

# Predicting EuroQol (EQ-5D) scores from the patient-reported outcomes measurement information system (PROMIS) global items and domain item banks in a United States sample

Dennis A. Revicki · Ariane K. Kawata ·  
Neesha Harnam · Wen-Hung Chen ·  
Ron D. Hays · David Cella

Accepted: 6 May 2009 / Published online: 27 May 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

## Abstract

**Background** Preference-based health index scores provide a single summary score assessing overall health-related quality of life and are useful as an outcome measure in clinical studies, for estimating quality-adjusted life years for economic evaluations, and for monitoring the health of populations. We predicted EuroQoL (EQ-5D) index scores from patient-reported outcomes measurement information system (PROMIS) global items and domain item banks.

**Methods** This was a secondary analysis of health outcome data collected in an internet survey as part of the PROMIS Wave 1 field testing. For this study, we included the 10 global items and the physical function, fatigue, pain impact, anxiety, and depression item banks. Linear regression analyses were used to predict EQ-5D index scores based on the global items and selected domain banks.

**Results** The regression models using eight of the PROMIS global items (quality of life, physical activities, mental health, emotional problems, social activities, pain, and fatigue and either general health or physical health items)

explained 65% of the variance in the EQ-5D. When the PROMIS domain scores were included in a regression model, 57% of the variance was explained in EQ-5D scores. Comparisons of predicted to actual EQ-5D scores by age and gender groups showed that they were similar. **Conclusions** EQ-5D preference scores can be predicted accurately from either the PROMIS global items or selected domain banks. Application of the derived regression model allows the estimation of health preference scores from the PROMIS health measures for use in economic evaluations.

**Keywords** Health preference scores · EQ-5D · PROMIS · Global health status · Health-related quality of life

## Introduction

Health outcomes assessment has advanced to a point where generic health-related quality of life (HRQL) measures are often used to examine the health status of populations and the effects of medical interventions. Generic HRQL profile measures, such as the SF-36 Health Survey [1], provide multiple health domains scores (e.g., physical function, mental health, pain, vitality, etc.), but not an overall index score. Preference-based health index scores provide a single summary score assessing overall health-related quality of life and are useful as an outcome measure in clinical studies, for estimating quality-adjusted life years for economic evaluations, and for monitoring the health of populations. Preference-based HRQL instruments provide information on the value of different health states and can be used to estimate health outcomes for cost-effectiveness analyses [2, 3].

---

D. A. Revicki (✉) · A. K. Kawata · N. Harnam · W.-H. Chen  
Center for Health Outcomes Research, United BioSource  
Corporation, 7101 Wisconsin Ave., Suite 600, Bethesda,  
MD 20814, USA  
e-mail: Dennis.Revicki@unitedbiosource.com

R. D. Hays  
Division of General Internal Medicine & Health Services  
Research, Department of Medicine, University of California  
at Los Angeles, Los Angeles, CA, USA

D. Cella  
Department of Medical Social Sciences, Northwestern  
University, Evanston, IL, USA

Several preference-based indexes have been developed, which includes the EuroQoL EQ-5D [4], the health utilities index (HUI) [5, 6], and the Quality of Well-Being Scale–Self Administered [7]. In addition, a preference-based score, the SF-6D, has been estimated from SF-36 items [8, 9]. Although each of these health indexes provides valuations on a 0 (dead) to 1 (perfect health) scale, they differ in health state classification systems, methods for preference assessment, and scoring algorithms. US normative data for these measures have been developed based on the National Health Measurement Study [2] and other national surveys [10, 11].

Previous studies have derived preference-based scores from generic HRQL profile measures [8, 9, 12, 13]. Lawrence and Fleishman [14] and Sullivan and Ghushchyan [15] discussed mapping the EQ-5D index from the SF-12 using nationally representative samples. Gray et al. [16] used regression analysis to explore the association between responses to the SF-12 and responses to each EQ-5D question, and found that both the US-based data (from the 2000 Medical Expenditure Panel Survey [MEPS]) and the UK-based data (from the 1996 Health Survey for England [HSE]) had similar demographic characteristics, as well as physical component summary (PCS) and mental component summary (MCS) scores, though HSE had higher mean EQ-5D index scores. Comparisons of EQ-5D index scores derived from the US and UK algorithms found that the US model predicted higher scores than the UK model for almost all EQ-5D health states, while the US model resulted in smaller gains in health preferences than the UK model for the large majority of simulated transitions between EQ-5D health states [17]. Mapping of the SF-12 to the EQ-5D index in a nationally representative US sample has also been conducted, though this was completed without the use of US population weights [18]. The use of mapping estimation methods allows patient preference scores to be derived from health status profiles based on the empirical relationship between these constructs, which is particularly useful when patient preference scores are unavailable.

Recently, the patient-reported outcomes measurement information system (PROMIS) project has developed several health domain item banks [19] and a short-form version of a global health questionnaire, a new generic health status measure based on the review of PROMIS item banks (Hays et al., submitted). Given that the PROMIS domain scores and global items are likely to see increased application in National Institutes of Health (NIH) and other studies, estimating health preference scores from the PROMIS measures will be useful for those studies where assessments of health preferences have not been included. These estimated health preference scores will be useful for economic analyses using data from studies that include the

PROMIS domain and global short-form instruments. The objective of this study was to estimate health preference scores based on the EQ-5D index, using selected PROMIS domain scores and the summary scores from the PROMIS global short form. We also compared the estimated health preference scores to EQ-5D index scores from several US national surveys by age and gender groups [2, 10].

## Methods

### Study design

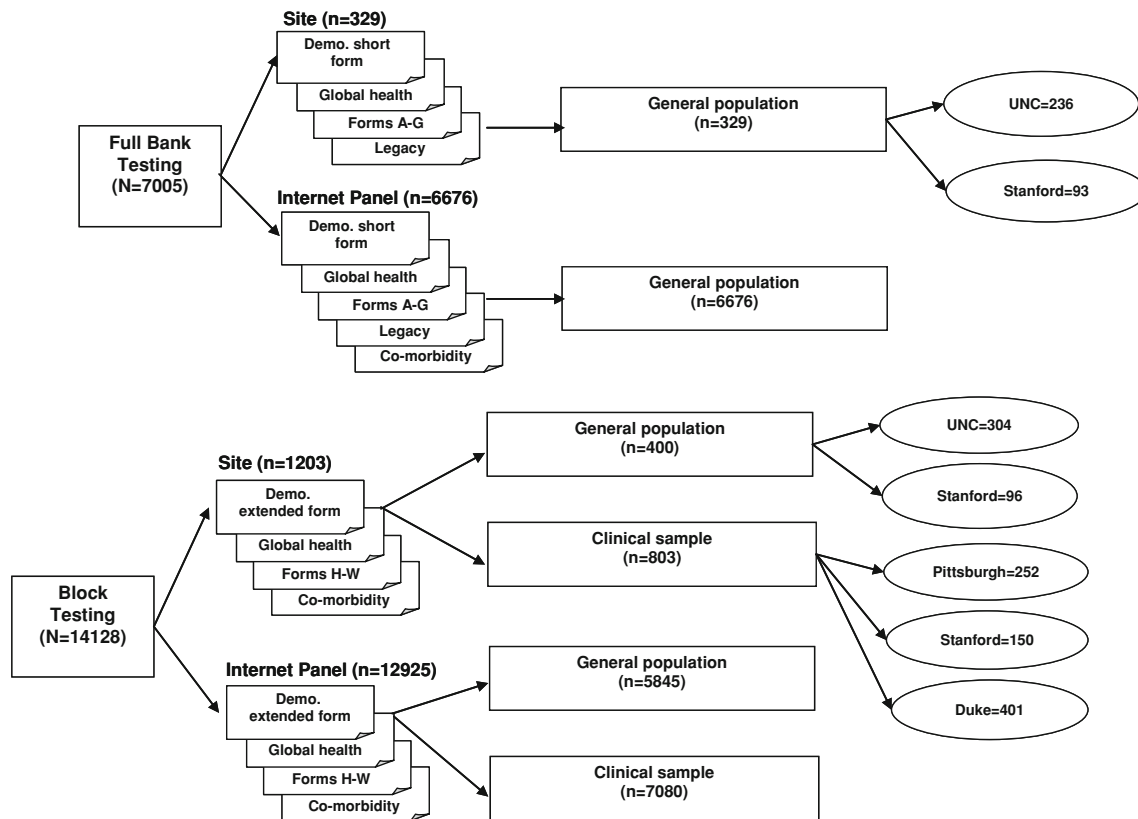
The PROMIS item banks were administered via web-based survey to a national internet panel maintained by Polimetrix (now YouGovPolimetrix; see <http://www.polimetrix.com>). The field test involved administering the item banks from five domains (i.e., pain, fatigue, physical functioning, social activities, emotional distress) to selected participants [19]. Some respondents were randomly assigned to administer different complete item banks, that is, all the items within a defined domain-specific bank, such as physical function or fatigue. Other respondents were randomly assigned to block-form item samples consisting of sets of seven consecutive items from each of 14 subdomains in the five PROMIS health domains.

### Study participants

The PROMIS sample was selected to be generally comparable to distributions of gender, age groups, race/ethnicity (white/African–American/Hispanic/other), and education (high school or less versus more than high school) based on the 2000 US census data (Liu et al., submitted). Study participants were identified from the Polimetrix internet panel and from selected clinical research centers. For the current study, the participants included subjects who administered the full item banks and the block data.

### Wave 1 sample

Because of the number of item banks being tested in Wave 1, a complex data-collection strategy was employed. This strategy included two arms and a total sample size of 21,133. A total of 19,601 subjects were recruited by Polimetrix, with the remaining 1,532 subjects recruited by PROMIS research sites (Fig. 1). In the full-bank testing arm, 7,005 persons from the general population were administered two of the 14, 56-item, subdomain-specific PROMIS item banks. In the block testing arm, 14,128 individuals administered randomly selected seven-item blocks measuring each of the 14 PROMIS-targeted subdomains. The PROMIS research sites and the Polimetrix sample included both community and



**Fig. 1** PROMIS wave 1 sample

clinical samples. The clinical samples included persons with heart disease ( $n = 1,156$ ), cancer ( $n = 1,754$ ), rheumatoid arthritis ( $n = 557$ ), osteoarthritis ( $n = 918$ ), psychiatric disorders ( $n = 1,193$ ), COPD ( $n = 1,214$ ), spinal cord injury ( $n = 531$ ), and other conditions ( $n = 560$ ).

## Measures

### EQ-5D

The EQ-5D is a preference-based instrument designed to measure generic health status across five dimensions of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, with three response levels (no problems, some problems, extreme problems) [4]. A unique EQ-5D health state is defined by combining one level from each of the five dimensions, and scores range from  $-0.109$  to  $1.0$ , with greater scores indicating better overall health. The calculation of the EQ-5D index scores was based on the valuation reported by Shaw et al. [13] that was derived from a large-scale survey of the US general population [10]. The EQ-5D also includes a single visual analogue scale (EQ-5D VAS) that was not used in this study.

### PROMIS global items

The PROMIS survey included ten global health items (Hays et al. submitted). One item was the general health question rating overall health on a poor-to-excellent scale. The remaining items covered quality of life, mental health, physical health (two items), pain, fatigue, social function (two items), and emotional distress. Based on these global items, Hays et al. (submitted) found evidence supporting two summary scores assessing physical and mental health. Mental health and physical health summary scores were developed from global items in factor and item response theory analyses conducted in the PROMIS Wave 1 sample. The PROMIS global items were administered to all participants in the Wave 1 sample (Fig. 1). The summary scores were calculated as sums of the relevant individual global items, and individual global item scores (untransformed) were included in the subsequent regression analyses.

### Domain item banks

The PROMIS initial item banks were developed based on the published literature, clinician review, and qualitative

research on patients with various health conditions (for more information, go to <http://www.nihpromis.org>). Existing domain-specific instruments were also reviewed for item content, and new items were developed for the PROMIS item banks [19, 20]. Content of the final set of physical function, fatigue, pain impact, anxiety, and depression items was revised based on the results of cognitive debriefing interviews [20]. For this study, we used the calibrated and available item banks measuring physical function, fatigue, pain impact, anxiety, and depression ([www.nihpromis.org](http://www.nihpromis.org)). The physical function item bank covered self-reported capability for upper extremity and lower extremity function. The fatigue bank was developed to cover both fatigue experience and impact. The pain impact domain included items on various impacts of pain on daily activities and function. The anxiety bank included various symptoms associated with anxiety, and the depression bank included items on depressed mood. Each item used a five- to six-level categorical response scale. The domain scores included in this analysis are T-scores derived from Theta scores from the item response theory calibrations. For the physical function domain scores, higher scores indicate better physical functioning. For the fatigue, pain impact, anxiety, and depression domain scores, higher scores indicate more severe impairment.

#### *Other measures*

Information on demographic characteristics was collected for the study participants (i.e., age, gender, race/ethnicity, education). Information was also collected on a number of chronic medical conditions in the Wave 1 sample. These chronic conditions were classified into groups of physical and mental health disorders.

#### *Statistical analysis*

A series of ordinary least squares (OLS) regression models were specified where EQ-5D index scores were predicted from different sets of PROMIS scores. First, three sets of regression models were performed using (1) all ten global items; (2) a subset of eight global items (reduced because of multicollinearity); and (3) a subset of eight global items (using alternative duplicative items). The Wave 1 analysis sample of 20,400 cases was separated into two randomly assigned split-half samples; the models were developed in the first sample and the analyses replicated in the second sample to confirm results. Second, we specified an OLS regression model using the PROMIS global item-based mental health and physical health summary scores to predict EQ-5D index scores in the block testing sample ( $n = 14,128$ ). Finally, a

regression analysis was performed including the T-scores for the PROMIS domain item banks for physical function, fatigue, pain impact, anxiety, and depression using the block design data. We selected these five domain banks because they (1) covered important patient-reported outcome constructs, including mental health and physical health; and (2) these item banks were calibrated and tested within the PROMIS project. Subjects were included in this analysis if they completed at least three items for each of the five relevant PROMIS domains and had an EQ-5D index score ( $n = 1,658$ ). The domain scores included in this analysis are T-scores derived from Theta scores from the item response theory calibrations.

We examined plots of residuals from the regression analyses and performed a Bland-Altman assessment of agreement [21] comparing the actual and predicted EQ-5D index scores. A range of agreement was defined as mean bias  $\pm$  2 standard deviation (SD) units. Intraclass correlation coefficients (ICCs) were calculated comparing actual and predicted EQ-5D scores.

Estimated EQ-5D index scores were also compared with actual EQ-5D scores based on the PROMIS general population sample by gender and age groups and by type of chronic medical condition ( $n = 2,722$ ; Liu et al. submitted). In addition, we compared the PROMIS estimated EQ-5D index scores to those reported in the Luo et al. [10] and Fryback et al. [2] studies by gender and age groups.

## **Results**

### *Sample demographic characteristics*

The PROMIS Wave 1 analysis sample consisted of 20,400 cases. A total of 733 cases were excluded; subjects with an average response time of less than 1 s per item ( $n = 573$ ) and/or ten consecutive items with response time less than half a second ( $n = 192$ ) were excluded from the analyses. The overall sample ( $n = 20,400$ ) was 52% female. The sample mean age was  $\sim$ 53 years: 12% were 18–29 years; 12% were 30–39 years; 16% were 40–49 years; 32% were 50–64 years; and 28% were 65 years or older. The racial/ethnic breakdown was 80% white, 9% African-American, 9% Hispanic or Latino, and 2% other races (Asian/Pacific Islanders or Native Americans). Educational attainment ranged from less than high school (3%) to college or above (44%), with 39% reporting some college and 16% a high school diploma. The general population sample used in some analyses is a subset of the PROMIS Wave 1 sample. Sample characteristics have been found to be consistent with those of the 2000 US census data (Liu et al., submitted).

## Regression analyses

To predict EQ-5D index scores, three regression models were run using the PROMIS data. Model 1 included the 10 global items as predictors of the EQ-5D index score and had an adjusted *R*-square of 0.65 (Table 1). Possible multicollinearity among the ten global items was examined using collinearity diagnostics (tolerance, variance inflation, condition index, and proportion of variance). Some multicollinearity was detected, particularly among the general health and physical health items. These items were highly correlated ( $r = 0.90$ ) and had low tolerance and high variance inflation factors. In Model 1, general health was a significant indicator ( $P = 0.0014$ ), but physical health was not significant ( $P = 0.526$ ). The social satisfaction item was not significant in Model 1 ( $P = 0.316$ ), and was dropped from subsequent regression models.

Two additional models were run to assess the effect of including general health and physical health in separate models (Table 1). Model 2 included the general health item, and Model 3 included the physical health item. All global items included in Models 2 and 3 were significant predictors of EQ-5D index scores, and overall model statistics for both indicated good fit. General health ( $P < 0.0001$ ) was a significant predictor in the absence of physical health (Model 2), while physical health ( $P = 0.0051$ ) was significant when general health was not included (Model 3). All models accounted for  $\sim 65\%$  of variability in EQ-5D index scores. The three models were repeated using the second split-half sample of Wave 1 data (*R*-square = 0.65). Results using the two separate samples were very similar, and there were no substantive differences.

The ICCs for each of the three models were 0.77. Review of residuals suggested reasonably good fit for all three models. For example, in Model 2 the mean residual was 0.0 (SE = 0.002), with 95% range limits of  $-0.17$  and  $0.14$  based on the distribution of residuals. The other two models had similar results. The Bland–Altman analyses indicated that the 95% limits of agreement between the predicted and actual EQ-5D scores ranged from  $-0.20$  to  $0.20$ . The largest differences were observed at the upper extremes of EQ-5D scores ( $>0.90$ ), and there was some evidence of overestimation in the lower range ( $<0.30$ ).

A regression model including only the mental health and physical health summary scores based on the ten global items accounted for 57% of the variance in EQ-5D index scores. The unstandardized regression coefficients for both the mental health summary ( $b = 0.031$ ,  $P < 0.0001$ ) and physical health summary ( $b = 0.137$ ,  $P < 0.0001$ ) scores were significant in this model.

The regression model including PROMIS physical function, fatigue, pain impact, anxiety, and depression scores resulted in an adjusted *R*-square of 0.57. Domains

were scored such that higher scores corresponded to higher levels of the attribute (e.g., better physical function, more fatigue). Results were in the expected directions for each domain. Regression coefficients for physical function ( $b = 0.0077$ ,  $P < 0.0001$ ), fatigue ( $b = -0.0021$ ,  $P < 0.0001$ ), pain impact ( $b = -0.0040$ ,  $P < 0.0001$ ), anxiety ( $b = -0.0023$ ,  $P < 0.0001$ ), and depression ( $b = -0.0022$ ,  $P < 0.0001$ ) were all statistically significant in the model.

The ICC for the PROMIS domain model was 0.73. Review of residuals suggested reasonably good fit for this model. The mean residual was 0.0 (SE = 0.003), with 95% range limits of  $-0.20$  and  $0.15$  based on the distribution of residuals. The Bland–Altman analysis indicated that the 95% limits of agreement between the predicted and actual EQ-5D scores ranged from  $-0.21$  to  $0.21$ . The largest differences were observed at lower EQ-5D scores ( $<0.40$ ).

## Comparison of predicted and actual EQ-5D index scores

In the PROMIS general population sample, there were few differences identified between actual and predicted EQ-5D index scores (Table 2). For the total sample, the actual mean EQ-5D score was 0.85 (SD = 0.16), compared with a predicted score of 0.85 (SD = 0.13). Differences in mean EQ-5D index scores by age and gender groups ranged from 0 to 0.02 points, with most (87%) deviations  $\leq 0.01$  points.

## EQ-5D index scores by medical conditions

Table 3 summarizes actual and predicted EQ-5D index scores by chronic conditions. As expected, subjects with both mental and physical conditions had the lowest EQ-5D preference scores (0.72–0.75). Subjects with no chronic conditions reported the best EQ-5D scores (0.92–0.94). Those with only physical or mental conditions had preference scores situated between subjects with none or both types of chronic conditions.

## Comparison of PROMIS-predicted and reported EQ-5D index scores from other studies

We compared predicted EQ-5D index scores from the PROMIS general population sample to gender and age groups reported in the Luo et al. [10] study (Table 4). In general, the preference scores were similar, although PROMIS sample females reported somewhat lower index scores and those aged 65 years and older reported higher index scores when compared with the gender and age groups in Luo et al. [10] study.

The PROMIS-predicted EQ-5D index scores were also compared with the National Health Measurement Study [NHMS] EQ-5D data reported by Fryback et al. [2]

**Table 1** Results of regression analyses predicting EQ-5D index scores from PROMIS global items

Variable	Model 1			Model 2			Model 3					
	N	P value for F	Adjusted R <sup>2</sup>	Estimate (SE)	N	P value for F	Adjusted R <sup>2</sup>	Estimate (SE)	N	P value for F	Adjusted R <sup>2</sup>	Estimate (SE)
Intercept	6,973	<.0001	0.6497		6,977	<.0001	0.6496		6,973	<.0001	0.6492	
Global01: general health				.19112 (.00658)****				.19150 (.00656)****				.19123 (.00658)****
Global02: quality of life				.00932 (.00292)**				.00786 (.00188)****				.00672 (.00204)**
Global03: physical health				.00541 (.00214)*				.00569 (.00202)**				.00527 (.00188)**
Global04: mental health				-.0018 (.00291)				.00816 (.00188)****				.00830 (.00188)****
Global05: social satisfaction				.00760 (.00195)***								
Global06: physical activities				.00173 (.00172)				.04486 (.00175)****				.04550 (.00175)****
Global07R: pain				.04501 (.00176)****				.02704 (.00064)****				.02713 (.00064)****
Global08R: fatigue				.02705 (.00064)****				.01268 (.00188)****				.01305 (.00188)****
Global09: social activities				.01285 (.00188)****				.00573 (.00193)**				.00613 (.00193)**
Global10R: Emotional problems				.00530 (.00199)**				.02535 (.00168)****				.02502 (.00168)****

Model 1 includes PROMIS global items 1–10; model 2 includes PROMIS global items 1–2, 4, 6–10; model 3 includes PROMIS global items 2–4, 6–10

\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$

**Table 2** Mean actual and predicted EQ-5D index scores by gender and age groups in PROMIS general population sample ( $n = 2,722$ )

Age groups	Males		Females		Total	
	Actual	Predicted	Actual	Predicted	Actual	Predicted
<35	0.89	0.89	0.88	0.86	0.88	0.87
35–44	0.87	0.86	0.85	0.84	0.86	0.85
45–54	0.87	0.86	0.79	0.79	0.82	0.82
55–64	0.84	0.85	0.81	0.82	0.82	0.83
65–74	0.86	0.88	0.84	0.85	0.85	0.86

(Table 5). Generally, the index scores were comparable for males (differences of 0.01 to 0.02 points) but varied more for females (differences of 0.01 to 0.08 points). The largest differences between the PROMIS and NHMS samples were for women aged 35–44 years (0.05 points) and 45–54 years (0.08 points). The preference scores for the older men and women were more comparable between the two samples.

## Discussion

We estimated EQ-5D index scores using the PROMIS global items and selected domain scores. Using different sets of the global items, we were able to account for 65% of the variance in preference scores. By comparison, about 57% of the variance in EQ-5D index scores was explained by the global item summary scores or selected PROMIS domain scores. These results are consistent with previous research in predicting health preference scores from HRQL profile measures [8, 14, 15, 18, 22]. For example, Lawrence and Fleishman [14] were able to explain 61–63% of the variance in EQ-5D index scores using SF-12 summary scores. Other researchers explained 58–63% of EQ-5D index scores using different HRQL measures [15, 18]. The availability of preference-based scores based on the PROMIS global items and domain scores enables potential application of these measures to population-based studies and economic evaluations. The main advantage of the PROMIS measures over other static health status measures is that the PROMIS domain item banks and scores allow flexibility in administration using either targeted short forms or computerized adaptive testing.

**Table 3** Mean actual and predicted EQ-5D index scores by disease classification in PROMIS general population Sample ( $n = 2,722$ )

Disease classification	<i>N</i>	Actual EQ-5D mean (SE)	Predicted EQ-5D mean (SE)
No physical/mental conditions	778	0.94 (0.004)	0.92 (0.005)
Physical conditions only	1,234	0.85 (0.004)	0.86 (0.005)
Mental conditions only	122	0.85 (0.013)	0.84 (0.006)
Physical and mental conditions	588	0.72 (0.008)	0.75 (0.006)

**Table 4** Comparing predicted PROMIS preference scores to Luo et al. [10]

Total sample	PROMIS Mean (SE)	Luo study Mean (SE)
	0.85 (0.01)	0.87 (0.01)
Gender		
Male	0.87 (0.01)	0.88 (0.01)
Female	0.83 (0.01)	0.86 (0.01)
Age groups		
18–44	0.86 (0.01)	0.91 (<0.01)
45–64	0.83 (0.01)	0.84 (0.01)
65+	0.86 (0.01)	0.81 (0.01)

**Table 5** Comparing predicted PROMIS preference scores to Fryback et al. [2]

Age groups	Males		Females		Total	
	Fryback	PROMIS	Fryback	PROMIS	Fryback	PROMIS
<35	NA	0.89	NA	0.86	NA	0.87
35–44	0.89	0.87	0.89	0.84	0.89	0.85
45–54	0.88	0.86	0.87	0.79	0.87	0.82
55–64	0.86	0.85	0.84	0.82	0.85	0.83
65–74	0.87	0.88	0.84	0.85	0.86	0.86

NA not available

The estimated EQ-5D index scores based on PROMIS global items were comparable to those directly assessed using the EQ-5D in this sample. Based on the Bland-Altman and other analyses, there was evidence of some overestimation for EQ-5D scores under 0.40; however, the ICCs indicated good agreement (0.77). Differences between the predicted and actual index scores were between 0 and 0.02 points by gender and age groups. Most of the observed deviations were less than 0.01 points. These findings are encouraging and suggest that the predicted EQ-5D index scores may be applied to future studies. More importantly, the predicted EQ-5D index scores varied by presence of physical or mental conditions and were most impaired in those with both mental and physical conditions. The predicted EQ-5D scores based on the PROMIS domains were also comparable to the actual measured EQ-5D scores, and demonstrated similar levels of agreement to the PROMIS global items.

The general pattern of predicted EQ-5D index scores by gender and age groups seen in the PROMIS sample was comparable to those in other recent studies [2, 10]. There is a general decline in index scores by age, although the oldest age group (65–74 years) showed a small increase in preference scores compared with those aged 55–64 years. These findings are consistent with the observed EQ-5D index scores reported in Fryback et al. [2].

There were few differences between the PROMIS sample and the Luo et al. [10] study sample on preference scores. However, in the PROMIS sample women reported somewhat lower index scores, and those aged 65 years and older reported higher index scores compared to those in Luo et al. [10] study. The Luo et al. [10] study used self-completion, as did the PROMIS study, and this may explain the comparability in mean scores. We found some differences by gender and age groups between the predicted EQ-5D index scores from the PROMIS sample and those in the NHMS [2]. The largest differences between the PROMIS and NHMS samples were for women aged 35–44 years and 45–54 years. The preference scores for the younger men and older men and women were comparable between the two samples. These observed differences may be due to different sampling strategies; the Fryback et al. [2] study over-sampled the elderly and ethnic minorities, while the PROMIS study attempted to recruit a representative national sample through an internet panel. Fryback et al. [2] weighted to account for this oversampling, but differences in response patterns and mode of administration (telephone interview vs. internet self-completion) may also have contributed to observed variability. Future research is needed to more carefully examine differences in the PROMIS-predicted EQ-5D index scores by ethnicity, gender, and age groups.

We recommend the PROMIS global item-based prediction equation as best for estimating EQ-5D scores if only one approach is considered. However, future application of these prediction equations depends on the incorporation of either the PROMIS global items or domain measures in future clinical and health services research studies. Given the flexibility of multi-domain short forms and computerized adaptive testing, the PROMIS domain item banks and domain scores may be very useful in clinical studies. The PROMIS global items have potential applications for large population-based and epidemiologic studies. The existing prediction equations allow flexibility to researchers depending on the PROMIS instruments included in their studies.

In general, if a researcher needs to include a preference-based health outcome measure in a study, the most recommended approach is to include one of the direct (i.e., time trade-off, standard gamble) or indirect (i.e., EQ-5D, HUI) measures of health preferences. As we have

demonstrated, it is possible to estimate a preference-based score using the PROMIS global items or domain scores in the absence of a preference-based instrument, for example, because of respondent burden or other issues. However, the researcher should recognize that this is a second-best approach and that primary data collection is recommended.

There are several limitations associated with these analyses. First, for analyses involving PROMIS global items, the ordinal nature of these measures may impact the coefficient estimation in the regression analyses. Second, the PROMIS data were all collected using a web-based survey, and there may be differences between the PROMIS sample and the US general population that may limit generalizability of these results. However, Liu et al. (submitted) found that the PROMIS sample was comparable in demographic characteristics and health status to samples from the US general population.

In summary, we predicted EQ-5D index scores based on the PROMIS global items and selected domain scores, and these predicted preference scores varied as expected by demographic characteristics and presence of mental or physical conditions in the PROMIS sample. The predicted index scores were generally comparable to other national samples by age and gender groups. Additional research is needed to further evaluate the validity of the predicted index scores and should also examine other possible approaches to mapping the PROMIS item banks, perhaps through item response theory analysis and the resultant theta scores or through health preference measures such as the EQ-5D, HUI, or direct utility measures. This study suggests that useful preference scores can be derived from the PROMIS measures, and these predicted EQ-5D index scores have applications in measuring the health of populations and estimating quality-adjusted life years for economic evaluations.

**Acknowledgments** This work was funded by the National Institutes of Health through the NIH Roadmap for Medical Research Cooperative Agreement (1U01-AR052177).

The Patient-Reported Outcomes Measurement Information System (PROMIS) is one of the National Institutes of Health (NIH) Roadmap initiatives to develop a computerized system measuring patient-reported outcomes in respondents with a wide range of chronic diseases and demographic characteristics. PROMIS was funded by cooperative agreements to a Statistical Coordinating Center (Evanston Northwestern Healthcare, PI: David Cella, PhD, U01R52177) and six Primary Research Sites (Duke University, PI: Kevin Weinfurt, PhD, U01AR52186; University of North Carolina, PI: Darren DeWalt, MD, MPH, U01AR52181; University of Pittsburgh, PI: Paul A. Pilkonis, PhD, U01AR52155; Stanford University, PI: James Fries, PhD, U01AR52158; Stony Brook University, PI: Arthur Stone, PhD, U01AR52170; and University of Washington, PI: Dagmar Amtmann, PhD, U01AR52171). NIH Science Officers on this project are Deborah Ader, PhD; Susan Czajkowski, PhD; Lawrence Fine, MD, DrPH; Louis Quatrano, PhD; Bryce Reeves, PhD; William Riley, PhD; and Susana Serrate-Sztejn, PhD. See the website at <http://www.nihpromis.org> for additional information on the PROMIS cooperative group.



**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Ware, J. E., Jr, & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, 30(6), 473–483. doi:10.1097/00005650-199206000-00002.
- Fryback, D. G., Dunham, N. C., Palta, M., Hanmer, J., Buechner, J., Cherepanov, D., et al. (2007). US norms for six generic health-related quality-of-life indexes from the national health measurement study. *Medical Care*, 45(12), 1162–1170. doi:10.1097/MLR.0b013e31814848f1.
- Gold, M. R., Patrick, D. L., & Torrance, G. W. (1996). Identifying and valuing outcomes. In M. R. Gold, D. L. Patrick, & G. W. Torrance (Eds.), *Cost-effectiveness in health and medicine*. New York: Oxford University Press.
- Brooks, R., Rabin, R., & de Charro, F. (2003). *The measurement and valuation of health status using EQ-5D: A European perspective*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Feeny, D., Torrance, G., & Furlong, W. (1996). Health utilities index. In D. Feeny, G. Torrance, & W. Furlong (Eds.), *Quality of life and pharmacoeconomics in clinical trials*. Philadelphia, PA: Lippincott-Raven Press.
- Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., et al. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care*, 40(2), 113–128. doi:10.1097/00005650-200202000-00006.
- Andresen, E. M., Rothenberg, B. M., & Kaplan, R. M. (1998). Performance of a self-administered mailed version of the quality of well-being (QWB-SA) questionnaire among older adults. *Medical Care*, 36(9), 1349–1360. doi:10.1097/00005650-199809000-00007.
- Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, 21(2), 271–292. doi:10.1016/S0167-6296(01)00130-8.
- Brazier, J. E., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care*, 42(9), 851–859. doi:10.1097/01.mlr.0000135827.18610.0d.
- Luo, N., Johnson, J. A., Shaw, J. W., Feeny, D., & Coons, S. J. (2005). Self-reported health status of the general adult U.S. population as assessed by the EQ-5D and Health Utilities Index. *Medical Care*, 43(11), 1078–1086. doi:10.1097/01.mlr.0000182493.57090.c1.
- Hanmer, J., Hays, R. D., & Fryback, D. G. (2007). Mode of administration is important in US national estimates of health-related quality of life. *Medical Care*, 45(12), 1171–1179. doi:10.1097/MLR.0b013e3181354828.
- Erickson, P. (1998). Evaluation of a population-based measure of quality of life: The health and activity limitation index (HALex). *Quality of Life Research*, 7(2), 101–114. doi:10.1023/A:1008897107977.
- Shaw, J. W., Johnson, J. A., & Coons, S. J. (2005). US valuation of the EQ-5D health states: Development and testing of the D1 valuation model. *Medical Care*, 43(3), 203–220. doi:10.1097/00005650-200503000-00003.
- Lawrence, W. F., & Fleishman, J. A. (2004). Predicting EuroQoL EQ-5D preference scores from the SF-12 health survey in a nationally representative sample. *Medical Decision Making*, 24(2), 160–169. doi:10.1177/0272989X04264015.
- Sullivan, P. W., & Ghushchyan, V. (2006). Mapping the EQ-5D index from the SF-12: US general population preferences in a nationally representative sample. *Medical Decision Making*, 26(4), 401–409. doi:10.1177/0272989X06290496.
- Gray, A. M., Rivero-Arias, O., & Clarke, P. M. (2006). Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Medical Decision Making*, 26(1), 18–29. doi:10.1177/0272989X05284108.
- Nan, L., Johnson, J. A., Shaw, J. W., & Coons, S. J. (2007). A comparison of EQ-5D index scores derived from the US and UK population-based scoring functions. *Medical Decision Making*, 27(3), 321–326. doi:10.1177/0272989X07300603.
- Franks, P., Lubetkin, E. I., Gold, M. R., Tancredi, D. J., & Jia, H. (2004). Mapping the SF-12 to the EuroQol EQ-5D index in a national US sample. *Medical Decision Making*, 24(3), 247–254. doi:10.1177/0272989X04265477.
- Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., et al. (2007). The patient-reported outcomes measurement information system (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. *Medical Care*, 45(5, Suppl 1), S3–S11. doi:10.1097/01.mlr.0000258615.42478.55.
- DeWalt, D. A., Rothrock, N., Yount, S., & Stone, A. A. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care*, 45(5, Suppl 1), S12–S21. doi:10.1097/01.mlr.0000254567.79743.e2.
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1(8476), 307–310.
- Revicki, D. A., & Kaplan, R. M. (1993). Relationship between psychometric and utility-based approaches to the measurement of health-related quality of life. *Quality of Life Research*, 2(6), 477–487. doi:10.1007/BF00422222.