Patient-Reported Outcomes

A Direct Comparison of the Measurement Properties of EQ-5D-5L, PROMIS-29+2 and PROMIS Global Health Instruments and EQ-5D-5L and PROPr Utilities in a General Population Sample

Fanni Rencz, PhD, Valentin Brodszky, PhD, Mathieu F. Janssen, PhD

A B S T R A C T

Objectives: We aimed to compare measurement properties of the 5-level version of EQ-5D (EQ-5D-5L) and 2 Patient-Reported Outcomes Measurement Information System (PROMIS) short forms, PROMIS-29+2 and PROMIS Global Health (PROMIS-GH-10), and of EQ-5D-5L and PROMIS-preference scoring system (PROPr) utilities.

Methods: A cross-sectional survey was conducted in a general population sample in Hungary (N = 1631). We compared the following measurement properties at the level of items, domains, and utilities, the latter using corresponding US value sets: ceiling and floor, informativity (Shannon's indices), agreement, convergent, and known-group validity. For the analyses, PROMIS items/domains were matched to EQ-5D-5L domains that cover similar concepts of health.

Results: The majority of PROMIS items showed enhanced distributional characteristics, including lower ceilings and higher informativity than the EQ-5D-5L. Good convergent validity was established between EQ-5D-5L and PROMIS domains capturing similar aspects of health. Mean EQ-5D-5L utilities were substantially higher than those of PROPr (0.864 vs 0.535). EQ-5D-5L utilities correlated moderately or strongly with PROPr (r = 0.61), PROMIS-GH-10 physical (r = 0.68), and mental health summary scores (r = 0.53). EQ-5D-5L utilities decreased with age, whereas PROPr utilities slightly increased with age. EQ-5D-5L utilities discriminated significantly better in 12/28 (ratio of F-statistics) and 18/26 (area under the receiver-operating characteristics curve ratio) known groups defined by age, self-perceived health status, and self-reported physician-diagnosed health conditions, including hypertension, diabetes, coronary heart disease, chronic kidney disease, and stroke.

Conclusions: This study provides comparative evidence on the measurement properties of EQ-5D-5L, PROMIS-29+2, and PROMIS-GH-10 and informs decisions about the choice of instruments in population health surveys for assessment of patients' health and for cost-utility analyses.

Keywords: EQ-5D-5L, generic preference-accompanied measures, PROMIS, psychometrics, utilities.

Introduction

Generic health status measures are increasingly used to inform decision making in public health and healthcare policy. These instruments are intended to assess general aspects of health without targeting any specific condition or treatment, and thus, they are appropriate for use across diverse patient groups and the general population. Specific type of generic health status measures are preference-accompanied measures, which have a set of health utilities (ie, value set) reflecting societal preferences for all health states possibly described by the instrument. In cost-utility analyses, health utilities are needed to estimate the impact of health interventions in terms of quality-adjusted life-years (QALYs). Examples of commonly used preference-accompanied measures include the EQ-5D, SF-6D, Health Utilities Index (HUI), Assessment of Quality of Life, and 15D.[1,2]

The Patient-Reported Outcomes Measurement Information System (PROMIS) initiative, launched by the National Institutes of Health in the United States in 2004, has developed item banks for over 100 different health domains.[3] The item development benefited from advanced psychometric methods, such as item response theory.[4] The item banks enable administration both through computerized adaptive testing and custom or fixed-lengths short forms, for example, PROMIS-29, -43, and -57 adult profile measures and the PROMIS Global Health (PROMIS-GH-10).[5,6] Seven PROMIS domains were selected to be included in descriptive system suitable for valuation,[7] and a value set, the PROMIS-Preference scoring system (PROPr), was developed for these based on standard gamble method.[8]

In recent years, a rapid growth has been seen in the utilization of PROMIS instruments in the United States, and its international use is also gradually expanding.[9] It is a commonly accepted approach to compare emerging instruments with well-established ones for which positive evidence of measurement properties is already available in multiple populations and contexts. The EQ-5D is the most widely used generic preference-accompanied measure that is also recommended in numerous pharmacoeconomic guidelines around the world.[10-13] There is a large body of evidence supporting its validity and reliability across a broad range of populations and conditions.[14,15] Originally the EQ-5D had 3 response levels in each domain (EQ-5D-3L) that was later extended to 5 (EQ-5D-5L).[16,17] In many populations, the EQ-5D-5L shows improved distributional characteristics, informativity, construct validity, and responsiveness.[15,18-20]

Taking into account the type, content, size, and purpose of the instrument, 2 existing PROMIS instruments are suitable for a comparison with the EQ-5D-5L as generic measure of health. The first one, PROMIS-29, is the shortest PROMIS adult profile measure, an extended version of which PROMIS-29+2 can be used to generate PROPr utilities.[5,8] The second one, PROMIS-GH-10, is a short form consisting of global items to evaluate health in general rather than focusing on specific elements of health.[6] Several studies have contrasted the measurement properties of PROMIS-29 and PROMIS-GH-10 or PROPr and EQ-5D utilities.[6,21-33] Nevertheless, earlier comparative studies had their own limitations, including (1) using the EQ-5D-3L when EQ-5D-5L was available,[30,31] (2) administering the PROMIS-29 and mapping cognitive function domain scores from other domain responses instead of using the PROMIS-29+2,[25,27,32,33] (3) focusing on only 1 or a limited number of measurement properties,[6,21-25,27-33] (4) providing a comparison of the characteristics of value sets without using any respondent data,[26] (5) relying on a value set developed before the introduction of the EuroQol Valuation Technology protocol,[22-24] or (6) testing the questionnaires in a limited number of languages (English, German, French, or Dutch).[6,21-25,27-33] Our objective is, therefore, to compare measurement properties of the Hungarian version of EQ-5D-5L with those of the PROMIS-29+2 and PROMIS-GH-10 instruments and EQ-5D-5L and PROPr utilities in Hungary.

Methods

Data Collection

An online cross-sectional survey was carried out involving a large general population sample in Hungary (n = 1700). Adults aged 18 years or over that gave their informed consent were eligible to participate in the study. Data were collected by a panel company in November 2020. Panel members received points as a reward for completing the survey. "Soft" quotas were used to obtain a broadly representative sample across age, gender, education, place of living, and region according to census data from the Hungarian Central Statistical Office. The Research Ethics Committee of the Corvinus University of Budapest granted approval for the survey (no. KRH/343/2020).

Measures

The Hungarian versions of EQ-5D-5L, PROMIS-29+2 v2.1, and PROMIS-GH-10 v1.2 were administered in a fixed order as part of a longer questionnaire (Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.002).[34] Participants were also asked about their age, gender, level of education, marital and employment status, place of living, region, and self-perceived

health status (5-point rating scale). Data on health conditions were recorded using a 2-step approach. First, using a predefined list of common chronic health conditions or chronic consequences of acute conditions, respondents were asked to report whether they experienced one or more conditions in the last 12 months. The items on the list were selected from the European Health Interview Survey supplemented by a few other prevalent conditions in the population.[35] In the second step, respondents required to indicate those conditions that had been diagnosed by a physician. There were no missing data on the survey because all questions were compulsory.

5-level version of EQ-5D

EQ-5D-5L is a preference-accompanied measure composed of 2 parts: a descriptive system and a visual analogue scale with endpoints of 0 (the worst imaginable health) and 100 (the best imaginable health).[17] The descriptive system covers 5 domains of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each consisting of 1 item. All 5 items have 5 severity-format response levels ("no problems" to "extreme problems"/"unable to") that together define overall 3125 health states. Both the descriptive system and EQ visual analogue scale use the day of completion (today) as a recall period. The US EQ-5D-5L value set was used to derive utilities that had been developed using the composite time trade-off method.[36] The theoretical range of EQ-5D-5L utilities was −0.573 to 1.

PROMIS measures

PROMIS-29+2, a preference-accompanied measure, is an extended version of the PROMIS-29 short form profile measure. PROMIS-29 covers the following 7 domains of health: physical function, anxiety, depression, fatigue, sleep disturbance, ability to participate in social roles and activities (hereafter referred to as social roles), and pain interference.[37] Each domain is represented by 4 items. In addition, the measure assesses pain intensity using a 0-10 numeric rating scale. In PROMIS-29+2, the 7 core domains are supplemented by an eighth domain of cognitive function (Cognitive Function Abilities v2.0 short form), which consists of 2 items. Six domains use a 7-day recall period, whereas the physical function and social roles domains do not specify a recall period. For all domains, the items use a 5-point response scale varying across levels of difficulty (ie, "without any difficulty" to "unable to do"), frequency ("never" to "always"), severity ("not at all" to "very much"), and global rating ("very poor" to "very good") formats. For all 8 domains, raw scores are computed by adding up respective item scores measured on a 4 to 20 scale for the 7 core domains and 2 to 10 for cognitive function. A higher score represents more of the concept being measured; thus, for scales of function (ie, physical function, social role, and cognitive function) a higher score corresponds to a better health status, and for all other domains, a higher score indicates worse health status. Utilities may be computed based on available responses because the 7 PROMIS domains needed for PROPr utilities are all covered by PROMIS-29+2.[8] Building on PROMIS-29+2, the total number of unique health states described by PROPr is 217 238 121, and the theoretical range of utilities is −0.022 to 0.954. Nevertheless, if using other PROMIS measures (eg, custom short forms for the 7 domains), it is possible for PROPr to reach 1.

PROMIS-GH-10 is a 10-item measure that includes ratings of 4 general items (health, quality of life, physical health, and mental health) and 6 other items capturing 5 domains of health: physical function (1 item), fatigue (1 item), pain (1 item, identical to the pain intensity scale of PROMIS-29+2), emotional distress (1 item), and social health (2 items).[6] Each question has 5 response choices,

Table 1. Ceiling, floor, informativity, and correlation between matched EQ-5D-5L and PROMIS-29+2 and PROMIS-GH-10 items.

EQ-5D-5L					PROMIS					EQ-5D-5L vs PROMIS			
Domain	Ceiling, %	Floor, %	Shannon's index (H')*	Shannon's evenness index (J')*	Item	Ceiling, %	Floor, %	Shannon's index (H')*	Shannon's evenness index (J')*	ΔCeiling, %[†]	ΔFloor, %[†]	Δ J'	r_s[‡]
MO	70.4	0.1	1.29 (1.23-1.36)	0.56 (0.53-0.58)	PF2 - going up and down stairs at a normal pace	65.1	1.4	1.48 (1.42-1.54)	0.64 (0.61-0.67)	−5.3[§]	1.3[§]	0.08[§]	−0.61
					PF3 - going for a walk of at least 15 minutes	83.0	1.4	0.94 (0.87-1.02)	0.41 (0.37-0.44)	12.6[§]	1.3[§]	−0.15[§]	−0.54
					Global06 - physical activities (eg, walking, climbing stairs)	58.4	1.5	1.6 (1.55-1.65)	0.69 (0.67-0.71)	−12[§]	1.5[§]	0.13[§]	−0.55
SC	92.5	0.2	0.5 (0.43-0.56)	0.21 (0.19-0.24)	-	-	-	-	-	-	-	-	-
UA	78.8	0.1	1.01 (0.95-1.08)	0.43 (0.41-0.46)	PF1 - vacuuming and yard work	77.4	1.2	1.11 (1.04-1.18)	0.48 (0.45-0.51)	−1.3	1.1[§]	0.05	−0.58
					PF4 - run errands and shop	84.3	0.9	0.87 (0.8-0.95)	0.38 (0.35-0.41)	5.5[§]	0.7[§]	−0.06	−0.49
					SR1 - doing regular leisure activities with others	50.9	2.8	1.8 (1.75-1.85)	0.77 (0.75-0.8)	−27.9[§]	2.6[§]	0.34[§]	−0.45
					SR2 - doing family activities	60.7	2.5	1.62 (1.56-1.68)	0.7 (0.67-0.72)	−18.1[§]	2.3[§]	0.26[§]	−0.42
					SR3 - doing usual work (include work at home)	50.2	1.6	1.76 (1.71-1.81)	0.76 (0.74-0.78)	−28.6[§]	1.5[§]	0.32[§]	−0.48
					SR4 - doing activities with friends	55.7	3.4	1.76 (1.71-1.82)	0.76 (0.74-0.78)	−23.1[§]	3.3[§]	0.32[§]	−0.45
					Global05 - satisfaction with social activities and relationships	14.6	6.3	2.13 (2.1-2.17)	0.92 (0.91-0.93)	−64.2[§]	6.2[§]	0.48[§]	−0.31
					Global09 - usual social activities and roles	15.9	3.8	2.07 (2.03-2.1)	0.89 (0.87-0.9)	−62.9[§]	3.7[§]	0.46[§]	−0.30
PD	56.2	0.5	1.5 (1.45-1.55)	0.64 (0.62-0.67)	P1 - pain interference with day to day activities	56.7	1.7	1.65 (1.6-1.71)	0.71 (0.69-0.74)	0.5	1.2[§]	0.07[§]	0.59
					P2 - pain interference with work around the home	59.8	2.8	1.63 (1.57-1.69)	0.7 (0.68-0.73)	3.6[§]	2.3[§]	0.06[§]	0.56
					P3 - pain interference with social activities	67.1	2.4	1.48 (1.41-1.54)	0.64 (0.61-0.66)	10.9[§]	1.9[§]	−0.01	0.53
					P4 - pain interference with household chores	60.4	2.4	1.6 (1.54-1.66)	0.69 (0.68-0.73)	4.2[§]	1.9[§]	0.04[§]	0.55
					Pain intensity 0-10	31.6	6.6	2.92 (2.87-2.97)	0.84 (0.83-0.86)	−24.6[§]	6.1	0.23[§]	0.63
					F1 – fatigued[‖]	33.7	3.4	1.99 (1.94-2.03)	0.86 (0.84-0.87)	−22.4[§]	2.9[§]	0.21[§]	0.44
					F2 - trouble starting things[‖]	44.5	3.6	1.89 (1.85-1.94)	0.81 (0.79-0.84)	−11.7[§]	3.1[§]	0.17[§]	0.47
					F3 - run-down[‖]	41.1	2.7	1.9 (1.85-1.94)	0.82 (0.8-0.84)	−15[§]	2.2[§]	0.17[§]	0.47
					F4 - fatigued on average[‖]	34.4	3.6	1.96 (1.92-2)	0.84 (0.82-0.86)	−21.8[§]	3.1[§]	0.2[§]	0.44
					Global08 - fatigue[‖]	24.0	1.0	1.85 (1.82-1.89)	0.8 (0.78-0.82)	−32.1[§]	0.6	0.15[§]	−0.49
					Global07 - pain	30.1	6.9	2.91 (2.86-2.96)	0.88 (0.83-0.86)	−26.1[§]	6.4	0.23[§]	0.63
AD	66.1	1.0	1.36 (1.30-1.42)	0.59 (0.56-0.61)	A1 - fearful	53.5	1.5	1.71 (1.65-1.76)	0.73 (0.71-0.76)	−12.6[§]	0.5	0.15[§]	0.63
					A2 - hard to focus on anything other than anxiety	68.7	1.1	1.38 (1.31-1.44)	0.59 (0.56-0.62)	2.6[§]	0.1	0.01	0.62
					A3 - overwhelming worries	57.3	2.5	1.68 (1.63-1.74)	0.73 (0.7-0.75)	−8.8[§]	1.5[§]	0.14[§]	0.61
					A4 - uneasy	40.6	2.6	1.92 (1.88-1.97)	0.83 (0.81-0.85)	−25.5[§]	1.6[§]	0.24[§]	0.62
					D1 - worthless	64.9	2.1	1.53 (1.47-1.6)	0.66 (0.63-0.69)	−1.2	1.2[§]	0.07[§]	0.57
					D2 - helpless	51.9	2.3	1.81 (1.76-1.86)	0.78 (0.76-0.8)	−14.2[§]	1.3[§]	0.2[§]	0.59
					D3 - depressed	68.2	1.9	1.45 (1.38-1.52)	0.62 (0.6-0.65)	2.1[§]	0.9[§]	0.04[§]	0.73
					D4 - hopeless	61.6	3.3	1.64 (1.57-1.7)	0.7 (0.68-0.73)	−4.5[§]	2.3[§]	0.12[§]	0.64
					Global04 - mental health (incl. mood and ability to think)	15.0	5.7	2.14 (2.11-2.17)	0.92 (0.91-0.93)	−51.1[§]	5.6[§]	0.33[§]	−0.55
					Global10 - emotional problems (anxious, depressed or irritable)	38.4	2.4	1.97 (1.93-2.01)	0.85 (0.83-0.87)	−27.7[§]	1.4[§]	0.26[§]	−0.65
-	-	-	-	-	C1 - being able to concentrate	41.2	6.1	1.87 (1.83-1.92)	0.81 (0.61-0.66)	-	-	-	-

continued on next page

Table 1. Continued

EQ-5D-5L					PROMIS					EQ-5D-5L vs PROMIS			
					C2 - being able to remember to do things	51.3	6.6	1.75 (1.7-1.81)	0.76 (0.66-0.71)	-	-	-	-
-	-	-	-	-	S1 - sleep quality	17.5	2.6	1.98 (1.94-2.02)	0.85 (0.83-0.87)	-	-	-	-
					S2 - refreshing sleep	11.4	16.2	2.21 (2.18-2.24)	0.95 (0.94-0.96)	-	-	-	-
					S3 - problem with sleep	42.2	3.3	1.95 (1.91-2)	0.84 (0.82-0.86)	-	-	-	-
					S4 - difficulty falling asleep	44.1	4.5	1.97 (1.93-2.01)	0.85 (0.83-0.87)	-	-	-	-
-	-	-	-	-	Global01 - health	10.2	5.2	2.07 (2.03-2.1)	0.89 (0.87-0.91)	-	-	-	-
					Global02 - quality of life	9.5	4.6	2.01 (1.97-2.05)	0.86 (0.85-0.88)	-	-	-	-
					Global03 - physical health	9.1	6.3	2.06 (2.02-2.1)	0.89 (0.87-0.9)	-	-	-	-
EQ VAS	7.2	2.0	5.6 (5.67-5.53)	0.84 (0.83-0.85)	-			-	-	-	-	-	-

AD indicates anxiety/depression; EQ-5D-5L, 5-level version of EQ-5D; MO, mobility; PD, pain/discomfort; PROMIS, Patient-Reported Outcomes Measurement Information System; SC, self-care; UA, usual activities; VAS, visual analogue scale.
*Confidence intervals for Shannon's indices were computed with 3000 bootstrap replications.
†The difference in ceiling and floor was tested by McNemar's test.
‡All Spearman's correlation coefficients were statistically significant ($P < .05$).
§Indicates a statistically significant difference between the EQ-5D-5L domain and its matched PROMIS item pair ($P < .05$).
‖Fatigue PROMIS items were matched to PD as fatigue was considered a form of discomfort.

with the exception of the pain scale. PROMIS-GH-10 uses severity, frequency, and global rating format response levels similar to PROMIS-29+2. Three items (emotional distress, fatigue, and pain) are assessed for the previous 7 days; in 6 items the recall period is unanchored (ie, 'in general'), and the physical function item has no recall period. Global physical and mental health summary scores can be computed, each using responses of the 4 corresponding items.[6]

Statistical Analyses

We observed some inconsistencies in the data and the decision was taken to exclude 69 respondents before the data analysis (Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.002).

The EQ-5D-5L and PROMIS measures were compared at the level of items, domains, and utilities.[38] The analytical strategy adopted here builds on previous studies on the comparative performance of different generic health status measures.[19,39-42] Whereas the EQ-5D-5L has a Hungarian value set,[43] a national value set for PROPr has so far been developed only in the United States.[8] For the sake of consistency, we decided to use US value sets for both measures to generate utilities.[8,36]

Item- and domain-level analyses

To study floor and ceiling, the distribution of responses classified into the worst and best possible health state (ie, profile) on one measure were examined across the items of the other measure(s). Floor and ceiling were also assessed for the individual items of all measures by computing the relative frequencies of respondents selecting the worst and best possible response options. PROMIS items were matched to EQ-5D-5L items that cover the same or similar concept of health based on the authors' expert opinion. We created 25 and 7 item pairs with the PROMIS-29+2 and PROMIS-GH-10, respectively (Table 1). Lower ceiling and higher floor were assumed for most PROMIS items because of their longer recall period, positively worded items and "non-severity" response scales.[44,45] The difference in ceiling and floor between the EQ-5D-5L and PROMIS measures was tested by

McNemar's test. P values $< .05$ were considered statistically significant.

Convergent and divergent validity across EQ-5D-5L and PROMIS items were tested by Spearman's correlations. We hypothesized moderate or strong correlations between items or domains aiming to capture similar aspects of health. Correlation coefficients were interpreted as very weak (< 0.20), weak (0.20-0.39), moderate (0.40-0.59), strong (0.60-0.79), and very strong (≥ 0.80).[46]

Informativity was analyzed by using Shannon's indices.[47,48] Shannon's index (H′) captures absolute informativity, whereas Shannon's evenness index (J′) measures the relative informativity of the distribution adjusted for the number of response categories.[39] The following formulas were used to compute Shannon's indices:

$$H' = - \sum_{i=l}^{L} p_i \, log_2 p_i$$

$$J' = \frac{H'}{log_2 L}$$

in which L denotes the number of response options in an item and p_i is the percentage of respondents choosing the ith response option.[39] Both for H′ and J′, a higher score indicates that more information is captured by the item. We calculated 95% confidence intervals (CIs) for H′ and J′ using a bootstrap method with 3000 replications.[49] We hypothesized that more variation would be captured by PROMIS items using milder and (often) positive frequency and global rating labels.

Comparison of utilities

Histograms were plotted to visually assess the distribution of the utilities. Ceiling and floor were detected by computing the relative frequencies of respondents scoring the highest and lowest score in the range of the value set. To assess efficiency of use of the utility scale, we used Shannon's indices as described above.

Nevertheless, the nearly 70 000-fold higher number of possible health states with the PROPr could limit the comparison. Thus, to create comparable categories, we divided the utility range using a fixed bin width of 0.05.[39]

The degree of agreement between EQ-5D-5L and PROPr utilities was determined based on an intraclass correlation coefficient (ICC) and a Bland-Altman plot. ICC was computed using a 2-way random-effects model based on an absolute agreement.[50] The ICC value was interpreted as follows: 0-0.39, poor; 0.40-0.59, fair; 0.60-0.74, good; and 0.75 to 1, excellent.[51] In the Bland-Altman plot, a mean difference close to 0 with 95% of the observations falling between the limits of agreement represented a good agreement.[52]

Convergent and divergent validity tests involved Pearson's correlation analysis across EQ-5D-5L and PROPr domains and utilities. We expected strong correlations between EQ-5D-5L and PROPr utilities.[22,24,25,27,32,33]

EQ-5D-5L and PROPr utilities were compared by using Wilcoxon signed-rank test. Both parametric (analysis of variance) and nonparametric (receiver-operating characteristics curve [ROC]) tests were used to assess whether the EQ-5D-5L or PROPr utility scores are able to discriminate across known groups of respondents based on age groups (18-24, 25-34, 35-44, 45-54, and 65+), self-perceived health status (excellent, very good, good, fair, and poor), and the presence of any self-reported physician-diagnosed health conditions (those with a certain condition vs "healthy" group). We hypothesized lower utilities with respect to increasing age, worse health status, and health conditions. An instrument's relative efficiency in distinguishing between groups was defined as the ratio of F-statistics in the analysis of variance or area under the ROC (AUROC) curve values. We used PROPr as a reference (in the denominator) to calculate relative efficiency, thus a ratio > 1 indicated that EQ-5D-5L was more efficient in distinguishing across groups. We calculated 95% CIs for ratios of F-statistics and AUROC values using 3000 bootstrap replications to test whether the ratios were significantly different from 1. All statistical analyses were performed with Stata 16 (StataCorp LLC, College Station, TX).

Results

Comparing the Descriptive Systems

Distribution, ceiling, and floor

Appendix 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.002 shows the demographic and health-related characteristics of the respondents. Of the 1631 participants, 658 (40.3%) classified into the best health state in the EQ-5D-5L, whereas this was 41 (2.5%) and 27 (1.7%) with the PROMIS 29+2 and PROMIS-GH-10, respectively. Of all theoretically possible health state profiles, respondents used 6.7% in the EQ-5D-5L and 0.0006% in the PROPr (with PROMIS-29+2) (Table 2). In the EQ-5D-5L, 71.2% of the sample reported to be in one of the 10 most common health state profiles, whereas this rate was merely 9.1% for PROPr (with PROMIS-29+2).

For those reporting the best possible health on the EQ-5D-5L, PROMIS items demonstrated some level of impairment (Fig. 1).

The ceiling for the 5 EQ-5D-5L items ranged between 56.2% (pain/discomfort) and 92.5% (self-care) (Table 1). In PROMIS-29+2, the minimum and maximum ceiling were 11.4% (refreshing sleep) and 84.3% (run errands and shop), whereas in PROMIS-GH-10, it ranged from 9.1% (physical health) to 58.4% (physical activities). Floors generally were small for all EQ-5D-5L and PROMIS items, with the highest proportion achieved by the positively phrased PROMIS-29+2 item, refreshing sleep (16.2%).

Of the 25 matched item pairs, PROMIS-29+2 showed lower ceilings in 15 cases (60.0%) and the EQ-5D-5L in 7 cases (28.0%) ($P <$ .05). EQ-5D-5L demonstrated lower floor in 22/25 pairs (88.0%) ($P <$ .05). PROMIS-GH-10 showed significantly lower ceilings in all 7 item pairs ($P <$ .05). EQ-5D-5L exhibited lower floor effects in 5 (71.4%) pairs ($P <$ .05). In the rest of the pairs, there was no statistically significant difference between EQ-5D-5L and PROMIS items.

Informativity

For the 5 EQ-5D-5L items, Shannon's evenness index (J') values ranged between 0.21 (self-care) and 0.64 (pain/discomfort) (Table 1). Range of J' values for PROMIS items was wider varying between 0.38 (run errands and shop) and 0.95 (refreshing sleep) for PROMIS-29+2 and between 0.69 (physical activities) and 0.92 (mental health) for PROMIS-GH-10.

Of the 25 matched item pairs, PROMIS-29+2 showed significantly better relative informativity in 19 pairs (76.0%) and the EQ-5D-5L in 1 pair (4.0%). The remainder pairs did not differ significantly. PROMIS-GH-10 demonstrated significantly better relative informativity in all 7 pairs.

Convergent and divergent validity

In line with our expectations, the majority of 25 PROMIS-29+2 and 7 PROMIS-GH-10 items demonstrated at least moderate correlations with their paired EQ-5D-5L items (Table 1). With regard to the domains, the strongest correlations across domains were observed between EQ-5D-5L anxiety/depression and PROMIS-29+2 anxiety and depression (Table 3). As expected, the sleep disturbance and cognitive function PROMIS-29+2 domains that have no corresponding EQ-5D-5L domains were (very) weakly correlated with all EQ-5D-5L domains, with the exception of the PROMIS-29+2 sleep disturbance versus EQ-5D-5L anxiety/depression. Similarly, the self-care EQ-5D-5L domain showed low correlations with all PROMIS domains with the exception of PROMIS-29+2 physical functioning and pain interference.

Comparison of EQ-5D-5L and PROPr Utilities

Distribution, ceiling floor, and informativity

The EQ-5D-5L utilities were negatively skewed, clustering at 1, whereas PROPr utilities showed an almost even distribution (Fig. 2). A considerably smaller range of the utility scale was used by PROPr than by the EQ-5D-5L (range 0.954 vs 1.318) (Table 2). PROPr utilities indicated a substantially lower ceiling than EQ-5D-5L (2.3% vs 40.3%). Overall 1.0% of all utilities were negative with the EQ-5D-5L, whereas none were with PROPr. When the utility range was divided into 0.05 bins, both the absolute and relative informativity were significantly higher for PROPr.

Agreement

EQ-5D-5L utilities were higher than PROPr utilities in 97.1% of respondents. The mean utility for the EQ-5D-5L was substantially higher than that of PROPr (0.864 vs 0.535, $P <$ .05). The ICC for agreement was fair (0.444 [95% CI −0.211 to 0.744]). The Bland-Altman plot indicated that 95.9% of the difference in utilities were between the limits of agreement (Fig. 2). The lack of agreement tended to increase at the lower end of the utility scale.

Convergent and divergent validity of EQ-5D-5L and PROPr utilities

Physical function, pain interference, and social role PROMIS-29+2 domains showed the strongest, whereas cognitive function and sleep disturbance the lowest, correlations with EQ-5D-5L utilities (Table 3). There was a strong correlation between

Table 2. Comparison of EQ-5D-5L and PROPr health state profiles and utilities.

Characteristics	EQ-5D-5L	PROPr
Health state profiles		
Theoretical number of health state profiles	3125	217 238 121
Observed number of health state profiles	208	1369
Proportion of health state profiles used, %	6.7	0.0006
Floor, %	0	0
Ceiling, %	40.3	2.3
Shannon's index (H')*	4.47 (4.31-4.62)	10.11 (10.05-10.17)
H' max	11.61	27.69
Shannon's evenness index (J')*	0.38 (0.37-0.40)	0.37 (0.36-0.37)
Utilities (US value sets)		
Theoretical range	−0.573 to 1	−0.022 to 0.954
Observed range	−0.318 to 1	0 to 0.954
Mean (SD) utility	0.864 (0.198)	0.535 (0.248)
Median (IQR) utility	0.940 (0.815-1)	0.552 (0.336-0.743)
Proportion of negative utilities, %	1.0	0
Utility range (with bin width of 0.05)		
Shannon's index (H')*	2.91 (2.82-3.00)	4.23 (4.21-4.25)
H' max	5.32	5.32
Shannon's evenness index (J')*	0.55 (0.53-0.56)	0.79 (0.79-0.80)

10 most common health state profiles	Profile[†]	Utility	n	%	Profile[‡]	Utility	n	%
	11111	1.000	658	40.3	10,4,4,4,20,4,20	0.954	38	2.3
	11112	0.943	113	6.9	10,4,4,4,20,6,20	0.910	24	1.5
	11121	0.940	110	6.7	10,4,4,4,20,5,20	0.935	21	1.3
	11122	0.883	89	5.5	8,4,4,4,20,6,20	0.827	13	0.8
	21121	0.844	55	3.4	10,4,4,4,20,7,20	0.902	12	0.7
	21111	0.904	47	2.9	10,4,4,4,20,8,20	0.899	10	0.6
	21221	0.776	25	1.5	2,4,4,4,20,12,20	0.285	9	0.6
	21122	0.787	24	1.5	10,4,4,4,20,9,20	0.889	7	0.4
	11123	0.817	22	1.3	10,4,5,4,20,6,20	0.869	7	0.4
	31121	0.818	18	1.1	10,4,5,4,20,7,20	0.861	7	0.4
	Total	-	1161	71.2	Total	-	148	9.1

EQ-5D-5L indicates 5-level version of EQ-5D; IQR, interquartile range; PROPr, Patient-Reported Outcomes Measurement Information System-Preference scoring system.
*Confidence intervals for Shannon's indices were computed with 3000 bootstrap replications.
[†]Domain order: mobility, self-care, usual activities, pain/discomfort, anxiety/depression.
[‡]Domain order: cognitive function, depression, fatigue, pain interference, physical function, sleep, social roles.

EQ-5D-5L and PROPr utilities. EQ-5D-5L utilities were moderately or strongly correlated with PROMIS-GH-10 summary scores.

### Known-group validity of EQ-5D-5L and PROPr utilities

There were large differences in utilities between EQ-5D-5L and PROPr, with a mean of 0.329 (range 0.176 [stroke] to 0.408 [dysmenorrhea, endometriosis]) (Table 4). Both EQ-5D-5L and PROPr were able to distinguish across the majority of the known groups. EQ-5D-5L discriminated significantly better in 12/28 (ratio of F-statistics) and 18/26 (AUROC ratio) known groups. In almost all of the remaining groups, the EQ-5D-5L demonstrated higher discriminatory power, although the difference between PROPr and EQ-5D-5L was insignificant. Mean EQ-5D-5L utilities decreased with age, whereas the reverse was true for mean PROPr utilities that increased with age.

Discussion

To our knowledge, this is the first study in the literature to compare the measurement properties of PROMIS-29+2 and EQ-

Figure 1. Self-reported health on PROMIS-29+2 and PROMIS Global Health in respondents with 11111 on the EQ-5D-5L (n = 658).
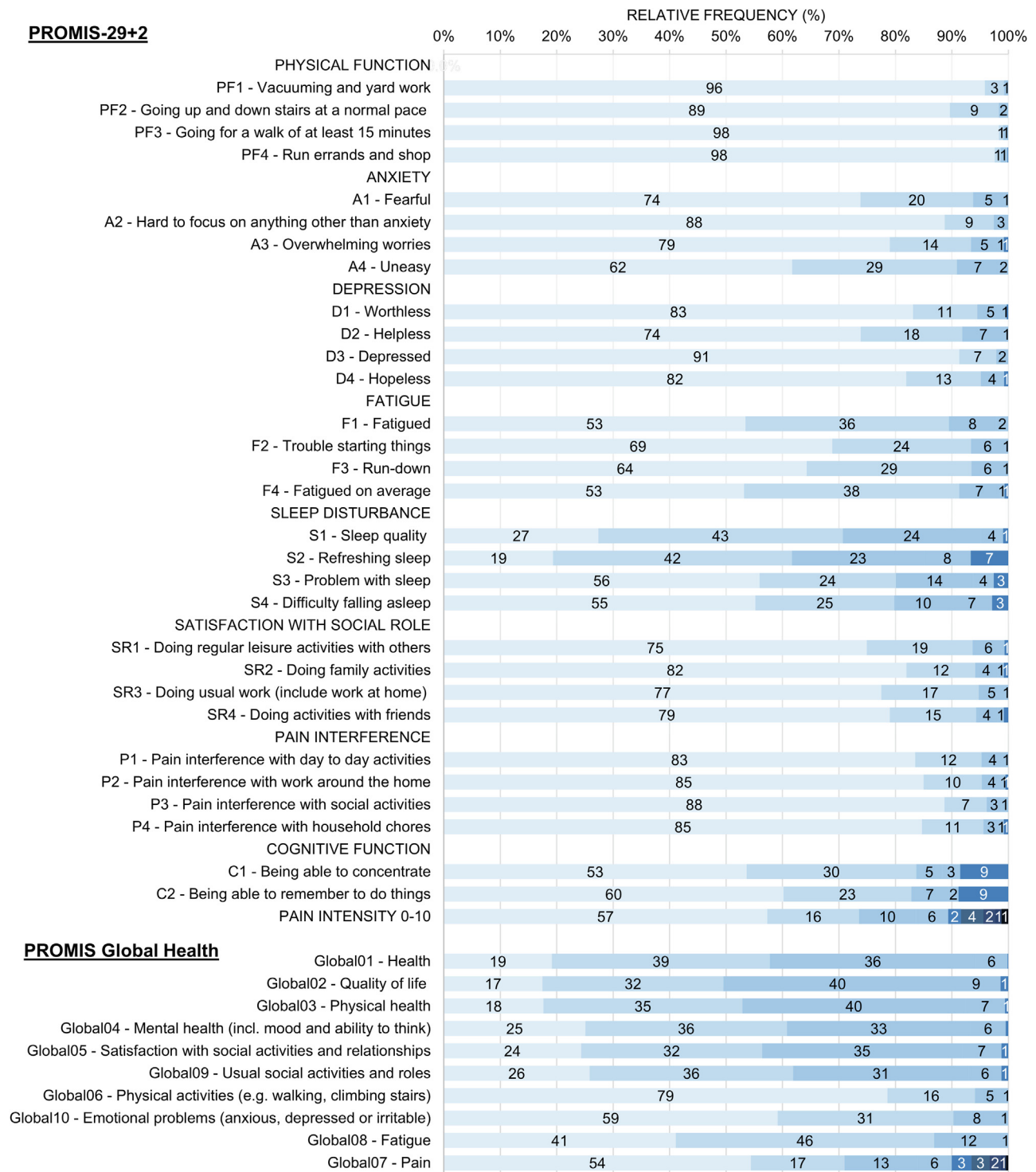
The lightest bars represent the best possible health status, while the darkest bars represent the worst possible health status. The order of items follows their order of appearance in the questionnaires.

5D-5L and also the first to compare EQ-5D-5L and PROPr utilities derived from PROMIS-29+2. Our study used another widely used PROMIS instrument, PROMIS-GH-10, that was also compared with the EQ-5D-5L. A further added value of this research is that, so far, to our knowledge, this has been the most extensive comparative study on these 2 PROMIS short forms and the EQ-5D-5L that included the comparisons of items, domains, and utilities.

One of the largest differences between the descriptive systems is that EQ-5D-5L is conceptualized around any departure from "full" or "normal" health, with the best health state indicating "no

Table 3. Convergent validity of the EQ-5D-5L, PROMIS-29+2, PROPr, and PROMIS-GH-10 domains and summary scores.

Measure		EQ-5D-5L						
		EQ VAS	Mobility	Self-care	Usual activities	Pain/discomfort	Anxiety/depression	EQ-5D-5L utility
PROMIS-29+2	Physical function	0.54	−0.64	−0.59	−0.62	−0.55	−0.31	0.70
	Anxiety	−0.40	0.17	0.19	0.32	0.38	0.71	−0.46
	Depression	−0.44	0.21	0.24	0.36	0.42	0.74	−0.52
	Fatigue	−0.47	0.29	0.26	0.41	0.52	0.58	−0.54
	Sleep disturbance	−0.40	0.21	0.17	0.28	0.38	0.43	−0.38
	Social role	0.52	−0.45	−0.37	−0.52	−0.54	−0.51	0.63
	Pain interference	−0.53	0.55	0.46	0.54	0.62	0.36	−0.66
	Cognitive function	0.18	−0.08	−0.10	−0.14	−0.12	−0.22	0.17
	Pain intensity scale (0-10)	−0.49	0.45	0.34	0.42	0.63	0.33	−0.56
	PROPr utility	0.55	−0.43	−0.32	−0.49	−0.55	−0.55	0.61
PROMIS-GH-10	Mental health summary score	0.52	−0.29	−0.24	−0.37	−0.46	−0.60	0.51
	Physical health summary score	0.66	−0.58	−0.43	−0.56	−0.67	−0.45	0.70

Note. For all EQ-5D-5L domains and PROMIS-29+2 Anxiety, Depression, Fatigue, Sleep disturbance, and pain interference domains and the pain intensity scale, higher scores reflect worse health status. For all other variables, higher scores refer to better health status.
EQ-5D-5L indicates 5-level version of EQ-5D; PROMIS, Patient-Reported Outcomes Measurement Information System; PROMIS-GH-10, PROMIS Global Health; PROPr indicates PROMIS-Preference scoring system; VAS, visual analogue scale.
All correlation coefficients are $P < .05$.

problems" in all domains. PROMIS instruments, however, are overall conceptualized around positive health, including frequency labels indicating that some problems are never experienced (both PROMIS measures) and positive labels, such as "excellent" and "very good" (PROMIS-GH-10). This is reflected by the largest differences in ceilings and in informativity for the PROMIS items compared with EQ-5D using these labels. Furthermore, the distributional characteristics of the PROMIS measures are further enhanced by including domains with more common and frequently occurring problems, such as sleep disturbance and fatigue, but also by including well-being items (eg, hopeless, worthless, social role).

Recall period is a further characteristic that may be responsible for the differences between EQ-5D-5L and PROMIS. The EQ-5D-5L asks the respondent to report a shorter recall period (today) that carries the risk of understating health problems if the respondent is not experiencing those on the day of completion. Many PROMIS-29+2 and PROMIS-GH-10 items use a longer recall period (past 7 days) that could further contribute to their lower ceiling and higher relative informativity. On the other hand, several lines of research have provided evidence that a longer recall period may result in reduced accuracy of self-reporting health problems.[45,53,54]

Although physical health is known to gradually decline with age, the literature is inconclusive regarding the pattern of change in mental and social health with different studies reporting declining, improving, or flat trajectories or a U-shaped curve over the adult lifespan.[55] General population reference values suggest
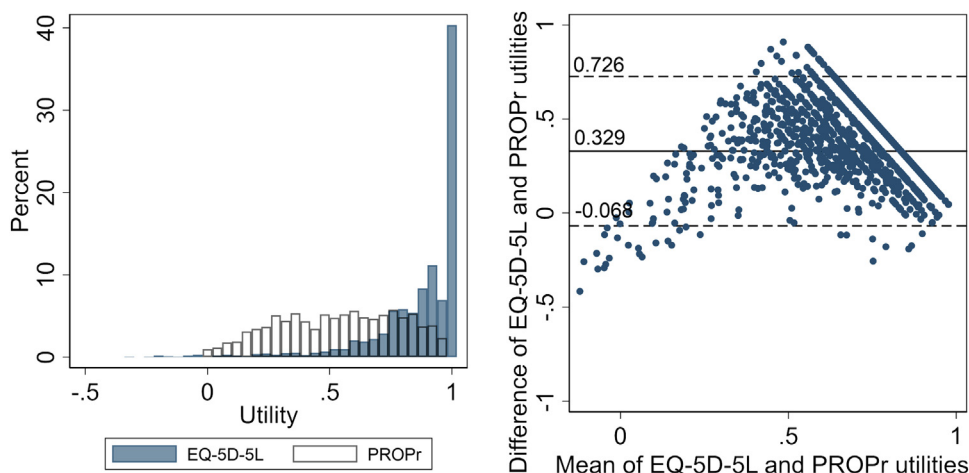
Figure 2. Characteristics of EQ-5D-5L and PROPr utilities.

Table 4. Known-group validity of the EQ-5D-5L and PROPr utilities.

Known groups	n	EQ-5D-5L			PROPr			EQ-5D-5L vs PROPr		
		Mean	SD	Mean difference*	Mean	SD	Mean difference*	Mean difference	Relative efficiency (F-test ratio [EQ-5D-5L/PROPr])[†]	AUROC ratio [EQ-5D-5L/PROPr] (95% CI)[†]
Age groups										
18-24	141	0.912	0.114	-	0.504	0.232	-	0.408	-	-
25-34	284	0.890	0.173	0.022	0.511	0.242	−0.007	0.379	5.873 (2.382-55.276)	n/a
35-44	295	0.895	0.154	−0.004	0.533	0.247	−0.021	0.362		
45-54	281	0.864	0.210	0.031	0.551	0.255	−0.019	0.313		
55-64	287	0.835	0.237	0.029	0.550	0.258	0.001	0.285		
65+	343	0.821	0.220	0.014	0.544	0.244	0.006	0.277		
Self-perceived health status										
Excellent	134	0.972	0.083	-	0.688	0.243	-	0.284	-	-
Very good	386	0.959	0.085	0.013	0.670	0.206	0.018	0.290	1.961 (1.572-2.449)	n/a
Good	658	0.909	0.110	0.051	0.559	0.212	0.110	0.349		
Fair	367	0.755	0.192	0.154	0.380	0.194	0.180	0.375		
Poor	86	0.395	0.327	0.360	0.171	0.147	0.209	0.224		
Physician-diagnosed health conditions[‡]										
Healthy	396	0.955	0.118	-	0.652	0.234	-	0.303	-	-
Allergies	284	0.829	0.211	0.126	0.470	0.244	0.182	0.360	1.012 (0.62-1.58)	1.054 (0.999-1.11)
Dysmenorrhea, endometriosis	52	0.821	0.244	0.134	0.413	0.252	0.239	0.408	0.914 (0.34-2.02)	0.933 (0.838-1.034)
Dermatological diseases	121	0.796	0.247	0.160	0.465	0.241	0.187	0.331	1.611 (0.897-2.687)	1.062 (0.993-1.132)
Cancer (including leukemia, lymphoma)	35	0.787	0.173	0.169	0.437	0.215	0.214	0.349	2.18 (0.81-5.062)	1.143 (1.042-1.243)
Hypertension	477	0.783	0.253	0.173	0.485	0.253	0.167	0.297	1.548 (1.084-2.284)	1.14 (1.091-1.194)
Diabetes	175	0.768	0.273	0.188	0.477	0.254	0.174	0.290	2.054 (1.27-3.278)	1.128 (1.064-1.202)
Asthma	103	0.765	0.259	0.190	0.398	0.222	0.254	0.367	1.207 (0.714-1.973)	1.037 (0.974-1.103)
Cataract	78	0.756	0.239	0.199	0.455	0.237	0.197	0.301	2.658 (1.419-5.005)	1.17 (1.092-1.264)
Glaucoma	23	0.752	0.165	0.204	0.389	0.218	0.263	0.363	2.206 (0.781-5.718)	1.122 (1.004-1.242)
GERD	165	0.751	0.238	0.204	0.394	0.234	0.258	0.357	1.294 (0.815-1.95)	1.074 (1.024-1.124)
Other visual disorders	221	0.750	0.271	0.206	0.404	0.254	0.247	0.345	1.139 (0.759-1.624)	1.061 (1.015-1.111)
Musculoskeletal diseases	491	0.746	0.243	0.209	0.419	0.232	0.233	0.328	1.112 (0.82-1.514)	1.118 (1.075-1.167)
Hyperlipidemia	232	0.746	0.253	0.209	0.419	0.231	0.232	0.327	1.35 (0.924-1.984)	1.11 (1.064-1.161)
Gastric or peptic ulcer	35	0.743	0.204	0.212	0.368	0.249	0.284	0.375	1.905 (0.833-3.981)	1.086 (0.988-1.203)
Bronchitis, emphysema, COPD	72	0.726	0.290	0.229	0.371	0.220	0.281	0.355	1.445 (0.834-2.448)	1.061 (0.985-1.145)
Arrhythmias	144	0.715	0.297	0.240	0.387	0.254	0.265	0.328	1.397 (0.926-2.044)	1.068 (1.017-1.126)
Hearing problems	96	0.711	0.327	0.244	0.434	0.275	0.218	0.278	2.288 (1.414-3.848)	1.141 (1.061-1.222)
Headache, migraine	135	0.697	0.298	0.258	0.339	0.224	0.313	0.358	1.104 (0.74-1.598)	1.039 (0.993-1.085)
Chronic kidney disease	26	0.688	0.320	0.268	0.417	0.201	0.235	0.271	3.622 (1.59-7.116)	1.122 (1.045-1.221)
Other cardiovascular diseases	63	0.661	0.320	0.294	0.362	0.249	0.289	0.299	2.209 (1.283-3.554)	1.122 (1.061-1.201)
Coronary heart disease	61	0.645	0.335	0.311	0.369	0.236	0.283	0.276	2.441 (1.459-3.879)	1.11 (1.049-1.172)
Anxiety	167	0.633	0.294	0.322	0.281	0.197	0.371	0.352	1.062 (0.779-1.446)	1.059 (1.028-1.094)
Urinary incontinence	64	0.619	0.348	0.336	0.354	0.263	0.298	0.266	2.498 (1.515-4.002)	1.111 (1.037-1.2)
Depression	127	0.585	0.310	0.370	0.247	0.189	0.405	0.338	1.238 (0.873-1.695)	1.052 (1.023-1.086)

Table 4. Continued

Known groups	n	EQ-5D-5L			PROPr			EQ-5D-5L vs PROPr		
		Mean	SD	Mean difference*	Mean	SD	Mean difference*	Mean difference	Relative efficiency (F-test ratio [EQ-5D-5L/PROPr])[†]	AUROC ratio [EQ-5D-5L/PROPr] (95% CI)[†]
Other psychiatric diseases	60	0.542	0.341	0.414	0.221	0.183	0.431	0.321	1.749 (1.079-2.679)	1.027 (0.976-1.075)
Stroke	23	0.515	0.397	0.440	0.339	0.252	0.313	0.176	5.043 (2.583-9.186)	1.106 (1.02-1.218)

Note. Columns may not add up because of rounding.
AUROC indicates area under the receiver-operating characteristics curve; COPD, chronic obstructive pulmonary disease; EQ-5D-5L, 5-level version of EQ-5D; GERD, gastroesophageal reflux disease; n/a, not applicable; PROPr, PROMIS-Preference scoring system.
*For age and self-perceived health status group, mean difference indicates the difference from the previous level and for health conditions it refers to the difference from the healthy group.
[†]Confidence intervals were calculated using 3000 bootstrap replications.
[‡]Sixty-six respondents reported a variety of other chronic diseases in an open-ended text box that are not included in any of the categories.

that utilities derived using most preference-accompanied measures (eg, EQ-5D-3L, EQ-5D-5L, SF-6D, and HUI) are able to detect the declining health gradient of age.[56-59] Nevertheless, PROPr utilities show a reverse age gradient, which might question their face validity. This trend was also observed in most subgroups of respondents with the exception of the 75+ age group in the US validation study of PROPr.[23] The most likely explanation for the reverse age gradient is that 2 of the 3 PROMIS domains associated with the largest disutilities in the PROPr value set (depression and fatigue) are also reversely associated with age in our sample. Face validity of PROPr utilities is further questioned by mean values being around 0.50, halfway the QALY scale between full health (1) and a state considered as bad as being dead (0), which can be considered a debatable value for a general population sample.

It is worth to compare our findings with those of previous work using patient or general population samples. At the level of domains, correlation coefficients from our study are in the range of those from previous work.[22,24,25,27,32,33] Nevertheless, the correlation coefficient between PROPr versus EQ-5D-5L was 0.61 in Hungary, whereas correlations ranged between 0.67 and 0.74 in the United States, Canada, Australia, the United Kingdom, Germany, and France.[22-25,27,32,33] In these studies, correlations between PROPr and HUI (range 0.60-0.70) and PROPr and SF-6D (range 0.71-0.79) were in line with what was observed between PROPr and EQ-5D-5L. The composition of the study samples (eg, general population, patients, and mixed sample of patients and the general population), language versions of the questionnaires, and to some extent, the slightly different instruments (eg, PROMIS-29+2, PROMIS-29, or PROMIS custom short forms) administered may also contribute to these differences.

EQ-5D-5L utilities covering a wider range of the utility spectrum appear to be more effective in generating differences between moderate/severe and mild health states. This finding, taken together with the abovementioned face validity problem with PROPr, suggests that PROPr is less suited to cost-utility analyses in which reflecting the full health–dead scale and being able to distinguish between health states are critical. Thus, using PROPr in cost-utility analyses may potentially detract from the validity and reliability of health technology assessment evidence and subsequent financial decisions. Future research is needed to explore how utility and QALY gains from treatment will vary between the 2 measures across patient populations.

Some limitations of this study need to be noted. First, we recruited a general population quota sample from an online panel that may be subject to selection bias, because digital literacy and internet access are requirements for registration. Future studies are recommended to use random sampling and paper-based questionnaires to preclude such bias. Second, data on severity within condition groups were not available; thus, we were not able to assess how well the 2 measures discriminate across severity groups. Third, the data collection was carried out during the second wave of the COVID-19 pandemic that might have affected health status and, subsequently, measurement properties of the instruments.[60,61] Compared with a pre-COVID general population survey conducted in 2018-2019 in Hungary using a similar online mode of administration, respondents in our sample reported slightly less problems in the mobility, usual activities, and pain/discomfort domains of the EQ-5D-5L.[62] Lastly, the order of the 3 health status measures within the survey was not randomized, which might have affected responses.

Conclusion

In conclusion, the PROMIS measures show enhanced distributional characteristics, including lower ceiling and higher informativity, and allow respondents to more easily self-report a wide range of health problems. Very large differences were observed in EQ-5D-5L versus PROPr utilities in overall mean values and by health condition groups but also according to basic demographics. In comparison with the EQ-5D-5L, PROPr utilities are less effective in distinguishing between most known groups by physician-diagnosed health conditions. Future research efforts are recommended to focus on comparing the measurement properties of the EQ-5D-5L and PROMIS measures across different patient populations and severity groups and to examine the impact of the choice of instrument on cost-utility estimates.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2023.02.002.

Article and Author Information

Author Affiliations: Department of Health Policy, Corvinus University of Budapest, Budapest, Hungary (Rencz, Brodszky); Section Medical Psychology and Psychotherapy, Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands (Janssen).

Correspondence: Fanni Rencz, PhD, Department of Health Policy, Corvinus University of Budapest, Fővám tér 8, H-1093, Budapest, Hungary. Email: fanni.rencz@uni-corvinus.hu

REFERENCES

1. Brazier J, Ara R, Rowen D, Chevrou-Severac H. A review of generic preference-based measures for use in cost-effectiveness models. *Pharmacoeconomics.* 2017;35(suppl 1):21–31.
2. Finch AP, Brazier JE, Mukuria C. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *Eur J Health Econ.* 2018;19(4):557–570.
3. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol.* 2010;63(11):1179–1194.
4. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care.* 2007;45(5 suppl 1):S22–S31.
5. Cella D, Choi SW, Condon DM, et al. Promis(®) adult health profiles: efficient short-form measures of seven health domains. *Value Health.* 2019;22(5):537–544.
6. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. *Qual Life Res.* 2009;18(7):873–880.
7. Hanmer J, Cella D, Feeny D, et al. Selection of key health domains from PROMIS(®) for a generic preference-based scoring system. *Qual Life Res.* 2017;26(12):3377–3385.
8. Dewitt B, Feeny D, Fischhoff B, et al. Estimation of a preference-based summary score for the patient-reported outcomes measurement information system: the PROMIS(®)-preference (PROPr) scoring system. *Med Decis Making.* 2018;38(6):683–698.
9. Evans JP, Smith A, Gibbons C, Alonso J, Valderas JM. The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS): a view from the UK. *Patient Relat Outcome Meas.* 2018;9:345–352.
10. Richardson J, Khan MA, Iezzi A, Maxwell A. Comparing and explaining differences in the magnitude, content, and sensitivity of utilities predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D multiattribute utility instruments. *Med Decis Making.* 2015;35(3):276–291.
11. Wang A, Rand K, Yang Z, Brooks R, Busschbach J. The remarkably frequent use of EQ-5D in non-economic research. *Eur J Health Econ.* 2022;23(6):1007–1014.
12. Kennedy-Martin M, Slaap B, Herdman M, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *Eur J Health Econ.* 2020;21(8):1245–1257.
13. Rencz F, Gulácsi L, Drummond M, et al. EQ-5D in Central and Eastern Europe: 2000-2015. *Qual Life Res.* 2016;25(11):2693–2710.
14. Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. *Qual Life Res.* 2021;30(3):647–673.
15. Buchholz I, Janssen MF, Kohlmann T, Feng YS. A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. *Pharmacoeconomics.* 2018;36(6):645–661.
16. Brooks R. EuroQol: the current state of play. *Health Policy.* 1996;37(1):53–72.
17. 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.
18. Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the standard EQ-5D three-level system with a five-level version. *Value Health.* 2008;11(2):275–284.
19. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *Pharmacoeconomics.* 2018;36(6):675–697.
20. Janssen MF, Buchholz I, Golicki D, Bonsel GJ. Is EQ-5D-5L better than EQ-5D-3L over time? A head-to-head comparison of responsiveness of descriptive systems and value sets from nine countries. *Pharmacoeconomics.* 2022;40(11):1081–1093.
21. Pan T, Mulhern B, Viney R, et al. Evidence on the relationship between PROMIS-29 and EQ-5D: a literature review. *Qual Life Res.* 2022;31(1):79–89.
22. Hanmer J. Cross-sectional validation of the PROMIS-Preference scoring system by its association with social determinants of health. *Qual Life Res.* 2021;30(3):881–889.
23. Hanmer J, Dewitt B. The development of a preference-based scoring system for PROMIS® (PROPr): a technical report-v1.4. December 2017. University of Pittsburgh. https://janelhanmer.pitt.edu/documents/technicalreportv1.4.pdf. Accessed March 15, 2023.
24. Hanmer J, Dewitt B, Yu L, et al. Cross-sectional validation of the PROMIS-Preference scoring system. *PLoS One.* 2018;13(7):e0201093.
25. Pan T, Mulhern B, Norman R, et al. *Are generic measures of health-related quality of life comparable? Unerstanding the measurement relationship between EQ-5D-5L, PROMIS-29 and PROPr.* 5th Academy Meeting of the EuroQol Group; March 2021.
26. Pan T, Mulhern B, Viney R, Norman R, Hanmer J, Devlin N. A comparison of PROPr and EQ-5D-5L value sets. *Pharmacoeconomics.* 2022;40(3):297–307.
27. Zhang J, Dewitt B, Tang E, et al. Evaluation of PROMIS preference scoring system (PROPr) in patients undergoing hemodialysis or kidney transplant. *Clin J Am Soc Nephrol.* 2021;16(9):1328–1336.
28. de Graaf JA, Visser-Meily JM, Schepers VP, et al. Comparison between EQ-5D-5L and PROMIS-10 to evaluate health-related quality of life 3 months after stroke: a cross-sectional multicenter study. *Eur J Phys Rehabil Med.* 2021;57(3):337–346.
29. Aghdaee M, Gu Y, Sinha K, Parkinson B, Sharma R, Cutler H. Mapping the Patient-Reported Outcomes Measurement Information System (PROMIS-29) to EQ-5D-5L. *Pharmacoeconomics.* 2023;41(2):187–198.
30. Oak SR, Strnad GJ, Bena J, et al. Responsiveness comparison of the EQ-5D, PROMIS Global Health, and VR-12 questionnaires in knee arthroscopy. *Orthop J Sports Med.* 2016;4(12):2325967116674714.
31. Shim J, Hamilton DF. Comparative responsiveness of the PROMIS-10 Global Health and EQ-5D questionnaires in patients undergoing total knee arthroplasty. *Bone Joint J.* 2019;101-b(7):832–837.
32. Klapproth CP, Fischer F, Merbach M, Rose M, Obbarius A. Psychometric properties of the PROMIS Preference score (PROPr) in patients with rheumatological and psychosomatic conditions. *BMC Rheumatol.* 2022;6(1):15.
33. Klapproth CP, Sidey-Gibbons CJ, Valderas JM, Rose M, Fischer F. Comparison of the PROMIS preference score (PROPr) and EQ-5D-5L index value in general population samples in the United Kingdom, France, and Germany. *Value Health.* 2022;25(5):824–834.
34. Rencz F, Janssen MF. Analyzing the pain/discomfort and anxiety/depression composite domains and the meaning of discomfort in the EQ-5D: a mixed-methods study. *Value Health.* 2022;25(12):2003–2016.
35. European Health Interview Survey (EHIS wave 3)—methodological manual. EUROSTAT. 10.2785/020714. Accessed June 11, 2021.
36. Pickard AS, Law EH, Jiang R, et al. United States valuation of EQ-5D-5L health states using an international protocol. *Value Health.* 2019;22(8):931–941.
37. Hays RD, Spritzer KL, Schalet BD, Cella D. PROMIS(®)-29 v2.0 profile physical and mental health summary scores. *Qual Life Res.* 2018;27(7):1885–1891.
38. Brazier J, Deverill M. A checklist for judging preference-based measures of health related quality of life: learning from psychometrics. *Health Econ.* 1999;8(1):41–51.

39. Bas Janssen MF, Birnie E, Bonsel GJ. Evaluating the discriminatory power of EQ-5D, HUI2 and HUI3 in a US general population survey using Shannon's indices. *Qual Life Res*. 2007;16(5):895–904.

40. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*. 2004;13(9):873–884.

41. Whitehurst DG, Bryan S. Another study showing that two preference-based measures of health-related quality of life (EQ-5D and SF-6D) are not interchangeable. But why should we expect them to be? *Value Health*. 2011;14(4):531–538.

42. McDool E, Mukuria C, Brazier J. A comparison of the SF-6Dv2 and SF-6D UK utility values in a mixed patient and healthy population. *Pharmacoeconomics*. 2021;39(8):929–940.

43. Rencz F, Brodszky V, Gulácsi L, et al. Parallel valuation of the EQ-5D-3L and EQ-5D-5L by time trade-off in Hungary. *Value Health*. 2020;23(9):1235–1245.

44. Khadka J, Gothwal VK, McAlinden C, Lamoureux EL, Pesudovs K. The importance of rating scales in measuring patient-reported outcomes. *Health Qual Life Outcomes*. 2012;10:80.

45. Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: challenges and potential solutions. *Curr Med Res Opin*. 2009;25(4):929–942.

46. Evans JD. *Straightforward Statistics for the Behavioral Sciences*. Belmont, CA: Thomson Brooks/Cole Publishing Co; 1996.

47. Shannon CE. The mathematical theory of communication. *Bell Syst Tech J*. 1948;27(3):379–423.

48. Shannon CE, Weaver W. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press; 1949.

49. Efron B. Bootstrap methods: another look at the jackknife. In: *Breakthroughs in Statistics*. Berlin, Germany: Springer; 1992:569–593.

50. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–428.

51. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284–290.

52. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.

53. Norquist JM, Girman C, Fehnel S, DeMuro-Mercon C, Santanello N. Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration. *Qual Life Res*. 2012;21(6):1013–1020.

54. Broderick JE, Schwartz JE, Vikingstad G, Pribbernow M, Grossman S, Stone AA. The accuracy of pain and fatigue items across different reporting periods. *Pain*. 2008;139(1):146–157.

55. Thomas ML, Kaufmann CN, Palmer BW, et al. Paradoxical trend for improvement in mental health with aging: a community-based study of 1,546 adults aged 21-100 years. *J Clin Psychiatry*. 2016;77(8):e1019–e1025.

56. Horsman JR, Furlong WJ, Feeny DH, Torrance G. Health utilities index (Hui®): population reference statistics. *Value Health*. 2013;16(7):A595.

57. Janssen MF, Szende A, Cabases J, Ramos-Goñi JM, Vilagut G, König HH. Population norms for the EQ-5D-3L: a cross-country analysis of population surveys for 20 countries. *Eur J Health Econ*. 2019;20(2):205–216.

58. Shiroiwa T, Fukuda T, Ikeda S, et al. Japanese population norms for preference-based measures: EQ-5D-3L, EQ-5D-5L, and SF-6D. *Qual Life Res*. 2016;25(3):707–719.

59. van den Berg BSf. 6d population norms. *Health Econ*. 2012;21(12):1508–1512.

60. Long D, Haagsma JA, Janssen MF, Yfantopoulos JN, Lubetkin EI, Bonsel GJ. Health-related quality of life and mental well-being of healthy and diseased persons in 8 countries: does stringency of government response against early COVID-19 matter? *SSM Popul Health*. 2021;15:100913.

61. Lapin BR, Tang WHW, Honomichl R, Hogue O, Katzan IL. Evidence of stability in patient-reported global health during the COVID-19 pandemic. *Value Health*. 2021;24(11):1578–1585.

62. Rencz F, Tamási B, Brodszky V, Ruzsa G, Gulácsi L, Péntek M. Did you get what you wanted? Patient satisfaction and congruence between preferred and perceived roles in medical decision making in a Hungarian national survey. *Value Health Reg Issues*. 2020;22:61–67.