

Quick Start **Guide**

SF-12v2[®] Health Survey



QualityMetric

We measure health.



INTRODUCTION

The SF-12v2® Health Survey (SF-12v2; Ware, Kosinski, Turner-Bowker, & Gandek, 2002; Ware et al., 2010) was developed from the 36-item SF-36v2® Health Survey (SF-36v2; Ware et al., 2007). The twelve SF-12v2 items are used to measure eight domains of health-related quality of life (HRQOL). The information obtained from the eight health domain scales is then aggregated to provide summary measures of the respondent's physical and mental health.

The purpose of this guide is to provide users with the basic information necessary to familiarize themselves with the SF-12v2 and to use the instrument properly. It is not intended to be a comprehensive report of existing information about the survey. Users of the SF-12v2 are encouraged to seek out additional information from the sources referred to throughout this guide, as well as from other sources.



DEVELOPMENT

The SF-12v2 was developed from the SF-36v2, the revised version of the SF-36® Health Survey (SF-36; Ware, Snow, Kosinski, & Gandek, 1993). Constructed to satisfy the minimum psychometric standards necessary for group comparisons, the SF-36 was first made available in “developmental” form in 1988 (Ware, 1988) and then in the “original” form in 1990 (Ware, 1990). The eight health domains measured by the SF-36 were selected from the 40 domains that were included in the Medical Outcomes Study (MOS; Stewart & Ware, 1992). Those chosen for the SF-36 represent the health domains most frequently measured by other widely used health surveys and those believed to be most affected by disease and health conditions (Ware, 1995; Ware et al., 1993). The SF-36 items represent multiple operational indicators of health, including behavioral function and dysfunction, distress and well-being, objective reports and subjective ratings, and both favorable and unfavorable self-evaluations of general health status (Ware et al., 1993).

To meet the need for an even briefer generic health survey, a subset of 12 items from the SF-36 were selected to yield Physical Component Summary (PCS) and Mental Component Summary (MCS) scores that were comparable to those from the parent survey. This abbreviated instrument, the SF-12® Health Survey (SF-12; Ware, Kosinski, & Keller, 1995), was later revised (resulting in the SF-12v2) during the development of the SF-36v2. In addition to other improvements, the SF-12v2 provides the ability to score the eight health domain scales.

HEALTH DOMAIN SCALES

The SF-12v2 includes one favorably scored scale for each of the eight health domains measured by the SF-36v2. The items that constitute each scale are taken directly from the SF-36v2.



PHYSICAL FUNCTIONING (PF)

The content of the two-item PF scale represents the levels of health-related limitations the respondent experiences in climbing stairs and performing moderate activities (e.g., vacuuming, golfing).

ROLE-PHYSICAL (RP)

The two-item RP scale covers role limitations related to physical health, including (a) limitations in the kind of work or other regular daily activities and (b) accomplishing less than the respondent would have liked.

BODILY PAIN (BP)

The BP scale consists of one item measuring the extent of pain interference on normal work activities.

GENERAL HEALTH (GH)

The one-item GH scale asks the respondent to rate his or her general health. Vitality (VT). This one-item measure asks the respondent to rate his or her vitality (i.e., having a lot of energy).

SOCIAL FUNCTIONING (SF)

The one-item SF scale assesses the frequency at which physical health or emotional problems interfere with normal social activities.

ROLE-EMOTIONAL (RE)

The two-item RE scale assesses how often emotional problems result in role limitations related to the amount of work or regular daily activities accomplished and the care with which they are performed.

MENTAL HEALTH (MH)

The two-item MH scale measures the frequency of the respondent's feeling: (a) calm and peaceful and (b) downhearted and depressed.

COMPONENT SUMMARY MEASURES

The PCS and MCS measures were developed to achieve a number of advantages in addition to reducing the eight-scale profile down to two component summary measures

without substantial loss of information. Theoretical advantages include a very large increase in the number of levels defined, smaller confidence intervals relative to each of the eight health domain scales, and the elimination of both floor and ceiling effects. A practical advantage is the reduction of the number of statistical comparisons required in an outcome study or clinical trial.

PHYSICAL COMPONENT SUMMARY (PCS)

This measure comprises the scores for all eight SF-12v2 health domain scales, yielding a single score that can be used as an overall measure of physical health.

MENTAL COMPONENT SUMMARY (MCS)

Also derived from the scores on all eight scales, this measure yields a single score that serves as an overall assessment of mental health.

AVAILABLE LANGUAGES

In addition to the original English-language version developed for use in the United States, the SF-12v2 has been translated or adapted into more than 97 languages.

Information about the translations and adaptations can be obtained at <http://www.qualitymetric.com> or by calling (800) 572-9394

PSYCHOMETRIC OVERVIEW

RELIABILITY

The SF-12v2 User's Manual (Ware et al., 2010) reports internal consistency reliability estimates of .91 for the PCS measure and .87 for the MCS measure. Health domain scale reliability estimates, obtained by correlating each SF-12v2 scale with the theta score for its corresponding item bank (a measure of alternate forms reliability), range from .64 to .86.



The 2-week test-retest reliability of the SF-12 PCS measure was .89 in the U.S. and .86 in the U.K. (Ware, Kosinski & Keller, 1995). Similarly, test-retest reliability coefficients of .76 and .77 were observed for the SF-12 MCS measure in the U.S. and U.K. samples, respectively. Because the SF-12v2 is considered an improved version of the SF-12, these estimates can be viewed as representing the lower-limit reliabilities that can be expected for the SF-12v2 measures.

VALIDITY

Numerous studies attesting to the validity of the SF-12v2 health domain scales and component summary measures are presented in both editions of the survey's User's Manual (Ware et al., 2002; Ware et al., 2010). Included in these manuals is evidence of construct validity, as demonstrated in findings from factor analyses, tests of convergent and discriminant validity, and known-groups comparisons; criterion validity, as revealed in correlations with other similar measures (concurrent validity), relationships with future events such as hospitalization (predictive validity), and inclusion in randomized controlled trials; and content validity, as supported by the inclusion of content representing the health domains most frequently measured in widely used health surveys and those believed to be most affected by disease and health conditions.

USES

EVALUATING AND MONITORING INDIVIDUAL PATIENTS IN CLINICAL PRACTICE

The SF-12v2 health domain scales and component summary measures 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. When administered at the beginning of an episode of care, the SF-12v2 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 then be compared to results obtained from one or more readministrations of the survey during the course of treatment, thus providing objective means of monitoring and later documenting the outcomes of the treatment. It is important to note that there are limitations to the use of the eight health domain scores with individual respondents. These limitations are addressed in a later section titled "Norm-Based Interpretation."

MONITORING POPULATIONS

Health plan administrators, employers, and researchers are continually challenged to find efficient and comprehensive ways of measuring the health of various populations. To this end, the SF-12v2's brevity lends itself to comprehensive population monitoring. Evidence of the effectiveness of the SF-12v2, as well as other members of the "Short Form family" of health surveys, in monitoring functioning and well-being, assessing disease burden, and comparing the health of different populations and patient groups can be found among the more than 14,000 SF-related publications (as of April 2010).

ESTIMATING THE BURDEN OF DISEASE

The SF-12v2 scales and measures are proving to be increasingly accepted, on the strength of validation studies to date, as valid health measures for the purpose of documenting disease burden. Estimates of the disease burden for 18 disease and impairment groups are available, calculated from the 1998 normative survey data. In addition, estimates of the unique effects of many chronic conditions on the SF-12v2 scales and measures are presented in tables found in the SF-12v2 User's Manual (Ware et al., 2010).



EVALUATING TREATMENT EFFECTS

Medical researchers conducting clinical trials increasingly recognize the need to define benefits more broadly than traditional clinical endpoints by including patient-reported outcomes (PRO) measures in clinical trials and other investigations. Thus, pharmaceutical companies frequently require the use of well-validated, documented, and accepted PRO measures that can capture differences in outcomes between alternative drugs, drugs versus placebos, and drug dosages over

relatively short periods of time. The SF-12v2, SF-36v2, and SF-36 are becoming widely recognized as leading PRO measures in clinical trials. Taken together, all of the Short Form surveys have been cited in a total of 1,899 published articles (as of March 2009) reporting randomized controlled trial results.

MANAGING DISEASE

Increasingly, disease management providers are incorporating PRO surveys into their measurement systems. Data from such surveys add significant value because they can improve risk prediction, enhance service planning and outcomes monitoring efforts, and ensure that program planning and evaluation efforts incorporate the patient's perspective. When used with one or more disease-specific measures, the SF-12v2 provides information that can help evaluate patients with common chronic conditions (e.g., asthma, congestive heart failure), as well as monitor and compare their outcomes over time.

PREDICTING RISK AND MAXIMIZING COST-EFFECTIVENESS

With increasing frequency, health plan administrators are recognizing that when generic HRQOL data from patients' self-assessments of physical and mental health are added to predictive models, their predictive power substantially improves, thus yielding information that helps providers better anticipate and manage health problems. The SF-12v2 can be used at baseline in risk stratification and, when repeated over time, for health outcomes monitoring. Moreover, because it measures a wide range of both physical and mental health

concepts, it can be used for risk prediction with most populations.

ADMINISTRATION

The SF-12v2 is simple to administer. Several general guidelines are provided here, including a summary of the most important "dos" and "don'ts" that the user should keep in mind when administering the survey.

ELIGIBILITY

The SF-12v2 is designed to be self-administered by individuals aged 18 years or older. Before giving a respondent a survey form, the examiner should determine the respondent's general reading ability. According to Microsoft® Word's readability determination, the SF-12v2 has a Flesch-Kincaid Grade Level score of 6.8 (Millhollon & Murray, 2001).

TIMING ADMINISTRATIONS

In clinical settings, the SF-12v2 should be administered before the respondent sees a healthcare provider so that the interaction does not influence the respondent's answers to the survey. If possible, the survey should also be administered before the respondent is asked other health questions or about concurrent illnesses, again so that any such discussion does not influence the respondent's answers. Because of the 4-week recall period that respondents must consider when answering many of the items on the SF-12v2's standard form, at least 4 weeks should elapse between administrations. The SF-12v2's acute form, which utilizes a 1-week recall period, is available for use when the longer interval is not possible or desired.

ADDRESSING PROBLEMS AND QUESTIONS

Respondents commonly ask questions or may display certain behaviors before, during, or after administration of health status assessments such as the SF-12v2. Table 1 identifies several DOs and DON'Ts based on common questions, behaviors, or circumstances that may occur during survey administration and offers suggestions for addressing them.

SCORING

Computer-based scoring services for the SF-12v2 are available through QualityMetric Incorporated or its licensed certified vendors. Scoring of the health domain scales involves: (a) recoding item response values, (b) summing recoded response values for all items in a given scale to obtain the scale raw score, (c) transforming the scale raw score to a 0-100 score, (d) transforming the 0-100 score to a z score, and (e) transforming the scale z score to a T score (mean = 50, standard deviation = 10). The PCS score is computed by: (a) multiplying each health domain z score by a scale-specific physical factor score coefficient, (b) summing the resulting products, and (c) converting the product total to a T score. The MCS score is computed in the same manner, instead using scale-specific mental factor score coefficients. The SF-6D Utility Index and a 6-month medical expenditure prediction measure can also be scored from SF-12v2 data, using algorithms different from those used to score the health domain scales and component summary measures.

INTERPRETATION

The SF-12v2 User's Manual (Ware et al., 2010) provides the information and guidelines needed to interpret SF-12v2 T scores using any or all of three general approaches: norm-based, content based, or criterion-based interpretation. The following sections offer brief summaries of each approach.

NORM-BASED INTERPRETATION

The SF-12v2 user can compare a respondent's obtained health domain scale and component summary measure T scores (or a group's mean T scores) to a variety of normative data, including general population norms, disease-specific norms, and norms specific to the respondent's age and gender. Some general guidelines concerning the use of U.S. general population norms are provided here.

In general, for individual respondents T scores of 45 or greater indicate at least average overall functioning or well-being in the health domain or overall health dimension – physical or mental – being assessed, as compared to the U.S. general population. For group-level data, T scores of 47

SF-36v2 ADMINISTRATION DOS & DON'TS

DOS

- ▶ DO introduce the SF-12v2 and explain the reasons for completing it and the importance and advantages for the respondent of doing so.
- ▶ DO have the respondents complete the items before they fill out any other health data forms and before they see their healthcare provider.
- ▶ DO be warm, friendly, and helpful
- ▶ DO request and encourage respondents to respond to all of the survey items.
- ▶ DO read and repeat a question and its response choices verbatim for respondents if they ask for clarification.
- ▶ DO tell respondents to answer items based on what they think each item means.
- ▶ DO have respondents complete the survey by themselves.
- ▶ DO inform respondents if they will be asked to fill out the same survey again.
- ▶ DO thank respondents for completing the survey

DON'TS

- ▶ DO NOT minimize the importance of the SF-36v2
- ▶ DO NOT discuss respondents' health, health data, or emotions with them before they complete the survey items.
- ▶ DO NOT force or command respondents to complete the survey.
- ▶ DO NOT accept incomplete SF-36v2 forms without first encouraging respondents to respond to any unanswered items.
- ▶ DO NOT change the wording of questions or response choices.
- ▶ DO NOT interpret or explain items.
- ▶ DO NOT allow spouses, family members, or friends to help respondents complete the items. Ideally, caregivers should not be present during this assessment.

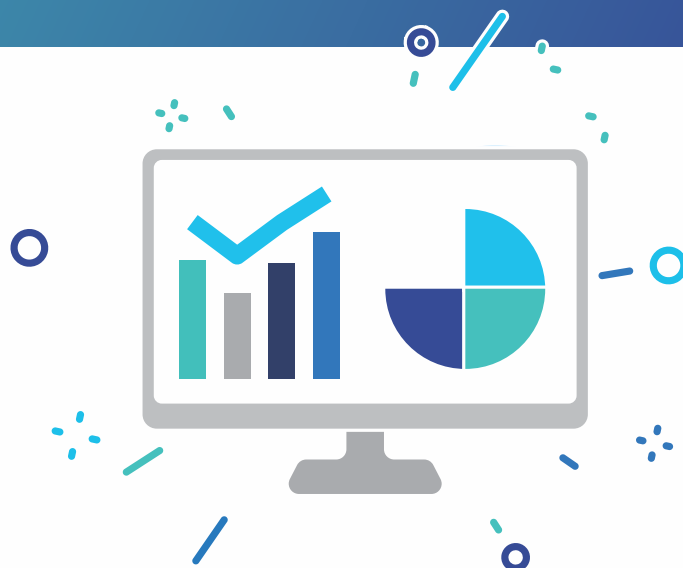


or greater should be considered at least average in relation to the U.S. general population. Individual respondent T scores that are less than 45 and group mean scores that are less than 47 indicate the presence of impaired functioning or well-being in the associated dimension.

When considering individual patient results, one can generally be confident that a patient whose health domain scale or component summary measure score falls more than 1 standard deviation (10 T-score points) below the population mean (50) is exhibiting significantly impaired functioning. Thus, scores less than 40 indicate significant impairment in that health domain or dimension.

Scores in the 40-to-44 range are below average and require further investigation, including consideration of the confidence interval around the score and the choice of age-, gender-, and/or disease-based norms, to determine the extent of limitations in functioning or well-being, if any. In general, the SF-12v2 PCS and MCS results for an individual respondent can be interpreted with confidence. The score for each of these measures is determined by the responses to all of the survey's 12 items, thus providing adequate reliability for interpretation at the patient level. However, unlike when SF-12v2 data from a large group of individuals are aggregated, health domain scores for individual respondents are not, in most cases, reliable enough to be used on their own. Exceptions include instances in which very large score differences are expected. For example, joint replacement surgery should lead to significant improvement (up to 15 T-score points) in the SF-12v2 PF score within a matter of weeks. That being said, although somewhat limited in their utility when interpreting results for an individual respondent, SF-12v2 health domain scores can nevertheless be used to flesh out a clearer picture of the respondent's health status than can be determined by PCS and MCS scores alone by reflecting the specific limitations that are contributing to each of the summary scores.

very large score differences are expected. For example, joint replacement surgery should lead to significant improvement (up to 15 T-score points) in the SF-12v2 PF score within a matter of weeks. That being said, although somewhat limited in their utility when interpreting results for an individual respondent, SF-12v2 health domain scores can nevertheless be used to flesh out a clearer picture of the respondent's health status than can be determined by PCS and MCS scores alone by reflecting the specific limitations that are contributing to each of the summary scores.



CONTENT-BASED INTERPRETATION

An interpretive approach based on analyses of the content of individual items is another strategy that can be used to understand the meaning of and interpret differences in component summary measure scores that fall between the extreme (highest and lowest) scores. This content-based approach involves examining the frequency distribution of specific responses to SF-12v2 items, across score levels of the component summary measures. For example, it is useful to know that more than 98% of the U.S. general population who score below 30 on the PCS measure are limited in climbing several flights of stairs.

In addition, more specific interpretations of scores obtained on the summary measures can be achieved by examining the response to each of the 12 SF-12v2 items. Item response analyses can be particularly useful when a score falls close to a cut-point score established by the user to make clinical decisions, such as the type of treatment (if any) required.

CRITERION-BASED INTERPRETATION

Criterion-based interpretation focuses on analyses of the relationships between the measure(s) in question and other variables, referred to as criteria, measured either concurrently or after a period of time. In terms of the SF-12v2, the established empirical strategy for evaluating the meaningfulness of scores is to link PCS and MCS T scores to important clinical benchmarks, such as the ability to work or healthcare utilization, and to show how differences in scores of a certain magnitude can help to predict important clinical and social events. 865 or by phone at (800) 572-9394.



RECOMMENDED SOURCES OF ADDITIONAL INFORMATION

Users of the SF-12v2 can refer to several sources for additional information about the development and use of the survey. Both editions of the SF-12v2 User's Manual (Ware et al., 2002; Ware et al., 2010) provide detailed descriptions of the survey's development and report results from initial and subsequently published psychometric studies. In addition, use of a search engine focused on clinical literature will identify published studies involving the SF-12v2 item that may be relevant to the intended use of the scores.

LICENSING AND CONTACT INFORMATION

QualityMetric Incorporated licenses the use of the SF-12v2 for both clinical and research purposes. Organizations and individuals wishing to use or reproduce any QualityMetric survey and/or any associated intellectual property for any purpose must obtain a license from QualityMetric.

For more information, go to qualitymetric.com, or contact QualityMetric in writing at 24 Albion Road, Building 400, Lincoln, RI 02865 or by phone at **(800) 572-9394**.



SELECTED BIBLIOGRAPHY

- Millhollon, M., & Murray, K. (2001). *Microsoft Word version 2002 inside out*. Redmond, WA: Microsoft Press
- Steward, A.L., & Ware, J.E., Jr. (Eds.). (1992). *Measuring functioning and well-being: The Medical Outcomes Study approach*. Durham, NC: Duke University Press.
- Ware, J.E., Jr. (1998). *How to score the revised MOS Short-Form health scales*. Boston, MA: Institute for the improvement of Medical Care and Health, New England and Medical Center.
- Ware, J.E., Jr. (1990). Measuring patient function and well-being: Some lessons from the Medical Outcomes Study. In K.A. Heitgoff & K.N. Lohr (Eds.), *Effectiveness and outcome in health care: Proceedings of an invitational conference by the Institute of Medicine Division of Health Care Services* (pp. 107-119). Washington, DC: National Academy Press
- Ware, J.E., Jr. (1995) The status of health assessment 1994. *Annual Review of Public Health*, 16, 327 - 354
- Ware, J.E., Jr., Kosinski, M., Bjorner, J.B., Turner-Bowker, D.M., Gandek, B., & Maruish, M.E., (2007) *User's manual for the SF-36v2 Health Survey* (2nd ed.). Lincoln, RI: QualityMetric, Incorporated.
- Ware, J.E., Jr., Kosinski, M., & Keller, S.D. (1995) *SF-12 How to score the SF-12 physical and mental health summary scales* (2nd ed.). Boston, MA: The Health Institute
- Ware, J.E., Jr., Kosinski, M., Turner-Bowker, D.M., & Gandek, B. (2002). *How to score Version 2 of the SF-12 Health Survey (with a supplement documenting Version 1)*. Lincoln, RI: QualityMetric, Incorporated.
- Ware, J.E., Jr., Kosinski, M., Gandek, B., Sundaram, M., Bjorner, J.B., Turner-Bowker, D.M., & Maruish, M.E. (2010) *Users manual for the SF-12v2 Health Survey* (2nd ed.). Lincoln, RI: QualityMetric, Incorporated.
- Ware, J.E., Jr., Snow, K.K., Kosinski, M., & Gandek, B. (1993). *SF-36 Health Survey manual and interpretation guide*. Boston, MA: The Health Institute

