Published in final edited form as:

Res Social Adm Pharm. 2018 November ; 14(11): 1080–1084. doi:10.1016/j.sapharm.2018.01.007.

Reliability and Validity of SF-12v2 Among Adults with Self-Reported Cancer

Naleen Raj Bhandari, MS¹, Niranjan Kathe, MS¹, Corey Hayes, PharmD, MPH^{1,2}, Nalin Payakachat, PhD¹

¹Division of Pharmaceutical Evaluation and Policy, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

²Psychiatric Research Institute, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

Abstract

Background: The SF-12v2 has been validated in general population and other conditions but the evidence is limited in Americans with cancer.

Objective: To assess reliability and validity of SF-12v2 among adults with self-reported cancer using the Medical Expenditure Panel Survey (MEPS).

Methods: Self-reported cancer participants (SCPs) were identified from 2003 MEPS. SF-12v2 was administered as part of self-administered questionnaires. Physical (PCS12) and mental (MCS12) component scores of SF-12v2 were evaluated for reliability (internal consistency, test-retest) and validity (convergent, discriminant, predictive, concurrent).

Results: 420 SCPs were identified with average age of 59.3 years (SE=0.9). 10.7% had multiple cancers (>1) and 52% had at least one OCC. Average PCS12 and MCS12 were 45.09 (SE=0.5) and 50.75 (SE=0.5). PCS12 and MCS12 demonstrated high internal consistency ($\alpha_{\text{PCS12}}=0.89$; $\alpha_{\text{MCS12}}=0.88$), acceptable test-retest reliability ($\text{ICC}_{\text{PCS12}}=0.82$; $\text{ICC}_{\text{MCS12}}=0.73$), strongly correlated with most of the expected EQ-5D domains ($r=0.51$ – 0.71), and demonstrated strong convergent validity on perceived health ($r=0.61$) and perceived mental health ($r=0.52$). PCS12 and MCS12 were able to discriminate between groups with and without physical/cognitive limitations. Only PCS12 was negatively correlated with number of OCCs.

Conclusion: The SF-12v2 is a reliable and valid instrument to quantify health-related quality of life among adults with self-reported cancer.

Address for Correspondence: Nalin Payakachat, PhD, Associate Professor, Division of Pharmaceutical Evaluation and Policy, University of Arkansas for Medical Sciences College of Pharmacy, 4301 W. Markham, Slot 522-4, Little Rock, AR 72205, Phone: (501) 686-7701, Fax: (501) 686-5156, npayakachat@uams.edu.

Conflict of Interest

Dr. Hayes was supported by the Translational Training in Addiction [1T32 DA 022981]. Dr. Payakachat received an honorarium for service as a paid consultant to Roche Ltd., service as a consultant to CBPartners, and ownership of stock in Pfizer. Others authors report no conflicts of interest.

Previous Presentation

Preliminary data presented as a poster at the Southern Pharmacy Administration Conference, June 24–26, 2016 at the University of Mississippi, Oxford, MS. Abstract published in: *Research in Social and Administrative Pharmacy*, 2016;12(4), e12.

Keywords

Cancer; Reliability and Validity; Quality of Life; SF-12; Psychometrics

INTRODUCTION

Health-related quality of life (HRQoL) has become an important health outcome in cancer care, clinical trials, and cancer research.^{1,2} Cancer can have significant impact on patients' health and wellbeing. Measuring HRQoL using the Medical Outcomes Study Short Form-12 (SF-12) has extensively increased in large national health surveys including the Medical Expenditure Panel Survey (MEPS).^{3,4} MEPS started to field version-2 of SF-12 (SF-12v2) in 2003. Since its inception, MEPS is used in research and medical-decision making by decision makers and health services researchers. While the SF-12v2 have been validated in the general U.S. population⁴ and among several disease conditions,^{5–14} they may not be applicable to people with cancer. Although some studies assessed psychometric properties of SF-12v2 in few cancer types,^{15–18} there is still limited information on psychometric properties of SF-12v2 in the Americans with cancer, especially using a nationally representative data such as MEPS. This study assessed psychometric properties on both reliability and validity of SF-12v2, using cross-sectional and longitudinal files, among adults with self-reported cancer using MEPS. Considering the research potential of MEPS database and the use of SF-12v2 in measuring HRQoL, a comprehensive evaluation of psychometric properties of SF-12v2 in adults with cancer is essential.

METHODS

Data Source

This study is based on MEPS 2003–04 data because of the availability of the EuroQol-5D-3L (EQ-5D-3L), which allowed us to compare and evaluate the validity of SF-12v2. In addition, 2003–04 was the last time when MEPS collected the EQ-5D-3L data. The MEPS collects information about the healthcare resource utilization (use and expenditures) and health status from a nationally representative, civilian non-institutionalized US sample.^{19,20} MEPS uses an overlapping panel design and collects data in five rounds of self-administered mail questionnaires (SAQ) and/or interviews over a two-year period.²⁰ The SAQ is administered in rounds 2 and 4 of each panel with approximately one year between the two SAQs.

Sample

Self-reported cancer participants aged ≥ 18 years were identified using Clinical Classification Codes (CCS): 011–045.²¹ Only participants who were in-scope for all five rounds and were eligible for both SAQs rounds 2 and 4 were included.

SF-12v2 is a 12-item self-reported HRQoL measure and reports summary scores for physical (PCS12) and mental components (MCS12).³ PCS12 and MCS12 scores range from 0 to 100 with mean of 50 and SD of 10, where higher scores indicate better physical and mental health.

Validating Measures

For validation of SF-12v2, EQ-5D-3L, perceived health, perceived mental health, number of other chronic conditions (OCC), and physical and cognitive limitations were used.

EQ-5D-3L is a self-reported HRQoL measure including 5-item health profile and a visual analog scale (EQ-5D-VAS).²² The EQ-5D health profile produces health preference score (EQ-5D-index) that ranges from −0.109 (worse than death) to 1 (perfect health).²² The EQ5D-VAS measures participants' current overall health that ranges from 0–100 (worst to best health). **Perceived Health, Perceived Mental Health, Physical and Cognitive Limitations** were reported directly from MEPS questions. **Number of OCC** was categorized as participants having none, 1–2, 3–4, and 5 or more. Further information on survey questions is available elsewhere.²³

Statistical Analysis

Weighted descriptive statistics were generated to provide national sample estimates. Ceiling and Floor effects of PCS12 and MCS12 were determined by percentages of participants with highest and lowest possible scores. All analyses were conducted using SAS v9.3 (SAS Institute Inc, Cary, NC). A graphical representation of methods is provided in Figure 1.

Reliability Internal consistency was assessed by Cronbach's alpha. A value 0.80 was considered highly reliable²⁴. Test-retest reliability was assessed using intra-class correlational coefficients (ICC), with the assumption that if a participant's condition did not change, responses on the SF-12v2 should not change from round 2 to round 4.⁴ ICCs 0.70 were considered acceptable.²⁵

Convergent and Discriminant Validity are the degree to which an instrument measures an intended concept. Using Spearman's rank correlation matrix between PCS12 and MCS12, and the EQ-5D items, perceived physical and mental health, we evaluated this validity. Correlations from 0.1–0.29 were considered weak, 0.3–0.49 as moderate, and 0.5 as strong.

Predictive Validity is the ability of an instrument to predict certain characteristics based on theoretical concepts. Self-reported physical and cognitive limitations in round 3 were predicted by using PCS12 and MCS12 scores from round 2 in two separate logistic regression models. **Concurrent Validity** – the ability to distinguish between theoretically known-groups was evaluated by using OCC in two separate general linear models.

RESULTS

Demographics

Of the total 560 adults with cancer, 420 were included in the analysis after excluding those <18 years (38), not in-scope for all five rounds (63), not eligible for either or both SAQ (33) and those with missing values for PCS12 and MCS12 (6). Weighted demographics are presented in Table 1. Average age of participants was 59.3 years (SE=0.9) with mostly females (57.4%) and Whites (93.4%). About one-third had skin cancer (32.7%), followed by breast cancer (9.9%), cancers of male genitals (9.0%) and colorectal cancer (3.6%). Further, 10.5% participants had more than one cancer. Nearly 52% reported presence of at least one

OCC. 24% and 7% reported physical and cognitive limitations. Average PCS12 and MCS12 scores were 45.09 (SE=0.5) and 50.75 (SE=0.5), with no evidence of floor or ceiling effects.

SF-12v2 Psychometrics

Reliability Both PCS12 ($\alpha_{\text{PCS12}}=0.89$) and MCS12 ($\alpha_{\text{MCS12}}=0.88$) demonstrated high internal consistency reliability, and acceptable test-retest reliability ($\text{ICC}_{\text{PCS12}}=0.82$; $\text{ICC}_{\text{MCS12}}=0.73$).

Convergent and Discriminant Validity PCS12 demonstrated high convergence with perceived health status ($r=0.61$) and hypothesized EQ-5D items (EQ-5D-index: $r=0.67$, mobility: $r=0.70$, usual activity: $r=0.71$, pain/discomfort: $r=0.67$, EQ-5D-VAS: $r=0.70$), except for the self-care ($r=0.36$). PCS12 showed moderate to weak convergence with perceived mental health ($r=0.38$) and EQ-5D anxiety/depression ($r=0.28$) indicating adequate discriminant validity. MCS12 converged moderate to high with hypothesized EQ-5D items (EQ-5D-index: $r=0.45$, anxiety/depression: $r=0.67$, EQ-5D-VAS: $r=0.51$) and perceived mental health ($r=0.52$). MCS12 demonstrated moderate to weak correlations with dissimilar EQ-5D items (mobility: $r=0.18$, self-care: $r=0.23$, pain/discomfort: $r=0.29$, usual activity: $r=0.37$) and perceived physical health ($r=0.42$) indicating adequate discriminant validity.

The overall correlation between PCS12 and MCS12 was weak ($r=0.17$), further differentiating the two latent concepts. All coefficients were statistically significant at $p<0.001$.

Predictive and Concurrent Validity Both PCS12 and MCS12 could discriminate between groups with and without physical/mental limitations, indicating high predictive validity. The average PCS12 was significantly lower in participants with physical limitations, compared to without limitation (31.23 vs. 48.18, $p<0.001$). Every 1-point increase in PCS12 score in round 2, a participant's likelihood of reporting a physical limitation in round 3 decreased by 12% (OR=0.88, 95%CI=0.85–0.90). Similarly, the average MCS12 was significantly lower in participants who reported mental limitations, compared to those without limitation (38.18 vs. 50.70, $p<0.001$). Every 1-point increase in MCS12 score in round 2, a participant's likelihood of reporting a cognitive limitation in round 3 decreased by 8% (OR=0.92, 95%CI=0.89–0.95). The number of OCCs was significantly and negatively correlated with PCS12 scores ($p<0.001$) (Figure 2). However, MCS12 scores were similar across the number of OCCs. The summary of psychometric properties is presented in Table 2.

DISCUSSION

SF-12v2, a generic HRQoL measure, is widely used in clinical trials and national surveys, facilitating the assessment and comparison of HRQoL across general as well as disease populations, however, with limited knowledge on its validity in cancer population, especially using a representative database. This study evaluated psychometric properties of SF-12v2 among nationally representative self-reported cancer adults, and found that SF-12v2 is reliable and valid in this population.

Absence of floor and ceiling effects was similar to MEPS general US population,⁴ the Tunisian population⁷ and the autistic population⁶ demonstrating no evidence of clustering of scores at lower or higher ends of scoring range, further facilitating the use of SF-12v2 to study improvements or decrements in HRQoL overtime in cancer population. With Cronbach's alpha >0.80 and ICCs >0.70 for PCS12 and MCS12, high reliability of SF-12v2 was evident, consistent with previous studies conducted among American-Chinese breast cancer (ACBC) patients¹⁶ and prostate cancer patients.²⁶ Acceptable ICCs demonstrated the stability in PCS12 and MCS12 scores of SF-12v2 with stable physical and mental health overtime. As opposed to previous study,¹⁶ the test-retest (one-year apart) reliability among the stable participants was shown to be acceptable in this study.

The weak correlation between PCS12 and MCS12 of SF-12v2 suggests that respective scales measured two different concepts, similar to Ware et al.'s two-factor SF-12v2 model in general US population and other studies.^{3,6,13,14,16,26} Further, the high convergence of PCS12 and MCS12 with the expected EQ-5D items infers that both these instruments measure similar latent concepts of perceived physical and mental health. Unlike the overall MEPS population,⁴ MCS12 scores correlated strongly with both anxiety/depression item of EQ-5D as well as perceived mental health, and PCS12 also reported better convergent validity. PCS12 revealed strong correlation with EQ-5D-index consistent with overall MEPS population.⁴ However, MCS12 had a moderate correlation with EQ-5D-index, possibly because the EQ-5D has more items focused on physical than mental aspect of HRQoL.

PCS12 significantly decreased with an increase in number of OCC in a dose-response relationship as noted in previous study.⁴ However, a similar pattern for MCS12 was not found in this study. Participants with one or two OCC reported higher mean MCS12 compared to those with no OCC. This inconsistency and insignificant difference could be partly due to participant's age differences as some of the comorbidities may be age-related and could distress differentially. Additionally, as these data were collected in different time points, response shift might also explain the discrepancy. This implies that a mental HRQoL assessment might have changed with time, probably due to recalibration of the tool, reprioritization or reconceptualization of the concept of mental HRQoL, without a change in objective circumstances.²⁷ Lastly, PCS12 and MCS12 demonstrated good predictive validity. Both component scores could predict future limitations, which implied the excellent predictive ability of SF-12v2 among cancer population.

This study results are limited as they are generalizable only to non-institutionalized adults with self-reported cancer. In addition, there is potential for recall bias due to self-reported nature of the data. Additionally, data on SF-12v2 and OCC – which was not comprehensive, were collected in different rounds. This plausibly affected the concurrent validity of MCS12 as the participants with more number of OCC may have re-conceptualized or reprioritized their definition of HRQoL as they get habituated to live with those conditions.

CONCLUSIONS

In conclusion, SF-12v2 (PCS12 and MCS12) performed well by demonstrating relatively high reliability, and adequate validity among adults with history of cancer in the US. This

study provides evidence to researchers and policy makers to support the use of SF-12v2 for quantifying HRQoL among those with cancer.

Bibliography

1. Fujisawa D, Inoguchi H, Shimoda H, et al. Impact of depression on health utility value in cancer patients. *Psychooncology*. 2016;25:491–495. [PubMed: 26283141]
2. Wedding U, Pientka L, Höffken K. Quality-of-life in elderly patients with cancer: A short review. *Eur J Cancer*. 2007;43:2203–2210. [PubMed: 17662595]
3. Ware J How to Score Version 2 of the SF-12 Health Survey (with a Supplement Documenting Version 1). Lincoln R.I.;Boston Mass.: QualityMetric Inc.;Health Assessment Lab; 2005.
4. Cheak-Zamora NC, Wyrwich KW, McBride TD. Reliability and validity of the SF-12v2 in the medical expenditure panel survey. *Qual Life Res*. 2009;18:727–735. [PubMed: 19424821]
5. Smedt D De, Clays E, Doyle F, et al. Validity and reliability of three commonly used quality of life measures in a large European population of coronary heart disease patients. *Int J Cardiol*. 2008;167:2294–2299.
6. Khanna R, Jariwala K, West-Strum D. Validity and reliability of the Medical Outcomes Study Short-Form Health Survey version 2 (SF-12v2) among adults with autism. *Res Dev Disabil*. 2015;43:51–60. [PubMed: 26151443]
7. Younsi M, Chakroun M. Measuring health-related quality of life: psychometric evaluation of the Tunisian version of the SF-12 health survey. *Qual Life Res*. 2014;23:2047–2054. [PubMed: 24515673]
8. Luo X, Lynn George M, Kakouras I, et al. Reliability, Validity, and Responsiveness of the Short Form 12-Item Survey (SF-12) in Patients With Back Pain. *Spine (Phila Pa 1976)*. 2003;28:1739–1745. [PubMed: 12897502]
9. Lim LL, Fisher JD. Use of the 12-item short-form (SF-12) Health Survey in an Australian heart and stroke population. *Qual life Res*. 1999;8:1–8. <http://www.ncbi.nlm.nih.gov/pubmed/10457733>. Accessed July 15, 2016. [PubMed: 10457733]
10. Lam CLK, Tse EYY, Gandek B. Is the standard SF-12 Health Survey valid and equivalent for a Chinese population? *Qual Life Res*. 2005;14:539–547. [PubMed: 15892443]
11. Jakobsson U, Westergren A, Lindskov S, Hagell P. Construct validity of the SF-12 in three different samples. *J Eval Clin Pract*. 2012;18:560–566. [PubMed: 21210901]
12. Montazeri A, Vahdaninia M, Mousavi SJ, et al. 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. [PubMed: 19758427]
13. 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*. 2017;5:22.
14. Kathe N, Hayes CJ, Bhandari NR, Payakachat N. Assessment of Reliability and Validity of SF-12v2 among a Diabetic Population. *Value Heal*. 11 2017.
15. Llewellyn CD, McGurk M, Weinman J. Head and neck cancer: To what extent can psychological factors explain differences between health-related quality of life and individual quality of life? *Br J Oral Maxillofac Surg*. 2006;44:351–357. [PubMed: 16105712]
16. Ashing-Giwa K, Lam CN, Xie B. Assessing health-related quality of life of Chinese-American breast cancer survivors: a measurement validation study. *Psychooncology*. 2013;22:704–707. [PubMed: 22302455]
17. Wilson TR, Alexander DJ, Kind P. Measurement of Health-Related Quality of Life in the Early Follow-Up of Colon and Rectal Cancer. *Dis Colon Rectum*. 2006;49:1692–1702. [PubMed: 17041750]
18. Annunziata MA, Muzzatti B, Giovannini L, et al. Is long-term cancer survivors' quality of life comparable to that of the general population? An Italian study. *Support Care Cancer*. 2015;23:2663–2668. [PubMed: 25646617]

19. Cohen JW, Cohen SB, Banthin JS. The medical expenditure panel survey: a national information resource to support healthcare cost research and inform policy and practice. *Med Care*. 2009;47:S44–50. [PubMed: 19536015]
20. United States Department of Health and Human Services, Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey. http://meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h147/h147doc.shtml. Published 2013 Accessed March 30, 2016.
21. Machlin S, Soni A, Fang Z. Understanding and Analyzing MEPS Household Component Medical Condition Data. http://meps.ahrq.gov/mepsweb/survey_comp/MEPS_condition_data.pdf. Accessed March 24, 2016.
22. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43:203–220. <http://www.ncbi.nlm.nih.gov/pubmed/15725977>. Accessed July 3, 2017. [PubMed: 15725977]
23. United States Department of Health and Human Services, Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey - Survey Questionnaire. https://meps.ahrq.gov/mepsweb/survey_comp/survey_questionnaires.jsp. Accessed March 30, 2016.
24. Mosier CI. On the reliability of a weighted composite. *Psychometrika*. 1943;8:161–168.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174. <http://www.ncbi.nlm.nih.gov/pubmed/843571>. Accessed June 28, 2016. [PubMed: 843571]
26. Hooper G, Gerber E. Measuring the Quality of Life in Men with Prostate Cancer. *Urol Nurs*. 2014;34:177–184. [PubMed: 25233619]
27. Blome C, Augustin M. Measuring Change in Quality of Life: Bias in Prospective and Retrospective Evaluation. *Value Heal*. 2015;18:110–115.

Figure 1.

Graphical representation of methods used in the psychometric evaluation of SF-12v2 among patients with self-reported cancer

Figure 2.
Concurrent Validity: PCS12/MCS12 vs. Number of Other Chronic Conditions

Table 1.

Demographic Characteristics

Characteristics	Value
Age, mean(SE)	59.3(0.9)
PCS12 Score, mean(SE)	45.09(0.5)
MCS12 Score, mean(SE)	50.75(0.5)
Sex, %(SE)	
Male	42.6(2.6)
Female	57.4(2.6)
Race, %(SE)	
White	93.4(1.0)
Black	3.8(0.8)
Other	2.8(0.8)
Highest Degree, %(SE)	
High School or Less	59.9(2.4)
Bachelor's Degree	19.9(1.8)
Master's or Doctorate Degree	12.8(1.8)
Other Degree	7.4(1.5)
Marital Status, %(SE)	
Married	64.1(2.4)
Never Married	8.6(1.6)
Separated/Divorced	13.7(1.9)
Widowed	13.6(1.7)
Employment Status, %(SE)	
Employed in RD 3	47.4(2.7)
Unemployed in RD 3	52.6(2.7)
Family Income (% FPL), %(SE)	
Poor/Negative (<100%)	9.0(1.3)
Near Poor (100% - <125%)	3.3(0.7)
Low Income (125% - <200%)	10.6(1.7)
Middle Income (200% - <400%)	27.9(2.9)
High Income (400%)	49.2(3.2)
Uninsured during 2003, %(SE)	
Yes	3.9(0.9)
No	96.1(0.9)
More than one cancer, % (SE)	10.7 (1.5)
Cancer Type, % (SE)	
Breast	9.9 (1.4)
Colorectal	3.6 (1.0)
Lung/Bronchus	1.9 (0.7)

Characteristics	Value
Skin	32.7 (2.7)
Uterus, Cervix, Ovary, other female genitals	3.7 (0.9)
Male genitals	9.0 (1.6)
Neoplasms of Unspecified nature/behavior	22.9 (2.3)
Other ^a	16.3 (1.8)
Number of Chronic Conditions, %(SE)	
None	31.4(2.2)
1–2	51.8(2.5)
3–4	14.9(1.7)
5	1.8(0.6)
Physical Limitation, %(SE)	24.3(2.1)
Work Limitation ^b , %(SE)	80.7(3.5)
Cognitive Limitation, %(SE)	6.8(1.0)

Unweighted N=420, sum of weights=90,100,450; PCS12: Physical Component Summary, MCS12: Mental Component Summary;

^aOther includes cancers of urinary organs, lymphatic and hematopoietic tissue, head and neck, bone, thyroid, brain and nervous system, secondary cancers, cancer without specification of site, and other gastrointestinal system;

^bOnly 90/420 participants were eligible to answer about work limitation in panel 8 survey.

Table 2.

Summary of Results for Internal Reliability, Test-retest Reliability, Convergent Validity, Discriminant Validity and Predictive Validity

Tests	PCS12	MCS12
Reliability		
Internal Consistency		
Cronbach's Alpha	0.89	0.88
Test-retest Reliability		
Intra-class Correlation Coefficient	0.82	0.73
Validity		
Construct Validity (Convergent and Discriminant), correlation coefficients (r)		
EuroQol Index (EQ-5D-index) *	0.67	0.45
EuroQol Mobility *	0.70	0.18
EuroQol SelfCare *	0.36	0.23
EuroQol Usual Activity *	0.71	0.37
EuroQol Pain/Discomfort *	0.67	0.29
EuroQol Anxiety/Depression *	0.28	0.67
EuroQol VAS *	0.70	0.51
Perceived Health *	0.61	0.42
Perceived Mental Health *	0.38	0.52
Criterion Validity		
<i>Predictive Validity, mean ($\pm$SD)</i>		
Physical Limitations *		
Yes	31.23 ($\pm$ 10.36)	
No	48.18 ($\pm$ 10.16)	
Cognitive Limitations *		
Yes		38.18 ($\pm$ 14.43)
No		50.70 ($\pm$ 10.09)
<i>Predictive Validity, Odds Ratio (95% Confidence Interval)</i>		
Physical limitations (round 3) predicted by PCS12 (round 2): 0.88 (0.85–0.90)		
Cognitive limitations (round 3) predicted by MCS12 (round 2): 0.92 (0.89–0.95)		

VAS: Visual Analog Scale; PCS12: Physical Component Summary; MCS12: Mental Component Summary;

* Statistically significant at $p < 0.001$