Development and validation of a mental health subscale from the Quality of Well-Being Self-Administered

Andrew J. Sarkin,
Health Services Research Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, CA, USA

Erik J. Groessl,
Health Services Research Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Jordan A. Carlson,
Health Services Research Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Steven R. Tally,
Health Services Research Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Robert M. Kaplan,
Department of Rehabilitation Medicine, National Institutes of Health, NIH Clinical Center, Bethesda, MD, USA

William J. Sieber, and
Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, CA, USA

Theodore G. Ganiats
Health Services Research Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, CA, USA

Andrew J. Sarkin: asarkin@ucsd.edu

Abstract

Purpose—The purpose of this study was to create and validate a mental health subscale for the Quality of Well-Being Self-Administered (QWB-SA).

Methods—The QWB-SA and other measures such as the Profile of Mood States (POMS), Medical Outcomes Study 36 Item Short Form (SF-36), EuroQOL 5D (EQ-5D), and Health Utilities Index Mark 2 (HUI) were administered to three samples: a general population (N = 3,844), a non-psychiatric medical population (N = 535), and a psychiatric population (N = 915).

Correspondence to: Andrew J. Sarkin, asarkin@ucsd.edu.

Independent expert ratings of which items represented the construct of mental health were used along with psychometric methods to develop and validate a 10-item QWB-SA mental health scale.

Results—The mental health scale demonstrated high internal consistency (Cronbach's alpha = 0.827–0.842) and strong correlations with other measures of mental health, such as the POMS ($r = -0.77$), mental health scale from the SF-36 ($r = 0.72$), EQ-5D mood item ($r = 0.61$), and HUI Emotion Scale ($r = 0.59$). It was not highly correlated with measures of physical health. Among the psychiatric population, the new mental health scale was moderately correlated with indicators of psychiatric problem severity.

Conclusions—It is now possible to report outcomes and relationships with mental health in studies that use the QWB-SA. This new mental health subscale can also be used with the large volume of previously collected data using the QWB-SA to examine the impact of illnesses and interventions on mental health-related quality of life.

Background

Generic health-related quality of life (HRQoL) measurement is useful for determining the broad impact of disease and illness on functioning and well-being [1] and for program evaluation activities such as determining outcomes and cost-effectiveness of interventions. Measures of HRQoL are usually considered to be either psychometrically based or preference-/utility-based [2]. Psychometrically based measures have subscales describing various dimensions of HRQoL. Preference-based measures typically summarize overall HRQoL in a single quantitative score ranging from 0 to 1.0, allowing for the calculation of quality-adjusted life years (QALYs), and facilitating cost-effectiveness analysis. Preference-based measures of HRQoL rarely have descriptive subscale scores. Yet, many users of preference-based scores express interest in using descriptive subscale scores from preference-based measures, especially for mental health.

Mental health has long been considered an important component of overall health [3] and is an especially important construct to include in HRQoL measurement because it affects longevity, quality of life, and physical functioning [4]. Indeed, many studies have underscored the importance of incorporating measures of mental health in health outcomes measurement [5, 6]. Mental disorders are typically associated with impairment in HRQoL comparable to other common health conditions, and interventions directed at improving HRQoL need to focus on recognition and treatment for mental disorders [7]. This paper focuses on a generic utility-based measure of health-related quality of life known as the Quality of Well-Being Self-Administered (QWB-SA). The QWB-SA has significant mental health content; however, unlike some other HRQoL measures, the QWB-SA previously did not offer a separate mental health subscale score.

There has been an increased interest among users of the QWB-SA and other HRQOL instruments in measuring HRQoL in patients with conditions such as depression [8–10], schizophrenia [11], psychosis [12], and other mental disorders [7, 13]. Also, there are a

growing number of integrated care programs combining mental and physical healthcare in response to evidence that people with mental illness have greatly increased morbidity and significantly shorter longevity than the general population [14].

Though it is important to include both physical and mental health constructs in generic HRQoL measurement, a mental health subscale would be a useful clinical and research tool [15]. For example, a mental health measure might help providers determine the impact of a therapy targeting mental health, or in the assessment of integrated care programs where a mental health component is being added or changed. A mental health score could also assess how a person's mental health is affected by disease, or how a person's mental health changes over time as a result of disease or intervention. Although it is usually desirable to include a measure of mental health, there are often concerns about response burden when respondents are assessed with too many different instruments. An attractive alternative is to derive scores for a mental health subscale from a single validated and widely used questionnaire, such as the QWB-SA. The purpose of this study was to derive and validate a mental health subscale from existing QWB-SA items and compare the subscale to other generic HRQoL and mental health measures.

Methods

Overview of methods

A QWB-SA mental health scale was developed using expert opinion, then validated psychometrically and compared to other generic HRQoL and mental health measures including the Short Form 36 (SF-36) mental health component, the SF-36 mental health scale, the SF-36 physical component, the SF-36 physical functioning scale, the EuroQoL 5D (EQ-5D) anxiety item, the Health Utilities Index Mark 2 (HUI2) emotion scale, the Profile of Mood States (POMS) depression-dejection scale, the POMS total mood disturbance, the Illness Management and Recovery scales (IMR), the Recovery Markers Questionnaire (RMQ), and the Mental Health Statistics Improvement Program (MHSIP).

The three studies reported here used two separate datasets from a project conducted by the Health Measurement Research Group and a third dataset from a large mental health system. Study 1 used data from the National Health Measurement Study (NHMS). The NHMS was designed to derive norms and compare indexes for six generic HRQoL measures [16] and provides data that can be generalized to the US population. Study 2 examined data from the Health Measurement in Patients Study (HMPS) that evaluated the same set of HRQoL measures among two clinical populations [17] and provided data to examine the psychometrics of the measure in clinical populations. Study 3 examined data collected from clients in a large county mental health treatment system to demonstrate known-groups validity and validate the measure for tracking outcomes in this and other mental health populations. Because there were different measures used and analyses varied slightly, the results from the three studies are presented separately. Details regarding samples and methods are presented in their respective sections.

Identifying QWB-SA items related to mental health—The QWB-SA was circulated to four experts in the field of HRQoL measurement and the QWB-SA for identification of

items related to the construct of mental health. Two of the raters were trained as clinical psychologists, one as a developmental psychologist, and one was a family physician who is an author of the QWB-SA. The expert raters worked independently of one another to rate each QWB-SA item as reflective of either (a) physical health, (b) mental health, or (c) undecided. Overall agreement among raters was assessed with one-way random effects intraclass correlation coefficients (ICC). The criteria for interpreting the strength of the ICC values were as follows: 0.60 good to excellent, 0.41–0.60 moderate, and   0.40 poor [18]. Overall agreement between raters was high (ICC = 0.77) across all QWB-SA items, and ten items were selected as measuring mental health based on a cutoff criterion of at least 50 % of raters endorsing the item as being a part of the construct of mental health. The 10 identified items were selected for the mental health scale and were further analyzed for their psychometric properties and validity as a mental health subscale.

Scores for the 10-item QWB-SA mental health scale ranged from 0 to 10 and represented the average number of these selected mental health symptoms experienced per day over the 3-day period reported in the QWB-SA. Traditional scoring of the QWB-SA involves preference weights, but they could not be used here because they are calibrated for the entire QWB-SA, and it is not possible to know whether social or functional limitations are due to mental health symptoms or other health issues.

Study 1

Participants

Study 1 used data from the NHMS, designed to derive norms and compare indices for six generic HRQoL measures [16]. Trained interviewers at the University of Wisconsin collected data from June 2005 through August 2006 on five HRQoL measures. A random digit dialed (RDD) telephonic interview methodology was utilized to obtain a sample of adults aged 35–89 years. The age range was designed to be representative of the older half of the general US population in 2005–2006. Older individuals (aged 65–89) and African Americans were oversampled to ensure adequate representation in the sample. When street addresses were available, advance letters were sent explaining the purpose of the study. The letters included a $2 pre-incentive intended to increase willingness to participate. Interview length averaged 40 min. Upon the completion of the telephone survey, participants were sent $25 compensation. Response rates were estimated to be between 45.6 and 56.3 %, and a total of 3,844 participants completed the study. Demographic characteristics for the final sample are presented in Table 1. For additional information on the data procedures for this dataset, please see Fryback et al., [16].

Measures

Quality of Well-Being Self-Administered (QWB-SA)—The QWB-SA is an HRQoL instrument designed for use in health policy analysis [4, 19–21]. It includes 73 items assessing symptoms and functioning over the past 3 days. Items are measured on a response scale where participants mark which days they had the symptom (yesterday, 2 days ago and/or 3 days ago) or mark "no days" if they did not have the symptom during the three-day time period. For the purpose of this study, we created a count of the number of days out of

the possible three that the symptom was present, and the scale ranged from 0 = no days, to 1 = one day, 2 = two days, or 3 = three days. This count was created for each of the 73 symptoms in the QWB-SA. These resulting count variables were treated as continuous variables in analyses, given that they represent frequency counts.

The reason for using the three-day period has a long empirical base. Early studies on the QWB systematically evaluated accuracy of recall periods for the reporting of symptoms. Using diary methods, it was noted that people do reasonably well in recollecting symptoms and functioning up to 1 week. Intervals longer than 1 week are associated with much greater error in the reporting of symptoms. Using reliability assessment methods, early QWB investigators recognized that there was substantial day-to-day variability in symptoms. A 1-day interval was much less reliable than a 1-week interval. However, a 3-day interval captured most of the information associated with a 1-week interval and the 3-day period was chosen as optimal [20, 21]. Symptom reporting on three separate days provides more reliable information than asking individuals to report on the block of 3 days together [22].

Short form 36 (SF-36)—The SF-36 is the most widely used measure in clinical trials assessing health-related quality of life. The SF-36 was designed to measure health status in the Medical Outcomes Study, for use in clinical practice and research, health policy evaluations, and general population surveys [23]. The SF-36 assesses function and symptoms over several time periods (i.e., "in a typical day" or "during the past 4 weeks"). This study analyzed data from four scales of the SF-36: the mental health component (composed of four subscales), mental health subscale, physical component (composed of four subscales), and physical functioning subscale. The SF-36 mental health component creates a score from 14 items across 4 scales: vitality, social functioning, role-emotion, and mental health. The SF-36 physical component creates a score from 21 items across 4 scales: physical functioning, role-physical, bodily pain, and general health. The mental health scale consists of 5 items, and the physical functioning scale consists of 10 items. The physical functioning scale has been shown to be a good measure of physical health, and the mental health scale has been shown to be a good measure of mental health [6, 24–26]. Brazier has developed a methodology that allows the conversion of SF-36 scores into a preference-based score [27]. Each scale and component uses a normative-based scoring [27] with a scale range between 0 and 100, where higher scores indicate higher levels of well-being.

EuroQol 5D (EQ-5D)—The EQ-5D is a HRQoL instrument designed for use in the clinical and economic evaluation of health care and in population health surveys [28]. The measure consists of five items asking about "your health today" each covering a different domain (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). The items are rated on a 3-point response scale ranging from 0 = no problems to 1 = some problems, or 2 = severe problems [29]. The current study used the anxiety/depression domain which consists of a single item for rating. Information of the EQ-5D is available at the following website: http://www.euroqol.org/.

Health Utilities Index Mark 2 (HUI)—The HUI is an HRQoL instrument designed to characterize health status and is widely used in clinical studies, population health surveys, and economic evaluations. The HUI consists of 15 items assessing functioning during the

past week. The HUI evaluates six attributes (sensation, mobility, emotion, cognition, self-care, and pain), with scores calculated from an algorithm derived from standard gamble assessments made by the general population [30]. The HUI2 emotion scale was used in this study and consists of five items with scores ranging from 0 to 1, where higher total scores indicate higher level of well-being. More information on the HUI can be found at the following website: http://www.healthutilities.com/.

Procedures

Analysis of psychometric properties and validity—Psychometric properties of the scale were evaluated by calculating the scale reliability coefficient Cronbach's alpha. To investigate concurrent validity, correlational analyses were used to compare each identified QWB-SA item and an identified QWB-SA mental health scale with the following: SF-36 mental health component, SF-36 mental health scale, EQ-5D anxiety scale, HUI2 emotion scale, POMS depression-dejection scale, and POMS total mood disturbance (TMD). To investigate discriminant validity, correlational analyses were used to compare an identified QWB-SA mental health subscale with the SF-36 physical component and SF-36 physical functioning scale. To investigate whether the QWB-SA mental health subscale was significantly more related to measures of mental health than measures of physical health, the relationships between the QWB-SA mental health subscale and other mental health measures were compared with the relationships between the QWB-SA mental health scale and physical health measures using confidence intervals. For Pearson correlation coefficients, 0.10 was considered small, 0.30 was considered moderate, and 0.50 was considered high [31]. The data were analyzed using Statistical Package for the Social Sciences (SPSS), version 16.0 [32].

Results

The 10-item subscale was compared to all possible 9-item subscales containing those items to confirm that no single item fit poorly and found that Cronbach's alpha for the 10-item subscale (alpha = 0.827) was larger than the Cronbach's alphas for any of the 9-item subscales. Pearson correlation coefficients for the 10-item QWB-SA mental health subscale with the mental health comparison measures are presented in Table 2. Most of these correlations were large ($p < 0.001$; absolute $r = 0.608$–0.698). Pearson correlation coefficients for the identified QWB-SA mental health scale with the physical health comparison measures were smaller and ranged from $−0.320$ to $−0.389$ ($p < 0.001$).

As evidence of discriminant validity, the relationship between the QWB-SA mental health subscale and the SF-36 mental health component was significantly greater than the relationship between the QWB-SA mental health subscale and the SF-36 physical component; $\beta = −0.034$, 95 % CI = $−0.035$ to $−0.033$, and $\beta = −0.015$, 95 % CI = $−0.017$ to $−0.014$, respectively. The relationship between the QWB-SA mental health scale and the SF-36 mental health scale was also significantly greater than the relationship between the QWB-SA mental health scale and the SF-36 physical functioning scale; $\beta = −0.020$, 95 % CI = $−0.021$ to $−0.020$, and $\beta = −0.007$, 95 % CI = $−0.008$ to $−0.007$, respectively.

Bivariate correlational analysis for the identified QWB-SA items is presented in Table 3 and indicated that each of the 10 items was significantly associated with each of the mental health comparison measures ($p < 0.001$; absolute $r = 0.250–0.591$). Pearson correlation coefficients for the 10 identified QWB-SA items with each comparison measure were as follows: $r = -0.121$ to -0.271 for the SF-36 physical component, $r = -0.172$ to -0.279 for the SF-36 physical functioning scale, $r = -0.280$ to -0.556 for the SF-36 mental health component, $r = -0.273$ to -0.591 for the SF-36 mental health scale, $r = 0.250–0.507$ for the EQ-5D anxiety/depression scale, and $r = -0.275$ to -0.520 for the HUI2 emotion scale. In particular, six items exhibited moderate to large correlations greater than absolute $r = 0.417$ with all of the mental health comparison measures. These six items were as follows: "spells of feeling nervous or shaky," "spells of feeling upset, downhearted, or blue," "excessive worry or anxiety," "feelings that you had little or no control over events in your life," "feelings of being lonely or isolated," and "feelings of frustration, irritation, or close to losing your temper."

Study 2

Participants

Study 2 participants were from the Health Measurement in Patients Study (HMPS) conducted by the Health Measurement Research Group. Adults over age 35 were recruited either prior to undergoing cataract extraction with lens replacement ($n = 376$) or prior to entering specialized heart failure management programs ($n = 159$). Participants in both study arms were required to have English language reading and writing abilities sufficient to understand the instructions and complete the questionnaires. For the cataract study component, exclusion criteria were traumatic cataract or visual impairment which precluded the ability to complete a large print version of the study materials. Cataract patients were also excluded if they were scheduled to undergo other simultaneous ophthalmological procedures. For the heart failure study component, exclusion criteria were a New York Heart Association Classification of class IV, myocardial infarction within the past 6 months, unstable angina, coronary artery bypass graft within the past 3 months, or being on a heart transplant list.

The majority of participants were recruited from medical centers at the University of California Los Angeles, the University of California San Diego, and the University of Wisconsin. Some additional participants for the cataract component were recruited from the University of Southern California. Recruitment took place at the medical centers during regularly scheduled appointments. Upon explanation of the study and successful consent, participants were given a measurement packet to take home and complete. Instructions directed participants to return the completed questionnaire in the included prepaid envelope within 7 days. Participants were compensated $20 upon the return of the completed questionnaire to the data management center. Although only the baseline (pre-cataract surgery or pre-heart failure program) questionnaire data were used for the current study, the HMPS was longitudinal, with additional follow-up questionnaires sent out at 1 and 6 months following cataract surgery or entry into a heart failure program. For additional information

on the data collection procedures for sample 2, or details on the outcomes of the longitudinal component of the study, please see Kaplan et al., [17].

Measures

The measures were the same as those used in study 1 with the addition of the POMS. The POMS was administered only to participants prior to undergoing cataract extraction with lens replacement at the University of California San Diego clinic ($n = 164$).

Profile of mood states (POMS)—The POMS contains 65 adjectives measuring six domains of affect or mood (tension-anxiety, depression-dejection, anger-hostility, vigor-activity, fatigue-inertia, and confusion-bewilderment [33]). Vigor-activity is the only domain of positive mood; therefore, it is subtracted when calculating total mood disturbance (TMD) scores. Items are rated on a 5-point intensity scale wherein $0 = $ not at all, $1 = $ a little, $2 = $ moderately, $3 = $ quite a bit, and $4 = $ extremely. The depression-dejection scale and TMD scores were used for the purpose of this study. The depression-dejection scale consists of 15 items and can range from 0 to 60. Scores for TMD can range from -32 to 200. For both the depression-dejection scale and TMD, higher scores correspond with more severe symptoms and dysfunction.

Procedures

Analysis of psychometric properties and validity—The statistical analyses were the same as in study 1.

Results

The same 10 identified items were further analyzed for their psychometric properties and validity as a mental health subscale. Again Cronbach's alpha for the 10-item subscale (alpha $= 0.842$) was larger than the Cronbach's alphas for any of the 9-item subscales. Pearson correlation coefficients for the 10-item QWB-SA mental health scale with the mental health comparison measures are presented in Table 2. Most of these correlations were large ($p < 0.001$; absolute $r = 0.591$–0.768). Pearson correlation coefficients for the identified QWB-SA mental health scale with the physical health comparison measures were smaller, -0.251 to -0.354 ($p < 0.001$).

As evidence of discriminant validity, the relationship between the QWB-SA mental health scale and the SF-36 mental health component was significantly greater than the relationship between the QWB-SA mental health scale and the SF-36 physical component; $\beta = -0.033$, 95 % CI $= -0.036$ to -0.029 and $\beta = -0.012$, 95 % CI $= -0.016$ to -0.008, respectively. The relationship between the QWB-SA mental health scale and the SF-36 mental health scale was also significantly greater than the relationship between the QWB-SA mental health scale and the SF-36 physical functioning scale; $\beta = -0.020$, 95 % CI $= -0.022$ to -0.019, and $\beta = -0.007$, 95 % CI $= -0.008$ to -0.005, respectively.

Bivariate correlational analysis for the identified QWB-SA items is presented in Table 3 and indicated that each of the 10 items was significantly associated with each of the mental health comparison measures ($p < 0.001$; absolute $r = 0.177$–0.694). Pearson correlation

coefficients for the 10 identified QWB-SA items with each comparison measure were as follows: $r = -0.107$ to -0.278 for the SF-36 physical component, $r = -0.163$ to -0.308 for the SF-36 physical functioning scale, $r = -0.194$ to -0.631 for the SF-36 mental health component, $r = -0.211$ to -0.669 for the SF-36 mental health scale, $r = 0.177$–0.588 for the EQ-5D anxiety/depression scale, $r = -0.221$ to -0.548 for the HUI2 emotion scale, $r = 0.313$–0.694 for the POMS depression-dejection scale, and $r = 0.321$–0.641 for the POMS TMD. The same six items from study 1 exhibited moderate to large correlations greater than absolute $r = 0.378$ with all of the mental health comparison measures. These six items were as follows: "spells of feeling nervous or shaky," "spells of feeling upset, downhearted, or blue," "excessive worry or anxiety," "feelings that you had little or no control over events in your life," "feelings of being lonely or isolated," and "feelings of frustration, irritation, or close to losing your temper."

Study 3

Participants

Study 3 participants were obtained from a large-scale assessment of mental health clients in San Diego County Adult and Older Adult Mental Health Services. Nine hundred and thirty-six surveys were completed. We obtained diagnosis information from the medical information system database. Only participants with the following primary diagnoses were included: 36.2 % were diagnosed with schizophrenia or schizoaffective disorder, 23.0 % were diagnosed with major depressive disorder, 15.0 % were diagnosed with bipolar disorder, 8.9 % were diagnosed with other depression disorder, 5.8 % were diagnosed with anxiety disorder, and 5.3 % were diagnosed with an unspecified psychotic disorder. Of these 936 participants, 21 were removed because there was no clear documentation of them having a severe mental illness.

Measures

Mental health statistics improvement program consumer survey (MHSIP)—The version of the MHSIP Consumer Survey used in this study consists of 36 item designed to assess the care of persons with mental illness and is widely used in public mental health systems [34]. Seven domains are assessed: general satisfaction, perception of access to services, perception of quality and appropriateness of care, perception of participation in treatment planning, perception of outcomes of services, perception of functioning, and perception of social connectedness. Each item is a declarative statement. Response options ranged on a 5-point scale from strongly agree to strongly disagree, where higher numbers corresponded with greater disagreement, and thus greater dysfunction. As of 2001, 38 states had implemented a version of the MHSIP consumer survey to assess consumer perception of care, with the number of items on the MHSIP versions ranging from 19 to 40 [35]. Reliability of the MHSIP was high in a pilot study (Cronbach's alpha = 0.95; [36]).

Illness Management and Recovery scales (IMR)—The Illness Management and Recovery (IMR) scales were developed by the Dartmouth Psychiatric Research Center to assess the intended outcomes of the Illness Management and Recovery program. Items for the IMR scales were generated by IMR practitioners and consumers with severe mental

illness to address the strategies targeted by the IMR program. The items were selected and reworded per clinician and consumer feedback [37, 38].

The IMR scale developers initially reported analyses of IMR client and clinician scale reliability and validity in *Measuring the Promise: A Compendium of Recovery Measures Volume II* [37]. Since then, there have been at least three additional articles that examined the psychometric properties of the IMR scales [39–41]. A three-factor structure for the IMR scales has been confirmed by two of these studies with some variation in the reported factors. The three factors used for this study were those published by Sklar et al. [41] and reflect the constructs of Symptom Management, Recovery Markers, and Substance Abuse with internal reliabilities of 0.76, 0.83, and 0.69. The scales showed adequate concurrent and criterion validity.

Recovery Markers Questionnaire (RMQ)—The Recovery Markers Questionnaire [42] is a free-standing subscale of the Recovery Enhancing Environment measure. The RMQ consists of 24 items, using five-point Likert-type response options that range from 1 'strongly agree' to 5 'strongly disagree.' Among other things, these recovery markers are intended to represent the client's motivation, health status, symptom control, connection with others, and whether they use their personal strengths, skills, and talents. The 24 items are averaged to create a single score that has high internal and face validity [42].

Procedures

Twice per year (May and November), adults aged 18 and older receiving psychological and behavioral treatment services are asked to complete an anonymous self-report questionnaire at their provider site during a 2-week data collection period. For this study, we used data collected in November 2010, at which time participants completed the QWB-SA mental health items and the MHSIP. Given that IMR and RMQ are assessed at other times, these were only used if they were given within 3 months of the survey that included QWB-SA mental health items.

In January 2010, the Behavioral Health Division of San Diego County's Health and Human Services Agency asked all Outpatient, Case Management, and Full Service Partnership programs to begin using the clinician version of the IMR scale to assess client recovery. The IMR scale was expected to be completed during each client's initial intake assessment and their treatment planning update visits occurring approximately every 6 months after intake. Academic researchers with experience in the evaluation of outcomes conducted trainings with clinical and administrative staff from each of these treatment facilities to instruct them on using the IMR scale. Treatment program staff members entered all IMR scale responses into the secure, online Health Outcomes Management System (HOMS). Programs that were unable to enter IMR scale responses directly into this system were asked to send their IMR scale assessments to the academic researchers via mail or fax. Upon receipt of these mailed/ faxed assessments, the researchers double-entered all IMR scale responses in HOMS, and checked all data for accuracy.

Surveys were matched to demographic and diagnosis data from a clinical database via a unique client identifier. Client demographic and clinical data were collected by program staff

during an intake assessment (i.e., when admitted into a program). For clients who initiated services at multiple programs, we randomly selected one assessment for matching. Previous analyses of these data have shown that diagnoses remain fairly consistent; clients who visit multiple programs tend to receive the same diagnosis from each program [43, 44]. Individuals for whom age, gender, race, or diagnosis data were missing were excluded from analyses. Only individuals with primary diagnoses falling under the following categories were included in the study: schizophrenia or other psychotic disorders, major depressive disorder, bipolar disorder, and anxiety disorders.

Results

For sample 3, Cronbach's alpha for the 10-item scale was 0.908, and removing any of the items to create a nine-item scale resulted in a lower Cronbach's alpha (0.892–0.905). The mean score was significantly higher than the other samples as shown in Table 2.

Pearson correlation coefficients for the 10-item QWB-SA mental health scale with the MHSIP, IMR, and RMQ measures are presented in Table 2. Most of these correlations with other measures of mental health or mental health improvements were large ($p < 0.001$; absolute $r = 0.373$–0.398). Pearson correlation coefficients for the identified QWB-SA mental health scale with the comparison measures for discriminant validity were smaller and ranged from -0.106 to -0.152 ($p < 0.001$).

A number of people had an assessment with the IMR and RMQ within 6 months of the survey period in which QWB-SA data were collected, and correlations with these instruments are presented in the last row of Table 2. Unlike the MHSIP, these data were not collected at the same time as the QWB-SA mental health items and the IMR were not self-report, so correlations would be expected to be lower. Correlations were higher with the measures of functioning, with QWB-SA mental health correlating 0.248 with IMR Management Scale and 0.368 with the RMQ total score. Correlations were much smaller with the IMR Substance and IMR Recovery scales that would be expected to be less directly related, respectively, 0.051 and 0.055.

Discussion

This study examines the validity of a new mental health subscale using existing items from the QWB-SA across three samples. The new subscale addresses the concern that the preference-based QWB-SA does not have descriptive subscales and does not assess mental health. The internal reliability of the new mental health subscale was good in all three samples. One reason it might have been higher in the sample with mental illness is that there was more variability within that sample, thus providing less restriction in the range of values. The 10-item scale also displayed superior inter-item consistency when compared to scales consisting of fewer items. While a scale consisting of fewer items may be ideal when response burden is an issue, these scales were not investigated further because the QWB-SA mental health scale is intended to be used only as part of the larger QWB-SA.

The QWB-SA mental health scale exhibited strong concurrent validity as evidenced by high correlations with other measures of mental health. Strong item scale correlations between

the QWB-SA items and comparison measures suggest that the identified items are measuring the construct of mental health. Each of the 10 items was related to the comparison measures, and the correlation between the 10-item scale and comparison measures was strong.

With the score derived from this new subscale, users of the QWB-SA can now assess mental health in addition to overall health status. Measuring mental health separately from total HRQoL increases both research and clinical utility of the QWB-SA by allowing treatment programs to assess mental health, how a person's mental health is affected by disease, or how a person's mental health changes over time. Indeed, many studies have underscored the importance of incorporating measures of mental health in health outcome programs [5, 6]. It is important to note that the authors support the use of the QWB-SA mental health scale in addition to the QWB-SA overall utility score, as both physical and mental health components are involved in HRQoL [1, 4, 45].

As expected, the cohort with mental illness reported lower mental HRQoL as measured by the QWB-SA mental health subscale. Most people even with very severe mental illness would not be expected to have all the symptoms; thus, the differences are not as extreme in absolute numbers, given the breadth of mental health concerns covered by the scale. While the scale captures the frequency of symptoms well, it does not capture additional severity beyond the pervasiveness of the symptom being present on all three measured days. The defined illness cohort (Study 2) also reported somewhat lower mental HRQoL on the QWB-SA mental health scale, as well as the other measures of physical and mental health.

Our results and conclusions are strengthened by testing in multiple samples, including a national RDD sample that is generalizable to the US population, a sample of individuals with impaired physical health, and a sample of individuals in treatment for mental health problems. The development of the mental health subscale for the QWB-SA will be ongoing and additional validation studies will help advance our understanding of the subscales properties.

Of course general limitations to measuring HRQoL by self-report discussed in the literature also apply to the present study. A specific limitation of our study is that the random national sample was conducted in people aged 35–89 years so results may not generalize to younger adults. The sample with severe mental illness was younger on average than the other samples, perhaps in part due to significantly reduced lifespan of people with severe mental health problems [14]. In addition, we do not know how well specific mental health diagnoses were represented in all three samples.

Other concerns include the differing time frames between the measures, and the separation of time between measures for Study 3. Although the different measures of mental health use different time frames based on their different theoretical basis, they are all assumed to be measuring the same underlying construct. Another limitation is not having more specific measures of mental health other than the POMS, SF-36, and EQ-5D for two of the samples. Thus, further validity work using groups with known mental health diagnoses and should be

examined in future studies. Additionally, concurrent validity of our scale with other measures of mental health should be examined.

Among the people with severe mental health problems, there may have selection biases that affected the representation of severity of problems within that population. On the one hand, those with the most severely disabling mental health problems might be unlikely to complete a survey, but those with the least severe problems have less frequent appointments and are thus less likely to complete a survey than those with moderate problems. Thus, there may be important differences between responders and non-responders in this study. Also, because of the self-report nature of questionnaire research, psychiatric symptoms such as mood or psychosis could skew perceptions of outcomes [46, 47]. Some of the comparison instruments were either not given at the same time or measured slightly different constructs than an ideal comparison measure would have. It is important to note that we did not investigate differences in mental health between psychiatric disorders. While only the more severe depression cases would be represented in our sample because of the nature of public services, a fuller range of schizophrenia severity is likely present. Further studies are needed to examine the ability of the QWB-SA mental health scale to assess and predict patient recovery in the mental health system while also providing a platform to monitor their physical health for an increasingly integrated healthcare system.

In summary, the QWB mental health subscale should be useful for existing data, ongoing studies, and future endeavors. Users of the QWB-SA already obtain responses to the mental health items and there is minimal burden for calculating the mental health score from existing QWB-SA questions, and indeed, the mental health score could be calculated retrospectively in existing databases allowing for new analyses. Although there may be aspects of mental health that were not represented, the goal of this study was to derive a subscale of mental health without changing the original instrument. Because the item scoring is not preference weighted, it is not intended to be a utility measure, but it may provide an additional useful index of mental health. Additionally, the full QWB-SA score now allows cost-utility analysis of the impact of mental health conditions.

References

1. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Annals of Internal Medicine. 1993; 118:622–629. [PubMed: 8452328]

2. Kaplan RM. The minimally clinically important difference in generic utility-based measures. Journal of Chronic Obstructive Pulmonary Disease. 2005; 2(1):91–97. [PubMed: 17136968]

3. World Health Organization. Constitution of the World Health Organization. Geneva: World Health Organization; 1949.

4. Kaplan RM, Anderson JP. A general health policy model: Update and applications. Health Services Research. 1988; 23:203–235. [PubMed: 3384669]

5. Bergner M. Measurement of health status. Medical Care. 1985; 23(5):696–704. [PubMed: 4010356]

6. McHorney CA, Ware JE, Raczek AE. 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. 1993; 31(3):247–263. [PubMed: 8450681]

7. Spitzer RL, Kroenke K, Linzer M, Hahn SR, Williams JB, deGruy FV, et al. Health-related quality of life in primary care patients with mental disorders. Results from the PRIME-MD 1000 study. Journal of the American Medical Association. 1995; 274(19):1511–1517. [PubMed: 7474219]

8. Pyne JM, Bullock D, Kaplan RM, Smith TL, Gillin C, Golshan S, et al. Health-related quality-of-life measure enhances acute treatment response prediction in depressed inpatients. Journal of Clinical Psychiatry. 2001; 62(4):261–268. [PubMed: 11379840]

9. Pyne JM, Patterson TL, Kaplan RM, Gillin JC, Koch WL, Grant I. Assessment of the quality of life of patients with major depression. Psychiatric Services. 1997; 48(2):224–230. [PubMed: 9021855]

10. Pyne JM, Patterson TL, Kaplan RM, Ho S, Gillin C, Golshan S, et al. Preliminary longitudinal assessment of quality of life in patient with major depression. Psychopharmacology Bulletin. 1997; 33(1):23–29. [PubMed: 9133748]

11. Pyne JM, Sullivan G, Kaplan R, Williams DK. Comparing the sensitivity of generic effectiveness measures with symptom improvement in persons with schizophrenia. Medical Care. 2003; 41(2): 208–217. [PubMed: 12555049]

12. Patterson TL, Kaplan RM, Grant I, Semple SJ, Moscona S, Koch WL, et al. Quality of well-being in late-life psychosis. Psychiatry Research. 1996; 63:169–181. [PubMed: 8878313]

13. Kaplan, RM. Health-related quality of life in mental health services evaluation. In: Miller, NE.; Madruger, KM., editors. Cost effectiveness psychotherapy: A guide for practitioners, researchers and policymakers. New York, NY: Oxford Press, Inc; 1996. p. 160-173.

14. Parks, J.; Svendsen, D.; Singer, P.; Foti, ME., editors. Morbidity and mortality in people with serious mental illness. Alexandria: National Association of State Mental Health Program Directors (NASMHPD) Medical Directors Council; 2006.

15. Ware JE, Gandek B. Overview of the SF-26 health survey and the international quality of life assessment (IQOLA) project. Journal of Clinical Epidemiology. 1998; 51(11):903–912.10.1016/j.bbr.03.031 [PubMed: 9817107]

16. Fryback DG, Dunham NC, Palta M, Hanmer J, Buechner J, Cherepanov D, et al. US norms for six generic health-related quality-of-life indexes from the national health measurement study. Medical Care. 2007; 45(12):1162–1170.10.1097/MLR.0b013e31814848f1 [PubMed: 18007166]

17. Kaplan RM, Tally S, Hays RD, Feeny D, Palta M, Fryback D. Comparing the responsiveness of five preference-based indexes in cataract and heart failure patients. Journal of Clinical Epidemiology. 2011; 64(5):497–506.10.1016/j.clinepi.2010.04.010 [PubMed: 20685077]

18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33:159–174. [PubMed: 843571]

19. Sieber WJ, Groessl EJ, David KM, Ganiats TG, Kaplan RM. Quality of Well-Being Self-Administered (QWB-SA) scale: User's manual. 2004 Copyrighted material.

20. Kaplan RM, Ganiats TG, Sieber WJ. The quality of well-being scale, self-administered (QWB-SA). 1996 Copyrighted material.

21. Kaplan RM, Sieber WJ, Ganiats TG. The quality of well-being scale: Comparison of the interviewer-administered version with the self-administered questionnaire. Psychological Health. 1997; 12:783–791.10.1080/08870449708406739

22. Andersen EM, Rothenberg BM, Kaplan RM. Performance of a self-administered mailed version of the quality of well-being (QWB-SA) Questionnaire among older adults. Medical Care. 1998; 38(9):1349–1360.

23. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical Care. 1992; 30:473–483. [PubMed: 1593914]

24. Ware, JE.; Snow, KK.; Kosinski, M.; Gandek, B. SF-36® health survey manual and interpretation guide. Boston, MA: New England Medical Center, The Health Institute; 1993.

25. Ware JE, Gandek B. the IQOLA Project Group. The SF-36® health survey: Development and use in mental health research and the IQOLA Project. International Journal of Mental Health. 1994; 23(2):49–73.

26. Ware, JE.; Kosinski, M.; Keller, SK. SF-36® physical and mental health summary scales: A user's manual. Boston, MA: The Health Institute; 1994.

27. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics. 2002; 21:271–292. [PubMed: 11939242]

28. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: Development and testing of the D1 valuation model. Medical Care. 2005; 43:203–220. [PubMed: 15725977]

29. Rabin R, de Charro F. EQ-5D: A measure of health status from the EuroQol group. Annals of Medicine. 2001; 33:337–343. [PubMed: 11491192]

30. Feeny, D.; Torrance, G.; Furlong, W. Health utilities index. In: Spilker, B., editor. Quality of life and pharmacoeconomics in clinical trials. Philadelphia, PA: Lippincott-Raven Press; 1996.

31. Cohen, J. Statistical power analysis for the behavioral sciences. 2. Hillsdale, NJ: Erlbaum; 1988.

32. SPSS Inc. SPSS base 16.0 for windows user's guide. Chicago: SPSS Inc; 2008.

33. McNair, DM.; Lorr, M.; Droppleman, LF. Profile of mood states. San Diego, CA: Educational and Industrial Testing Service; 1992.

34. Center for Mental Health Services. The final report of the mental health statistics improvement project (MHSIP) task force on a consumer-oriented mental health report card. Rockville, MD: Center for Mental Health Services; 1996.

35. NASMHPD Research Institute, Inc. State profile highlights: Performance measures system: Consumer survey 2000–2001. Jun. 2002 Available at: http://www.nri-inc.org/projects/Profiles/Profiles01/03MHSIP2001.pdf

36. Minsky S, Lloyd P. The use of unique client identifiers in state mental health service systems. Administration and Policy in Mental Health Services Research. 1996; 23(3):231–252.10.1007/BF02108322

37. Mueser, KT.; Gingerich, S.; Salyers, MP.; McGuire, AB.; Reyes, RU.; Cunningham, H. Illness management and recovery (IMR) scales. In: Campbell-Orde, T.; Chamberlin, J.; Carpenter, J.; Leff, HS., editors. Measuring the promise: A compendium of recovery measures. Vol. 2. Cambridge, Mass: Evaluation Center at Human Services Research Institute; 2005. p. 124-132.

38. Mueser KT, Meyer PS, Penn DL, Clancy R, Clancy DM, Salyers MP. The illness management and recovery program: Rationale, development, and preliminary findings. Schizophrenia Bulletin. 2006; 32:32–43.10.1093/schbul/sbl022

39. Hasson-Ohayon L, Roe D, Kravetz S. The psychometric properties of the illness management and recovery scale: Client and clinician versions. Psychiatry Research. 2008; 160(2):228–235.10.1016/j.psychres.2007.06.013 [PubMed: 18599126]

40. Salyers MP, Godfrey JL, Mueser KT, Labriola S. Measuring illness management outcomes: A psychometric study of clinician and consumer rating scales for illness self management and recovery. Community Mental Health Journal. 2008; 43(5):459–480.10.1007/s10597-007-9087-6 [PubMed: 17514504]

41. Sklar M, Sarkin A, Groessl E, Gilmer T. The psychometric properties of the Illness Management and Recovery scale in a large American public mental health system. Psychiatry Research. 201210.1016/j.psychres.2012.03.013

42. Ridgway, P.; Press, A. [February 1, 2008] Assessing the recovery commitment of your mental health service: A user's guide for the developing recovery enhancing environments measure (DREEM). 2004. Retrieved from http://64.233.169.104/search?q=cache:mB9Ag810-SMJ:www.outcomemeasures.csip.org.uk/silo/files/dreem-pdf+DREEM,+ridgway,+press&hl=en&ct=clnk&cd=3&gl=us

43. Gilmer T, Stefancic A, Ettner S, Manning W, Tsemberis S. Effect of full-service partnerships on homelessness, use and costs of mental health services, and quality of life among adults with serious mental illness. Archives of General Psychiatry. 2010; 67(6):645–652. [PubMed: 20530014]

44. Folsom D, Lindamer L, Montross L, Hawthorne W, Golshan S, et al. Diagnostic variability for schizophrenia and major depression in a large public mental health care system dataset. Psychiatry Research. 2006; 144(2–3):167–175. [PubMed: 16979244]

45. Hays RD, Stewart AL. The structure of self-reported health in chronic disease patients. Psychological Assessment: A Journal of Consulting and Clinical Psychology. 1990; 2:22–30.10.1037/1040-3590.2.1.22

46. Atkinson M, Zibin S, Chuang H. Characterizing quality of life among patients with chronic mental illness: A critical examination of the self-report methodology. The American Journal of Psychiatry. 1997; 154(1):99–105. [PubMed: 8988966]

47. Winter L, Lawton MP, Langston CA, Ruckdeschel K, Sando R. Symptoms, affects, and self-rated health: Evidence for a subjective trajectory of health. Journal of Aging and Health. 2007; 19(3): 453–469.10.1177/0898264307300167 [PubMed: 17496244]

Table 1

Demographic characteristics of the three samples

Characteristics	Sample 1		Sample 2		Sample 3	
	N	%	N	%	N	%
Gender						
Male	1,641	42.7	261	48.8	469	51.3
Female	2,203	57.3	274	51.2	446	48.7
Age						
18–34	–	–	–	–	322	35.2
35–44	642	16.7	29	5.4	170	18.6
45–64	1,510	39.3	215	40.2	411	44.9
65–89	1,692	44.0	291	54.4	12	1.3
Race						
White	2,562	66.7	451	84.3	503	55.0
Black	1,086	28.3	28	5.2	121	13.2
Other	178	4.6	30	5.6	280	30.6
Missing	18	0.5	26	4.9	11	1.2
Education						
< High school	464	12.1	39	7.3	–	–
High school	1,159	30.2	104	19.4	–	–
Some post-high school	856	22.3	161	30.1	–	–
4-yr college degree	1,341	34.9	212	39.6	–	–
Missing	24	0.6	19	3.6	–	–
Total	3,844	100	535	100	915	100

Table 2

Rater agreement, scale reliability coefficients, and correlation coefficients between the identified 10-item QWB-SA mental health scale and other measures

Samples	Concurrent measures of validity								Discriminant Measures	
	Cronbach's alpha	Mean (S.D.)	SF-36 MH component[b]	SF-36 MH Scale[b]	EQ-5D anxiety/depression Item	HUI2 emotion scale[b]	POMS depression-dejection scale[a]	POMS TMD[a]	SF-36 physical component[b]	SF-36 physical functioning scale[b]
Sample 1	0.827	1.06 (1.74)	−0.663**	−0.698**	0.608**	−0.631**	—	—	−0.320**	−0.389**
Sample 2	0.842	1.31 (1.83)	−0.690**	−0.723**	0.609**	−0.591**	0.767**	0.768**	−0.251**	−0.354**

Samples	Concurrent measures of validity						Discriminant measures				
	Cronbach's alpha	Mean (S.D.)	IMR management scale[c]	Recovery Markers Quest[d]	MHSIP functioning	MHSIP outcomes	MHSIP social	IMR recovery scale[c]	IMR substance scale[c]	MHSIP access	MHSIP quality of services	MHSIP participation
Sample 3	0.908	2.52 (2.28)	−0.248**	−0.368**	−0.397**	−0.398**	−0.373**	−0.055	−0.051	−0.152**	−0.106*	−0.118**

Sample 1 $n = 3,844$, Sample 2 $n = 535$, Sample 3 $n = 915$

Overall agreement between raters ICC = 0.78

Absolute agreement between raters = 60%

[*]
$p < 0.01$;

[**]
$p < 0.001$

[a] Scale score was calculated using normative-based weighted scoring

[b] $n = 164$

[c] $n = 519$

[d] $n = 405$

Table 3

Rater agreement and correlation coefficients between the 10 identified QWB-SA items and other measures

Item content	% Agree mental health	% Agree undecided	% Agree physical health	Sample	SF-36 physical component[a]	SF-36 physical functioning scale[a]	SF-36 MH component[a]	SF-36 MH scale[a]	EQ-5D anxiety/depression scale	HUI2 emotion scale[a]	POMS depression-dejection scale[b]	POMS TMD[b]
Trouble falling asleep or staying awake	50	50	0	Sample 1	-0.271^{**}	-0.266^{**}	-0.298^{**}	-0.328^{**}	0.289^{**}	-0.277^{**}	–	–
				Sample 2	-0.205^{**}	-0.210^{**}	-0.327^{**}	-0.330^{**}	0.218^{**}	-0.248^{**}	0.370^{**}	0.410^{**}
Spells of feeling nervous or shaky	75	25	0	Sample 1	-0.229^{**}	-0.279^{**}	-0.439^{**}	-0.473^{**}	0.418^{**}	-0.436^{**}	–	–
				Sample 2	-0.194^{**}	-0.262^{**}	-0.486^{**}	-0.497^{**}	0.398^{**}	-0.378^{**}	0.411^{**}	0.440^{**}
Spells of feeling upset, downhearted, or blue	100	0	0	Sample 1	-0.204^{**}	-0.270^{**}	-0.556^{**}	-0.591^{**}	0.506^{**}	-0.520^{**}	–	–
				Sample 2	-0.122^{**}	-0.226^{**}	-0.631^{**}	-0.669^{**}	0.588^{**}	-0.548^{**}	0.652^{**}	0.641^{**}
Excessive worry or anxiety	100	0	0	Sample 1	-0.207^{**}	-0.278^{**}	-0.525^{**}	-0.568^{**}	0.507^{**}	-0.503^{**}	–	–
				Sample 2	-0.107^{*}	-0.203^{**}	-0.557^{**}	-0.603^{**}	0.555^{**}	-0.512^{**}	0.592^{**}	0.606^{**}
Feelings that you had little or no control over events in your life	100	0	0	Sample 1	-0.199^{**}	-0.260^{**}	-0.502^{**}	-0.521^{**}	0.447^{**}	-0.476^{**}	–	–
				Sample 2	-0.208^{**}	-0.308^{**}	-0.553^{**}	-0.566^{**}	0.458^{**}	-0.412^{**}	0.694^{**}	0.634^{**}
Feelings of being lonely or isolated	100	0	0	Sample 1	-0.191^{**}	-0.265^{**}	-0.490^{**}	-0.503^{**}	0.427^{**}	-0.454^{**}	–	–
				Sample 2	-0.136^{**}	-0.234^{**}	-0.585^{**}	-0.556^{**}	0.448^{**}	-0.432^{**}	0.684^{**}	0.630^{**}
Feelings of frustration, irritation, or close to losing your temper	100	0	0	Sample 1	-0.149^{**}	-0.184^{**}	-0.419^{**}	-0.440^{**}	0.366^{**}	-0.417^{**}	–	–
				Sample 2	-0.172^{**}	-0.217^{**}	-0.412^{**}	-0.461^{**}	0.430^{**}	-0.410^{**}	0.411^{**}	0.433^{**}
Confusion, difficulty understanding written spoken word, significant memory loss	50	50	0	Sample 1	-0.201^{**}	-0.226^{**}	-0.280^{**}	-0.273^{**}	0.250^{**}	-0.285^{**}	–	–
				Sample 2	-0.119^{**}	-0.163^{**}	-0.194^{**}	-0.211^{**}	0.177^{**}	-0.221^{**}	0.313^{**}	0.347^{**}
Thoughts or images you could not get out of your mind	100	0	0	Sample 1	-0.121^{**}	-0.172^{**}	-0.369^{**}	-0.376^{**}	0.334^{**}	-0.339^{**}	–	–
				Sample 2	-0.107^{*}	-0.173^{**}	-0.394^{**}	-0.359^{**}	0.285^{**}	-0.327^{**}	0.316^{**}	0.327^{**}
A loss of appetite or over-eating	50	0	50	Sample 1	-0.234^{**}	-0.249^{**}	-0.289^{**}	-0.301^{**}	0.275^{**}	-0.275^{**}	–	–
				Sample 2	-0.278^{**}	-0.297^{**}	-0.357^{**}	-0.360^{**}	0.318^{**}	-0.306^{**}	0.386^{**}	0.389^{**}

Sample 1 $n = 3{,}844$, Sample 2 $n = 535$

*
$p < 0.01$;

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

$**$
$p < 0.001$

[a] Scale score was calculated using normative-based weighted scoring

[b] $n = 164$