EQ-HWB-S and EQ-HWB as an outcome measure in mental health and well-being: a comparison to the EQ-5D-5L, EQ-5D-3L and the Short Warwick-Edinburgh Mental Well-being Scale

EuroQol Plenary 2022 Working Paper

Authors: Maja Kuharic[1], Andrea L. Monteiro[2], A. Simon Pickard[1]

Corresponding author: Maja Kuharic, Department of Pharmacy Systems, Outcomes and Policy. College of Pharmacy, University of Illinois at Chicago. 833 South Wood Street (MC 871). Chicago, IL 60612 USA. E-mail: mkuhar2@uic.edu

Affiliations:

1 Department of Pharmacy Systems, Outcomes and Policy. College of Pharmacy, University of Illinois at Chicago. 833 South Wood Street (MC 871). Chicago, IL 60612

2 Department of Pharmacy Practice. College of Pharmacy, University of Illinois at Chicago. 833 South Wood Street (MC 871). Chicago, IL 60612

UIC UNIVERSITY OF ILLINOIS CHICAGO

Abstract

Objectives: The extent to which generic measures are sensitive to changes in the kinds of symptoms, functioning, and quality of life that are of relevance to capturing mental health and well-being has been the subject of debate. The EQ-HWB is intended to broadly capture aspects of health and well-being that may be missed by existing generic measures as well as compared to other measures intended to capture mental well-being such as the Short Warwick-Edinburgh Mental Well-Being Scale (S-WEMWBS). Thus, this study aims to examine content overlap and compare measurement properties and discriminative ability of EQ-HWB/EQ-HWB-S in relation to the EQ-5D-5L, EQ-5D-3L, and S-WEMWBS with respect to mental health conditions.

Methods: An online panel of US-based respondents completed a survey that included the EQ-HWB Version 1.1, EQ-5D-5L, EQ-5D-3L, and S-WEMWBS. In addition to item-level analysis, the EQ-HWB/EQ-HWB-S was scored using a non-preference-based scoring approach under development where two sets of items were combined to match the items of the experimental version of the EQ-HWB and 3 proposed subscales (psychosocial, pain/discomfort and activities). US-based value sets were applied to generate EQ-5D-3L and EQ-5D-5L index scores. The analysis examined content overlap qualitatively and the strength of correlation between related items/constructs EQ-HWB, EQ-HWB-S, and S-WEMWBS. Discriminative ability of measures using known group comparisons (KGC) was performed using effect sizes (ES) and analysis of variance F-ratios based on any self-reported mental health problem, clinical depression, and general anxiety disorder (GAD).

Results: The dataset included a total of 903 participants, including 172 participants who self-reported any mental health problems, 113 clinical depression, and 98 generalized anxiety disorder. Most content overlap measured with Jaccard Index was found between EQ-HWB-S and S-WEMWBS (33%). Strong associations ($r_s > 0.5$) were found between conceptually overlapping/related items of S-WEMWBS and EQ-HWB such as: "Thinking Clearly" and "Concentrating and Thinking Clearly", "Dealing with Problems" and "Cope", "Relaxed" and "Anxious". The EQ-HWB-psychosocial LSS and EQ-HWB-S LSS tended to better discriminate than S-WEMWBS summary score based on any MH problem (F-ratio: 2.22, 95% CI 1.61-2.89 and 2.13, 95% CI 1.51-3.00 respectively), clinical depression (F-ratio: 2.24, 95% CI 1.64-3.02 and 2.08, 95% CI 1.48-2.92) and GAD (F-ratio: 2.13, 95% CI 1.51-2.99 and 2.23, 95% CI 1.53-3.28). For KGC based on MH conditions, EQ-HWB-psychosocial LSS and EQ-HWB-S exhibited very large effect sizes (i.e. >1.2) across all conditions, while other measures (EQ-5D-5L, EQ-5D-3L, and S-WEMWBS) exhibited large ES (i.e. >0.8).

Conclusion: Initial evidence supports the validity of the EQ-HWB and EQ-HWB-S as outcome measures in mental health and well-being. All measures demonstrated discriminative ability, with the EQ-HWB psychosocial LSS and EQ-HWB-S tending to outperform both the SWEBMS and EQ-5D-5L/EQ-5D-3L in terms of known groups based on mental health, highlighting its future potential as a measure of mental health using a psychometrically derived summary score.

Introduction

Mental health (MH) is an integral and essential component of health [1]. According to the National Institute on Mental Health, nearly one-in-five U.S. adults live with a mental illness [2, 3]. Mental disorders are increasingly recognized as leading causes of disease burden worldwide [4] and represent 23% of the total cause of disability, higher than cancer and coronary heart disease [5]. They accounted for 654.8 million estimated cases in 1990 and 970.1 million cases in 2019, corresponding to an increase of global prevalence by 48.1% between 1990 and 2019, with no evidence of a global reduction in the burden since 1990 [6]. Collaborators of Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) called in 2019 to reduce the burden of mental disorders, and coordinate the delivery of effective prevention and treatment programs by governments and the global health community [6]. Therefore, prioritizing mental health alongside other care interventions are important considerations for decision-makers.

The extent to which generic measures are sensitive to changes in the kinds of symptoms, functioning, and quality of life that are of interest to people with MH problems has been the subject of debate [7]. There has been a shift in MH service policy from an emphasis on treatment focused on reducing symptoms, based on a narrow notion of health and disease, to a more holistic approach that takes into consideration well-being, recovery, social functioning, and quality of life (QOL) [7, 8]. Such shift in mental policy necessitates that appropriate outcome measures are in place. However, few such measures are standardized and routinely collected across MH services [8]. Furthermore, in that context, there is an ongoing debate about how and with which instruments the benefits of mental healthcare interventions could be adequately measured and valued [9, 10]. Moreover, there is an issue of comparability of interventions across different patient groups and the subsequent allocation of resources useful in decision making. A recent study by Krugten et al [11] assessed the content validity and the suitability of existing QOL instruments for use in economic evaluations of MH problems. The study highlighted the multitude of available QOL instruments used in people with MH problems and indicated that none of the available QOL instruments fully cover the dimensions previously found to be important in people with MH problems. Furthermore, the adequacy of often used generic health-related QOL (HRQL) instruments, such as the EQ-5D and the 36-item Short-Form Health Survey (SF-36), has been questioned in the context of MH [9, 10]. More specifically, some have suggested that these instruments, in certain situations, lack the sensitivity to sufficiently reflect the impact of MH problems on QOL [12]. The EQ-5D, for example, appears to perform well in mild to moderate MH conditions [13, 14] but exhibited weak correlations with severe MH problems such as schizophrenia [10]. Some argue that this may be because these commonly used

QOL instruments have been developed primarily for people with a physical illness, thereby limiting the coverage of dimensions perceived important to the QOL of people with MH problems [11, 15]. Hence, the debate in this area relates both to the sensitivity of existing HRQL instruments, and also to the scope of relevant outcomes (i.e., potentially broadening the evaluative space) [11, 16].

EQ Health and wellbeing (EQ-HWB) measures have been developed internationally for evaluating interventions in health, public health, and social care including the impact on patients, social care users, and carers to broadly capture aspects of health and well-being that may be missed by existing generic measures [17-19]. It was modified to include aspects beyond health and extended to include other factors that may have a direct or indirect impact on physical and MH symptoms, such as being a carer [17, 18]. During EQ-HWB development, initial psychometric evidence on item-level showed that most items were able to discriminate well between those with and without an identified MH condition in all six countries (UK, US, Argentina, Germany, Australia, and China) [18]. The psychometric performance of the current experimental EQ-HWB version on a scale level in people with MH problems and how the EQ-HWB compares to other measures intended to capture mental well-being such as the Short Warwick-Edinburgh mental well-being scale (S-WEMWBS) is currently unknown. Given that the EQ-HWB/EQ-HWB-S was developed to include domains that are of interest to patients with MH symptoms among others, evidence on the ability of the EQ-HWB to capture MH issues in comparison to other measures will inform future use of the measure. This study aims to examine and compare the measurement properties of EQ-HWB/EQ-HWB-S in relation to the EQ-5D-5L, EQ-5D-3L, and S-WEMWBS with respect to MH conditions focusing on: (1) content overlap between EQ-HWB/EQ-HWB-S and EQ-5D with mental well-being measure S-WEMWBS, and (2) psychometric properties of EQ-HWB/EQ-HWB-S in relation to the EQ-5D-5L, EQ-5D-3L, and S-WEMWBS. Properties were examined in terms of response distributions, convergent validity, and discriminative ability in patients with any mental health condition, including more common conditions - clinical depression and generalized anxiety disorder (GAD). Evidence of the validity of EQ-HWB/EQ-HWB-S in patients with MH problems can support its application in future research and inform its use in evaluating benefits of MH interventions.

Methods

Data Source

This study was conducted on secondary analysis of cross-sectional data collected from an adult sample of US-based cancer survivors (n=403) and members of the general population (n=500) between August and September 2019 during the psychometric stage of Extending the QALY (E-QALY) project. Respondents in this dataset were recruited from an internet panel and quota-sampled on age, gender, and race to support comparability to the general population of the United States (US). Data collection was approved by the institutional review board at the University of Illinois at Chicago (IRB# 2019-0184) and all respondents provided informed consent.

The US-arm E-QALY survey, in addition to the measures described below, included questions on sociodemographics and common chronic conditions. Each respondent was asked about gender, age, marital status, education level, caregiver status and burden, and social care utilization. Health questions included the self-reported list of any long-standing physical or MH conditions. Cancer survivors were also asked to provide information about their cancer history.

Measures

EQ-5D consists of a health state classification system with five dimensions of mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. There are two versions: the 3-level (3L) version which has three severity levels for each dimension (no problems, moderate problems, and extreme problems) and the more recently developed 5-level (5L) version with five levels of severity (no problems, slight problems, moderate problems, severe problems, unable to/extreme problems) [20]. The EQ-5D is also comprised of a Visual Analog Scale (EQ VAS) that assesses the respondent's self-rated health on a vertical scale from zero (worst health you can imagine) to 100 (the best health you can imagine) [21]. EQ-5D index values were obtained through a scoring function from a set of population-based preference weights from the US valuation study for the 3L [22] and 5L [23].

EQ Health and Well Being (EQ-HWB) [17, 18] is composed of a 25-item profile measure (EQ-HWB) and a 9-item classifier developed for valuation purposes to generate utility values (EQ-HWB-S) [18] embedded in the profile version. The measures conceptually include items related to 32 subdomains grouped into 7 high-level domains: activity, relationships, cognition, self-identity, autonomy, feelings, and physical sensations. Responses to items are either a five-point Likert frequency scale (i.e., not at all, only occasionally, some of the time, often, most or all of the time), severity scale (i.e., mild, slight, moderate, severe, very severe OR not at

all, a little bit, somewhat, quite a bit, very much) or difficulty scale (i.e., no difficulty, slight, some, a lot of, unable). The dataset used in this study contains EQ-HWB Experimental version 1.1. The analytic dataset combined two sets of items to match the current EQ-HWB Experimental version 1.2 by a proposed algorithm: "Get around Inside"/'Get around outside", and "Thinking clearly"/"Concentrating" [24]. EQ-HWB-S is scored using a non-preference-based scoring approach under development that supports the use of a level summary score (LSS) [24]. EQ-HWB is scored in one of the proposed scoring approaches under development in which EQ-HWB is divided into three subscale LSSs: psychosocial, pain/discomfort, and activities with items "See" and "Hear" excluded from an LSS and are to be analyzed separately [24].

Short Warwick-Edinburgh Mental Wellbeing Scale (S-WEMWBS) [25, 26] is a short version of the longer-version scale Warwick–Edinburgh Mental Wellbeing Scale (WEMWBS) developed to identify the level of positive MH and wellbeing. It can be used for the evaluation of projects, programs, and policies that aim to improve mental wellbeing [25]. Most items represent aspects of psychological and eudemonic well-being, and few cover hedonic well-being or affect [27]. S-WEMWBS, contains items on optimism, usefulness, feeling relaxed, thinking clearly, dealing with problems, feeling close to others, and being able to make up one's own mind [26]. The seven statements are positively worded with five response categories from 'none of the time' to 'all of the time'. A raw score to interval scale transformation of S-WEMWBS scores has been developed using Rasch Analysis, giving a possible range from 7 to 35 with higher scores reflecting greater overall mental well-being [26].

Analysis

The analysis followed methods recommended by the COnsensus-based Standards for the selection of health status Measurement Instruments checklist on evaluating measurement (COSMIN) and quality criteria for measurement properties of health status questionnaires [28, 29]. Statistical significance was, $p < 0.05$, with all analyses conducted using SAS Version 9.4. (SAS Institute Inc.) Cary, North Carolina and RStudio Version 2021.09.

Patterns of responses

We conducted descriptive analyses to study the response pattern to the EQ-HWB, EQ-HWB-S, S-WEMWBS, EQ-5D-5L and EQ-5D-3L. More specifically, we evaluated acceptability by computing the percentage of missing data for each item. A proportion of missing values greater than 5% was defined *a priori* as a marker of potential problems, as a higher proportion of missing values may indicate the item's lower interpretability and acceptability [28]. We examined floor and ceiling effects at the score and item

4

level by calculating the proportion of participants scoring at the lowest and highest possible levels. At the item level, a floor or ceiling effect was considered if at least 50% of the respondents scored at the minimum or maximum level [30]. At the scale level, an effect was indicated if at least 15% of the respondents scored at the lowest or highest summary score which may indicate that measures may be insensitive to deteriorations/improvements over time or differences between groups [28].

Content analysis

The S-WEMWBS, EQ-HWB, EQ-HWB-S, EQ-5D items were subject to qualitative content analysis [31]. The objective was to evaluate content to identify similarities and differences through cross-comparison of S-WEMWBS with EQ-HWB, EQ-HWB-S and EQ-5D [32, 33]. Items were considered as the same item content across measures as long as they were (a) similar items, which meant that they have identical wording or common concept, such as "Thinking clearly" and "Thinking clearly/concentrating" and (b) oppositely worded, such as "Anxious" and "Relaxed", following criteria used in a previous study evaluating content overlap among depression scales [33]. Based on the analysis of items we calculated statistical content overlap between PROMs using the Jaccard Index [34]. The Jaccard index measures the degree of similarity between two sets of data with a range from 0 (no overlap) to 1 (full overlap), and is calculated as the number of shared items divided by the total number of items in any two PROMs [34]. The Jaccard index was expressed as a percentage (multiplied by 100). The analysis was conducted independently by two researchers (ALM and MK) and was continuously discussed in the research group until an agreement was reached (i.e. the number of shared items).

Construct validity

Construct validity is assessed in light of the fact that there is no gold standard for the measurement of HRQL in MH. The goal was to examine two related empirical tests of convergent validity and known group differences.

Convergent validity included exploring correlations between items and summary statistics of the S-WEMWBS and EQ-HWB, EQ-HWB-S, EQ-5D-5L and 3L. Pearson's correlation coefficients were used to study bivariate associations and correlations of the summary scores. The *a priori* hypothesis was that the correlation between the S-WEMWBS and EQ-HWB-psychosocial subscale and EQ-HWB-S LSS will be strong in strength and negative in direction (worse health and wellbeing on EQ-HWB and better positive mental well-being on S-WEMWBS). The degree of association between individual items of the S-WEMWBS and EQ-HWB/EQ-HWB-S were examined with the Spearman correlation coefficients (r_s). The *a priori*

hypothesis was that the correlation between the conceptually related domains of S-WEMWBS such as "Thinking Clearly", "Close to other people", "Relaxed", "Optimistic about future", and "Dealing with problems well", will at least moderate in strength with related EQ-HWB/EQ-HWB-S items "Concentrating/Thinking Clearly", "Lonely", Anxious", "Look forward", "Control" respectively (lower health and well-being and higher positive mental well-being). Correlation coefficients were interpreted according to Cohen's guidelines, ie, "strong" (≥0.51), "moderate" (0.31-0.50), "weak" (0.11-0.30), and "none" (0-0.10) [35].

Discriminative validity using known-groups comparison (KGC) was performed to test the sensitivity of the EQ-HWB-S/EQ-HWB subscales, S-WEMWBS, EQ-5D-5L/3L in the ability to capture expected differences between subgroups based on having any MH conditions, clinical depression, GAD or any other mental condition [36]. Based on the literature [37] the *a priori* hypothesis was that EQ-HWB/EQ-HWB-S LSSs will be higher (indicating worse health and well-being) in patients with a self-reported MH condition. For each known group, we calculated the mean scores of the measures, one-way ANOVA significance tests, and Cohen's effect sizes (ES). ES was used to quantify the magnitude of the difference between each predefined known group. The magnitude of the ES was interpreted to the Cohen's thresholds: small (0.2-0.49), medium (0.5-0.79), large (0.8-1.19) and very large (1.20≥) [38]. The result of the performance comparison between EQ-HWB-S/EQ-HWB, S-WEMWBS, EQ-5D-5L, and EQ-5D-3L was expressed as relative efficiency which is the ratio of F statistics between 2 measures. Bias-corrected 95% confidence intervals were calculated by bootstrapping [39].

Sample size

The sample size estimation for the E-QALY study was based on recommendations regarding the minimum sample size considered adequate for the development of factor analysis and item response theory models [40]. There are no general criteria for the required sample size in a validation study. A sample size greater than 100 respondents is generally recommended [41]. For the purposes of this analysis, our sample size of 172 respondents was deemed sufficient.

Results

Sample characteristics

The dataset included a total of 903 participants, with data collected from the general population (n=503) and cancer survivors (n=400). Of the 903 total participants, 172 (19%) participants self-reported having any MH condition and were included in this analysis. Out of those who reported having MH condition, 113 self-reported having clinical depression, 98 GAD, and 52 other MH conditions. The average age of participants with MH condition was 46.2 (SD 15.1) and 65% were female. The age and sex distribution of MH respondents differ from US general population in which MH respondents are younger and female. Such characteristics are consistent with the literature on depression and GAD, in which higher prevalence is generally observed in females [42, 43] and in younger populations [44]. The most common chronic condition among the MH population was cancer (35%) and 77% reported having any physical health problem. Those with self-reported MH conditions had an overall mean EQ-5D-5L index score of 0.56 (SD 0.31), EQ-5D-3L index score of 0.64 (SD 0.21), EQ-HWB psychosocial LSS was 46 (SD 15.44), EQ-HWB-S 24.86 (SD 8.22) and S-WEMWBS 20.05 (SD 4.94), with scores indicating worse health than the general US population on all measures. A detailed breakdown of the respondents' characteristics for the general population, and subsamples of respondents with any MH problems, clinical depression, GAD, and other MH conditions are presented in **Table 1**.

Content analysis

Comparison and classification of the total 37 items in 5 measures revealed considerable content overlap in items representing mental well-being in EQ-HWB/EQ-HWB-S and S-WEMWBS. Out of 7 S-WEMWBS items, the content of 6 items overlapped with EQ-HWB ("Control", "Concentrate/Think Clearly", "Hope", "Lonely", "Stigma", "Self-worth" and "Anxious"), 4 with EQ-HWB-S ("Control", "Concentrate/Think Clearly", "Lonely", "Anxious"), and 1 with EQ-5D ("Anxious/Depressed"). EQ-HWB-S had the highest degree of overlap as measured by the Jaccard index with S-WEMWBS (33%), followed by EQ-HWB (23% overlap), and EQ-5D (9% overlap). The results are not unexpected, given that EQ-HWB aims to capture a broader impact on health and well-being, including MH. Specific overlap among instruments is presented in **Figure 1.a, 1.b, 1.c**.

Distribution of responses

Missing data

The percentage of missing items among participants with MH condition across EQ-HWB/EQ-HWB-S, EQ-5D-5L/3L, and S-WEMWBS were generally low, ranging from 0% to 4.07%, averaging 0.63%. The item with the most missing data was EQ-HWB/EQ-HWB-S item "Concentrate/Think clearly" with 4.07% missing data. However, this item is a set of combined two items from the Experimental version of EQ-HWB 1.1. to reflect 1.2. version and missingness of responses might be overestimated.

Floor and Ceiling effects

On the item level, the proportion of participants who chose the lowest response option varied between 7.04% for the S-WEMWBS item "Relaxed" to 70.83% for EQ-5D-3L item "Self-care" while the proportion who chose the highest response option varied from 0% for the EQ-HWB/EQ-HWB-S item on "Get around Inside/Outside" to 27.22% on the EQ-HWB/EQ-HWB-S item "Exhausted". No ceiling effects (more than 50% at the highest level) were noted, but floor effects (more than 50% at the lowest level) were observed for the EQ-HWB item "Unsafe", EQ-5D-3L/5L items "Self-care" and EQ-5D-3L item "Mobility". On the scale level, the proportion of respondents who scored the lowest summary score varied from 0% for EQ-HWB/EQ-HWB-S to 2.35% for S-WEMWBS. The proportion who scored at the highest level varied from 0% for EQ-HWB/EQ-HWB-S to 4.22% on EQ-5D-3L. Floor effect (more than 15% at the lowest level) was observed for the EQ-HWB activities subscale (19.20%) but ceiling effects (more than 15% at the highest level) were not observed for any measure summary or index score (**Table 2.a and 2.b.).**

Construct Validity

Convergent Validity

EQ-HWB item "Cope" had strong associations ($r_s > 0.5$) with most S-WEMWBS items (six out of seven in total), followed by items "Accepted" and "Feel good" which were associated with a total of four S-WEMWBS items. As expected, strong associations were found between conceptually overlapping/related items of S-WEMWBS and EQ-HWB such as the S-WEMWBS item "Think clearly" with EQ HWB item "Concentrate/Think clearly" ($r_s = 0.55$); S-WEMWBS item "Relaxed" and EQ-HWB item "Anxious" ($r_s = 0.51$); S-WEMWBS item "Optimistic about future" and EQ-HWB item "Look Forward" ($r_s = 0.52$); S-WEMWBS item "Close to other people" and EQ-HWB item "Lonely" ($r_s = 0.58$) in addition to being

8

associated with three additional items: "Unsupported" ($r_s = 0.56$), "Accepted" ($r_s = 0.63$) and "Feel good" ($r_s = 0.57$). The only exception from our hypothesis was the S-WEMWBS item "Dealing with problems well" was not strongly correlated with EQ-HWB item "Control", however, it did show strong correlations with item "Cope" ($r_s = 0.58$) and moderate with "Control" ($r_s = 0.43$). All the correlations were statistically significant at the p<0.05 level. On the other hand, EQ-5D-5L item "Anxious" exhibited moderate association with S-WEMWBS items, while all other items were moderately or weakly correlated with S-WEMWBS items ($r_s = 0.11 - 0.30$), with the exception of EQ-5D item "Self-care" and S-WEMWBS item "Dealing with problems well" ($r_s = 0.35$). For detailed results, please see **Table 3.a., 3.b. and 3.c.**

At the summary score level, as expected, EQ-HWB psychosocial LSS was strongly correlated with S-WEMWBS ($r_s = -0.70$) meaning worse psychosocial status on EQ-HWB is associated with worse mental well-being, followed by EQ-HWB-S summary score ($r_s = -0.62$). On the other hand, EQ-5D index scores were only moderately correlated with S-WEMWBS summary scores ($r_s = 0.4$ and $r_s = 0.45$ for 5L and 3L respectively), suggesting a broader measurement concept of EQ-HWB may be measuring a concept more related to mental well-being in addition to health. Results for summary scores are shown in **Table 3.d**.

Known Group Validity

All measures were able to detect known group differences between those with and without any MH problem, clinical depression, GAD, and other MH problems with large ($d \geq 0.8$) to very large ($d \geq 1.2$) effect sizes. As expected with our hypothesis, the EQ-HWB-psychosocial LSS and EQ-HWB-S LSS tended to be able to better discriminate than S-WEMWBS summary score based on having any MH condition (F-ratio: 2.22, 95% CI 1.61-2.89 and 2.13, 95% CI 1.51-3.00 respectively), clinical depression (F-ratio: 2.24, 95% CI 1.64-3.02 and 2.08, 95% CI 1.48-2.92) and GAD (F-ratio: 2.13, 95% CI 1.51-2.99 and 2.23, 95% CI 1.53-3.28). For KGC based on MH conditions, EQ-HWB and EQ-HWB-S exhibited very large ES (i.e. d d ≥ 1.2) across all conditions, while other measures (EQ-5D-5L, 3L, and S-WEMWBS) exhibited large ES (i.e. d d ≥ 0.8) to very large ES in GAD. Specifically, the EQ-HWB-psychosocial LSS showed the largest ES for patients with/without any MH condition (d = 1.34) and clinical depression (d = 1.37). The results indicated EQ-HWB-S and EQ-HWB psychosocial LSS may be better in discriminating as compared to EQ-5D measures and S-WEMWBS. Results of the KGC analysis are presented in **Table 4.**

Discussion

Overview

This study is the first to examine and compare content overlap and measurement properties, including the discriminative ability of EQ-HWB/EQ-HWB-S in relation to the EQ-5D-5L, EQ-5D-3L, and S-WEMWBS with respect to MH conditions, specifically depression, GAD and other MH conditions. Given that EQ-HWB/EQ-HWB-S was developed to include aspects of life that are of interest to people with MH problems among others, understanding the appropriateness of the measures in these conditions is of significance for future EQ-HWB development.

The result of this study indicates that EQ-HWB/EQ-HWB-S perform well across groups with MH conditions, including clinical depression, GAD, and other MH conditions. EQ-HWB/EQ-HWB-S measures are acceptable for patients with MH condition and there is substantial overlap between the content of EQ-HWB and S-WEMWBS, with a high degree of convergence and similar response patterns observed across most items, primarily those conceptually related. This high degree of convergence lends support to both measures' EQ-HWB/EQ-HWB-S construct validity as a measure of MH and wellbeing [45]. Additionally, we found all five measures were able to discriminate between respondents with any MH condition, clinical depression, GAD, and other MH conditions. Nevertheless, the EQ-HWB psychosocial LSS and EQ-HWB-S appeared to outperform the S-WEMWBS, EQ-5D-5L/3L descriptive system in the ability to discriminate between known groups in MH. Our hypothesis that the measures will display construct validity in common MH problems was supported.

The results of the content analysis revealed that EQ-HWB/EQ-HWB-S address potentially relevant aspects of MH and wellbeing with domains covering concepts of Control, Concentration, Lonely and Anxious, in addition to EQ-HWB Hope Stigma, Self-worth. This suggests EQ-HWB has important additional content coverage for measuring MH and wellbeing, as compared to EQ-5D which is more focused on physical HRQL. Even though EQ-5D might lack domains specifically related to MH and wellbeing, it outperforms MH-specific measure S-WEMWBS in the ability to distinguish among patients with/without MH conditions. The KGV evidence for the psychometric validity of the EQ-5D in common MH problem patient samples is consistent with previous empirical work in mild depression and anxiety samples [46-48]. Indeed, in our study, EQ-5D 3L and 5L summary scores distinguished between patient groups with any MH, clinical depression, and GAD with large to very large ES, which is consistent with studies finding MH-related conditions can have a profound impact on overall health and well-being [49]. For example, previous research in clinical depression supported high convergent validity between two measures even

in cases they capture very different domains only minimally related [33], as we see with EQ-5D and S-WEMWBS.

One of the core characteristics of MH conditions is it affects all aspects of life [2]. Impairments and limitations in cognitive, emotional, and motivational function caused by MH problems may lead to disability and loss of QOL [50]. For example, clinical depression does not only cause feelings of sadness for a long period of time, losing energy and interest in doing things. Depression can actually change the ability to think, impairs attention and memory, as well as information processing and decision-making skills [51]. Previous research has suggested that none of the available QOL instruments used in people with MH problems fully cover seven dimensions previously found to be important in people with MH problems: well-being and ill-being; relationships and belonging; activity; self-perception; autonomy; hope and hopelessness; physical health [11]. EQ-HWB focusing on well-being along with QOL dimensions, presents here an additional advantage over more MH-focused scales, with the assessment of a wide range of relevant QOL dimensions in MH which seems crucial in capturing the heterogeneity of the MH problems.

Strengths and limitations

The limitation of the present study is that it was performed on a cross-sectional sample of US respondents, which restricts its application to this other groups of people around the world. Samples from online panels may be systematically different from those recruited through alternative modes of data collection. A well-known bias of US online panels is it tends to be disproportionately white and unrepresentative of minorities [52-54]. To mitigate all these potential for bias, we applied quotas on age, sex, ethnicity, and race to improve generalizability to the general US population. The lack of participants who are unable to self-report their QOL may contribute to issues of bias and equity, in the context of the evaluation of health and social care interventions. While these issues limit generalizability, this study was not intended to be strictly representative of the US population. Furthermore, the cross-sectional nature of this study limited our ability to explore the instruments' sensitivity to changes over time, which is an important aspect of the performance of measures. However, the ability to discriminate among respondents will likely translate into a relatively higher sensitivity and responsiveness to improvements on the HRQL continuum [19]. Finally, scoring of EQ-HWB subscales and EQ-HWB-S is done using the LSS approach under development while other scoring methods (such as the preference-based weights for the EQ-HWB-S) are still being developed. Nevertheless, previous research also argues that the differences between different MAUIs utilities are primarily attributable to differences in the instruments' descriptive systems [55], and the results of this study lend support to the proposed LSS scoring approach.

Conclusion

Initial evidence supports the validity of the EQ-HWB and EQ-HWB-S instruments as outcome measures in MH and well-being in the US population. All measures demonstrated discriminative ability, with the EQ-HWB psychosocial LSS and EQ-HWB-S tending to outperform both the S-WEMWBS and EQ-5D-5L/EQ-5D-3L. This highlights EQ-HWB's future potential as a measure of MH that encompasses a broader range of generic outcomes, including direct and indirect impacts on both, MH and wellbeing. Further research on the validity of EQ-HWB for use in various mental health conditions will aid users in selecting among HRQL measures for clinical trials, economic evaluation, and other applications.

Table 1. Patient Characteristics

Characteristic		General US population (n=904)	Any Mental Health Condition (n=172)	Clinical Depression (n=113)	Generalized Anxiety Disorder (n=98)	Other Mental Health Conditions (n=52)
		n (%)	n (%)	n (%)	n (%)	n (%)
Age	mean (SD)	53.8 (17.45)	46.22 (15.09)	46.42 (15.98)	44.34 (14.78)	39.92 (11.98)
Age Group	18-24	58 (6.45)	13 (7.6)	12 (10.62)	9 (9.18)	4 (7.84)
	25-34	110 (12.24)	32 (18.71)	20 (17.7)	21 (21.43)	17 (33.33)
	35-44	112 (12.46)	36 (21.05)	23 (20.35)	20 (20.41)	11 (21.57)
	45-54	124 (13.79)	31 (18.13)	15 (13.27)	19 (19.39)	11 (21.57)
	55-64	173 (19.24)	35 (20.47)	25 (22.12)	20 (20.41)	7 (13.73)
	65-74	239 (26.59)	21 (12.28)	16 (14.16)	7 (7.14)	1 (1.96)
	75-84	81 (9.01)	3 (1.75)	2 (1.77)	2 (2.04)	0 (0)
	85+	2 (0.22)	0 (0)	0 (0)	0 (0)	0 (0)
Gender	Male	463 (51.27)	59 (34.3)	37 (32.74)	30 (30.61)	18 (34.62)
	Female	436 (48.28)	112 (65.12)	75 (66.37)	68 (69.39)	34 (65.38)
	Other	4 (0.44)	1 (0.58)	1 (0.88)	0 (0)	0 (0)
Race	White	770 (85.27)	154 (89.53)	101 (89.38)	88 (89.8)	46 (88.46)
	Black	90 (9.97)	16 (9.3)	11 (9.73)	7 (7.14)	6 (11.54)
	Indian	5 (0.55)	1 (0.58)	1 (0.88)	1 (1.02)	0 (0)
	Native American	21 (2.33)	9 (5.23)	8 (7.08)	5 (5.1)	3 (5.77)
	Other Asian	18 (1.99)	2 (1.16)	2 (1.77)	1 (1.02)	0 (0)
	Other (not listed)	22 (2.44)	4 (2.33)	3 (2.65)	3 (3.06)	3 (5.77)
	Pacific Islander	3 (0.33)	0 (0)	0 (0)	0 (0)	0 (0)
Hispanic	Yes	113 (12.61)	29 (16.86)	17 (15.04)	17 (17.35)	7 (13.46)
Education	Did not finish high school	20 (2.21)	2 (1.16)	2 (1.77)	2 (2.04)	2 (3.85)
	Completed high school or equivalent	177 (19.6)	40 (23.26)	24 (21.24)	25 (25.51)	13 (25)
	Completed some college	308 (34.11)	67 (38.95)	47 (41.59)	40 (40.82)	23 (44.23)
	Completed a Bachelor's degree	229 (25.36)	44 (25.58)	27 (23.89)	24 (24.49)	10 (19.23)
	Completed a Professional or Graduate deg	169 (18.72)	19 (11.05)	13 (11.5)	7 (7.14)	4 (7.69)
Social care user	Yes	86 (9.54)	29 (16.86)	15 (13.27)	19 (19.39)	6 (11.54)
Disability	Yes	102 (11.3)	44 (25.58)	35 (30.97)	26 (26.53)	14 (26.92)
Carer	Yes	196 (21.88)	51 (29.82)	32 (28.57)	29 (29.9)	15 (28.85)

Hours spent as carer	1 to 19	92 (50.27)	17 (35.42)	13 (43.33)	9 (33.33)	4 (26.67)
	20 to 49	48 (26.23)	15 (31.25)	8 (26.67)	11 (40.74)	3 (20)
	50 or more	43 (23.5)	16 (33.33)	9 (30)	7 (25.93)	8 (53.33)
Any physical health problem	Yes	576 (63.79)	133 (77.33)	91 (80.53)	79 (80.61)	35 (67.31)
Cancer Diagnosis	Yes	400 (44.3)	61 (35.47)	39 (34.51)	33 (33.67)	13 (25)
Clinical Depression	Yes	113 (12.51)	113 (65.7)	113 (100)	61 (62.24)	27 (51.92)
Generalized Anxiety Disorder	Yes	98 (10.85)	98 (56.98)	61 (53.98)	98 (100)	25 (48.08)
Other Mental Conditon	Yes	52 (5.76)	52 (30.23)	27 (23.89)	25 (25.51)	52 (100)
HRQL and well-being						
EQ-HWB psychosocial LSS	mean (SD)	33.04 (14.33)	46.00 (15.44)	48.00 (15.88)	48.40 (15.43)	49.53 (16.12)
EQ-HWB pain/ discomfort LSS	mean (SD)	8.83 (3.63)	11.28 (3.84)	11.26 (3.85)	11.84 (3.67)	11.02 (4.02)
EQ-HWB activities LSS	mean (SD)	4.97 (2.50)	6.41 (2.76)	6.64 (2.72)	6.62 (2.67)	6.38 (2.91)
EQ-HWB-S LSS	mean (SD)	18.06 (7.71)	24.86 (8.22)	25.7 (8.24)	26.54 (8.07)	25.84 (8.83)
EQ-5D-5L index score	mean (SD)	0.76 (0.26)	0.56 (0.31)	0.55 (0.26)	0.51 (0.32)	0.55 (0.36)
EQ-5D-3L index score	mean (SD)	0.79 (0.19)	0.64 (0.21)	0.62 (0.21)	0.6 (0.22)	0.64 (0.25)
EQ VAS	mean (SD)	72.27 (19.38)	59.19 (23.22)	57.23 (22.7)	56.28 (23.22)	60 (25.44)
Short-WEMWBS	mean (SD)	24.07 (6.37)	20.05 (4.94)	19.49 (4.37)	19.31 (4.79)	19.19 (5.12)

HRQL = Health Related Quality of Life, SWEMWBS = Short Warwick Edinburgh Mental Well-being Scale, EQ-HWB = EQ Health and well-being, EQ-HWB-S = EQ Health and well-being short form.

EQ-HWB subscales and EQ-HWB-S have been scored as level summary score, EQ-5D-5L and EQ-5D-3L index score with utility value set for the US. More details on scoring in Methods. EQ-HWB and EQ-HWB-S higher score indicates worse HRQL, while S-WEMWBS, EQ-5D-5L, EQ-5D-3L and EQ VAS higher scores indicates better HRQL. EQ-HWB-psychosocial level summary score can range 16-80 with higher scores represent worse psychosocial health and well-being. EQ-HWB-pain/discomfort level summary score can range 4-20 with higher scores represent worse pain/discomfort health. EQ-HWB-activities level summary score can range 3-15 with higher scores represent worse activities health and well-being. EQ-HWB-S level summary score can range 9 – 45, with higher scores indicating worse HRQL. S-WEMWBS was scored using Rasch analysis scoring and can range 7 - 35 with higher scores indicatin better health and mental well-being. EQ-5D-5L index score (US value set) can range –0.573 (55555) to 1 (11111). EQ-5D-3L index score (US value set) can range –0.103 (33333) to 1 (11111). EQ VAS score is rated on a scale 0-100.
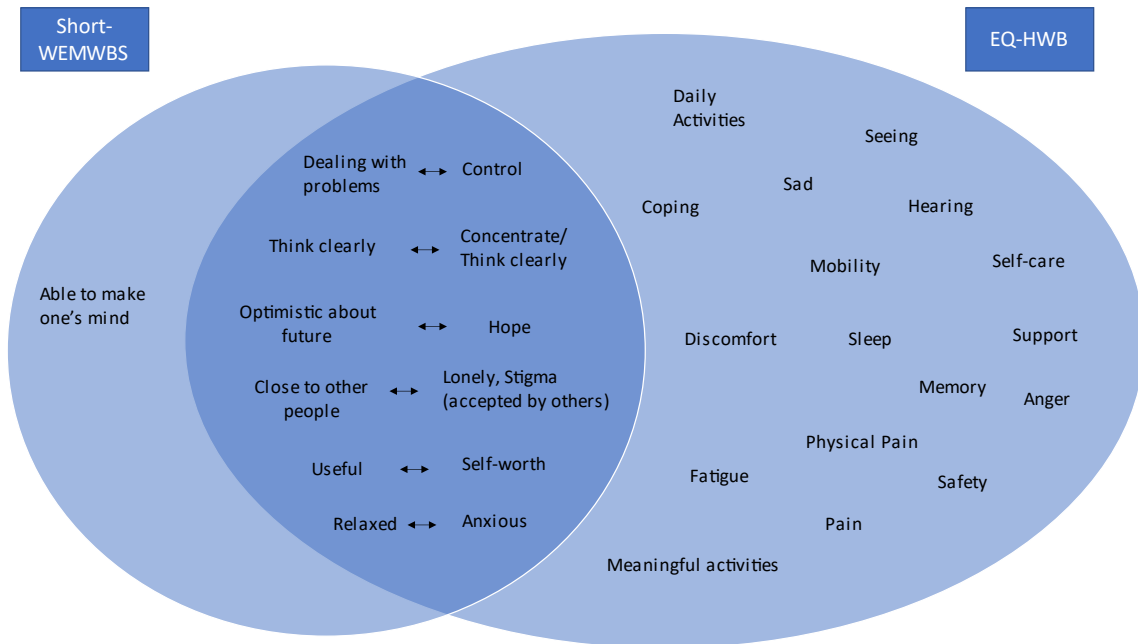
Figure 1.a. Item content overlap of the EQ-HWB and S-WEMWBS as measured by the Jaccard index.
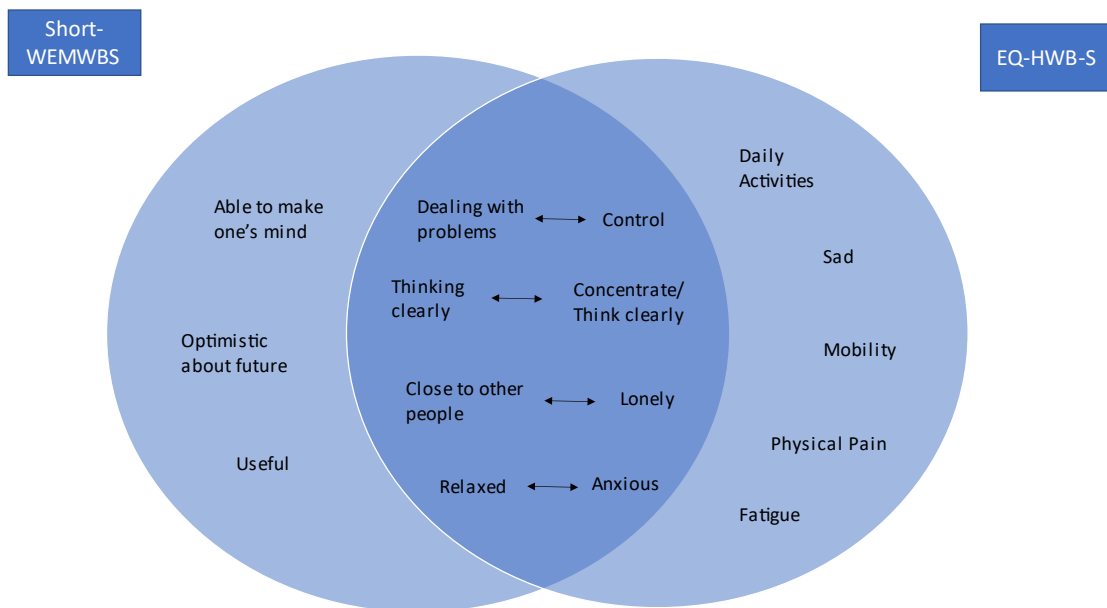
Figure 1.b. Item concent overlap of the EQ-HWB-S and S-WEMWBS as measured by the Jaccard index.
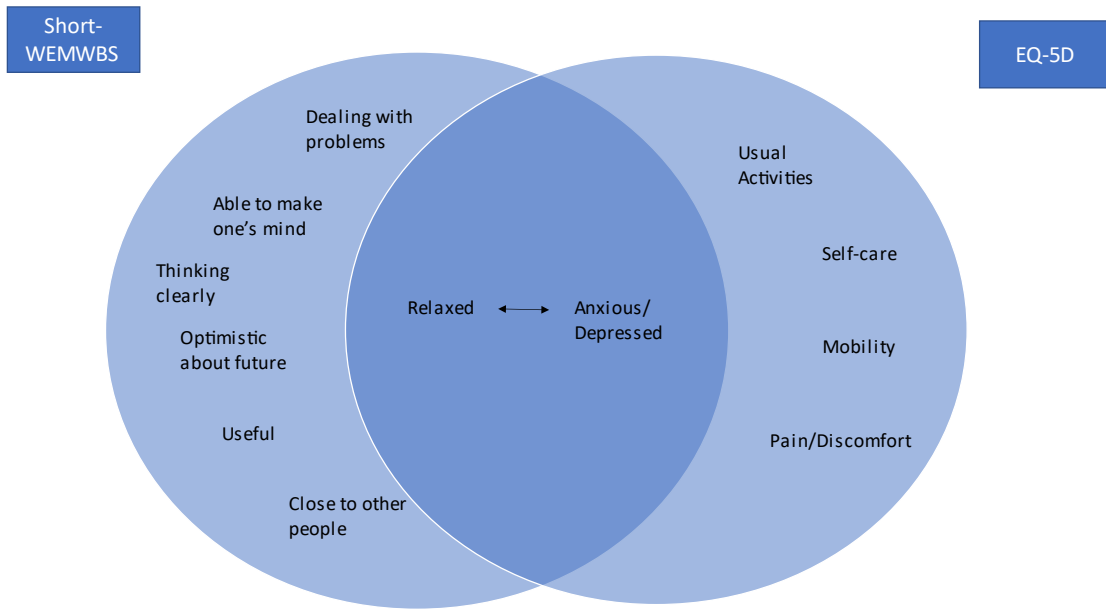
Figure 1.c. Item content overlap of the EQ-5D and S-WEMWBS as measured by the Jaccard index.

Note: The Jaccard index measures similarity between two sets of data with a range from 0 (no overlap) to 1 (total overlap), and is calculated as the number of shared items divided by the total number of items in any two PROMs.

*EQ-5D and S-WEMWBS Jaccard Index =1/11=0.09 * 100 = 9.09%*

*EQ-HWB-S and S-WEMWBS Jaccard Index = 4/12= 0.33 * 100 = 33.3%*

*EQ-HWB and S-WEMWBS Jaccard Index = 6/26= 0.23 * 100 = 23.07%*

Table 2.a. Response distributions, acceptability, floor and ceiling effects of the EQ-HWB, EQ-HWB-S, S-WEMWBS, EQ-5D-5L, EQ-5D-3L in self-reported mental health condition patients (n=172)

Domain	Abbreviation for Item	Item	%	%	%	%	%	%
EQ-HWB			No difficulty	Slight difficulty	Some difficulty	A lot of difficulty	Unable	Missing
Seeing	See	How well can you see (using your glasses or contact lenses if they are needed)?	46.51	27.33	19.77	5.81	0.58	0.00
Hearing	Hear	How well can you hear (using hearing aids if you normally wear them)?	65.7	19.19	8.14	6.4	0.58	0.00
Mobility and daily activity	Get around Inside/Outside	How well were you able to get around inside your home *	41.86	24.42	20.93	12.79	0	0.00
		How well were you able to get around outside your home *						0.00
	Day Activities	How well were you able to do your day to day activities	31.58	30.41	22.81	12.28	2.92	0.58
Personal needs/self-care	Difficult wash	How difficult is it for you to wash, toilet, dress yourself, eat or care for your appearance?	38.37	20.93	32.56	8.14		0.00
			None of the time	Only occasionally	Some of the time	Often	Most or all of the time	
Sleep	Sleep	I had problems with my sleep	11.11	16.96	19.88	25.73	26.32	0.58
Fatigue	Exhausted	I felt exhausted	10.65	16.57	18.34	27.22	27.22	1.74
Lonely	Lonely	I felt lonely	22.81	21.64	19.88	18.13	17.54	0.58
Support	Unsupported	I felt unsupported by other people	28.65	23.39	21.64	14.04	12.28	0.58
Memory	Remember	I had trouble remembering	12.87	26.9	24.56	25.15	10.53	0.58
Concentrate	Concentrate /Think Clearly	I found it hard to concentrate *	10.3	24.24	23.64	35.15	6.67	4.07
		I had trouble thinking clearly *						
Anxious	Anxious	I felt anxious	14.2	20.71	20.12	24.85	20.12	1.74
Safety	Unsafe	I felt unsafe	51.46	19.3	14.04	7.6	7.6	0.58
Anger	Frustrated	I felt frustrated	11.18	15.88	22.35	36.47	14.12	1.16
Happy	Sad	I felt sad	15.79	23.98	18.71	24.56	16.96	0.58
Hope	Look Forward	I felt that I had nothing to look forward to	23.39	23.39	21.05	19.88	12.28	0.58
Control	Control	I felt I had no control over my day to day life	26.9	24.56	22.22	15.79	10.53	0.58
Coping	Cope	I felt unable to cope with my day to day life	17.44	25	29.65	20.93	6.98	0.00
Stigma	Accepted	I felt accepted by others	19.77	25.58	26.16	18.02	10.47	0.00
Self-worth	Feel Good	I felt good about myself	16.86	18.6	29.07	20.35	15.12	0.00
Meaningful activities	Do Wanted	I could do the things I wanted to do	18.6	20.93	27.91	22.09	10.47	0.00

Pain	Pain (frequency)	How often did you experience physical pain	6.43	28.65	22.22	17.54	25.15	0.58
	Pain (severity)	I had no/ mild/ moderate/ severe/ very severe physical pain.	9.3	40.12	29.65	15.12	5.81	0.00
Discomfort	Discomfort (severity)	I had no/ mild/ moderate/ severe/ very severe physical discomfort	18.02	36.05	27.91	15.12	2.91	0.00
	Discomfort (frequency)	How often did you experience physical discomfort e.g. feeling sick, breathless, itching etc. (but not including pain)	15.88	27.65	23.53	19.41	13.53	0.00

Short-WEMWBS			All of the time	Often	Some of the time	Rarely	None of the time	Missing
	Thinking Clearly	I've been thinking clearly	10.53	30.41	34.5	18.71	5.85	0.58
	Feeling Useful	I've been feeling useful	9.3	19.77	36.05	27.33	7.56	0.00
	Dealing with Problems	I've been dealing with problems well	9.3	20.93	38.37	24.42	6.98	0.00
	Able to make up mind	I've been able to make up my own mind about things	25.58	25.58	29.07	12.21	7.56	0.00
	Feeling Optimistic	I've been feeling optimistic about the future	6.98	20.93	36.05	26.16	9.88	0.00
	Relaxed	I've been feeling relaxed	7.02	17.54	35.09	29.82	10.53	0.58
	Feeling Close	I've been feeling close to other people	8.72	19.77	34.88	25	11.63	0.00

EQ-5D-5L			No problems	Slight problems	Moderate problems	Severe problems	Extreme / Unable	Missing
		Mobility	46.2	22.22	24.56	5.26	1.75	0.58
		Self-care	64.71	22.94	11.18	1.18	0	1.16
		Usual Activities	37.65	31.76	18.82	10	1.76	1.16
		Anxiety/Depression	15.48	19.64	29.76	20.83	14.29	2.33
		Pain	15.98	33.14	32.54	11.83	6.51	1.74

EQ-5D-3L			No problems	Some problems	Extreme/ Unable			Missing
		Mobility	51.48	46.75	1.78			0.00
		Self-care	70.83	28.57	0.6			2.33
		Usual Activities	41.07	55.95	2.98			2.33
		Anxiety/Depression	14.88	57.74	27.38			2.33
		Pain	23.08	61.54	15.38			1.74
							AVERAGE	= 0.63

Table 2.b. Floor and ceiling effect of the EQ-HWB, EQ-HWB-S, S-WEMWBS, EQ-5D-5L, and EQ-5D-3L summary scores in self-reported mental health condition patients (n=172)

Summary score	Possible score range	Observed score range	Floor (%)	Ceiling (%)
EQ-HWB psychosocial LSS	16 – 80	18 – 80	0 %	1.92 %
EQ-HWB pain/discomfort LSS	4 – 20	4 – 20	2.34 %	1.76 %
EQ-HWB activities LSS	3 – 15	3 – 15	19.20 %	0.58 %
EQ-HWB-S LSS	9 – 45	10 – 42	0 %	0 %
EQ-5D-5L index score	−0.573 (55555) to 1 (11111)	(-0.44) - 1	0 %	2.99 %
EQ-5D-3L index score	−0.103 (33333) to 1 (11111)	0.04 - 1	0 %	4.22 %
EQ VAS	0 – 100	0 - 100	0.62 %	1.23 %
Short-WEMWBS	7 – 35	7 - 35	2.35 %	2.35 %

HRQL = Health-Related Quality of Life, S-WEMWBS = Short Warwick Edinburgh Mental Well-being Scale, EQ-HWB = EQ Health and well-being, EQ-HWB-S = EQ Health and well-being short form.

Note: Items in **bold** are part of EQ-HWB-S

* items have been combined to reflect subsequent EQ-HWB versions. More details on combining items in methods.

The floor is defined as the lowest possible scale score (%). Ceiling as highest possible scale score (%).

EQ-HWB, EQ-HWB-S, higher score indicates worse HRQL, while S-WEMWBS, EQ-5D-5L, EQ-5D-3L, and EQ VAS higher score indicates better HRQL. More details on scoring in methods.

Table 3.a. Correlation matrices between the EQ-HWB-S and Short-WEMWBS items

EW-HWB-S / S-WEMWBS	Inside/ Outside	Day activities	Exhausted	Lonely	Concentrate/ Think clearly	Anxious	Sad	No control	Physical pain (S)
I've been thinking clearly	0.16 *	0.29	0.3	0.34	0.55	0.38	0.45	0.31	0.27
I've been feeling useful	0.24 *	0.31	0.39	0.48	0.36	0.37	0.46	0.43	0.32
I've been dealing with problems well	0.23 *	0.31	0.34	0.44	0.38	0.45	0.45	0.43	0.24
I've been able to make up my own mind about things	0.32	0.32	0.24	0.38	0.41	0.27	0.37	0.36	0.19
I've been feeling optimistic about the future	0.23	0.29	0.4	0.46	0.33	0.38	0.46	0.43	0.34
I've been feeling relaxed	0.24 *	0.34	0.47	0.54	0.45	0.51	0.47	0.47	0.31
I've been feeling close to other people	0.22 *	0.26 *	0.36	0.58	0.41	0.4	0.45	0.41	0.3

Note, all p <.0001, except results with * p <.05 and results with ** = not statistically significant. S-WEMWBS items were reversed to reflect the direction of the EQ-HWB. (S) = severity

Table 3.b. Correlation matrices between the remaining EQ-HWB items not part of EQ-HWB-S and Short-WEMWBS items

EQ-HWB / S-WEMBWS	See	Hear	Difficult wash	Sleep	Unsupported	Remember	Unsafe	Frustrated	Look forward	Cope	Accepted	Feel good	Do wanted	Physical pain (F)	Physical discomfort (S)	Physical discomfort (F)
I've been thinking clearly	.15	.05	.19	.30	.35	.43	.37	.29	.38	.57	.42	.42	.45	.16	.25	.23
I've been feeling useful	.08	.05 *	.23	.25	.43	.26	.26	.41	.45	.58	.52	.62	.49	.23	.20	.19
I've been dealing with problems well	.02	-.05 *	.26	.24	.42	.27	.33	.41	.48	.58	.42	.48	.48	.21	.20	.17
I've been able to make up my own mind about things	.23	.19	.34	.12	.30	.31	.30	.31	.42	.50	.43	.39	.40	.06	.21	.21
I've been feeling optimistic about the future	.06	.02 **	.22	.25	.43	.28	.28	.46	.52	.55	.60	.66	.46	.21	.23	.15
I've been feeling relaxed	.15	-.11 *	.20	.34	.50	.31	.30	.53	.52	.53	.59	.58	.44	.23	.26	.25
I've been feeling close to other people	.13	-.02	.23	.28	.56	.28	.34	.40	.49	.46	.63	.57	.43	.13	.28	.21

Note, all p <.0001, except results with * p <.05, results with ** = not statistically significant. S-WEMWBS items were reversed to to reflect the direction of the EQ-HWB. S= severity, F= frequency.

3.c. Correlation matrices between the EQ-5D-5L and S-WEMWBS items

Short-WEMWBS	EQ5D5L					EQ5D3L				
	Mobility	Self-care	Usual activities	Pain	Depression/ Anxiety	Mobility	Self-care	Usual activities	Pain	Depression/ Anxiety
I've been thinking clearly	0.1 *	0.25	0.21	0.2	0.34	0.17	0.3	0.17	0.23	0.24
I've been feeling useful	0.24	0.26	0.27	0.26	0.39	0.22	0.21	0.18	0.3	0.34
I've been dealing with problems well	0.15 *	0.35	0.26	0.15	0.4	0.16 *	0.26	0.14 *	0.13 *	0.36
I've been able to make up my own mind about things	0.09 *	0.21	0.14	0.18	0.3	0.2	0.27	0.21	0.21	0.29
I've been feeling optimistic about the future	0.18	0.23	0.22	0.22	0.44	0.17	0.18	0.19	0.23	0.36
I've been feeling relaxed	0.19	0.26	0.29	0.27	0.47	0.17	0.18	0.18	0.31	0.45
I've been feeling close to other people	0.18	0.26	0.23	0.23	0.36	0.24	0.21	0.19	0.23	0.32

Note, all p <.05, except results with * = not statistically significant. S-WEMWBS items were reversed to to reflect the direction of the EQ-5D.

3.d. Correlation matrices between the S-WEMWBS summary score and EQ-HWB, EQ-HWB-S and EQ-5D.

	EW-HWB (psychosocial)	EQ-HWB (pain/discomfort)	EQ-HWB (activities)	EQ-HWB-S	EQ5D5L	EQ5D3L	EQ VAS	Short-WEMWBS
EW-HWB psychosocial subscale	1							
EQ-HWB pain/discomfort subscale	0.48	1						
EQ-HWB activities subscale	0.54	0.63	1					
EQ-HWB-S	0.94	0.61	0.7	1				
EQ-5D-5L	-0.56	-0.7	-0.76	-0.69	1			
EQ-5D-3L	-0.61	-0.63	-0.67	-0.67	0.8	1		
EQ VAS	-0.57	-0.55	-0.55	-0.57	0.67	0.6	1	
Short-WEMWBS	-0.7	-0.32	-0.35	-0.62	0.4	0.45	0.45	1

Note, all p <.001

	0 – 0.10 none
	0.11 – 0.30 weak
	0.31 – 0.50 moderate
	0.51≥ strong

Table 4: Comparison of measures based on known groups comparisons for self-reported mental health conditions

Variable for KGV Instrument		Any mental health problem (Yes/No)		Clinical depression (Yes/No)		Generalized anxiety disorder (Yes/No)		Other mental condition (Yes/No)	
			95% CL		95% CL		95% CL		95% CL
EQ-HWB-S	ES	1.33	1.10-1.55	1.32	1.05-1.59	1.41	1.09-1.68	1.16	0.76-1.53
EQ-HWB-psychosocial	ES	1.34	1.09-1.58	1.37	1.08-1.65	1.37	1.09-1.67	1.3	0.89-1.69
EQ-HWB-pain/discomfort	ES	1.01	0.8-1.21	0.85	0.61-1.09	1.05	0.78-1.3	0.79	0.44-1.15
EQ-HWB-activities	ES	0.76	0.55-1.00	0.80	0.56-1.08	0.86	0.58-1.14	0.62	0.20-0.99
EQ-5D-5L	ES	1.19	1.44-0.93	1.11	1.39-0.81	1.33	1.64-0.99	0.92	1.39-0.45
EQ-5D-3L	ES	1.2	1.47-0.96	1.14	1.41-0.87	1.27	1.57-0.98	0.91	1.34-0.48
S-WEMWBS	ES	0.91	1.08-0.74	0.92	1.09-0.73	0.94	1.14-0.74	0.91	1.18-0.64
EQ-HWB-S	F-value	196.15		141.57		146.19		48.22	
EQ-HWB-psychosocial	F-value	204.25		152.45		139.83		61.12	
EQ-HWB-pain/discomfort	F-value	113.42		58.55		81.27		22.45	
EQ-HWB-activities	F-value	65.19		51.5		54.43		13.82	
EQ-5D-5L	F-value	158.59		100.36		129.55		30.65	
EQ-5D-3L	F-value	160.3		105.57		118.75		30.18	
S-WEMWBS	F-value	92.1		68.09		65.52		30.06	
EQ-HWB-S/ S-WEMWBS	F-ratio	2.13	1.51-3.00	2.08	1.48-2.92	2.23	1.53-3.28	1.60	0.80-2.92
EQ-HWB (psychosocial)/ S-WEMWBS	F-ratio	2.22	1.61-2.98	2.24	1.64-3.02	2.13	1.51-2.99	2.03	1.07-3.55
EQ-5D-5L/ S-WEMWBS	F-ratio	1.72	1.1-2.72	1.47	0.88-2.44	1.98	1.19-3.26	1.02	0.27-2.60
EQ-5D-3L/ S-WEMWBS	F-ratio	1.74	1.13-2.69	1.55	0.96-2.47	1.81	1.12-2.91	1.00	0.32-2.36

ES - Cohen's d effect size

	0 – 0.19 none
	0.20 – 0.49 small
	0.50 – 0.79 medium
	0.80 – 1.19 large
	1.20 ≥ very large

References

1. Organization, W.H., *Mental health: strengthening our response.* 2018.
2. Administration., S.A.a.M.H.S. *Living well with serious mental illness. (Last updated: 03/03/2022) Retrieved from https://www.samhsa.gov/serious-mental-illness (05/30/2022).*
3. Health, N.I.o.M. *Mental Illness.* 2022 June 11,2022]; Available from: https://www.nimh.nih.gov/health/statistics/mental-illness.
4. Patel, V., et al., *The Lancet Commission on global mental health and sustainable development.* The Lancet, 2018. **392**(10157): p. 1553-1598.
5. Hewlett, E. and K. Horner, *Mental health analysis profiles (MhAPs): England.* 2015.
6. *Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019.* Lancet Psychiatry, 2022. **9**(2): p. 137-150.
7. Connell, J., A. O'Cathain, and J. Brazier, *Measuring quality of life in mental health: are we asking the right questions?* Soc Sci Med, 2014. **120**: p. 12-20.
8. Ruddick, F., *Promoting mental health and wellbeing.* Nursing Standard (through 2013), 2013. **27**(24): p. 35.
9. Brazier, J., *Is the EQ–5D fit for purpose in mental health?* The British Journal of Psychiatry, 2010. **197**(5): p. 348-349.
10. Papaioannou, D., J. Brazier, and G. Parry, *How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in schizophrenia? A systematic review.* Value in Health, 2011. **14**(6): p. 907-920.
11. van Krugten, F.C.W., et al., *Instruments to assess quality of life in people with mental health problems: a systematic review and dimension analysis of generic, domain- and disease-specific instruments.* Health Qual Life Outcomes, 2021. **19**(1): p. 249.
12. Saarni, S.I., et al., *Quality of life of people with schizophrenia, bipolar disorder and other psychotic disorders.* The British Journal of Psychiatry, 2010. **197**(5): p. 386-394.
13. Lamers, L., et al., *Comparison of EQ-5D and SF-6D utilities in mental health patients.* Health economics, 2006. **15**(11): p. 1229-1236.
14. Mulhern, B., et al., *Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D.* The British Journal of Psychiatry, 2014. **205**(3): p. 236-243.
15. Brazier, J., et al., *A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures.* Health Technol Assess, 2014. **18**(34): p. vii-viii, xiii-xxv, 1-188.
16. Makai, P., et al., *Quality of life instruments for economic evaluations in health and social care for older people: a systematic review.* Social science & medicine, 2014. **102**: p. 83-93.
17. Brazier, J., et al., *The EQ Health and Wellbeing: Overview of the Development of a Measure of Health and Wellbeing and Key Results.* Value Health, 2022.
18. Peasgood, T., et al., *The psychometric testing of potential items for the HWB .* Value in Health, Themed Issue on EQ-HWB, 2021.
19. Monteiro, A.L., M. Kuharic, and A.S. Pickard, *A Comparison of a Preliminary Version of the EQ-HWB Short and the 5-Level Version EQ-5D.* Value Health, 2022. **25**(4): p. 534-543.
20. Gudex, C., *The descriptive system of the EuroQOL instrument,* in *EQ-5D concepts and methods: a developmental history.* 2005, Springer. p. 19-27.

21. Rabin, R., M. Oemar, and M. Oppe, *EQ-5D-3L User Guide Basic Information on How to Use the EQ-5D-3L Instrument.* Rotterdam: EuroQol Group, 2011.
22. Shaw, J.W., J.A. Johnson, and S.J. Coons, *US valuation of the EQ-5D health states: development and testing of the D1 valuation model.* Med Care, 2005. **43**(3): p. 203-20.
23. Pickard, A.S., et al., *United States Valuation of EQ-5D-5L Health States Using an International Protocol.* Value Health, 2019. **22**(8): p. 931-941.
24. You-Shan Feng, T.K., Tessa Peasgood, Lidia Engel, Brendan Mulher, John Brazier, A. Simon Pickard, *Scoring the EQ-HWB and the EQ-HWB-S: Can We Do It Without Value Sets? A Non Parametric Item Response Theory Analysis of Three Datasets*, in *EuroQol Plenery 2021.* 2021.
25. Tennant, R., et al., *The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation.* Health Qual Life Outcomes, 2007. **5**: p. 63.
26. Stewart-Brown, S., et al., *Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey.* Health and Quality of Life Outcomes, 2009. **7**(1): p. 15.
27. Stewart-Brown, S., et al., *Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey.* Health Qual Life Outcomes, 2009. **7**: p. 15.
28. Terwee, C.B., et al., *Quality criteria were proposed for measurement properties of health status questionnaires.* J Clin Epidemiol, 2007. **60**(1): p. 34-42.
29. Gagnier, J.J., et al., *COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures.* Qual Life Res, 2021. **30**(8): p. 2197-2218.
30. Cappelleri, J., et al., *Psychometric analysis of the Three-Factor Eating Questionnaire-R21: results from a large diverse sample of obese and non-obese participants.* International journal of obesity, 2009. **33**(6): p. 611-620.
31. Elo, S. and H. Kyngäs, *The qualitative content analysis process.* Journal of advanced nursing, 2008. **62**(1): p. 107-115.
32. Skogestad, I.J., et al., *Lack of content overlap and essential dimensions – A review of measures used for post-stroke fatigue.* Journal of Psychosomatic Research, 2019. **124**: p. 109759.
33. Fried, E.I., *The 52 symptoms of major depression: Lack of content overlap among seven common depression scales.* Journal of Affective Disorders, 2017. **208**: p. 191-197.
34. Levandowsky, M. and D. Winter, *Distance between sets.* Nature, 1971. **234**(5323): p. 34-35.
35. Cohen, J., *A power primer.* Psychol Bull, 1992. **112**(1): p. 155-9.
36. Fayers, P.M. and D. Machin, *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. 2013: John Wiley & Sons.
37. Hansson, L., *Quality of life in depression and anxiety.* International Review of Psychiatry, 2002. **14**(3): p. 185-189.
38. Fritz, C.O., P.E. Morris, and J.J. Richler, *Effect size estimates: current use, calculations, and interpretation.* Journal of experimental psychology: General, 2012. **141**(1): p. 2.
39. Kirby, K.N. and D. Gerlanc, *BootES: An R package for bootstrap confidence intervals on effect sizes.* Behavior Research Methods, 2013. **45**(4): p. 905-927.
40. Jiang, S., C. Wang, and D.J. Weiss, *Sample Size Requirements for Estimation of Item Parameters in the Multidimensional Graded Response Model.* Front Psychol, 2016. **7**: p. 109.
41. Fayers PM, M.D.Q.o.l.a., analysis and interpretation. Chichester: Wiley; 2000.
42. Albert, P.R., *Why is depression more prevalent in women?* J Psychiatry Neurosci, 2015. **40**(4): p. 219-21.
43. Vesga-López, O., et al., *Gender differences in generalized anxiety disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC).* J Clin Psychiatry, 2008. **69**(10): p. 1606-16.

44. Barker, M.M., et al., *Prevalence and Incidence of Anxiety and Depression Among Children, Adolescents, and Young Adults With Life-Limiting Conditions: A Systematic Review and Meta-analysis.* JAMA Pediatr, 2019. **173**(9): p. 835-844.

45. Streiner, D.L., G.R. Norman, and J. Cairney, *Health measurement scales: a practical guide to their development and use*. 2015: Oxford University Press, USA.

46. Mulhern, B., et al., *Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D.* Br J Psychiatry, 2014. **205**(3): p. 236-43.

47. Brazier, J., *Measuring and valuing mental health for use in economic evaluation.* J Health Serv Res Policy, 2008. **13 Suppl 3**: p. 70-5.

48. Lamers, L.M., et al., *Comparison of EQ-5D and SF-6D utilities in mental health patients.* Health Econ, 2006. **15**(11): p. 1229-36.

49. Nouri, F., et al., *How different domains of quality of life are associated with latent dimensions of mental health measured by GHQ-12.* Health and Quality of Life Outcomes, 2021. **19**(1): p. 255.

50. Investigators, E.M., et al., *Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project.* Acta psychiatrica scandinavica, 2004. **109**: p. 21-27.

51. Shilyansky, C., et al., *Effect of antidepressant treatment on cognitive impairments associated with depression: a randomised longitudinal study.* Lancet Psychiatry, 2016. **3**(5): p. 425-35.

52. Kennedy, C., et al., *Evaluating online nonprobability surveys.* Pew Research, 2016.

53. Duffy, B., et al., *Comparing data from online and face-to-face surveys.* International Journal of Market Research, 2005. **47**(6): p. 615-639.

54. Baker, R., et al., *Research synthesis: AAPOR report on online panels.* Public Opinion Quarterly, 2010. **74**(4): p. 711-781.

55. Richardson, J., A. Iezzi, and M.A. Khan, *Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and 'micro-utility' effects.* Quality of Life Research, 2015. **24**(8): p. 2045-2053.