5. Skip a survey item
6. Review previous item responses
7. Discontinue and later return to the administration, with or without saving already entered responses
8. Submit survey responses for database entry and/or scoring and terminate the administration upon completion.

9. Display error messages when errors (e.g., entry of responses outside of allowable range) occur during the performance of any of the listed tasks.

Note that the respondent should be required to enter identifying information (name and/or ID number) and any other non-survey information that may be required by the user (e.g., gender, age). In addition, the following recommendations from Dillman et al. (2009) should be considered:

- Design the software in a manner that takes into consideration the technological capabilities of the intended respondents (e.g., familiarity and/or experience with taking online surveys, devices on which the survey will be administered).
- Carefully consider whether the Short Form items should be presented one-by-one or as a continuous form in which the respondent moves from item to item by scrolling up and down the form.

Accordingly, Short Form software developed and currently marketed by QualityMetric employs the continuous form format, as it best replicates the paper form mode of administration. However, QualityMetric has qualitative evidence that the single-item presentation is well accepted across a few chronic condition samples, and it will be incorporating single-item displays into products that are currently in development. Also, users may find it beneficial to use the single-item presentation format under specific circumstances. Examples of such circumstances include surveys that are programmed for administration by PDA or other handheld device, or if the Short Form survey is being administered as part of a lengthy survey in which other surveys or individual items are included. Dillman et al. present a particularly good discussion of the arguments for and against each type of item presentation format.

- If the Short Form survey is either administered alone or as part of a combined battery of other survey instruments or...
questions, include informative and interesting welcome screens and a closing screen, which is presented after the last item has been administered.

- Be prudent in the use of any audiovisual capabilities, especially ones in which there is limited research regarding their effects, and avoid the presentation of any unnecessary information or other “clutter.”

- In addition to any other quality assurance testing that might be performed, test the software on all platforms, browsers, and user settings that might be employed for the administration of the survey.

Dillman and Smyth (2007) and Dillman et al. (2009) provide additional recommendations that those developing Web-based Short Form administration capabilities may wish to consider.

A checklist to help guide the development of certified Short Form Internet, desktop, or handheld device modes of administration is presented in Appendix A.3.

### Administration by IVRS

#### Mode-Specific Considerations

In general, interactive voice response systems (IVRS) allows for the gathering of information using a telephone. Survey administration via IVRS typically involves the presentation of prerecorded survey questions that the respondent answers orally (on systems utilizing voice recognition software) or, more commonly, by using the telephone keypad to select a numbered, multiple-choice response option or to give a numeric response, such as that pertaining to age or the frequency of a behavior or event.

Administration of the Short Form surveys using IVRS entails some of the same considerations that must be addressed when using the telephone interview or Internet administration format. In particular, the standardized interview scripts developed especially for IVRS administration should be used for recording the administration of the survey. Again, the reader is referred to Dillman (2000) and Dillman et al. (2009) for recommendations for the development and implementation of IVRS survey administration capabilities.

#### Certification Standards

The approved Short Form IVRS interview scripts available from QualityMetric should serve as the basis for the administration of the Short Form survey via IVRS. Generally, when using a voice or keypad response entry method, respondents must be able to engage in the same types of survey completion behaviors as they would if they were completing a survey through a live interview or a visually oriented mode (e.g., paper form, Internet). Thus, IVRS software must allow for and the recorded administration must instruct the respondent how to perform the following tasks:

1. Enter the respondent’s name or identification number
2. Enter an item response
3. Change an item response
4. Have instructions and items repeated
5. Skip an item
6. Submit survey responses for database entry and/or scoring and terminate the administration upon completion unless data submission automatically occurs upon completion of the survey.

Recorded error messages must also be developed and programmed, error messages that are then presented when errors occur. Systems with voice recognition/response entry capabilities (instead of or in addition to key entry response capabilities) should provide error messages for unclear voice responses (e.g., “Did I hear you say . . . ?”), offer access to human assistance at appropriate times (e.g., “Do you wish to speak with an operator?”), and allow the respondent to carry out functions by voice command (e.g., “Repeat the question”).

Although not required, the licensed vendor may also wish to include functionality that will allow the respondent to review previously entered item responses and/or discontinue and later return to the administration with or without saving already entered responses.

A checklist to help guide the development of certified Short Form IVRS modes of administration is presented in Appendix A.4.
Administration by Translated Version of a Short Form Survey

Mode-Specific Considerations

There are several circumstances that require a Short Form survey to be administered in a language other than the language in which the surveys were developed (i.e., U.S. English). One common example is their use in a clinical drug trial being conducted in several non-English speaking countries. Another example is use for clinical purposes in the U.S. where a significant percentage of the population being assessed does not have English as its primary language (e.g., Hispanic populations in Florida, California, and Texas). QualityMetric makes each of the Short Form surveys available to licensees in a great number of languages and adaptations. For example, at the time of the publication of this guide, the paper form version of the SF-36v2 was available in 110 languages or English language adaptations. The numbers of translations and adaptations for paper form versions of the other Short Form surveys are similar. Thus, in most instances, QualityMetric will meet the user’s needs for one or more translations of a Short Form survey. When this is not the case, the user can contract with QualityMetric for the development of the desired language or adaptation.

Certification Standards

In order to be certified and licensed for use, each translation of a Short Form survey must have been developed through the same standard process that QualityMetric uses when translating or adapting its surveys. This process builds on the methodology developed by the International Quality of Life Assessment (IQOLA) Project (Bullinger et al., 1998) and follows guidelines proposed by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force for Translation and Cultural Adaptation (Wild et al., 2005) and the Scientific Advisory Committee (SAC) of the Medical Outcomes Trust (SAC, 2002). The methods and results from the translations and adaptations studies of the SF-36 that were conducted for the IQOLA Project are described in a series of articles published in a special issue of the Journal of Clinical Epidemiology (in particular, see Gandek & Ware, 1998a; Gandek & Ware, 1998b). QualityMetric currently requires a translation or English-language adaptation of a Short Form survey to undergo the following development process:

1. Two independent forward translations of the source (U.S. English) questionnaire are developed by translation consultants, who are native speakers of the target language and reside in the country for which the translation is being developed.
2. The resulting forward translations are reconciled by the consultants in order to develop one preliminary forward translation.
3. A third translator, who is bilingual in the target language and English, then translates this preliminary translation back into English.
4. The backward translation is reviewed to check for content equivalence with the source questionnaire by reviewers in the U.S. who have expertise in translation development and in measurement. Potential discrepancies between the source and translated versions are discussed by the reviewers and the primary in-country consultant until all issues are resolved and a second preliminary forward translation is developed.
5. The translation then is tested for face validity and administrative feasibility under supervised self-administration, with at least five individuals who are native speakers of the target language and live in the country for which the translation is being developed. These individuals differ in sociodemographic characteristics (e.g., age, gender, education) and are representative of the general population. (Unlike the case with the generic Short Form surveys, when a disease-specific questionnaire is being translated or adapted, the representative population is one whose
members have the disease in question.) Any items that are identified as problematic (e.g., difficult to answer, unclear, not understood correctly) are discussed among the backward translation reviewers and in-country translators until all issues are resolved.

6. The final, formatted translation then goes through proofreading by at least one, and generally two, native speakers of the target language.

This same process is required for all translations and adaptations, regardless of the mode in which the translated version of the survey is administered.

Effects of Modes of Administration on Short Form Survey Results

When selecting a mode of administration for Short Form surveys, vendors and researchers licensed by QualityMetric should carefully consider effects of administration mode on the results of HRQOL instruments in general and, specifically, on the results of the Short Form surveys that have been demonstrated in the literature. It is beyond the scope of this guide to provide a comprehensive review of administration mode effects for all HRQOL instruments. However, two recent large studies investigating the effects of administration modes on HRQOL survey results are worth noting.

The first is a study by Hanmer, Hays, & Fryback, (2007) that compared findings from four large-scale, nationally representative studies employing two or more HRQOL measures (i.e., the EuroQol EQ-5D, Health Utilities Index Mark 3 [HUI3], the commonly used general health rating item with 5 response choices [EVGFP], and a visual analog scale [VAS] rating of current health). One of three modes of administration—live telephone interview survey, self-administered mailed survey, and self-administered paper survey with an interviewer present—was used in each study. The total sample across all four studies included more than 35,000 U.S. and Canadian adults. Findings indicated that when mode of administration was the same, age- and gender-stratified scores were generally consistent across the EQ-5D, HUI3, and VAS. An age effect was noted on the EQ-5D, HUI3, and EVGFP for those over 70 years, with telephone-interviewed participants reporting better health status than self-report paper form participants. However, female and older participants reported worse health status.

In another study, Gwaltney, Shields, and Shiffman (2008) conducted a meta-analysis of 65 published studies that investigated the equivalence of paper and electronic versions of a variety of HRQOL measures, including seven SF-36 studies. In some studies, a PDA was employed as the electronic mode, whereas the other studies employed a PC or laptop computer. Mean HRQOL scores for the paper and electronic versions were not significantly different (average mean difference was 0.2% of the scale range). Thirty of the 32 studies reporting correlations between paper form and computerized assessments had average correlations greater than .75 and the weighted summary correlation between modes was .90. In the four studies reporting paper-paper test-retest reliability and paper-computer concordance, the average correlations (.88 and .91, respectively) did not differ significantly, nor did the average PDA-paper correlation (.91) differ significantly from the average PC-paper correlation (.90). Although age was found to be negatively related to the paper form-electronic correlations, the trend was very small and the correlations for the oldest age groups were greater than .75. Overall, Gwaltney et al. concluded that the two modes of administration produce equivalent HRQOL scores.

Several studies involving the Short Form surveys have demonstrated that different methods of administration may affect survey results. The findings from a few of the key studies are summarized in the following section and in Table 4.2.

Studies have shown that responses to the SF-36 tend to be more favorable when data are collected by face-to-face or telephone interview (McHorney, Kosinski, & Ware, 1994; Ware, Kosinski, & Keller, 1994). In a randomized trial
### Table 4.2

**Effects of Method of Data Collection on Short Form Results: Findings from Selected Studies**

<table>
<thead>
<tr>
<th>Investigation</th>
<th>Sample</th>
<th>N</th>
<th>Short Form Survey</th>
<th>Methods of Data Collection</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bliven, Kaufman, &amp; Spertus (2001)</td>
<td>Cardiology clinic outpatients</td>
<td>66</td>
<td>RAND-36</td>
<td>Paper form, touch-screen Internet</td>
<td>Both interclass correlation coefficients (ICC) and Pearson correlations for matching scales from the two modes were significant, suggesting no systematic variation by mode.</td>
</tr>
<tr>
<td>Buskirk &amp; Stein (2008)</td>
<td>Cancer survivors</td>
<td>140 &amp; 155</td>
<td>SF-36</td>
<td>Telephone interview, mail out/mail back (MO/MB)</td>
<td>With each sample using only one or the other administration mode, all scales from both modes achieve a Cronbach’s alpha ≥ .70, with alphas for the MO/MB scales generally higher, to as much as .11; overall, mean unadjusted scale scores were found to be higher for the phone mode, and the multivariate effect of mode was seen in higher RP, VT, and MH scores for the phone mode.</td>
</tr>
<tr>
<td>Caro, Caro, Caro, Wouters, &amp; Juniper (2001)</td>
<td>Asthma outpatient</td>
<td>68</td>
<td>SF-36</td>
<td>Paper form, electronic diary</td>
<td>Concordance for identical responses to items across the two administration formats ranged from 59% to 91%, with almost half achieving 80% or higher; health domain scale ICCs ranged from .83 to .97, with no consistent variation being noted.</td>
</tr>
<tr>
<td>Hanscom, Lurie, Homa, &amp; Weinstein (2002)</td>
<td>Low back pain patients</td>
<td>15, 815 &amp; 3,574</td>
<td>SF-36</td>
<td>Paper form, laptop computer</td>
<td>Data quality for the computer responders was found to be better than that for the paper-form responders from several perspectives, including missing value rates for health domain scales, component summary measures, and overall survey;</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Study</th>
<th>Description</th>
<th>N</th>
<th>Short Form</th>
<th>Mode of Administration</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jorngarden, Wettergen, &amp; von Essen (2006)</td>
<td>Swedish citizens covered by civil registration, age 13-24 years</td>
<td>585</td>
<td>SF-36</td>
<td>MO/MB, telephone interview</td>
<td>Cronbach’s alphas for telephone interviews ranged from .62 (BP) to .86 (PF), while ranging from .77 (RP) to .91 (PF) for the MO/MB administration; males scored significantly higher than females on 5 of the 10 SF-36 measures from the telephone interview and on 7 of the 10 measures on the MO/MB form; while the 16-19 year-old subsample did not differ significantly from the 20-23 year-old subsample on any of the 10 measures by either mode of administration, the 13-15 year-old sample scored significantly higher than the 16-19 year-olds on the telephone RE and the 20-23 year-olds on the telephone VT, SF, and MCS</td>
</tr>
<tr>
<td>Lungenhausen, Lange, Maier, Schaub, Trampisch, &amp; Endres (2007)</td>
<td>Randomly sampled patients taking part in the German Acupuncture Trials</td>
<td>823</td>
<td>SF-12</td>
<td>MO/MB, telephone interview</td>
<td>Mean MO/MB MCS scores were significantly lower than telephone interview scores (mean difference = 3.5 T-score points) when compared to the mean MO/MB PCS T-score difference of 1.8 (considered to be within the range of equivalence for this study); administration order effects were also noted for MCS</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
<th>Sample Size</th>
<th>Survey Tool</th>
<th>Data Collection Method</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lyons, Wareham, Lucas, Price, Williams, &amp; Hutchings (1999)</td>
<td>British general medicine, urology, endocrinology, and gastroenterology outpatients</td>
<td>210</td>
<td>SF-36</td>
<td>MO/MB, face-to-face interview</td>
<td>MO/MB scores were generally lower than interview scores, regardless of order of administration, with the difference being significant ($p &lt; .05$) for the GH, MH, PF, RE, and VT scales.</td>
</tr>
<tr>
<td>McHorney, Kosinski, &amp; Ware (1994)</td>
<td>U.S. general population</td>
<td>1,682 &amp; 782</td>
<td>SF-36</td>
<td>MO/MB, face-to-face interview, telephone interview</td>
<td>Results revealed a lack of equivalence between responses to MO/MB surveys and those from personal interviews administered by phone. Average scores for the SF-36 MCS measure were 2.43 T-score points higher for those interviewed by telephone versus by MO/MB. Underlying this difference in MCS scores were significant differences for all health domain scales except GH.</td>
</tr>
<tr>
<td>Perkins &amp; Sanson-Fisher (1998)</td>
<td>Australian random community sample</td>
<td>418 &amp; 421</td>
<td>SF-36</td>
<td>Telephone interview, MO/MB</td>
<td>Data collection costs were lower for the telephone mode; significantly higher overall consent rate was achieved with telephone mode, with younger respondents being more likely to refuse participation via MO/MB mode and older respondents more likely to refuse administration via telephone. The rate of missing responses was higher with the mail mode; health ratings were generally more favorable during the telephone administration. Cronbach’s alpha coefficients for the RP, VT, SF, and RE scales were found to differ significantly by administration method, with the higher coefficients being obtained</td>
</tr>
</tbody>
</table>

(continued on next page)
Table 4.2
Effects of Method of Data Collection on Short Form Results: Findings from Selected Studies (Continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Sample Size</th>
<th>Administration</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ravens-Seiberer, Erhart, Wetzel, Krugel, &amp; Brambosch (2008)</td>
<td>Two groups of randomly sampled German adults with children aged 8-17</td>
<td>899 &amp; 791</td>
<td>SF-8</td>
<td>Telephone interview, MO/MB</td>
</tr>
<tr>
<td>Ryan, Corry, Attewell, &amp; Smithson (2002)</td>
<td>Healthy high school and university students and staff, senior citizen club members, and chronic pain patients in Australia</td>
<td>101</td>
<td>SF-36</td>
<td>Paper form, electronic (unspecified)</td>
</tr>
</tbody>
</table>
| Saleh, Radosevich, Kassim, Moussa, Dykes, Bottolfson, & Orthopedic (knee/hip) patients | 87   | SF-36 | MO/MB, palmtop computer | No significant differences in mean squares, $SD$s, floor or ceiling percentages, or ICC retest correlations were found (coefficients (continued on next page)
<table>
<thead>
<tr>
<th>Study</th>
<th>Setting</th>
<th>Sample Size</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>et al. (2002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suris, Borman, Lind, &amp; Kashner (2007)</td>
<td>VA medical and psychiatric patients</td>
<td>97</td>
<td>SF-36</td>
</tr>
<tr>
<td>Ware, Kosinski, DeBrota, Andrejasich, &amp; Bradt (1995)</td>
<td>Patients recruited at ambulatory care facilities and nonmedical business work sites</td>
<td>525</td>
<td>MO/MB, telephone interview, IVRS</td>
</tr>
<tr>
<td>Ware, Kosinski, Dewey, &amp; Gandek (2001)</td>
<td>U.S. general population</td>
<td>750, 768, &amp; 907</td>
<td>MO/MB, Internet, telephone interview</td>
</tr>
</tbody>
</table>

Preliminary results revealed no differences in data quality or tests of scaling assumptions across the three administration methods. Average PCS scores did not differ by method. However, average MCS scores were more favorable (by 1.8 NBS points, \( p < .01 \)) with the personal telephone interview compared to both self-administered and IVRS-administered surveys. The latter two methods did not differ.

Five health domain scales differences (PF, BP, GH, VT, and MH) were significant, with higher average T-scores for phone interviews (1.2–3.75 points) in comparison (continued on next page)
Table 4.2

Effects of Method of Data Collection on Short Form Results: Findings from Selected Studies (Continued)

<table>
<thead>
<tr>
<th>Study Reference</th>
<th>Patient Group</th>
<th>Sample Size</th>
<th>Mode</th>
<th>Method Details</th>
<th>Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weinberger, Oddone, Samsa, &amp; Landsman (1996)</td>
<td>Outpatient veterans</td>
<td>172</td>
<td>SF-36</td>
<td>MO/MB, face-to-face interview, telephone interview</td>
<td>With one exception, all 8 scales achieved an acceptable level of internal consistency (.70); average scale test-retest correlations for the 3 modes ranged from .79 to .83; scale score differences between modes were not significant; with a few exceptions, scale ceiling and floor percentages were approximately the same; average scale correlations between modes of administrations ranged from .74 to .81.</td>
</tr>
<tr>
<td>Wilson, Kitas, Carruthers, Reay, Skan, Harris, et al. (2002)</td>
<td>British rheumatology outpatients</td>
<td>51</td>
<td>SF-36</td>
<td>Paper-and-pencil, desktop computer</td>
<td>Cronbach’s alphas for all but the BP and SF scales ranged from .83 to .94 for the paper version and from .84 to .95 for the computer version, while Spearman’s $p$ for BP and SF ranged from .68 to .97 (all $p$’s &lt; .01); scale score correlations for the two modes ranged from .80 (SF) to .96 (PF); 67% of the patients preferred the computerized version, with significantly</td>
</tr>
</tbody>
</table>

(continued on next page)
Table 4.2  
Effects of Method of Data Collection on Short Form Results: Findings from Selected Studies (Continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Population Description</th>
<th>Sample Size</th>
<th>Measure</th>
<th>Data Collection Method</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood &amp; McLauchlan (2006)</td>
<td>Elderly British patients who underwent total hip arthroplasty</td>
<td>90</td>
<td>SF-36</td>
<td>MO/MB, face-to-face interview</td>
<td>The interview sample’s RP and RE scores were significantly higher ( p' \text{s} &lt; .01 ) than those for the MO/MB sample.</td>
</tr>
</tbody>
</table>

- more of the younger group (< 47 years old) than the older group having this preference.
conducted during the norming of the SF-36, McHorney et al. found a lack of score equivalence on some domains for mail and personal telephone interview modes. Average MCS T scores for telephone interview respondents were 2.43 points higher ($\pm 0.3$, $p < .001$) than scores for MO/MB respondents. This difference is nearly one fourth of a standard deviation, a noteworthy amount. In other terms, the effect of data collection method on MCS T scores is approximately one fourth the impact of a depressive disorder. Underlying this difference in MCS T scores were significant differences for seven of the eight health domain scales (all but GH). There was no effect on the PCS measure.

In a randomized cross-over study, Ware, Kosinski, DeBrota, Andrejasich, and Bradt (1995) examined the effect of three SF-36 administration methods on patient acceptance, cost and quality of data, mean scale scores, test-retest and internal consistency reliability, and empirical validity. Participants recruited at ambulatory care facilities and non-medical business work sites ($N = 525$) were randomly assigned to complete their first SF-36 by personal interview over the telephone, self-administration through the mail, or IVRS. Two-weeks later, half of the patients completed the survey again using the same administration method, while the other half were randomly assigned to another method. Preliminary results revealed no differences in data quality or tests of scaling assumptions across the three administration methods. Average PCS T scores did not differ by mode. However, average MCS T scores were higher (indicating better health) by 1.8 points ($p < .01$) in the telephone interview as compared to both mail and IVRS surveys. The latter two methods did not differ. Results from the Ware, Kosinski, DeBrota, et al. (1995) study indicate that SF-36 mental health scales administered by personal telephone interview should not be directly compared with those administered by other methods without adjustment for the effect of data collection method.

Because of the impact of data collection methods, as demonstrated in studies of the SF-36, and the common practice of varying data collection methods within and between studies, studies of the data collection methods were replicated and extended during the norming of the SF-8 (Ware et al., 2001). All studies also included the SF-36v2 to replicate previous analyses of SF-36 data. Moreover, the norming studies were expanded to include Internet administrations of both the SF-8 and SF-36v2 surveys ($N = 768$). Internet responses were compared with those obtained from personal telephone interview ($N = 750$) and MO/MB self-administered forms ($N = 907$).

Ware et al. (2001) found that for the SF-36v2 health domain scales and component summary measures, the pattern of differences in average scores between the telephone interview and MO/MB groups were not unlike the pattern observed in previous studies, although the obtained differences tended to be somewhat smaller. Also, differences were apparent in both the PCS and MCS measures. Mean scale T scores for the telephone interview respondents were significantly higher on five of the eight SF scales (PF, BP, GH, VT, MH) than for mail respondents. PCS and MCS T scores were also significantly higher for telephone interview respondents ($1.68$, $p < .001$, and $1.38$, $p < .01$, respectively). No significant differences in SF-36v2 scale or component summary measure T scores were found between Internet and mail respondents.

It is important to note that the Ware et al. (2001) studies of data collection methods involved general population samples of convenience in which study participants were not randomized to data collection methods. Further, because respondents differed substantially in their characteristics and response rates across methods, it was necessary to adjust for these differences using regression methods, as was done in previous studies. Despite attempts to thoroughly adjust for all measured differences in respondent characteristics, these regression-based estimates of the effects of data collection methods may be biased (see Ware et al. [2001] for additional information).

Given these findings, the possible effects of administrative mode should be considered in all studies involving the Short Form surveys. For example, the National Committee for Quality Assurance (NCQA) subtracted 1.9 T-score points from the PCS scores and 4.5 T-score
points from the MCS scores derived from Medicare Health Outcomes Survey results obtained from SF-36 telephone interview surveys before combining those survey results with self-administered survey results (NCQA, 2004). For studies of the elderly being treated under Medicare, one should consider the recommendations published by NCQA for correcting PCS and MCS T scores obtained from telephone administration of the SF-36. General population findings, which included the Medicare population (McHorney, Kosinski, & Ware, 1994), support these recommendations. Further studies are needed to determine whether different adjustments are warranted for PCS or MCS T scores and to determine if adjustments are warranted for nonelderly, general population scores as well.

In light of the findings of these studies and those from studies summarized in Table 4.2, one should be aware that the method by which Short Form survey data are collected may impact the obtained results. Ideally, data collection should always be limited to one method if the data are to be aggregated or when an individual’s results are to be compared to results obtained from later administration(s) or results from another individual or group of individuals. When data collection methods do vary within a sample or when results are compared across samples assessed using different methods, the effect of the methods used should be evaluated using some of the methods employed in the studies reported in Table 4.2, and the results should be interpreted with due caution.
References


References


Lungenhausen, M., Lange, S., Maier, C., Schaub, C., Trampisch, H. J., & Endres, H. G. (2007). Randomised controlled comparison of the Health Survey Short Form (SF-12) and the Graded Chronic Pain Scale (GCPS) in telephone interviews versus self-administered questionnaires. Are the results equivalent? BMC Medical Research Methodology, 7, 50.


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, 551–567.


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, 247–263.


Ware, J. E., Jr. (2007). *SF-36 Health Survey*. In M. E. Maruish (Ed.), *User’s manual for the SF-36 Health Survey (2nd ed.)*. Lincoln, RI: QualityMetric Incorporated.


Ware, J. E., Jr. (1988). *How to score the revised MOS Short-Form health scales*. Boston: Institute for the Improvement of Medical Care and Health, New England Medical Center.


References


Appendix
Checklists for Development of Certified Modes of Administration
Appendix A.1
Checklist for the Development of a Certified Mode of Administration: Paper Form

General Requirements

___ Provide a means for identifying the survey respondent.
___ Maintain the exact wording of survey instructions and items.
___ Maintain the order of Short Form item presentation.
___ Administer Short Form survey items together.
___ Administer generic Short Form survey content first.
___ Exercise caution when administering only select items from the Short Form surveys (see Chapter 3).
___ Display the appropriate copyright notice provided by QualityMetric as part of any Short Form administration.

Paper Form Mode-Specific Requirements

___ Simplicity in design (e.g., no grid lines, selective use of emphasis, no unnecessary repetition of instructions)
___ Consistency in design (e.g., maintain graphical structure)
___ Organization in design (e.g., completion instructions placed close to items)
___ Naturalness in design (e.g., format items to create natural flow of reading)
___ Clarity in design (e.g., use of graphical guides, use of 8- to 12-point serif font, number each page)
___ Attractiveness in design (e.g., asymmetric composition, include cover page)
A Guide to the Development of Certified Modes of Short Form Survey Administration
Appendix A.2
Checklist for the Development of a Certified Mode of Administration: Mail-Out Form

General Requirements

____ Provide a means for identifying the survey respondent.
____ Maintain the exact wording of survey instructions and items.
____ Maintain the order of Short Form item presentation.
____ Administer Short Form survey items together.
____ Administer generic Short Form survey content first.
____ Exercise caution when administering only select items from the Short Form surveys (see Chapter 3).
____ Display the appropriate copyright notice provided by QualityMetric as part of any Short Form administration.

Mail-Out Form Mode-Specific Requirements

____ Simplicity in design (e.g., no grid lines, selective use of emphasis, no unnecessary repetition of instructions)
____ Consistency in design (e.g., maintain graphical structure)
____ Organization in design (e.g., completion instructions placed close to items)
____ Naturalness in design (e.g., format items to create natural flow of reading)
____ Clarity in design (e.g., use of graphical guides, use of 8- to 12-point serif font, number each page)
____ Attractiveness in design (e.g., asymmetric composition, include cover page)
____ Examiner contact information and mail-back address provided on form or in cover letter
Appendix A.3
Checklist for the Development of a Certified Mode of Administration: Internet, Desktop, or Handheld Device

General Requirements

____ Provide a means for identifying the survey respondent.
____ Maintain the exact wording of survey instructions and items.
____ Maintain the order of Short Form item presentation.
____ Administer Short Form survey items together.
____ Administer generic Short Form survey content first.
____ Exercise caution when administering only select items from the Short Form surveys (see Chapter 3).
____ Display the appropriate copyright notice provided by QualityMetric as part of any Short Form administration.

Internet, Desktop, or Handheld Device

Mode-Specific Requirements

____ Item presentation consistent with paper-form item presentation
____ Sans-serif fonts preferred
____ All instructions necessary for survey completion displayed
____ Entry of identifier (e.g., name, ID number, birth date) required
____ Software allows respondent to perform same survey completion activities as paper form administration (e.g., change an item response, skip a survey item, have instructions repeated)
____ Display appropriate error messages
____ Avoid presentation of unnecessary information and other “clutter”
____ Quality assurance testing performed on all platforms that might be utilized by the licensee
Appendix A.4
Checklist for the Development of a Certified Mode of Administration: Interactive Voice Response Systems

General Requirements

___ Provide a means for identifying the survey respondent.
___ Maintain the exact wording of survey instructions and items.
___ Maintain the order of Short Form item presentation.
___ Administer Short Form survey items together.
___ Administer generic Short Form survey content first.
___ Exercise caution when administering only select items from the Short Form surveys (see Chapter 3).
___ Display the appropriate copyright notice provided by QualityMetric as part of any Short Form administration.

IVRS Mode-Specific Requirements

___ Approved, standardized IVRS interview script employed
___ All instructions necessary for survey completion presented
___ Audio file allows respondent to perform same survey completion activities as paper form administration (e.g., change an item response, skip a survey item, have instructions repeated)
___ Error messages recorded and appropriately cued