VALUE IN HEALTH 21 (2018) 432–440

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Assessment of Reliability and Validity of SF-12v2 among a Diabetic Population

Niranjan Kathe, MS[1], *Corey J. Hayes, PharmD, MPH*[1,2], *Naleen Raj Bhandari, MS*[1], *Nalin Payakachat, PhD*[1,*]

[1]*Division of Pharmaceutical Evaluation and Policy, University of Arkansas for Medical Sciences College of Pharmacy, Little Rock, AR, USA;* [2]*Division of Health Services Research, Psychiatric Research Institute, University of Arkansas for Medical Sciences, Little Rock, AR, USA*

ABSTRACT

Objectives: To validate the Medical Outcomes Study Short Form version 2 (SF-12v2) in diabetic patients. **Methods:** Adults with self-reported diabetes from the Medical Expenditure Panel Survey (2011–2013) were identified. Reliability (internal consistency and test-retest) and validity (construct, concurrent, criterion, and predictive) of the SF-12v2 were assessed. The SF-12v2 consists of two normalized composite scores: the physical component summary score (PCS12) and the mental component summary score (MCS12). Confirmatory factor analysis was conducted to assess the instrument structure. Concurrent (convergent and discriminant) validity was assessed by a multitrait-multimethod matrix using the Patient Health Questionnaire, the Kessler Scale, and perceived health and mental health questions. The predictive validity was assessed by estimating future limitations. The concurrent validity was tested by comparing the MCS12, PCS12, and utility scores (six-dimensional health state short form) across comorbidity scores. **Results:** The final sample comprised 2214 diabetic patients with mean normalized (population mean 50; range 0–100) PCS12 and MCS12 scores of 40.81 (standard error 0.33) and 49.82 (standard error 0.26), respectively. The PCS12 and MCS12 scores showed good internal consistency (Cronbach α: PCS12 0.85; MCS12 0.83) and acceptable test-retest reliability (intraclass correlation coefficient: PCS12 0.72; MCS12 0.63) and produced acceptable goodness-of-fit indices (normed fit index 0.95; comparative fit index 0.95; root mean square error of approximation 0.11 [95% confidence interval 0.1017–0.1188]). The PCS12 and MCS12 were moderately correlated with perceived health and perceived mental health. The MCS12 was highly correlated with the Patient Health Questionnaire and the Kessler Scale. Both the PCS12 and the MCS12 could predict the future health limitations. The PCS12, MCS12, and utility scores demonstrated sensitivity to the presence of comorbidity scores. **Conclusions:** The SF-12v2 is a valid generic instrument for measuring quality of life in diabetic patients.

Keywords: confirmatory factor analysis, diabetes, SF-6D, SF-12v2, validation.

Introduction

Diabetes mellitus is a chronic illness affecting about 10% of the US population [1]. Diabetic patients are prone to developing other comorbidities, namely, heart diseases, dyslipidemia, depression, and co-complications such as diabetic nephropathy, diabetic retinopathy, glaucoma, diabetic foot ulcers, and amputations [2]. The comorbidity burden and chronic nature of the disease lead to poor health-related quality of life (HRQOL) [3]. In addition to clinical end points, secondary end points such as HRQOL are vitally important in the management of diabetes to measure overall improvements in well-being. Studies assessing the impact of comorbidities and functional impairment associated with diabetes on HRQOL have used generic HRQOL instruments [4,5]. The Medical Outcomes Study Short Form version 2 (SF-12v2) is one such widely used generic HRQOL instrument. The instrument generates two summary scores including the physical component summary score (PCS12) and the mental component summary score (MCS12).

The SF-12v2 has been tested for reliability and validity in a general population during its development stage by Ware et al. [6]. The instrument demonstrated high internal consistency, test-retest reliability, construct validity, and criterion validity, which were consistent with previous findings using the short form 36 health survey (SF-36) [6]. Cheak-Zamora et al. [7] also reported similar results using a nationally representative database. Nevertheless, there were a few disagreements in the results, such as poor concurrence of the MCS12 with mental health measures.

Preliminary results from the study were presented as a podium presentation at Southern Pharmacy Administration Conference, 2016, Oxford, MS. The abstract for the presentation was published in *Research in Social and Administrative Pharmacy* Volume 12, Issue 4, July–August 2016, pages e11–e12, ISSN 1551-7411, http://dx.doi.org/10.1016/j.sapharm.2016.05.027.

Conflicts of interest: C. J. Hayes was supported by the Translational Training in Addiction (grant no. 1T32 DA 022981). No conflicts of interest exist for any of the authors.

* *Address correspondence to:* Nalin Payakachat, Division of Pharmaceutical Evaluation and Policy, University of Arkansas for Medical Sciences College of Pharmacy, 4301 West Markham, Slot 522, Little Rock, AR 72205.

E-mail: npayakachat@uams.edu

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2017.09.007

Other studies have evaluated the SF-12v2 in different disease populations [8–11] and found results similar to those reported by Ware et al. [6] and Cheak-Zamora et al. [7]. Effective use of the instrument in specific disease populations requires thorough psychometric evaluation, which the present studies lack.

Despite HRQOL being an important end point in the evaluation of interventions for diabetic patients, the SF-12v2 has not been validated in a nationally representative diabetic population. Therefore, the aim of this study was to assess the reliability (test-retest and internal consistency), structural validity, construct validity, and criterion validity of the SF-12v2 among diabetic patients. On the basis of previous studies by Ware et al. [6] and Cheak-Zamora et al. [7], this study hypothesizes that the SF-12v2 would demonstrate adequate reliability and validity in a diabetic population.

Research Design and Methods

Data Source

This study used a retrospective longitudinal cohort design as well as data from the publicly available survey data, the Medical Expenditure Panel Survey (MEPS). The MEPS is a nationwide representative survey of noninstitutionalized civilians, families, and providers [12,13]. The survey gathers data on the experience, use, and costs of various health care services. The MEPS consists of medical, household, and insurance components. The household component is a subsample of the previous year's National Health Interview Survey population. The survey collects information at the individual and household levels. It involves a panel design, collecting information through interviews over 2 years in five rounds using computer-assisted telephone interviewing, which is administered to one of the interviewing members of the household who records the responses on behalf of all the household members taking part in the survey. An additional self-administered questionnaire (SAQ) is mailed separately in rounds 2 and 4 of the panel for each eligible member of the household to complete individually. The major components of the SAQ include the SF-12v2, the Patient Health Questionnaire (PHQ-2), and the Kessler Scale (K-6) [14,15]. Other important components of the interview included the diabetes-specific survey questionnaire; questions on cognitive limitations, social limitations, perceived health, perceived mental health, and chronic conditions; and demographic information. Longitudinal files of panels 16 and 17 were used for this study, which encompass the years from 2011 to 2013 and include 36,435 study participants.

Study Sample

Only adults 18 years or older and having a self-reported diabetes diagnosis were included in the study. The sample was required to be within the scope for all five rounds and eligible for the SAQ in rounds 2 and 4. It was also required to have nonmissing responses for all the SF-12v2 items in these two rounds. The study sample identification is shown in Figure 1.

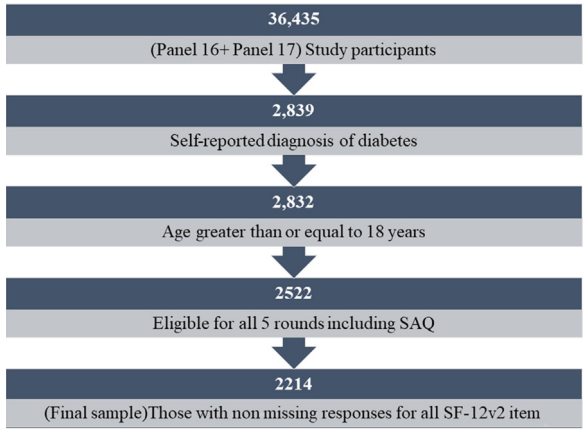

Fig. 1 – Selection process for the final analytical sample. SAQ, self-administered questionnaire; SF-12v2, Medical Outcomes Study Short Form version 2.

Measures

Medical Outcomes Study Short Form version 2

The SF-12v2 is a shortened version of the SF-36 [16,17]. Similar to the SF-36, it consists of items that evaluate eight subscales pertaining to HRQOL, namely, role limitations due to physical restrictions (RP), physical functioning (PF), bodily pain (BP), general health compared with others (GH), vitality (VT), social functioning (SF), role limitations due to emotional issues (RE), and mental health (MH) [17–19]. The responses for questions on GH, BP, and MH were reverse-coded to correspond with the direction of the summary scores. The scores from items on PF, RP, RE, and MH were combined and transformed to a Z score (range of 0–100) to create respective scales using methods from previous studies [6]. The PCS12 was calculated by combining and normalizing RP, GH, BP, and PF scales, whereas the MCS12 was calculated by combining and normalizing RE, MH, RE, SF, and VT scales. Normalized summary scores are generated from the items in each component. The scores range from 0 to 100, where higher scores indicate better health.

Six-dimensional health state short form

The six-dimensional health state short form (SF-6D) score is a preference-based single utility score. The SF-6D utility scores can be estimated using PCS12, MCS12, and age- and sex-specific norms [20]. The SF-6D utility scores calculated as such were used to estimate the disutility associated with diabetes-related comorbidities, namely, diabetes with complications, cardiac diseases, diabetes, and kidney and eye problems caused by diabetes.

Comorbidity Scores

We used the clinical classification codes from the medical condition file in the MEPS and condition-specific weights to calculate the PCS12- and MCS12-specific comorbidity indices [21]. The health-related quality of life comorbidity index (HRQOL-CI) was used for testing the concurrent validity. For estimating concurrent validity, the comorbidity scores were categorized on the basis of the distribution of comorbidity scores obtained for our final sample population. Scores for the PCS12 were grouped into the following score categories: 0, 1 to 2, 3 to 4, 5 to 7, and 8 or more, whereas scores for the MCS12 were grouped into the following score categories: 1, 2, 3, and 4 or more.

Diabetes-Related Comorbidities

The MEPS priority condition questions on cardiac diseases (namely, myocardial infarction, stroke, coronary heart disease, and other heart conditions) and arthritis were used to record the presence of the respective disease. Responses from the diabetes care survey were used to record the presence of diabetes-related kidney and eye problems. Furthermore, the medical conditions file was used to identify patients with diabetes-related complications using Clinical Diagnosis Classification code 50. Binary (yes/no) variables were created for each comorbidity. The diabetes-related comorbidities were used in

concurrent validity to measure SF-6D utility scores among different disease subpopulations.

Other Validating Measures

Perceived health and perceived mental health

For evaluation of convergent validity, we used two questions from the survey that asked the interviewing member the following: "In general, compared with other people of the same age, would you say your (health)/mental health is excellent, very good, good, fair, or poor?" Responses were reverse-coded to align with the directionality of the SF-12v2. The responses from rounds 2 and 4 were used in our study because it was concurrently administered with the SF-12v2. The perceived health and the perceived mental health were used to obtain stable populations for test-retest reliability.

Patient Health Questionnaire

The PHQ-2 score was generated from two questions that recorded incidences of depression and anhedonia with a 2-week recall period. Patients were asked in rounds 2 and 4 "Over the past 2 weeks, how often have you been bothered by any of the following problems: little interest or pleasure in doing things/feeling down, depressed, or hopeless?" The scores ranged from 0 to 6, with higher scores indicating higher depression severity. The PHQ-2 score was used to test the construct validity.

The Kessler Scale

The survey asked the members "During the past 30 days, about how often did you feel nervous, hopeless, restless or fidgety, so sad that nothing could cheer you up, that everything was an effort/worthless?" The sample participants were asked to respond on a scale from 0 to 4 for each item, 0 being "none of the times" and 4 being "all of the time." The scale score ranged from 0 to 24, with higher scores indicating higher depression severity. The K-6 score was also used to test the construct validity.

Health limitations

On a binary response scale (yes/no), the survey also asked the sample participants to record any limitations they experienced in the previous 3 months. The interviewing member of the household was asked "Is anyone in the family limited in any way in the ability to work at a job/do housework or go to school because of an impairment or a physical or mental health problem?" The responses were recorded for all the household members in the panel. The health limitation variable was used to assess the predictive ability of the PCS12 and the MCS12.

Social and cognitive limitations

On a binary response scale (yes/no), the sample participants were asked to record any social or cognitive limitations experienced by the household members in the previous 3 months. The corresponding responses were collected for all the members of the household participating in the survey.

Demographic characteristics and other variables

The demographic characteristics used were age, years since receiving a diagnosis of diabetes, sex, race/ethnicity, marital status, employment status, region, education in years, and insurance status. The number of years since diabetes diagnosis was defined as the difference between the age at diabetes diagnosis and the age during the first round of the interview.

Statistical Analysis

Univariate analysis was used to explore the PCS12 and MCS12 scores of the diabetic population. Flooring and ceiling effects were evaluated by determining the percentage of sample participants with the highest and the lowest extreme scores [22]. Means and SDs were calculated for continuous variables, and frequencies and percentages were reported for categorical variables. In addition to reporting the sample mean, we also derived the national estimates using longitudinal weights, adjusting for the complex sampling design of the MEPS.

To assess the measures for reliability and validity, the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist guidelines were reviewed. The COSMIN checklist is an international collaborative effort to summarize the necessary measurements and analyses required for establishing the reliability and validity of an instrument. On the basis of the COSMIN checklist and previous research by Cheak-Zamora et al. [7], reliability and validity were estimated as shown in Figure 2.

Reliability

Internal consistency reliability is a measure of how well each item relates to other items in the scale and also to the scale score [23]. The subscales GH, RP, PF, and BP were tested for correlation with the PCS12, and MH, RP, SF, VT, and RE were tested for their correlation with the MCS12. The internal consistency was measured using Cronbach α; a cutoff score of 0.7 or more represents acceptable internal consistency reliability.

Furthermore, test-retest reliability was assessed, which estimates the stability of responses over time. Test-retest reliability was evaluated using the intraclass correlation coefficient (ICC) between the PCS12/MCS12 score obtained from round 2 (baseline) and that from round 4 (follow-up). Test-retest reliability was assessed only among those sample participants with stable (same) responses to perceived health and perceived mental health at baseline and follow-up.

Validity

The COSMIN checklist recommends testing patient-reported instruments on three key areas of validity: content, construct, and criterion. This study included the components in accordance with the checklist. Nevertheless, because the SF-12v2 is an established instrument measuring the content/face, validity testing of the instrument was not conducted. Cross-cultural validity was not applicable because the instrument was in English; also, responsiveness was not included because no particular event or intervention was conducted between the two measurements.

Construct validity

As per the COSMIN recommendation, the construct validity included convergent and discriminant validity (hypothesis testing) and confirmatory factor analysis (structural validity).

Convergent and discriminant validity. Construct validity, which consists of both convergent validity and discriminant validity, tests how well the instrument measures the intended construct. For the purpose of testing construct validity, the instrument is correlated with other measures quantifying a similar construct. In our case, the PCS12 and MCS12 scores were tested with responses from the PHQ-2 and K-6 questions on perceived health and perceived mental health from the same round. The Spearman rank correlation coefficients were used to assess the convergent and discriminant validity because of the ordinal nature of certain items. The PCS12 was expected to be convergent and well

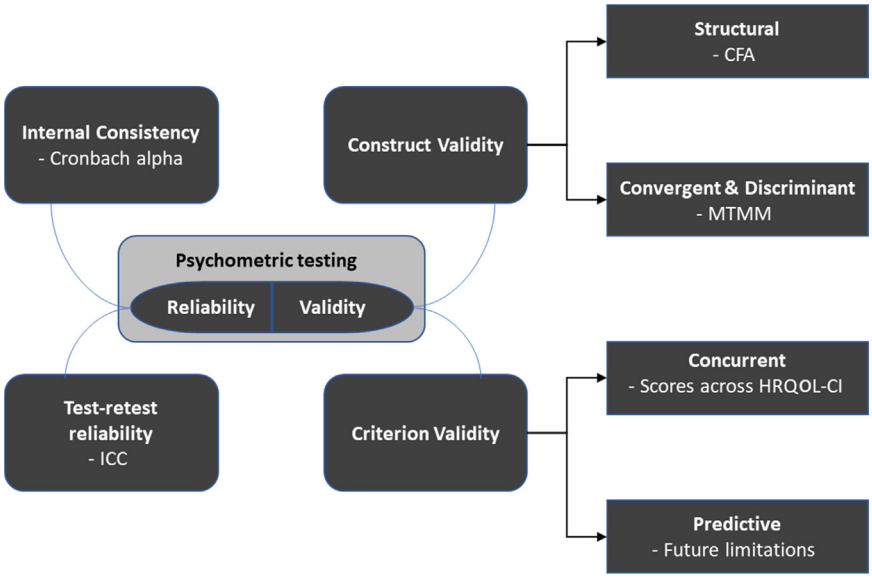

Fig. 2 – Overview for psychometric evaluation of the SF-12v2. The reliability domain of psychometric testing involved internal consistency measured using Cronbach α and test-retest reliability for stable diabetic patients (having the same response to perceived health questions) measured using the ICC. The validity domain of psychometric testing included construct and criterion validity. Concordance with the theoretical structure was used to establish structural validity. Convergent and discriminant validity were assessed by assessing correlation with other measures. Predictive and concurrent validity were measured using known-group comparison. Future limitations refer to social, cognitive, physical, and work limitations from round 3. CFA, confirmatory factor analysis; HRQOL-CI, health-related quality of life comorbidity index; ICC, intraclass correlation coefficient, K-6, Kessler Severity of Depressive Symptom score; MTMM, multitrait multimethod matrix (measures used are K-6, PHQ-2, perceived physical health, and perceived mental health); PHQ-2, Patient Health Questionnaire; SF-12v2, Medical Outcomes Study Short Form version 2.

correlated ($r > 0.5$) with perceived health, whereas it was expected to be divergent and poorly correlated ($r < 0.4$) with the PHQ-2 and K-6 scores on perceived mental health. Similarly, the MCS12 was expected to be convergent and well correlated with perceived mental health and the PHQ-2 and K-6 scores, whereas it was expected to be divergent and poorly correlated with perceived health [7].

Confirmatory factor analysis. In addition to reliability and validity, this study tested the instrument structure of the SF-12v2. The SF-12v2 and the SF-36 share a great deal of similarities in their instrument structure. Both scales yield PCS12 and MCS12 composite scores, representing physical and mental health, respectively. To confirm the consistency of the instrument structure in a diabetic population, confirmatory factor analysis (CFA) was performed using a two-factor model [6,24]. Compared with exploratory factor analysis, CFA tests the theoretical structure of the instrument and determines how well the instrument model fits the empirical data. RP, PF, and BP were theorized to load on PCS12, whereas RE, SF, and MH were theorized to load on MCS12 [6,25,26]. Domains of vitality and general health were allowed to load on both the components on the basis of previous studies [27]. Also, the two factors were proposed to be correlated [27]. Goodness-of-fit indices included the comparative fit index (CFI), goodness of fit (GFI), standardized root mean square (SRMR), and normed fit index (NFI). CFA, GFI, and CFI values more than 0.9 are considered acceptable, whereas SRMR values less than 0.08 are considered acceptable [28].

Criterion validity

Criterion validity estimates the extent to which the instrument correlates to a relevant outcome. It is further divided into predictive and concurrent validity. Concurrent validity correlates current outcomes, whereas predictive validity helps predict future outcomes on the basis of the current measure.

Concurrent validity. To establish concurrent validity, the PCS12 and MCS12 scores were analyzed for people with differing levels of HRQOL-CI. The PCS12 and MCS12 comorbidity scores were compared with their respective component scores. The mean PCS12 scores were compared across PCS comorbidity score categories 0, 1 to 2, 3 to 4, 5 to 7, and 8 or more, whereas the mean MCS12 scores were compared across MCS comorbidity score categories 0, 1, 2, 3, and 4 or more, using one-way analysis of variance. The categories were chosen on the basis of the distribution of scores in the study sample. In addition, the pairwise Tukey test assessed the ability of the PCS12 and the MCS12 to distinguish between people with respective component comorbidity scores of 0 and greater than 0.

The effects of different comorbidities on the SF-6D utility scores were also assessed. The SF-6D scores were compared between those with and without comorbidities. A separate regression analysis was conducted and mean scores were compared with the mean scores for each comorbidity, including diabetes complications, cardiac diseases, and kidney and eye problems associated with diabetes.

Predictive validity. We used logistic regression analysis to assess the predictive validity for both the component scores. The PCS12 from round 2 was used to predict work and physical limitation responses in round 3, whereas the MCS12 from round 2 was used to predict cognitive and social limitations in round 3.

Results

Sample Descriptive

The demographic characteristics of the final sample of 2214 participants are presented in Table 1. On average, sample participants were 59.7 years old and had 12.1 years of education. Most of the sample participants were non-Hispanic white, female, married, employed, and had insurance during year 1. The population had an average PCS12 comorbidity score of 3.9 (standard error [SE] = 0.08) and an MCS12 comorbidity score of 0.66 (SE = 0.03). More than half (59.1%) of the patients managed their diabetes through oral medications, whereas 21.7% of the patients required add-on insulin therapy. The mean MCS12 and PCS12 scores were 49.82 (SE = 0.26) and 40.81 (SE = 0.33), respectively, with no evidence of ceiling or flooring effects.

Reliability

Cronbach α values for the PCS12 and the MCS12 were 0.85 and 0.83, respectively, indicating high internal consistency. The ICC for PCS12 was 0.73, demonstrating high test-retest reliability. The ICC for MCS12 was 0.62, demonstrating adequate test-retest reliability.

Validity

Construct validity

Convergent and discriminant validity. A Spearman rank correlation matrix was constructed for the two component scales, the PHQ-2 score, and the K-6 score for perceived health and perceived mental health (Table 2). As per our hypothesis, the PCS12 was well correlated with perceived health ($r = 0.49$), but poorly correlated with perceived mental health ($r = 0.29$), PHQ-2 score ($r = 0.34$), and K-6 score ($r = 0.39$). Conversely, the MCS12 was moderately correlated with perceived health ($r = 0.40$) and perceived mental health ($r = 0.43$), but strongly correlated with K-6 score ($r = 0.67$) and PHQ-2 score ($r = 0.73$).

Confirmatory factor analysis. The results from the CFA of the two-factor model are shown in Figure 3. The proposed two-factor model in CFA produced adequate goodness-of-fit indices (GFI = 0.95, NFI = 0.95, CFI = 0.95, and SRMR = 0.03). The theoretical model proposed exhibited good conformity with the empirical data. The Wald test did not suggest deletion of any of the paths and covariance because of nonsignificance. No additional improvements in model fit (χ^2) were suggested by Lagrange multiplier statistics.

Criterion validity

Predictive validity. Increase in the PCS12 was associated with decreased likelihood of reporting future work limitations (odds ratio [OR] = 0.885; 95% confidence interval [CI] = 0.895–0.875) and physical limitations (OR = 0.887; 95% CI = 0.897–0.877). Similarly, increase in the MCS12 was associated with decreased likelihood of reporting social limitations (OR = 0.941; 95% CI = 0.952–0.93) and cognitive limitations (OR = 0.929; 95% CI = 0.940–0.918).

Concurrent validity. The PCS12 decreased significantly with increasing physical comorbidity score, where sample participants reporting a comorbidity score of 0 had a PCS12 of 50.2 and those with a physical comorbidity score of 8 or more had a PCS12 of 30.2. The difference in the PCS12 scores was significant for any value of physical comorbidity score beyond 0. The MCS12 also followed a similar trend, where sample participants with a mental health comorbidity score of 0 had an MCS12 of 50.2 and those with a comorbidity score of 4 or more had an MCS12 of 44.6. As seen in Figure 4A,B, a dose-response relationship was observed between component scores and comorbidity scores, and although the change in PCS12 scores was much steadier compared with that in MCS12 scores, it was still significant.

The differences in the SF-6D utility scores were compared among those with and without diabetes-related comorbidities. As presented in Table 3, all comorbidities were associated with reduction in the SF-6D utility scores. Of the comorbidities tested, kidney diseases were associated with the greatest disutility. The SF-6D preference-based utility scores were also found to be responsive to the presence of comorbidities.

Discussion

The present study is the first to perform a psychometric evaluation of the SF-12v2, a widely used generic HRQOL instrument, among a generalizable diabetic population in accordance with the COSMIN guidelines. The results from this study provide strong evidence in the favor of using the SF-12v2 in diabetic patients. The SF-12v2 demonstrated good internal consistency and adequate test-retest reliability over a 1-year period. Nevertheless, the MCS12 showed only moderate test-retest reliability. Even though MCS12's performance is not in complete agreement with the stated hypothesis, it was deemed acceptable on the basis of results from previous literature [6,7,16,29]. In further evaluation, no ceiling or flooring effects were observed, even though the diabetic population was expected to be sicker than the general population. The CFA confirmed the accordance of the theoretical factor structure with data from a generalizable diabetic population. The MCS12 and PCS12 showed good convergence with other instruments measuring similar constructs, and they both demonstrated the ability to predict future health limitations. They also showed a concordant dose-response relationship with HRQOL-CIs, which are objective measures of disease severity. In addition, the instrument also demonstrated its usefulness in measuring the preference-based utility and was sensitive in measuring clinically meaningful changes in health status associated with diabetes-related comorbidities.

Internal consistency, test-retest reliability, convergent validity, and discriminant validity testing were conducted on the basis of a previous study by Cheak-Zamora et al. [7] Unlike Cheak-Zamora et al., this study used the PHQ-2 and the K-6, which are stronger mental health measures than the perceived mental health question and the EuroQol five-dimensional questionnaire (EQ-5D) anxiety/depression domain, to measure the convergent and discriminant validity of the MCS12 [7]. The responses from these mental health questionnaires strengthened the study analysis because they allowed a thorough evaluation of the construct validity of the MCS12. Nevertheless, the panels used in this study did not include a concurrently administered EQ-5D instrument. Hence, the convergent validity assessment of the PCS12 was limited to only a comparison with the perceived health question. This study also assessed the factor structure of the SF-12v2, which was necessary so as to have reasonable confidence in the stability of the instrument [30]. The study did not evaluate the emergence of any new factor structure using exploratory factor analysis. Rather, this study assessed how well the data from a representative diabetic population fit the existing instrument structure [27,31–33]. Instead of using one-way analysis of variance to evaluate the predictive validity of the PCS12 and the MCS12, this study deviated from the Cheak-Zamora et al. [7] design by using logistic regression models to predict the response to health limitations (in future) on the basis of the PCS12/MCS12 from round 2. The study also differed in its approach to assess

Table 1 – Sample characteristics.

Characteristic	Weighted (N = 3,233,915)		Unweighted (N = 2,214)	
	Mean/percent	SEM*/SE percent	Mean/frequency	SD/percent
Age (y)	59.39	0.39	57.79	13.87
Years since diabetes diagnosis*	9.15	0.27	8.96	8.74
Body mass index	31.87	0.25	31.82	8.02
PCS12	40.81	0.33	40.89	12.02
MCS12	49.82	0.26	48.78	10.98
SF-6D	0.72	0.01	0.71	0.15
Years of education	12.04	0.14	11.19	4.75
Region				
Northeast	16.89	1.42	330	14.91
Midwest	22.50	1.35	406	18.34
South	41.47	1.50	945	42.68
West	19.14	1.14	533	24.07
Sex				
Female	50.85	1.36	1,164	52.57
Male	49.15	1.33	1,050	47.43
Race				
Hispanic	15.69	1.45	613	27.69
Non-Hispanic white	61.31	1.77	840	37.94
Non-Hispanic black	15.09	1.17	558	25.20
Non-Hispanic Asian	4.79	0.54	148	6.68
Non-Hispanic other/mixed races	3.15	0.64	55	2.48
Marital status				
Never married	12.35	0.83	337	15.22
Currently married	59.36	1.34	1,556	55.06
Divorced/widowed	28.29	1.15	2,214	29.72
Employment status				
Not employed	56.70	1.50	1,288	58.18
Employed	43.30	1.50	926	41.82
Therapeutic intensity				
Diet and exercise	8.43	0.85	163	7.75
Insulin only	10.66	0.87	190	9.03
Oral therapy	59.18	1.47	1,590	58.82
Oral + insulin therapy	21.73	1.18	513	21.03
Insured during the year				
Yes	91.84	0.63	1,932	87.26
No	8.161	0.63	282	12.74
Comorbidity score				
HRQOL-CI-MCS12	0.62	0.03	0.57	0.90
HRQOL-CI-PCS12	3.33	0.08	2.93	2.86
Diabetes with complications†				
Yes	1.86	0.04	28	1.26
No	98.14	0.24	2,186	38.74
Cardiac diseases				
Yes	26.91	0.14	507	22.90
No	73.09	0.21	1,707	77.10
Eye problems				
Yes	16.26	0.10	391	17.66
No	83.74	0.22	1,823	82.34
Kidney problems				
Yes	7.09	0.06	178	8.04
No	92.91	0.24	2,036	91.96

Note. The weighted means and proportions take into account the complex survey design of the MEPS.
HRQOL-CI-MCS12, health-related quality of life comorbidity index (mental component score); HRQOL-CI-PCS12, health-related quality of life comorbidity index (physical component score); MCS12, mental component score; MEPS, Medical Expenditure Panel Survey; PCS12, physical component score; SE, standard error; SEM, standard error of the mean; SF-6D, six-dimensional health state short form (preference-based utility score for SF-12v2); SF-12v2, Medical Outcomes Study Short Form version 2.
* Years since diabetes = (Age in round 1 – Age at diabetes diagnosis).
† Diabetes with complication = Clinical Diagnosis Classification code of 50.

Table 2 – Spearman rank correlation coefficients for convergent and discriminant validity.

Measure	Perceived mental health	Perceived physical health	PHQ-2	K-6
PCS12	0.29	0.49*	−0.34	−0.39
MCS12	0.43†	0.40	−0.67*	−0.73*

Note. Multitrait multimethod matrix method assesses the correlation between the component scores of perceived physical health, perceived mental health, and depression scales (K-6 and PHQ-2).
K-6, Kessler Severity of Depressive Symptom score; MCS12, mental component score; PCS12, physical component score; PHQ-2, Patient Health Questionnaire.
* Convergent correlations significantly greater than discriminant correlations at $P \leq 0.001$.
† Convergent correlation coefficients not significantly different (PCS12 compared with MCS12) at $P = 0.05$.

the concurrent validity. In the assessment of the concurrent validity, this study used an HRQOL-CI as opposed to disease counts or generic comorbidity indices that have been used in previous studies. The HRQOL-CI approach is an improvement over previous work because it is specific in predicting the HRQOL. Finally, unlike previous studies, the present analysis also calculated the preference-based utility using age, sex, and PCS12 and MCS12 scores. Derived SF-6D utility scores from the SF12-v2 have demonstrated better performance in previous studies compared with the EQ-5D utility scores, providing further support for the use of the SF12-v2 and the SF-6D [34].

The current findings are in agreement with findings from the SF-12v2 validation in the general US population [7]; there were, however, some key differences. Only moderate correlation existed between the PCS12 and the perceived health and between the MCS12 and the perceived mental health. Previous studies, however, have reported similar results, indicating the limitation of the perceived health questions in capturing the complexity of the components [6,7]. The scale imbalance of the perceived health questions is likely to be responsible for the poor correlation. Also, the strong correlations between MCS12, K-6, and PHQ-2 scores confirmed the inadequacy of the perceived mental health question. Unfortunately, a study by Maurischat et al. [33] is the only study testing the SF-12v2 in a diabetic population. The study, however, has limited generalizability because it used data from 343 participants from a rehabilitation center in Germany and assessed only the structural validity of the instrument. Other

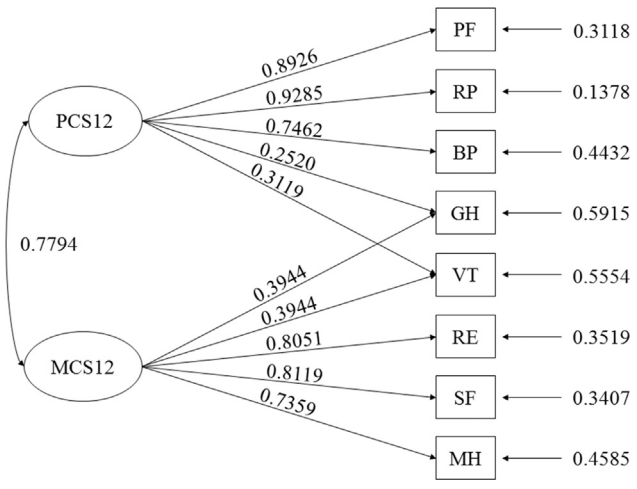

Fig. 3 – Confirmatory factory analysis: two-factor model for the SF-12v2. The factor structure tested in confirmatory factory analysis and standardized coefficients obtained from the model are presented. All coefficients are significant at $P \leq 0.001$. The coefficients were in agreement with previous reports. BP, bodily pain; GH, general health; MCS12, mental component score; MH, mental health; PCS12, physical component score; PF, physical functioning; RP, role physical; SF, social functioning; SF-12v2, Medical Outcomes Study Short Form version 2; RE, role emotional; VT, vitality.

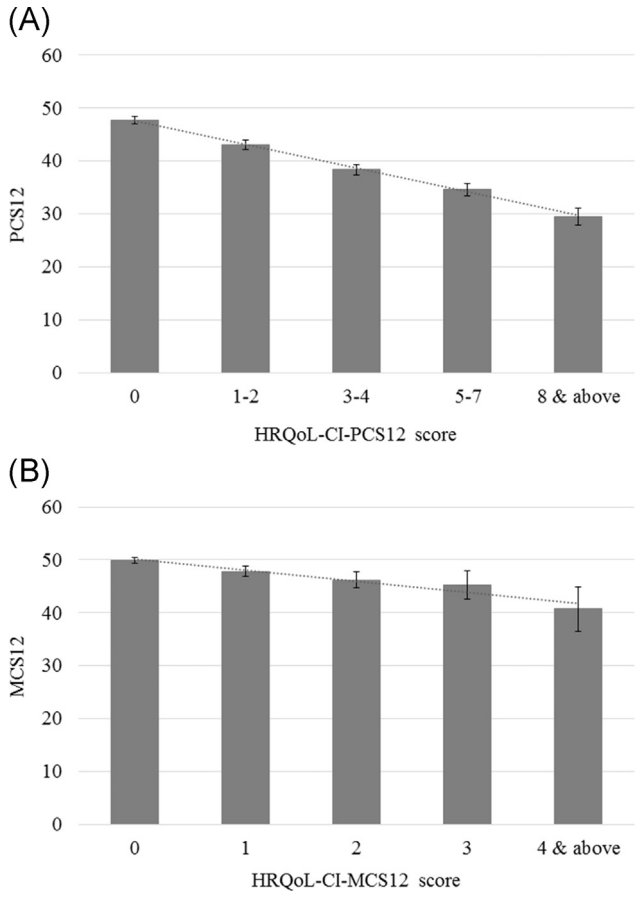

Fig. 4 – (A) PCS12 change with increasing physical comorbidity scores. (B) MCS12 change with increasing mental health comorbidity scores. The MCS12 and PCS12 scores were compared across HRQOL-CI-MCS12 and HRQOL-CI-PCS12. Both the component scores demonstrated concurrent validity because a decreasing trend across increasing levels of comorbidity was observed. HRQOL-CI-MCS12, health-related quality of life comorbidity index (mental component score); HRQOL-CI-PCS12, health-related quality of life comorbidity index (physical component score); MCS12, mental component score; PCS12, physical component score.

Table 3 – Comparison of SF-6D utility scores across different subpopulations.

Subpopulation	SF-6D utility scores		
	Comorbidity absent	Comorbidity present	Disutility
Diabetes with complications	0.7259	0.6390	0.0869*
Cardiac disease†	0.7397	0.6824	0.0573*
Kidney disease due to diabetes	0.7327	0.6141	0.1186*
Eye problem due to diabetes	0.7358	0.6649	0.0709*

MCID, minimal clinically important difference; SF-6D, six-dimensional health state short form (preference-based utility score for SF-12v2); SF-12v2, Medical Outcomes Study Short Form version 2.
* $P < 0.001$; Disutility > MCID (0.041).
† Cardiac disease = myocardial infarction/stroke/coronary heart disease.

studies have compared the performance of the SF-12v2 with that of diabetes-specific instruments [35,36]. Although disease-specific instruments are more sensitive for diabetic patients, they have limited scope of comparison across different disease states and in utility assessment [37–39].

This study addresses the key concern of the lack of holistic psychometric testing of the SF-12v2 in a generalizable diabetic population. The study demonstrates that the psychometric properties of the SF-12v2 are similar between diabetic patients and the general population. The instrument can also be used to derive the preference-based utility scores among the diabetic population for use in economic analyses. The SF-12v2 is a widely used HRQOL instrument and is administered within the MEPS, allowing for nationally representative assessments of HRQOL among US diabetic patients. The study findings can be a useful guide for HRQOL instrument selection and advance future research in diabetes and HRQOL.

The study acknowledges some important limitations. The test-retest reliability was likely affected by the weak correlation of perceived health questions with component scores. Sensitivity analysis was performed to confirm whether weak correlation affected the study findings. A well-correlated measure of mental health status, the PHQ-2, was used to identify patients with a stable (same) PCS12 score. The resultant ICC value obtained was not significantly different from our base case (0.65 compared with 0.62, both suggesting only adequate test-retest reliability). Although we did not have a good measure of physical health analogous to the PHQ-2 (for MCS12), we performed a similar sensitivity analysis for the PCS12 using those with stable responses for the physical health limitation question. The physical health limitation question had a better correlation than the perceived health question (0.56 compared with 0.49). Nevertheless, the resultant ICC value obtained for the PCS12 was also 0.73 (similar to our base-case analysis). Thus, the adequate test-retest reliability is likely because of the time difference between the two measurements. Limitations also exist in the self-reported nature of the data, creating the potential for recall bias. In addition, response shift is possible because of the time lag between the administrations of the SAQ and the panel interview session when other information is collected. Even though the time lag is short (6–8 months), it is likely to affect the concurrent validity. The comorbidity scores used to assess the concurrent validity were not extensive, with a large percentage of patients having a score of 0. Despite these limitations, the instrument performed well in demonstrating sufficient validity and reliability.

Conclusions

The study demonstrated evidence of reliability and validity of the SF-12v2 in measuring HRQOL in diabetic adults. Therefore, the SF-12v2 can be used with confidence to quantify HRQOL among diabetic patients.

REFERENCES

[1] US Department of Health and Human Services. National Diabetes Statistics Report: estimates of diabetes and its burden in the United States, 2014. 2014: Available from: http://198.246.124.29/diabetes/pdfs/data/2014-report-acknowledgments.pdf. [Accessed November 23, 2015].
[2] Gary TL, Crum RM, Cooper-Patrick L, et al. Depressive symptoms and metabolic control in African-Americans with type 2 diabetes. Diabetes Care 2000;23:23–9.
[3] Natarajan S, Nietert P. Hypertension, diabetes, hypercholesterolemia, and their combinations increased health care utilization and decreased health status. J Clin Epidemiol 2004;57:954–61.
[4] Nichols L, Barton P. Diabetes, minor depression and health care utilization and expenditures: a retrospective database study. Cost Eff Resour Alloc 2007;5:1.
[5] Egede L, Zheng D, Simpson K. Comorbid depression is associated with increased health care use and expenditures in individuals with diabetes. Diabetes Care 2002;25:464–70.
[6] Ware JE, Kosinski M, Keller SD. How to Score the SF-12 Physical and Mental Health Summary Scales (3rd ed.). Lincoln, RI: QualityMetric, 1998.
[7] Cheak-Zamora N, Wyrwich K, McBride T. Reliability and validity of the SF-12v2 in the medical expenditure panel survey. Qual Life Res 2009;18:727–35.
[8] Kontodimopoulos N, Pappa E, Niakis D, Tountas Y. Validity of SF-12 summary scores in a Greek general population. Health Qual Life Outcomes 2007;5:1.
[9] Jayasinghe UW, Proudofoot J, Barton CA, et al. Quality of life of Australian chronically-ill adults: patient and practice characteristics matter. Health Qual Life Outcomes 2009;7:50.
[10] Gandhi SK, Salmon JW, Zhao SZ, et al. Psychometric evaluation of the 12-Item Short Form Health Survey (SF-12) in osteoarthritis and rheumatoid arthritis clinical trials. Clin Ther 2001;23:1080–98.
[11] Hayes CJ, Bhandari NR, Kathe N, Payakachat N. Reliability and validity of the Medical Outcomes Study Short Form-12 Version 2 (SF-12v2) in adults with non-cancer pain. Healthcare (Basel) 2017;5:22.
[12] Cohen J, Monheit A, Beauregard K, Cohen S. The Medical Expenditure Panel Survey: a national health information resource. Inquiry 1996;33:373–89.
[13] Cohen S. Design strategies and innovations in the medical expenditure panel survey. Med Care 2003;41:3–5.
[14] Kroenke K, Spitzer R, Williams J. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003;41:1284–92.
[15] Moran M. Kessler Psychological Distress Scale (K6). Available from: http://www.statisticssolutions.com/kessler-psychological-distress-scale-k6/. Accessed April 12, 2016.
[16] Ware JE, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34:220–33.
[17] Gandek B, Ware JE, Aaronson NK, et al. Cross-validation of item selection and scoring for the SF-12 health survey in nine countries: results from the IQOLA Project. J Clin Epidemiol 1998;51:1171–8.
[18] Lacson E, Xu J, Lin S, Dean S. A comparison of SF-36 and SF-12 composite scores and subsequent hospitalization and mortality risks in long-term dialysis patients. Clin J 2010;5:252–62.
[19] Jenkinson C, Layte R, Jenkinson D. A shorter form health survey: Can the SF-12 replicate results from the SF-36 in longitudinal studies? J Public Health Med 1997;19:179–86.

[20] Hanmer J. Predicting an SF-6D preference-based score using MCS and PCS scores from the SF-12 or SF-36. Value Health 2009;12:958–66.
[21] Mukherjee B, Ou H, Wang F, Erickson SR. A new comorbidity index: the health-related quality of life comorbidity index. J Clin Epidemiol 2011;64:309–19.
[22] Khanna R, Jariwala-Parikh K, West-Strum D. Health-related quality of life and its determinants among adults with autism. Res Autism Spectr Disord 2014;8:157–67.
[23] Nunnally J, Bernstein I, Berge J. Psychometric Theory. New York, NY: McGraw-Hill, 1967.
[24] Montazeri A, Vahdaninia M, Mousavi SJ, et al. The 12-item medical outcomes study short form health survey version 2.0 (SF-12v2): a population-based validation study from Tehran, Iran. Health Qual Life Outcomes 2011;9:12.
[25] Cunningham W, Nakazono T, Tsai K. Do differences in methods for constructing SF-36 physical and mental health summary measures change their associations with chronic medical conditions and utilization. Qual Life Res 2003;12:1029–35.
[26] Hays R, Morales L. The RAND-36 measure of health-related quality of life. Ann Med 2001;33:350–7.
[27] Fleishman JA, Selim AJ, Kazis LE. Deriving SF-12v2 physical and mental health summary scores: a comparison of different scoring algorithms. Qual Life Res 2010;19:231–41.
[28] Schreiber JB, Nora A, Stage FK, et al. Reporting structural equation modeling and confirmatory factor analysis results: a review. J Educ Res 2006;99:323–38.
[29] Ware JE, Kosinski M, Keller SK. SF-36 Physical and Mental Summary Scales: S User's Manual. Boston, MA: The Health Institute, 1994.
[30] Ellis C, Grubaugh AL, Egede LE. Factors associated with SF-12 physical and mental health quality of life scores in adults with stroke. J Stroke Cerebrovasc Dis 2010;22:309–17.
[31] Montazeri A, Vahdaninia M, Mousavi SJ, Omidvari S. The Iranian version of 12-Item Short-Form Health Survey (SF-12): factor structure, internal consistency and construct validity. BMC Public Health 2009;9:341.
[32] Montazeri A, Goshtasebi A, Vahdaninia M, Gandek B. The Short Form Health Survey: translation and validation study of the Iranian version. Qual Life Res 2005;14:875–82.
[33] Maurischat C, Herschbach P, Peters A, Bullinger M. Factorial validity of the Short-Form 12 (SF-12) in patients with diabetes mellitus. Psychol Sci Q 2008;50:7–20.
[34] Bharmal M, Thomas J III. Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. Value Health 2006;9:262–71.
[35] Sundaram M, Kavookjian J, Patrick JH, et al. Quality of life, health status and clinical outcomes in type 2 diabetes patients. Qual Life Res 2007;16:165–77.
[36] Vileikyte L, Peyrot M, Bundy C, et al. The development and validation of a neuropathy- and foot ulcer-specific quality of life instrument. Diabetes Care 2003;26:2549–55.
[37] Ortiz Z, Shea B, Garcia D, Boers M. The responsiveness of generic quality of life instruments in rheumatic diseases: a systematic review of randomized controlled trials. J Rheumatol 1999;26:210–6.
[38] Coons S, Rao S, Keininger D, Hays R. A comparative review of generic quality-of-life instruments. Pharmacoeconomics 2000;17:13–35.
[39] Wiebe S, Guyatt G, Weaver B, Matijevic S. Comparative responsiveness of generic and specific quality-of-life instruments. J Clin Epidemiol 2003;56:52–60.