ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Comparative-Effectiveness Research/HTA

The Use of Generic Patient-Reported Outcome Measures in Emergency Department Surveys: Discriminant Validity Evidence for the Veterans RAND 12-Item Health Survey and the EQ-5D

Jae-Yung Kwon, PhD, RN, Lena Cuthbertson, MEd, Reg. OT (BC), Richard Sawatzky, PhD, RN

ABSTRACT

Objectives: This study aimed to compare discriminant validity evidence of 2 generic patient-reported outcome measures (PROMs), the Veterans RAND 12-Item Health Survey (VR-12) and level 5 of EQ-5D (EQ-5D-5L), for use in emergency departments (EDs).

Methods: Data were obtained via a cross-sectional survey of 5876 patients in British Columbia (Canada) who completed a questionnaire after visiting an ED in 2018. We compared the extent to which the VR-12 and the EQ-5D-5L distinguished among groups of ED patients with different levels of comorbidity burden and self-reported physical and mental or emotional health status. Multivariable logistic regression was used to evaluate the ability of the 2 PROMs to identify patients presenting with a mental health (MH) condition.

Results: All the measures produced small effect sizes (ESs) for discriminating comorbidity levels (R^2 range: 0.00 [VR-12 mental component summary {MCS}] to 0.10 [VR-12 physical component summary score]). The EQ-5D visual analog scale offered the largest ES for discriminating self-reported physical health (R^2 = 0.48), whereas the MCS, the VR-12 MH domain, and the EQ-5D-5L anxiety/depression dimension had the largest ESs for discriminating self-reported mental or emotional health (R^2 = 0.42, 0.40, and 0.38, respectively). The MCS produced a medium ES (R^2 = 0.42) along with the VR-12 utility score (R^2 = 0.27) compared with the EQ-5D-5L index (R^2 = 0.19). Having a MH condition was predominantly identified by the MCS (Pratt index = 0.56).

Conclusions: The VR-12 PROM provides a more comprehensive measurement of MH than the EQ-5D-5L, which is important to inform healthcare service needs for patients who present in EDs with MH challenges.

Keywords: comorbidity, discriminant validity, EQ-5D, Veterans RAND 12-Item Health Survey.

VALUE HEALTH. 2022; 25(12):1939–1946

Introduction

Emergency departments (EDs) provide critical first-line healthcare for a broad spectrum of patients in need of urgent care.[1] The past decade has seen increased workload and overcrowding in EDs.[2,3] In the current fast-paced nature of patient management, patient-reported signs and symptoms may often be overlooked.[4] One way to efficiently capture patients' perspective of health is through the use of patient-reported outcome measures (PROMs). Two widely used generic PROMs (ie, not condition specific) that have potential for routine collection and use in EDs are the Veterans RAND 12-Item Health Survey (VR-12[5,6]) and level 5 of EQ-5D (EQ-5D-5L[7]).[8] The VR-12 (derived from the 12-Item Short Form Health Survey [SF-12][9]) includes 12 items, which are summarized into a physical component summary (PCS) and mental component summary (MCS) scores for the measurement of physical and mental health (MH).[6] The EQ-5D is a family of instruments that use a preference-based descriptive system for valuing health by deriving a utility score from individuals' responses to 5 questions measuring severity in mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.[10]

Previous studies have shown that the EQ-5D and the VR-12 (along with SF-12) have been used to discriminate between patients with different severity levels of chronic conditions (eg, arthritis, pain, cardiovascular, and cancer[11–15]) and MH.[16] These findings showed that these instruments were not interchangeable. For example, the SF-12 component scores were able to differentiate patients with and without chronic conditions when similar patients reported no problems on the EQ-5D dimensions,[12,14,15] suggesting that the SF-12 might be more suitable for measuring the health of populations with morbidity. For MH, both the EQ-5D and the SF-6D (preference-based measure derived from the SF-12) discriminated between severity subgroups, with slightly larger health gains for the EQ-5D for subgroups with the highest severity

of MH problems,[16] yet no studies have examined discriminative validity evidence of the VR-12 and the EQ-5D-5L specifically in the ED context. Evidence on discriminant validity can help inform the selection of a PROM and the interpretation of its scores for patient-centered decision making within the ED population.

Although both the VR-12 and the EQ-5D have widely been used to assess health burden of general populations,[14,15] they are fundamentally different from a theoretical perspective. As part of the "SF family of instruments" (including versions 1 and 2 of the SF-36 and SF-12), the VR-12 is based on psychometric measurement theory for obtaining summary scores based on responses from individuals to multiple questions about the presence, frequency, or intensity of their physical and MH status. Responses to individual questions are then aggregated to create a global summary scale. Psychometric health measures have been widely used to assess treatment outcomes and to compare the effectiveness and performance of healthcare programs and delivery systems.[17] In contrast, the EQ-5D is based on economic utility theory, which is designed to facilitate the calculation of quality adjusted life-years required for conducting cost-utility analysis and to help with resource allocation decisions.[18] A utility is a specific type of health assessment that ranges from 0 (anchored to dead) to 1 (anchored as complete health) with values < 0 indicating a health state worse than dead. The health states are determined from individuals' responses to 5 questions about the severity of particular health dimensions. Each health state is then assigned a utility value, which conceptually reflects societal preferences for different health states based on the different levels of risk that people in a general population are willing to take to reach a better health status.[19]

As part of an initiative to enhance public accountability and to support local and system level quality improvement, the British Columbia (BC) Office of Patient-Centred Measurement implements provincially coordinated sector surveys, including in the ED sector in BC. The more recent 2018 ED Sector Survey marks the first time that generic PROMs, such as the VR-12 and the EQ-5D, were included in data collection.[20] These data offer a unique opportunity to examine how well the VR-12 can discriminate among patients who present to EDs, how it compares with the EQ-5D-5L, and how the summary scores of these instruments perform in assessing the health burden in the ED population. Accordingly, the overarching goal of this research was to compare the distributions of the summary scores and the discriminant validity evidence for the VR-12 and EQ-5D-5L instruments with respect to indicators of both physical and MH status. In particular, our objectives were to compare their distributional properties, concordance, and ability to discriminate among groups of patients differentiated by levels of comorbidity burden, self-reported physical and mental or emotional health status, and having a MH condition.

Methods

Data Sources

This was a secondary analysis of cross-sectional data from the 2018 ED Sector Survey[21] in BC, Canada. The data were provided by the BC Office of Patient-Centred Measurement, which is mandated by the BC Ministry of Health and the 7 health authorities to both measure and monitor the quality and safety of the healthcare system in the province from the perspective of those who receive services. A full description of the recruitment and data collection process is published elsewhere.[20] Patients were recruited from across the 108 BC EDs between January 1 and March 31, 2018. Census samples (survey invitations sent to all discharged patients) were conducted for facilities with fewer than 350 ED discharges whereas all other facilities were simple randomly sampled to achieve a balanced number of completions from each facility. All patients selected were contacted within 3 weeks after discharge from the ED to reduce recall bias. Patients were recruited by an invitation letter and the surveys were completed either by phone interview (with a surveyor specialized in administering health study surveys) or online (by self-logging onto the study URL using their unique survey ID provided in the invitation letter). Although more than 99% of the sample of participants completed the questionnaire in English, there was an option for non-English speakers to complete in a different language (including French, German, Spanish, Chinese [traditional or simplified], and Korean) with a surveyor who was fluent in the preferred language. From this original patient recruitment (N = 43 887), a total of 15 604 patients (35.6%) completed the survey. To obtain comprehensive comorbidity information, respondents were linked to the National Ambulatory Care Reporting System (NACRS),[22] a data file from 29 EDs that reported data to NACRS, resulting in a total of 8063 patients.

Measures

The VR-12 includes items with 3- to 6-point response options (eg, all of the time to none of the time) and is derived from the Veterans RAND 36 Item Health Survey, which was adapted from the RAND Short-Form 36 Version 1.0 (see website: https://www.bu.edu/sph/about/departments/health-law-policy-and-management/research/vr-36-vr-12-and-vr-6d/).[5,6] The 12 items were selected based on the most effective approximations of the 8 domain scores without administering the additional 24 questions. These domain scores include physical functioning (PF), role limitations due to physical problems, bodily pain, general health (GH) perceptions, energy and vitality, social functioning, role limitations due to emotional problems, and MH. For this study, the VR-12 questions asked patients to reflect on a 1-week recall period. PCS and MCS scores were transformed to a t-score metric with a mean of 50 and a standard deviation of 10 based on 2019 Canadian norms.[23] Mode effect adjustments were applied to account for social desirability bias resulting from phone versus online survey completion. In addition, the VR-12 utility score was obtained based on responses to 8 of the VR-12 items[24] and generated based on the Canadian general population preferences ranging from −0.59 for worst to 1.0 for best health states.[23]

The EQ-5D-5L is a preference-based descriptive system that provides a utility score based on a single question for mobility, self-care, usual activities, pain/discomfort, and anxiety/depression dimensions and includes a visual analog scale (VAS) (EuroQol VAS [EQ-VAS]) using a vertical scale ranging from 0 to 100 (worst to best imaginable) (see website: https://eq-5d-demo.euroqol.org/?).[10] The EQ-5D initially had 3 levels and was revised to 5 response levels (5L): 1 ("no problems"), 2 ("slight problems"), 3 ("moderate problems"), 4 ("severe problems"), and 5 ("extreme problems") to describe greater range of health status.[10] The EQ-5D-5L describes 3125 distinct health states, with 11111 representing the best and 55555 the worst possible health states. The EQ-5D-5L index scores were calculated using a health utility algorithm based on the Canadian general population preferences, which ranged from −0.15 (worst) to 0.95 (best health state).[25]

Comorbidities were assessed using the Charlson comorbidity index (CCI) as a proxy measure of the patients' overall disease burden.[26] The CCI assesses the comorbid status of patients based on 17 conditions in the International Classification of Diseases, with each condition adjusted for the relative risk of 1-year mortality. Patients were classified into 4 groups based on the above administrative health data sources: no comorbidity (CCI score = 0),

mild (CCI scores of 1-2), moderate (CCI scores of 3-4), and severe (CCI scores ≥ 5).

Physical and mental or emotional health status were measured using 2 items that were not part of the PROMs. One item from the ED Patient Experiences with Care[27] asked "In general, how would you rate your overall physical health?"; the other from a global rating question[28] "In general, how would you rate your overall mental or emotional health?" Responses were provided on 5-point scale from "excellent" to "poor." An indicator of having a MH condition was derived based on the Canadian Emergency Department Information System classification of presenting complaints for MH problems and substance use[29] (supplemented with diagnosis of depression from the linked data set).[30] Sociodemographic variables included age, gender, highest level of education, and ethnicity.

Statistical Analysis

All data sources were housed within the Secure Research Environment of Population Data BC (a multiuniversity data resource that offers researchers a centralized secure location for the access and processing of large collections of health data). Statistical analyses were conducted using R statistical software version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria)[31] for preparing the data, SAS statistical software version 9.4 (SAS Institute Inc, Cary, NC)[32] for calculating the CCI, and Mplus statistical software version 8.3 (Muthén & Muthén, Los Angeles, CA)[33] for multiple imputation and modeling.

From the linked NACRS cohort (n = 8063), patients who did not complete any items in the VR-12 or the EQ-5D were excluded (n = 2187) for a total of 5876 patients. Descriptive statistics were computed with means and standard deviations or frequency counts and percentages. To address missing data (3.2% total), multiple imputation (Bayesian estimation method) was used with clustering at the facility level to account for correlation among patients. Auxiliary variables included sociodemographic and survey items to increase the accuracy of the imputed values. Based on recommendations,[34] 20 imputed data sets were created for analyses.

The distributional properties were assessed using the scatterplot and histograms whereas Spearman correlational analyses were used to assess concordance between the VR-12 (domains, PCS, MCS, and utility) and the EQ-5D-5L (dimensions, index, and VAS) scores to account for ordinal/nonnormal data (eg, dimensions of the EQ-5D-5L). Discriminant validity of the VR-12 (domains, PCS, MCS, and utility) and the EQ-5D-5L (dimensions, index, and VAS) scores was assessed based on "known-groups" comparisons to examine the extent each instrument distinguished between patients based on the level of comorbidity burden (not reported, mild, moderate, severe) and self-reported physical and mental or emotional health status (poor, fair, good, very good, and excellent). Ordinal logistic regression was used to compare the groups clustered by facility[35] and obtain effect sizes (ESs) describing the magnitudes of differences based on the pseudo R-squared, with ES interpreted as small (0.10), medium (0.30), and large (0.50) based on Cohen's criterion.[36]

To further examine discriminant validity with respect to MH concerns, multivariable logistic regression was conducted to compare the extent to which the 2 measures explain variation in MH conditions. Summary scores and demographics (gender, age, ethnicity, education, CCI) were selected following recommended purposeful variable selection approach.[37] Variables having a univariate test at the alpha cutoff of 0.25 were initially included in the multivariate analysis. In the iterative process of variable selection, covariates were retained for the final multivariable model if they attained an alpha level of 0.1 and change in parameter estimates was greater than 20% (only ethnicity was removed). Initial variables not selected for the original multivariate model were then reintroduced to determine whether there were substantial changes (> 20%) in any of regression coefficients. ESs were based on Pratt index,[38] which quantifies the relative contributions of each variable to the overall explained variance of the multivariable logistic regression model (based on likelihood-ratio R-squared).

We hypothesized that the VR-12 summary and utility scores as well as the EQ-5D-5L index score and VAS would be lower with higher levels of comorbidity burden. We expected the VR-12 summary and utility scores as well as the EQ-5D-5L index score and VAS to be higher in patients with higher ratings of physical and mental or emotional health status. Finally, because of greater coverage of the items, we expected the VR-12 to better discriminate patients having a MH condition than the EQ-5D-5L.

Results

The average age of the respondent sample was 57.4 years (SD 19.9). Most were white (70.1%) and women (54.7%). The sample had some level of college education and a few reported severe comorbidity burdens. For 5836 of the respondents (99.3%), the surveys were administered in English whereas the rest (40 respondents; 0.7%) were administered in other languages (see Table 1).

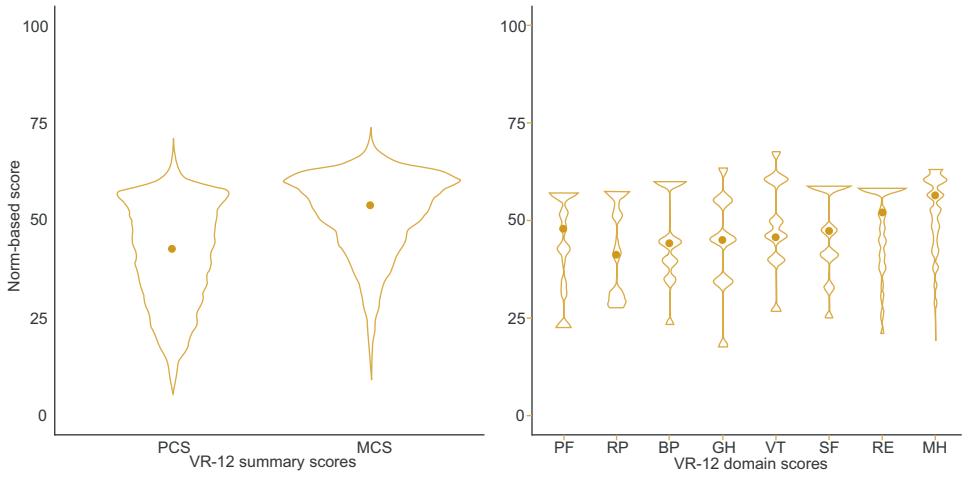
The stacked bar graphs show the distributions of the VR-12 items (see Appendix 1 Fig. 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.07.016) and the EQ-5D-5L

Table 1. Patients' characteristics.

Variables	Total (n = 5876, 100%) f (%)
Age (n = 5872), mean (SD)	57.4 (19.9)
13-17	149 (2.5)
18-64	3314 (56.4)
≥ 65	2409 (41.0)
Ethnicity (n = 5876)	
White	3925 (66.8)
Asian	868 (14.8)
Indigenous	324 (5.5)
Other	759 (12.9)
Gender (n = 5872)	
Women	3217 (54.7)
Highest level of education (n = 5638)	
< High school	1096 (18.7)
High school	1348 (23.9)
College	1529 (27.1)
Undergraduate	966 (17.1)
Postgraduate	699 (12.4)
CCI (n = 5876), mean (SD)	1.0 (2.0)
0 (no documented comorbidities)	3909 (66.5)
1-2 (mild)	1085 (18.5)
3-4 (moderate)	518 (8.8)
≥ 5 (severe)	364 (6.2)
English version instrument administration	5836 (99.3)
Non-English instrument administration	40 (0.7)

Note. f indicates frequency. Asian included South Asian, Southeast Asian, Chinese, Korean, and Japanese. Other ethnicity included Latin American, black, multiple ethnicity and other ethnicity, prefer not to answer, or do not know. CCI indicates Charlson comorbidity index.

Figure 1. Violin plot of the VR-12 summary and domain scores.



Note: based on 20 imputed data sets. BP indicates bodily function; GH, general health; MCS, mental component summary; MH, mental health; PCS, physical component summary; RE, role emotional; RF, role functioning; RP, role physical; SF, social functioning; VR-12, Veterans RAND 12-Item Health Survey; VT, vitality.

responses (see Appendix 1 Fig. 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.07.016). For the VR-12 items, a third of patients reported poor or fair GH (34%), and more than half reported MH challenges (eg, 53% feeling energetic only some to none of the time). For the EQ-5D-5L items, most patients (ranging from 65% to 91%) reported having no to slight problems for all items.

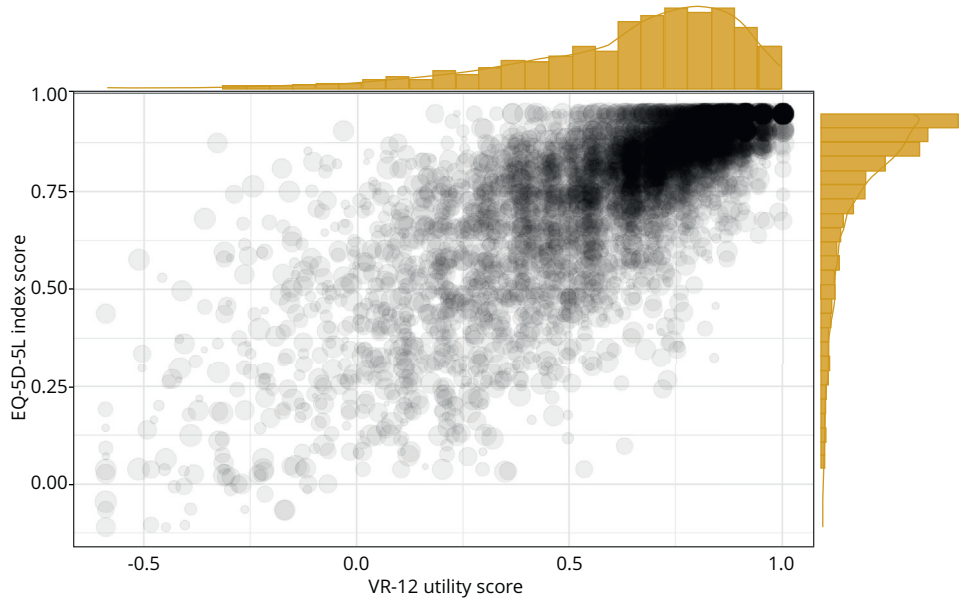
Figure 1 illustrates the distributions of the VR-12 summary and domain scores. The PCS was lower than the MCS, indicating that patients who presented to BC EDs predominantly experienced physical rather than mental impairments. In general, the PCS and MCS were relatively normally distributed (mean [SD] = 41.29 [12.79], range = 5.25-71.17; skewness = −0.39 and kurtosis = −0.78 for PCS; mean [SD] = 51.79 [10.18], range = 8.02-73.65; skewness = −0.94 and kurtosis = 0.49 for MCS), with some outliers observed at the lower end for MCS. Among the 8 domain scales, the role limitations due to physical problems had the broadest interquartile range whereas the GH had the highest range.

Table 2. VR-12 and EQ-5D-5L correlations.

	PCS	MCS	PF	RP	BP	GH	VT	SF	RE	MH	VR-12 utility	Mobility	Self-care	Usual act.	Pain	Anxiety	EQ-5D-5L index
MCS	0.13																
PF	0.86	0.17															
RP	0.84	0.31	0.68														
BP	0.75	0.30	0.53	0.60													
GH	0.61	0.35	0.51	0.48	0.40												
VT	0.57	0.58	0.50	0.55	0.46	0.52											
SF	0.48	0.68	0.44	0.54	0.46	0.40	0.52										
RE	0.28	0.76	0.35	0.44	0.37	0.36	0.43	0.50									
MH	0.21	0.85	0.26	0.34	0.36	0.36	0.51	0.53	0.53								
VR-12 utility	0.73	0.65	0.69	0.73	0.72	0.54	0.71	0.75	0.58	0.68							
Mobility	−0.66	−0.21	−0.64	−0.56	−0.48	−0.43	−0.43	−0.39	−0.31	−0.25	−0.55						
Self-care	−0.46	−0.24	−0.45	−0.44	−0.38	−0.31	−0.32	−0.36	−0.31	−0.24	−0.48	0.49					
Usual act.	−0.70	−0.32	−0.64	−0.68	−0.57	−0.46	−0.52	−0.52	−0.39	−0.35	−0.68	0.63	0.54				
Pain	−0.63	−0.29	−0.51	−0.53	−0.69	−0.42	−0.43	−0.41	−0.33	−0.34	−0.60	0.53	0.38	0.57			
Anxiety	−0.18	−0.63	−0.23	−0.27	−0.29	−0.33	−0.39	−0.44	−0.55	−0.62	−0.50	0.23	0.22	0.31	0.33		
EQ-5D-5L index	0.71	0.44	0.66	0.67	0.66	0.52	0.57	0.57	0.49	0.48	0.76	−0.75	−0.57	−0.79	−0.82	−0.54	
EQ-VAS	0.60	0.39	0.54	0.55	0.46	0.64	0.56	0.46	0.39	0.41	0.62	−0.50	−0.38	−0.55	−0.49	−0.37	0.62

Note. Based on Spearman correlation.
Act. indicates activities; BP, bodily function; EQ-5D-5L, level 5 of EQ-5D; EQ-VAS, EuroQol visual analog scale; GH, general health; MCS, mental component summary; MH, mental health; PCS, physical component summary; PF, physical functioning; RE, role emotional; RF, role functioning; RP, role physical; SF, social functioning; VR-12, Veterans RAND 12-Item Health Survey; VT, vitality.

Figure 2. Relationship between the VR-12 utility and the EQ-5D-5L index score.

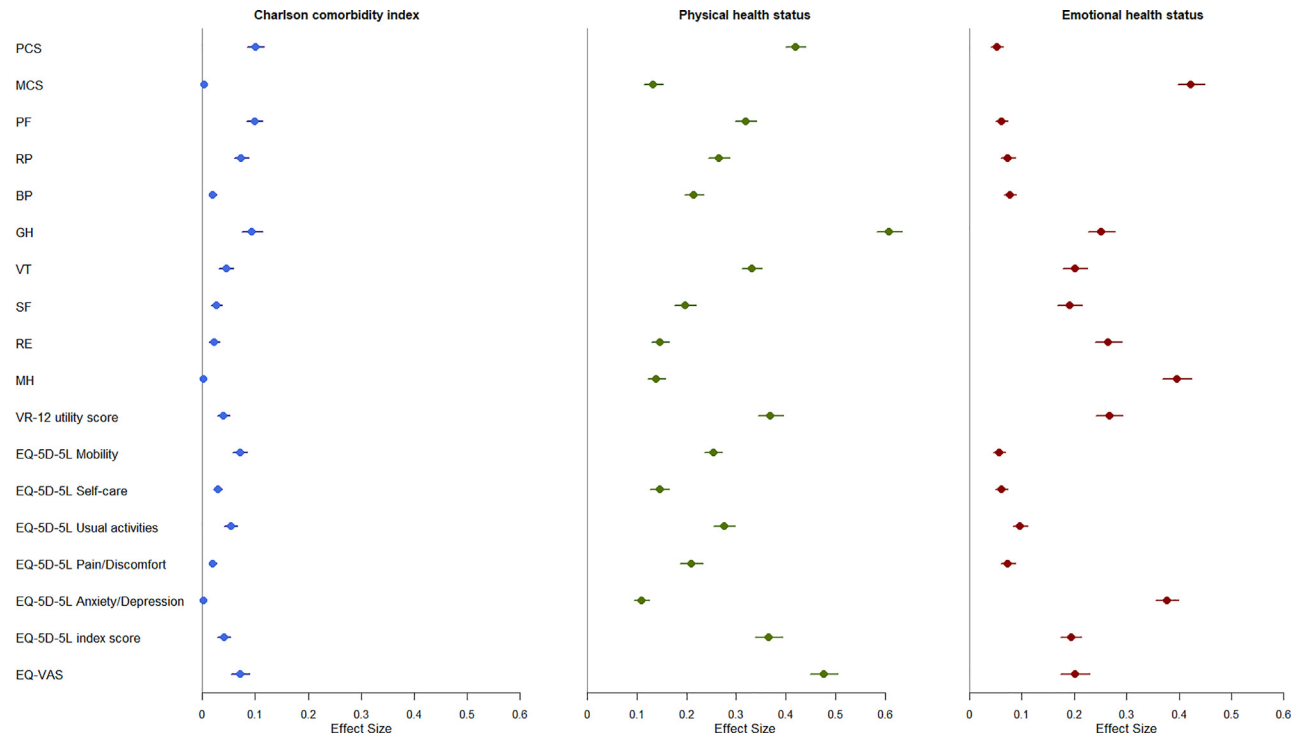


Note: based on 20 imputed data sets. EQ-5D-5L indicates level 5 of EQ-5D; VR-12, Veterans RAND 12-Item Health Survey.

Concordance Between the VR-12 and the EQ-5D-5L

Between the 2 instruments, the highest correlation was for the VR-12 utility score and the EQ-5D-5L index score (r = 0.76) (see Table 2). The EQ-5D-5L index score had higher correlation with the PCS (r = 0.71) than the MCS (r = 0.44), whereas the MCS had higher correlation with the EQ-5D-5L anxiety/depression (r = −0.63) than other dimensions (range: r = −0.21 [mobility] to −0.32 [usual activity]). The EQ-VAS had higher correlations with the GH domain, VR-12 utility score, and PCS (r = 0.64, 0.62, and 0.60, respectively).

Figure 3. Discriminant validity effect sizes for comorbidity and physical and emotional health.



Effect size = pseudo R-squared based on logistic regression models. Error bars = 95% confidence intervals. BP indicates bodily function; EQ-5D-5L, level 5 of EQ-5D; EQ-VAS, EuroQol visual analog scale; GH, general health; MCS, mental component summary; MH, mental health; PCS, physical component summary; PF, physical functioning; RE, role emotional; RP, role physical; SF, social functioning; VR-12, Veterans RAND 12-Item Health Survey; VT, vitality.

Table 3. Adjusted odds ratios for comparing patients with and without mental health conditions.

Variable	Has mental health conditions (n = 377)	Has no mental health conditions (n = 5499)	Bivariate OR (95% CI)	Adjusted OR (95% CI)	Pratt index
Age, mean (SD)	49.20 (20.7)	57.98 (19.7)	0.98 (0.97-0.98)	0.98 (0.98-0.99)	0.12
Women, %	55.4	54.7	1.03 (0.85-1.26)	-	-
White, %	62.9	67.1	-	-	-
Asian, %	14.3	14.8	0.96 (0.68-1.37)	-	-
Indigenous, %	9.5	5.2	1.91 (1.32-2.78)	-	-
Other ethnicity, %	13.3	12.9	1.03 (0.78-1.38)	-	
Less than high school, %	24.9	19.1	1.40 (1.14-1.73)	1.23 (0.94-1.60)	0.01
High school, %	22.8	24.0	0.94 (0.73-1.20)	0.85 (0.60-1.20)	0.00
College, %	28.4	27.0	-	-	-
Undergraduate, %	16.2	17.3	0.93 (0.69-1.25)	0.79 (0.58-1.09)	0.00
Postgraduate, %	7.7	12.7	0.57 (0.38-0.86)	0.66 (0.44-1.01)	0.01
CCI, mean (SD)	2.24 (2.3)	1.98 (2.0)	1.06 (1.02-1.10)	1.09 (1.05-1.13)	0.02
PCS, mean (SD)	42.25 (13.4)	41.22 (12.8)	1.01 (1.00-1.02)	1.02 (1.00-1.04)	0.03
MCS, mean (SD)	42.58 (12.3)	52.44 (9.7)	0.93 (0.92-0.93)	0.93 (0.91-0.95)	0.56
VR-12 utility, mean (SD)	0.47 (0.4)	0.64 (0.3)	0.19 (0.14-0.25)	2.55 (1.02- 6.39)	0.00
EQ-5D-5L index, mean (SD)	0.67 (0.3)	0.78 (0.2)	0.12 (0.07-0.18)	0.12 (0.05-0.26)	0.24

Note. Adjusted model R-square = 21%.
CI indicates the 95% confidence interval; CCI, Charlson comorbidity index; EQ-5D-5L, level 5 of EQ-5D; MCS, mental component summary; OR, odds ratio; PCS, physical component summary.

Further comparison of the marginal distributions of the VR-12 utility and the EQ-5D-5L index score showed that the VR-12 utility score was skewed to the right (mean [SD] = 0.63 [0.28], range = −0.59 to 1.00; skewness = −1.29, kurtosis = 1.72) whereas the EQ-5D-5L index score showed even more skewness and kurtosis (mean [SD] = 0.77 [0.20], range = −0.15 to 0.95; skewness = 2.42, kurtosis = 2.42) with most scoring near the perfect range (mode = 0.95) (see Fig. 2). In contrast, the EQ-VAS was relatively normally distributed (mean [SD] = 67.86 [20.37], range = 1-100; skewness = −0.87, kurtosis = 0.46) (not shown).

Discriminant Validity of VR-12 and EQ-5D-5L

Patients with a higher comorbidity burden had lower VR-12 domain, summary, and utility scores and more problems on the EQ-5D-5L dimensions with lower EQ-5D-5L index and VAS scores (see Fig. 3 and Appendix 1 Tables 1-3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.07.016 for details).

Across the groups defined by severity of comorbidity burden, the ESs for the VR-12 summary scores were 0.10 for the PCS and 0.00 for the MCS, whereas the utility score was 0.04. The ESs for the VR-12 domain scores ranged from 0.00 (MH) to 0.10 (PF). The ES for the EQ-5D-5L index score was 0.04 and the VAS score was 0.07. The ESs for the EQ-5D-5L dimensions ranged from 0.00 (anxiety/depression) to 0.07 (mobility).

Across the groups defined by different levels of physical health, the ESs for the VR-12 summary scores were 0.42 for the PCS and 0.13 for the MCS, whereas the utility score was 0.37. The ESs for the VR-12 domain scores ranged from 0.14 (MH) to 0.61 (GH). The ES for the EQ-5D-5L index score was 0.37 and for the VAS score was 0.48. The ESs for the EQ-5D-5L dimensions ranged from 0.11 (anxiety/depression) to 0.28 (usual activities).

Across the groups defined by different levels of mental or emotional health, the ESs for the VR-12 summary scores were 0.05 for the PCS and 0.42 for the MCS, whereas the utility score was 0.27. The ESs for the VR-12 domain scores ranged from 0.06 (PF) to 0.40 (MH). The ES for the EQ-5D-5L index score was 0.19 and the VAS score was 0.20. The ESs for the EQ-5D-5L dimensions ranged from 0.06 (mobility and self-care) to 0.38 (anxiety/depression).

Variable importance ES based on multivariable logistic regression revealed that most of the explained variance (21%) could be attributed to the MCS (Pratt index = 0.56), followed by the EQ-5D-5L index score (0.24), adjusted for age, education, and CCI (see Table 3).

Discussion and Conclusion

To the best of our knowledge, this is the first study to compare findings arising from the use of the VR-12 and the EQ-5D-5L in EDs. We found that the EQ-5D-5L index score was more strongly correlated with the VR-12 PCS than the MCS, whereas only the EQ-5D-5L anxiety/depression dimension was strongly correlated with the MCS compared with the rest of the EQ-5D-5L dimensions. This result suggests that the EQ-5D-5L predominantly measures physical health status and, to a lesser extent, MH status. We also found the VR-12 component summary and utility scores to be more normally distributed than the EQ-5D-5L index score, which was mostly in the perfect score range, and suggest a ceiling effect. The discriminant validity results showed that the ESs for both the VR-12 PCS and the EQ-5D-5L index scores were small for discriminating comorbidity levels whereas the EQ-VAS showed a large ES for discriminating different levels of physical health. Nevertheless, the VR-12 MCS was found to be more effective than

the EQ-5D-5L index score in discriminating between patients with different levels of self-reported mental or emotional health status and having a MH condition.

There may be several reasons for the observed differences between the 2 instruments. The VR-12 captures 8 health dimensions that assess the impact of or the interference with activities of daily life, whereas the EQ-5D-5L measures the level of severity for 5 dimensions (each based on a single item). In the assessment of MH, the VR-12 MCS is comprised of 4 domains (total of 6 items) (role emotional [2 items], vitality [1 item], MH [2 items], social functioning [1 item]) with 3- to 6-point response options (eg, all of the time to none of the time), whereas the EQ-5D-5L index score has only 1 anxiety/depression item on a 5-point scale (eg, no problem to extreme problem). Although the anxiety/depression item on the EQ-5D-5L had high correlation with the VR-12 MCS, an increase in the number of items to assess MH and response options would lead to a more descriptive and sensitive instrument. For example, a patient who might be a bit anxious might indicate "no problem" on the EQ-5D-5L anxiety/depression item, whereas on the VR-12 the same patient may indicate "a little of the time" to the item asking "have you felt calm and peaceful?" and "a little of the time" to the item "have you felt downhearted and blue?." In addition, the recall periods of the 2 instruments differ; the VR-12 asks respondents to reflect on a past period of time (in our study, we used a recall period of 1 week) and the EQ-5D-5L asks respondents to rate their current level of health "today." Taken together, these differences may partly explain the greater variability and lower ceiling effects of the VR-12 relative to the EQ-5D-5L, which may in turn explain improved ability of the VR-12 to discriminate among groups of patients differentiated by levels of comorbidity burden, self-reported physical and, in particular, mental or emotional health status, and having a MH condition.

The EQ-5D-5L index score reflects the values that "society" places on different health states (based on economic utility theory) with negative values that theoretically correspond to health states worse than dead, which may not entirely be accounted for by measures of physical and MH concerns.[39] Interestingly, the VR-12 utility score was able to discriminate between patients with different levels of comorbidity burden and physical health status comparable with the EQ-5D-5L index score (and even better for mental or emotional health status). As such, our findings suggest that even though the VR-12 was not originally developed based on utility theory, the VR-12 utility scores nonetheless provide results that are comparable with those based on the EQ-5D-5L.

In addition to the discriminant validity evidence from this study, other aspects of validity should be considered. Notably, the content of the VR-12 results in a more comprehensive picture of "mental or emotional health," whereas the EQ-5D includes only a single question about the severity of anxiety/depression. Furthermore, because the VR-12 and the EQ-5D-5L yield different (although related) information about respondent's health (based on different underlying measurement theories), clinicians, researchers, and policy makers are advised to consider the appropriateness of the information obtained from these PROMs[40] for their selection and use in ED settings.

A limitation of this study is that the patient-reported data included in our analysis represented respondents from only 29 of 108 BC EDs (see Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.07.016). Although this may limit generalizability of our results (given that this patient cohort may differ from those that do not report NACRS data, such as small-volume EDs in BC), the linkage allowed us to obtain comprehensive comorbidity information for each respondent. Another possible limitation and area for further research is how other ethnic groups may have influenced the results of the study given that most of the cohort was white. Finally, because the PROMs data are self-reported, it is also possible that those who did not respond to either the VR-12 or the EQ-5D-5L may have reported better or worse MH; nonetheless, our analysis shows that these instruments are not interchangeable, particularly in distinguishing patients with self-reported emotional or MH status and having a MH condition.

Results of this analysis suggest that the EQ-5D-5L may be a less descriptive and sensitive measure of MH status compared with the VR-12. Thus, the use of the EQ-5D-5L in ED settings may result in a less than optimal assessment of MH related concerns where patients would more likely report "no problem" on the EQ-5D-5L, which could lead to missed opportunities to intervene. The VR-12 is, therefore, recommended for use in EDs in which there is interest in identifying patient populations with MH challenges. Our findings indicate these 2 measurement instruments are clearly not interchangeable; this is important information for researchers, healthcare decision makers, and policy makers to consider when selecting a generic PROM for use in EDs.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2022.07.016.

Article and Author Information

Accepted for Publication: July 20, 2022

Published Online: August 30, 2022

doi: https://doi.org/10.1016/j.jval.2022.07.016

Author Affiliations: School of Nursing, Trinity Western University, Langley, BC, Canada (Kwon, Sawatzky); School of Nursing, University of Victoria, Victoria, BC, Canada (Kwon); Office of Patient-Centred Measurement, British Columbia Ministry of Health, Vancouver, BC, Canada (Kwon, Cuthbertson); BC SUPPORT Unit, Patient-Centred Measurement Methods Cluster, Vancouver, BC, Canada (Kwon, Cuthbertson, Sawatzky); Evaluation and Outcome Sciences, Providence Health Care Research Institute, Vancouver, BC, Canada (Sawatzky); Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden (Sawatzky).

Correspondence: Jae-Yung Kwon, PhD, RN, School of Nursing, HSD Bldg A402A, PO Box 1700 STN CSC, Victoria, BC V8W 2Y2, Canada. Email: jykn1@uvic.ca

Author Contributions: *Concept and design:* Kwon, Sawatzky, Cuthbertson
Acquisition of data: Sawatzky, Cuthbertson
Analysis and interpretations of data: Kwon, Sawatzky
Drafting of the manuscript: Kwon, Sawatzky, Cuthbertson
Critical revision of the paper for important intellectual content: Kwon, Sawatzky, Cuthbertson
Statistical analysis: Kwon, Sawatzky
Obtaining funding: Cuthbertson, Sawatzky
Administrative, technical, or logistic support: Cuthbertson
Supervision: Sawatzky, Cuthbertson

Conflict of Interest Disclosures: The authors reported no conflicts of interest.

Funding/Support: This study was supported by the BC SUPPORT Unit (Patient-Centred Measurement Methods Cluster) as part of the Postdoctoral Fellowship program between Trinity Western University and the BC Ministry of Health's Office of Patient-Centred Measurement. This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program in support of Dr Sawatzky's Research Chair in Person-Centred Outcomes at Trinity Western University.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and

interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: The authors thank Dr Lara Russell for preliminary analyses informing this project, Dr Nick Bansback and Dr Logan Trenaman for their research to develop the utility scoring algorithms, and the BC Office of Patient-Centred Measurement for collecting the PROMs data and supporting the project.

REFERENCES

1. Pines JM, Hilton JA, Weber EJ, et al. International perspectives on emergency department crowding. *Acad Emerg Med*. 2011;18(12):1358–1370.
2. Rasouli HR, Esfahani AA, Nobakht M, et al. Outcomes of crowding in emergency departments; a systematic review. *Arch Acad Emerg Med*. 2019;7(1):e52.
3. Brennan JJ, Chan TC, Hsia RY, Wilson MP, Castillo EM. Emergency department utilization among frequent users with psychiatric visits. *Acad Emerg Med*. 2014;21(9):1015–1022.
4. Abar B, Holub A, Lee J, DeRienzo V, Nobay F. Depression and anxiety among emergency department patients: utilization and barriers to care. *Acad Emerg Med*. 2017;24(10):1286–1289.
5. Kazis LE, Lee A, Spiro A, et al. Measurement comparisons of the Medical Outcomes Study and Veterans SF-36® Health Survey. *Health Care Financ Rev*. 2004;25(4):43–58.
6. Kazis LE, Miller DR, Skinner KM, et al. Applications of methodologies of the Veterans Health Study in the VA Healthcare System: conclusions and summary. *J Ambul Care Manage*. 2006;29(2):182–188.
7. EQ-5D-5L user guide: basic information on how to use the EQ-5D-5L instrument. EuroQol Group. https://euroqol.org/publications/user-guides/. Accessed February 22, 2022.
8. PROMs background document. Canadian Institute for Health Information. https://www.cihi.ca/sites/default/files/proms_background_may21_en-web_0.pdf. Accessed February 22, 2022.
9. Ware J, 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(3):220–233.
10. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
11. Conner-Spady BL, Marshall DA, Bohm E, Dunbar MJ, Noseworthy TW. Comparing the validity and responsiveness of the EQ-5D-5L to the Oxford hip and knee scores and SF-12 in osteoarthritis patients 1 year following total joint replacement. *Qual Life Res*. 2018;27(5):1311–1322.
12. De Smedt D, Clays E, Annemans L, De Bacquer D. EQ-5D versus SF-12 in coronary patients: are they interchangeable? *Value Health*. 2014;17(1):84–89.
13. Tawiah AK, Al Sayah F, Ohinmaa A, Johnson JA. Discriminative validity of the EQ-5D-5 L and SF-12 in older adults with arthritis. *Health Qual Life Outcomes*. 2019;17(1):68.
14. Johnson JA, Coons SJ. Comparison of the EQ-5D and SF-12 in an adult US sample. *Qual Life Res*. 1998;7(2):155–166.
15. Johnson JA, Pickard AS. Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. *Med Care*. 2000;38(1):115–121.
16. Lamers LM, Bouwmans CAM, van Straten A, Donker MCH, Hakkaart L. Comparison of EQ-5D and SF-6D utilities in mental health patients. *Health Econ*. 2006;15(11):1229–1236.
17. Sutherland HJ, Till JE. Quality of life assessments and levels of decision making: differentiating objectives. *Qual Life Res*. 1993;2(4):297–303.
18. Revicki DA, Kaplan RM. Relationship between psychometric and utility-based approaches to the measurement of health-related quality of life. *Qual Life Res*. 1993;2(6):477–487.
19. Feeny DH, Torrance GW. Incorporating utility-based quality-of-life assessment measures in clinical trials: two examples. *Med Care*. 1989;27(3 Suppl):S190–S204.
20. R.A. Malatest & Associates. 2018 emergency department sector survey: technical report. BCPCM. https://www.bcpcm.ca/tools-and-resources. Accessed February 22, 2022.
21. British Columbia Patient-Centred Measurement Working Group: Emergency Department Sector Survey 2018 data set. Data available. Population Data BC. http://www.popdata.bc.ca/data. Accessed April 8, 2021.
22. British Columbia Ministry of Health: National Ambulatory Care Reporting System data set (NACRS). Data available. Population Data BC. http://www.popdata.bc.ca/data. Accessed April 8, 2021.
23. Bansback N, Trenaman L, Mulhern B, et al. Estimation of a Canadian preference-based scoring algorithm for the Veterans RAND 12-Item Health Survey: a population survey using a discrete-choice experiment. *CMAJ Open*. 2022;10(3):E589–E598.
24. Selim AJ, Rogers W, Qian SX, Brazier J, Kazis LE. A preference-based measure of health: the VR-6D derived from the Veterans RAND 12-Item Health Survey. *Qual Life Res*. 2011;20(8):1337–1347.
25. Xie F, Pullenayegum E, Gaebel K, et al. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Med Care*. 2016;54(1):98–105.
26. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43(11):1130–1139.
27. Weinick RM, Becker K, Parast L, et al. Emergency department patient experience of care survey: development and field test. *Rand Health Q*. 2014;4(3):5.
28. Ahmad F, Jhajj AK, Stewart DE, Burghardt M, Bierman AS. Single item measures of self-rated mental health: a scoping review. *BMC Health Serv Res*. 2014;14:398.
29. Bullard MJ, Musgrave E, Warren D, et al. Revisions to the Canadian emergency department triage and acuity scale (CTAS) guidelines 2016. *CJEM*. 2017;19(S2):S18–S27.
30. Tonelli M, Wiebe N, Fortin M, et al. Methods for identifying 30 chronic conditions: application to administrative data [published correction appears in *BMC Med Inform Decis Mak*. 2019;19(1):177]. *BMC Med Inform Decis Mak*. 2015;15:31.
31. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org. Accessed February 22, 2022.
32. SAS/ACCESS® 9.4 interface to ADABAS. SAS Institute Inc. https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/acadbas/titlepage.htm. Accessed February 22, 2022.
33. Muthén LK, Muthén BO. Mplus User's Guide. 8th ed. Los Angeles, CA: Muthén & Muthén; 1998-2017.
34. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci*. 2007;8(3):206–213.
35. Kim S. ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Commun Stat Appl Methods*. 2015;22(6):665–674.
36. Cohen J. *Statistical Power Analysis for the Behavioral Science*. Hillsdale, NJ: L Erlbaum Associates; 1988.
37. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York, NW: Wiley; 2000.
38. Thomas DR, Hughes E, Zumbo BD. On variable importance in linear regression. *Soc Indic Res*. 1998;45(1/3):253–275.
39. Gandhi M, Rand K, Luo N. Valuation of health states considered to be worse than death—an analysis of composite time trade-off data from 5 EQ-5D-5L valuation studies. *Value Health*. 2019;22(3):370–376.
40. Hawkins M, Elsworth GR, Nolte S, Osborne RH. Validity arguments for patient-reported outcomes: justifying the intended interpretation and use of data. *J Patient Rep Outcomes*. 2021;5(1):64.