

Mapping the EQ-5D Index from the SF-12: US General Population Preferences in a Nationally Representative Sample

Patrick W. Sullivan, PhD, Vahram Ghushchyan, PhD

Background. Previous mapping algorithms estimating EQ-5D index scores from the SF-12 were based on preferences from a UK community sample. However, preferences based on the general US population are most appropriate for cost-effectiveness analyses done from the societal perspective in the United States. **Objective.** To provide a mapping algorithm for estimating EQ-5D index scores from the SF-12 based on a nationally representative sample and using preferences based on the general US population. **Methods.** The Medical Expenditure Panel Survey (MEPS) 2002 and 2000 data were used as independent derivation and validation sets to estimate the relationship between SF-12 scores and EQ-5D index scores, controlling for sociodemographic characteristics and comorbidity burden. Prediction equations for end-users who only have access to SF-12 scores were derived and compared. The empirical performance of censored least

absolute deviations (CLAD), Tobit, and ordinary least squares (OLS) analytic methods were compared by calculating the mean prediction error in the validation set. **Results.** The fully specified CLAD model resulted in the lowest mean prediction error, followed by OLS and Tobit. The CLAD prediction equation based only on SF-12 scores performed better than the fully specified OLS and Tobit models. **Conclusion.** The current research provides an algorithm for mapping EQ-5D index scores from the SF-12. This algorithm may provide analysts with an avenue to obtain appropriate preference-based health-related quality-of-life scores for use in cost-effectiveness analyses when only SF-12 data are available. **Key words:** health-related quality of life; SF-12; EQ-5D; cost-utility analysis; cost-effectiveness analysis; utility; econometric methods. (*Med Decis Making* 2006; 26:401–409)

BACKGROUND

Health status measures provide a numeric score representing a profile of health status across several domains, such as physical and mental health. The SF-12 and its longer form, the SF-36, are examples of generic health status measures. Although these

measures provide important information about health-related quality of life (HRQL), they do not incorporate preferences for health states and cannot be used in cost-effectiveness analyses. In contrast, preference-based HRQL instruments are appropriate for the calculation of quality-adjusted life-years (QALYs) in cost-effectiveness analysis (CEA).^{1,2} QALYs rely on preference weights as the metric by which to reflect the HRQL impact of different health states.

Preference-based HRQL instruments can be separated into 3 categories: direct preference measures, multiattribute health status classification systems (MAHSCS), and estimation or mapping methods.³ Examples of direct preference measures include the standard gamble (SG),⁴ time tradeoff (TTO),⁵ and rating scales such as the Visual Analog Scale (VAS).⁶ MAHSCS offer a practical and theoretically appealing means of incorporating community-based preferences that is consistent with the underlying theory of CEA and QALYs.⁷ The EQ-5D, Health Utilities Index (HUI), and Quality

Received 19 March 2005 from University of Colorado School of Pharmacy, Pharmaceutical Outcomes Research Program, Denver, Colorado. This research was done without external funding. The authors would like to thank 2 anonymous reviewers for helpful comments. Revision accepted for publication 8 May 2006.

Address correspondence to Patrick Sullivan, PhD, University of Colorado School of Pharmacy, Pharmaceutical Outcomes Research Program, 4200 East Ninth Avenue, Box C238, Denver, CO 80262; telephone: (303) 315-1560; fax: (303) 315-1797; e-mail: Patrick.Sullivan@UCHSC.edu.

DOI: 10.1177/0272989X06290496

of Well-being (QWB) Scale are examples of MAHSCS with empirically based, community-derived preference functions.⁷⁻¹⁰ However, only the EQ-5D has a preference scoring function based on the US general population.

Like MAHSCS, mapping estimation methods rely on the theoretical relationship between health status and preferences to derive preference estimates.³ Mapping methods enable analysts to estimate preference-based HRQL scores when MAHSCS or directly elicited preferences are not feasible or available. The SF-12 is a valid and reliable measure of generic health status and has been used extensively.¹¹ In cases where analysts would like to conduct a cost-effectiveness analysis and preference-based HRQL scores are not available but SF-12 scores are, it is practical to use a mapping algorithm to estimate EQ-5D index scores from SF-12 scores. Previous studies have developed mapping algorithms to estimate EQ-5D index scores from the SF-12 in nationally representative US populations.^{12,13} However, the preferences underpinning the EQ-5D index scores were based on a community sample in the United Kingdom.¹⁴ UK-based TTO scores have been shown to be significantly different from US-based TTO scores,¹⁵ implying that the EQ-5D index scores may also differ. Recently, preferences have been derived for the EQ-5D index based on a community sample in the United States.¹⁶ These scores would be more appropriate for cost-effectiveness analyses done from the societal perspective in the United States.¹ The purpose of this research is to provide a mapping algorithm for estimating EQ-5D index scores from the SF-12 based on a nationally representative US population and using the recently available US community-based preferences.¹⁶

METHODS

MEPS and HRQL Instruments

The Medical Expenditure Panel Survey (MEPS) is a nationally representative survey of the US civilian noninstitutionalized population with oversampling of Hispanics and blacks, collecting detailed information on demographic characteristics, health conditions, health status, use of medical care services, health insurance coverage, income, and employment.¹⁷ The current research used the 2000 and 2002 MEPS public use data.¹⁷ The 2000 and 2002 data were used to provide a larger sample and because they each contain data on unique individuals that do not overlap (unlike MEPS 2001 data, which have panels

that overlap with 2000 and 2002) but are equally nationally representative. The SF-12 and EQ-5D were administered via the self-administered questionnaire (SAQ), a paper-and-pencil questionnaire administered to adults aged 18 years or older. If an individual was unable to respond to the SAQ, the questionnaire was completed by a proxy. In the MEPS 2000 data set, there were 14,286 individuals with valid MCS-12, PCS-12, and EQ-5D index scores and 23,647 in 2002, respectively. The 2000 and 2002 medical conditions files were also linked to the full-year consolidated data file to estimate the comorbidity burden of each individual. In the medical conditions files, 693 three-digit ICD-9 codes were mapped by professional coders from the medical conditions reported by survey respondents.¹⁸ Clinical classification categories (CCC) were derived in MEPS from clinically meaningful combinations of ICD-9 codes into 259 mutually exclusive categories.^{17,19} The number of chronic conditions (NCC) was calculated as the total number of chronic CCC codes reported by an individual. Only chronic (lasting >1 year) conditions were included to ensure that the condition was experienced while the SAQ was administered.

The EQ-5D consists of a 5-item descriptive system that measures 5 dimensions of health status (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with 3 levels per dimension (no problem, some problems, and extreme problems). There are 243 unique health states deriving from the combination of all possible dimensions and levels of the 5 questions. Preferences for these health states are mapped using a multiattribute value function (MAVF). The scoring algorithm for the EQ-5D_{index} descriptive system used in this research is based on US community preferences.¹⁶

From this scoring algorithm, EQ-5D_{index} scores are calculated based on responses to the 5-item questionnaire. The EQ-5D has been used as an HRQL and preference measure in a wide variety of diseases and conditions in more than 600 publications, and its construct validity, reliability, and responsiveness have been documented extensively.^{20,21}

The SF-12 Health Survey is a general health status instrument with 12 questions producing 2 summary scores, the Physical Component Summary (PCS-12) and the Mental Component Summary (MCS-12). These summary scales are scored such that higher scores represent better physical and emotional function, respectively, and are standardized so that the mean score is 50 and standard deviation is 10 in the general population.¹¹

Econometric Issues in Estimating Preference-Based HRQL Scores

EQ-5D index scores in MEPS exhibit a ceiling effect, with a significant number of respondents rating themselves in full health.^{22,23} This has been shown for the HUI as well.²⁴ Ignoring the bounded nature of preference-based HRQL scores and using ordinary least squares (OLS) will result in biased and inconsistent estimates.²⁵ The Tobit model provides a consistent and efficient alternative for estimating such data.^{25,26} The Tobit model specification with upper censoring at 1.0 is defined as follows:

$$y_i^* = \beta'x_i + u_i,$$

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* < 1.0 \\ 1.0 & \text{if } y_i^* \geq 1.0 \end{cases},$$

where y_i is a latent measure of the preference score, x_i is a vector of independent variables affecting preference scores, β is a vector of unknown parameters, u_i are residuals that are independently and normally distributed with mean 0 and a common variance σ^2 , and y_i is the actual preference score as measured by the EQ-5D.

However, in the face of heteroskedasticity or non-normality, the Tobit model also produces biased estimates.²⁵ In contrast, because the censored least absolute deviations (CLAD) estimator does not depend on distributional or homoskedasticity assumptions of the errors and is robust to censoring, it produces consistent estimates even in the face of heteroskedasticity, nonnormality, and censoring.²⁷ Based on these theoretical limitations that have been confirmed in the empirical econometric literature,^{24,28,29} the current research estimates EQ-5D index scores using CLAD, Tobit, and OLS.

To test for heteroskedasticity and normality in the Tobit model, we used the likelihood ratio test described by Petersen and Waldman³⁰ and the Hausman test, respectively.²⁵ All analyses were conducted in STATA. CLAD was performed in STATA using an existing user written add-in³¹ as well as additional programming to incorporate sample weights. Programs were written in STATA to perform the Hausman and the likelihood ratio tests.

Analysis

The MEPS sample design includes stratification, clustering, multiple stages of selection, and over-sampling of minority populations.³² Unless otherwise

noted, we incorporate MEPS sampling weights in all analyses, ensuring nationally representative estimates. The CLAD algorithm used in this research controls for MEPS sampling weights and produces nationally representative point estimates. However, unlike OLS and Tobit, CLAD does not make explicit assumptions about error structure. The result is that the MEPS variance structure cannot be formally included when using CLAD. As a result, bootstrapping based on the MEPS variance structure was used to calculate robust standard errors.

Independent derivation and validation data sets were used to compare the empirical performance of CLAD, Tobit, and OLS. The MEPS 2002 data were used as the derivation data set to regress EQ-5D index scores on independent variables. The MEPS 2000 data were then used as the validation data set to predict EQ-5D index scores using the regression coefficients from the derivation data set. Any EQ-5D index scores predicted >1 were truncated at 1.0 to remain consistent with the scale bounds. The absolute value of the difference between the predicted and actual EQ-5D index score (prediction error) was then calculated for each individual in the validation data set. The total prediction error, mean prediction error, and 95% confidence intervals were then calculated based on the prediction error in the validation data set (MEPS 2000) to measure the relative performance of the 3 analytic methods. CLAD was then used on the final data set (pooled MEPS 2000 and 2002) to produce the mapping algorithm for future use.

The approach to model specification was guided by an attempt to include all relevant variables that have been shown to have a theoretically meaningful and/or statistically significant relationship with EQ-5D index scores in previously published multivariate models. Based on previous research^{22,23} and mapping algorithms,^{12,13} the following variables were considered: PCS-12, MCS-12, PCS-12², MCS-12², age, sex, race, ethnicity, income, education, and NCC. The correlation of these variables was then examined. To avoid multicollinearity, variables with correlations ≥ 0.8 were omitted from the model specification. Although some variables that are not statistically significant may be retained with this approach and end-users likely do not have access to information for all of these variables, it was considered important to have a correct model specification to obtain unbiased estimates. After estimating the full model, alternatives are derived to enable mapping based only on MCS-12 and PCS-12 scores for end-users. In the final recommended algorithm, the

Table 1 Sociodemographic Characteristics, Mean Age, Chronic Conditions, EQ-5D_{index}, and PCS-12 and MCS-12 Scores of US Adults in the Final 2000 and 2002 MEPS Sample

	Number	%	Mean Age (Years)	Mean NCC	Mean EQ-5D	Mean PCS-12	Mean MCS-12
MEPS sample	37,933	100	45.4	1.9	0.868	49.3	51.2
Age group							
18–29	7875	21.2	23.7	0.7	0.922	53.4	51.4
30–39	7747	20.2	34.6	1.1	0.902	52.3	51.1
40–49	7916	20.8	44.4	1.6	0.872	50.3	50.8
50–59	6123	16.2	54.2	2.4	0.843	48	50.9
60–69	3868	10.1	64.2	3.2	0.825	45.4	52.2
70–79	2896	7.5	74.3	4	0.791	41.2	52.1
≥80	1508	4.0	83.6	4.4	0.738	36.8	50.8
Sex							
Female	20,594	51.9	46.0	2.2	0.855	48.5	50.4
Male	17,339	48.1	44.7	1.5	0.881	50.1	52.1
Race							
White	30,894	83.6	46.1	2	0.868	49.3	51.3
Black	5144	11.2	42.1	1.5	0.86	49	51
American Indian	359	0.8	42.1	1.9	0.834	48	49.3
Other race	1536	4.4	41.8	1.4	0.893	50.2	51.5
Ethnicity							
Hispanic	7935	11.1	39.3	1.2	0.879	50.2	50.9
Education							
No degree	9014	17.4	46.9	2.1	0.819	45.9	49.8
High school diploma	18,964	51.4	44.9	1.9	0.864	49	51.2
Other degree	2457	7.0	44.6	1.9	0.883	50.3	51.9
Bachelor's degree	5034	16.3	44.0	1.7	0.907	52	52
Master's or doctorate degree	2464	7.8	49.1	1.9	0.909	51.9	52.4
Family income (as % of poverty line)							
Poor/negative (<100%)	5029	9.7	43.4	2.1	0.789	45.3	47.5
Near poor (100%–124%)	1846	3.8	48.1	2.4	0.801	45.3	48.1
Low income (125%–199%)	5712	12.4	46.8	2.1	0.837	47.1	49.9
Middle income (200%–399%)	12,038	31.9	44.5	1.8	0.873	49.5	51.6
High income (≥400%)	13,308	42.2	45.9	1.8	0.897	51.1	52.5

Note: MEPS, Medical Expenditure Panel Survey; PCS-12, Physical Component Summary; MCS-12, Mental Component Summary; NCC, number of chronic conditions.

coefficients for the intercept, MCS-12, PCS-12, and MCS-12·PCS-12, were taken from the fully specified model results and a second constant was then derived to substitute information about sociodemographic variables that may not be available. This approach was taken to ensure that end-users with only MCS-12 and PCS-12 information could map EQ-5D index scores while ensuring that consistent estimates were obtained from the fully specified model.

RESULTS

Table 1 provides descriptive statistics for the MEPS 2000 and 2002 sample used for the final regressions. Sociodemographic characteristics, mean age, number of chronic conditions, EQ-5D index, and PCS-12 and MCS-12 scores are reported for the 2000 and 2002

MEPS sample. Mean EQ-5D index and PCS-12 scores decline by increasing age category. EQ-5D index, PCS-12, and MCS-12 scores appear lower for females than males, are lower for blacks and American Indians compared to whites, are higher for other races and Hispanics (except MCS-12) compared to whites and non-Hispanics, and are lower for lower levels of educational attainment and poverty status compared to higher levels. The mean NCC reported increases consistently by age category, is greater for females than males, and generally declines for higher levels of educational attainment and income.

The correlation of considered variables resulted in elimination of only 2 variables. Both PCS-12² and MCS-12² had extremely high correlation coefficients (0.99) with the PCS-12 and MCS-12, respectively, and were excluded from the model specification. All

Table 2 Comparison of Mean Prediction Error for OLS, Tobit, and CLAD Estimators and Previous Prediction Algorithms in the Validation Set (MEPS 2000)

US Preferences ^a			
Analytic Method	Mean Prediction Error	Standard Error Prediction Error	95% Confidence Interval Prediction Error
OLS ^b	0.0746	0.00087	0.0729–0.0763
CLAD ^b	0.0726	0.00095	0.0707–0.0745
Tobit ^b	0.0765	0.00113	0.0742–0.0787
CLAD WA ^c (PCS-12 and MCS-12 only)	0.0749	0.00091	0.0731–0.0767
CLAD MPE ^d (PCS-12 and MCS-12 only)	0.0744	0.00092	0.0726–0.0762

Note: Prediction error: absolute value of the difference between the predicted and actual EQ-5D index score for each individual in the validation data set. MEPS, Medical Expenditure Panel Survey; PCS-12, Physical Component Summary; MCS-12, Mental Component Summary; OLS, ordinary least squares; CLAD, censored least absolute deviations.

a. Prediction error for the predicted v. actual EQ-5D index score based on the US general population preferences.

b. Fully specified model.

c. CLAD weighted average—regression coefficients from the derivation sample for all variables except PCS-12 and MCS-12 scores were multiplied by the prevalence estimates in Table 1 to derive the prediction equation.

d. CLAD minimized prediction error—the effect of all variables except PCS-12 and MCS-12 scores were summarized by a constant that minimized the prediction error. This is the recommended prediction equation for end-users who only have access to PCS-12 and MCS-12 scores. This method uses the following prediction equation: EQ-5D = 0.057867 + 0.010367*PCS-12 + 0.00822*MCS-12 – 0.000034*PCS-12*MCS-12 – 0.01067.

other variables had acceptable correlations and were included in the final model specification.

The homoskedasticity assumption (likelihood ratio test = 506; $P < 0.001$) and the normality assumption (Hausman test statistic = -2851; $P < 0.001$) were rejected for the classic Tobit model, suggesting that Tobit estimates are likely biased. As with many health status measures in population health surveys, it appears that the EQ-5D index is not normally distributed and exhibits a significant ceiling effect (46% reported EQ-5D_{index} = 1.0). Given these factors, the CLAD estimates are theoretically the only consistent estimates of EQ-5D index scores.

The mean and total prediction error are compared across individuals (Table 2) to measure the relative empirical performance of CLAD, Tobit, and OLS in predicting actual EQ-5D index scores in the validation sample. CLAD resulted in the lowest mean prediction error. Surprisingly, in this experiment, OLS performed better than Tobit. The mean prediction error incorporates MEPS sampling weights and is nationally representative. The actual and predicted EQ-5D index scores are displayed graphically in Figure 1. Predicted scores <0.4 were extremely rare and are not shown. Of note is the fact that the minimum possible decrement from full health (1.0) is 0.14, resulting in no scores between 0.86 and 1.0 on the EQ-5D index.

To address the practical dilemma faced by end-users who do not have access to additional variables other than PCS-12 and MCS-12 scores, alternative prediction equations based only on PCS-12 and MCS-12

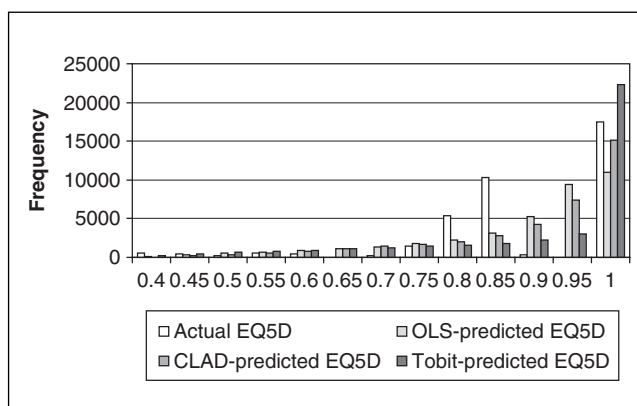


Figure 1 Actual, ordinary least squares (OLS), censored least absolute deviations (CLAD), and Tobit predicted EQ-5D index scores in the 2000 and 2002 Medical Expenditure Panel Survey (MEPS) sample.

scores were developed. To address this practical necessity, previous algorithms provide model specifications restricted to PCS-12 and MCS-12 scores alone.^{12,13} In the current approach, it was considered important not to estimate a model restricted to only PCS-12 and MCS-12 scores because it may result in model misspecification. Hence, our approach uses the results of the full model specification to ensure accurate estimates of the impact of PCS-12 and MCS-12 scores. The first alternative is to use the prevalence estimates provided in Table 1 along with the coefficients for each respective variable (i.e., coefficient for

near poor multiplied by the proportion of near poor in Table 1) from the full regression to estimate a “weighted average” composite effect of these variables. The CLAD “weighted average” (CLAD WA) approach results in a constant of -0.01647 for CLAD, -0.00677 for OLS, and -0.03132 for Tobit in the final sample.

The second alternative was to take advantage of the specification of the full model and derive a constant that minimized the absolute prediction error. The full model specification can be characterized as follows:

$$\text{EQ-5D} = \alpha + \gamma_i + \eta_i + \varepsilon_i,$$

where α = the intercept, γ_i = the available SF-12 variables, η_i = all other (non-SF-12) variables, and ε = the error term:

$$\gamma_i = \delta_1 \text{PCS-12}_i + \delta_2 \text{MCS-12}_i + \delta_3 \text{PCS-12}_i \cdot \text{MCS-12}_i,$$

$$\eta_i = \beta_1 \text{NCC}_i + \beta_2 \text{Age}_i + \beta_3 \text{Male}_i + \beta_4 \text{Hispanic}_i + \beta_5 \text{Black}_i + \beta_6 \text{AmericanIndian}_i + \beta_7 \text{Other}_i + \beta_8 \text{Income2}_i + \beta_9 \text{Income3}_i + \beta_{10} \text{Income4}_i + \beta_{11} \text{Income5}_i + \beta_{12} \text{Degree2}_i + \beta_{13} \text{Degree3}_i + \beta_{14} \text{Degree4}_i + \beta_{15} \text{Degree5}_i.$$

Because the end-user has limited information about variables in η_i , we want to offer a constant that will minimize the absolute prediction error. $\hat{\eta}$ is the solution to the following optimization problem:

$$\text{Min}_{\hat{\eta}} |EQ5D_i - \alpha - \gamma_i - \hat{\eta}|.$$

Note that α and γ_i are not estimated in this equation. We already have consistent estimates of α and γ_i from the results of the fully specified regression (and these coefficients were used in the minimization problem above). This method ensures that the regression coefficients from the full model specification are used. $\hat{\eta}$ is a constant representing the composite effect of all non-SF-12 variables. We estimated $\hat{\eta}$ to be -0.01067 . Table 2 shows that both CLAD WA and CLAD minimized prediction error (CLAD MPE) methods result in good empirical prediction, and CLAD MPE performs better than the fully specified OLS regression in the validation set.

Regression results for the final model specification in the entire data set (MEPS 2000 and 2002) are shown in Table 3. CLAD, Tobit, and OLS regressions are estimated to compare results empirically. The statistical significance of the coefficients in Table 3

varies across methods for Hispanic, other race, income, and education. Pseudo- R^2 for OLS and Tobit were calculated as $1 - \text{LL}_{\text{constant}}/\text{LL}_{\text{full}}$, whereas pseudo- R^2 for CLAD was calculated as $1 - (\text{sum of raw deviations}/\text{sum of absolute deviations})$. $\text{LL}_{\text{constant}}$ is the log likelihood of the regression of EQ-5D index scores on a constant, and LL_{full} is the log likelihood of the regression of EQ-5D index scores on the full regression model. The sum of the raw deviations is similar to the sum of the absolute deviations for the regression of EQ-5D index scores on a constant, whereas the sum of absolute deviations in the denominator is for the full model. Although all 3 measures of pseudo- R^2 are measuring something similar (the amount of variance explained by the full model compared to a model including no explanatory variables), they are not comparable across methods. CLAD is not based on maximum likelihood (ML) estimation, and hence a pseudo- R^2 based on log likelihood cannot be computed. Although OLS does not use ML, a log likelihood can be calculated and compared to Tobit. However, OLS ignores 46% of the EQ-5D index scores that are clustered at 1.0 and therefore explains a much smaller amount of variance than Tobit (which estimates the variance of a latent variable with possible scores =1.0 and thus results in a much larger pseudo- R^2). Given the lack of comparability of the pseudo- R^2 values, the comparison of prediction error (Table 2) may be a more appropriate measure of the relative performance of the 3 methods for the purpose of mapping closest to actual EQ-5D index scores.

DISCUSSION

The purpose of this research was to develop an algorithm to allow mapping of SF-12 scores to EQ-5D index scores based on a nationally representative sample. There are 2 published algorithms to map EQ-5D index scores from the SF-12 in nationally representative data sets.^{12,13} The current research differs in several important respects from the previous research. First, the mapping algorithm provided in this research is based on a nationally representative US sample and maps to the EQ-5D index values based on the general US population,¹⁶ providing the first mapping algorithm of preference-based EQ-5D index scores that meet the criteria of the Panel on Cost-Effectiveness in Health and Medicine (PCEHM) for reference-case cost-effectiveness analyses from the societal perspective in the United States. The US and UK community-based TTO scores have been shown to differ significantly, implying that EQ-5D index scores may also differ.¹⁵ Although very useful

Table 3 Final Model Specification Results in the 2000 and 2002 MEPS Sample

	OLS Coefficient	Standard Error	CLAD ^{a,b} Coefficient	Standard Error ^c	Tobit Coefficient	Standard Error
PCS-12	0.012416***	0.000410	0.010367***	0.000788	0.005901***	0.000438
MCS-12	0.01032***	0.000395	0.00822***	0.000814	0.004191***	0.000421
PCS-12*MCS-12	-0.000073***	0.000008	-0.000034*	0.000015	0.000121***	0.000009
NCC	-0.005085***	0.000468	-0.005838***	0.001077	-0.007495***	0.000636
Age	-0.000331***	0.000042	-0.0003***	0.000063	-0.000892***	0.000079
Male	-0.004914***	0.001256	-0.002958*	0.001351	-0.00468*	0.002333
Hispanic	0.005688**	0.001902	0.002022	0.001816	0.012855***	0.003401
Black	-0.002183	0.001921	-0.001439	0.002409	-0.002323	0.003491
American Indian	-0.008085	0.009279	-0.005967	0.008769	-0.004776	0.013405
Other race	0.010217**	0.002942	0.003037	0.003063	0.018278**	0.006425
Near poor	0.009712*	0.004475	0.005712	0.004755	0.014206*	0.006324
Low income	0.014765***	0.002970	0.00682	0.003566	0.018878***	0.004479
Middle income	0.014942***	0.002906	0.006256*	0.003151	0.016858***	0.004156
High income	0.016543***	0.002829	0.006298*	0.003133	0.020348***	0.004104
High school degree	0.003612	0.001899	0.002066	0.002071	0.001345	0.003169
Other degree	0.005612	0.002951	0.005353	0.003294	0.005116	0.005562
Bachelor's degree	0.012431***	0.002433	0.008268**	0.002599	0.021885***	0.004580
Master's or doctorate degree	0.015208***	0.002604	0.011652**	0.003212	0.026864***	0.005135
Intercept	-0.079759***	0.019981	0.057867	0.043367	0.158734***	0.021086
Pseudo-R ²	0.580 ^c	—	0.405	—	0.923	—

Note: MEPS, Medical Expenditure Panel Survey; PCS-12, Physical Component Summary; MCS-12, Mental Component Summary; OLS, ordinary least squares; CLAD, censored least absolute deviations; NCC, number of chronic conditions.

a. The CLAD prediction equation is recommended for end-users who have data on all variables.

b. CLAD minimized prediction error (CLAD MPE) is recommended for end-users who only have access to MCS-12 and PCS-12 scores. The average effect of all variables except PCS-12 and MCS-12 scores is summarized by a constant = -0.01067. This method uses the following prediction equation: EQ-5D = 0.057867 + 0.010367·PCS-12 + 0.00822·MCS-12 - 0.000034·PCS-12·MCS-12 - 0.01067.

c. Bootstrap estimates of the standard errors and P values based on the MEPS sample variance structure.

*P < 0.05. **P < 0.01. ***P < 0.001.

to understanding the relationship between the SF-12 health status instrument and the EQ-5D index, previous algorithms mapped to EQ-5D index values based on a community sample in the United Kingdom¹⁴ and do not meet the PCEHM criteria for the United States.¹ Second, the current research takes a different analytic approach, addressing the characteristics of the data (i.e., 46% censoring, nonnormality, and heteroskedasticity) and comparing different econometric approaches.

This mapping algorithm may be useful to researchers who need preference-based HRQL scores when only SF-12 scores are available. In these cases, analysts could use the algorithm from this research to derive EQ-5D index scores from available values of the SF-12.

Some comments on the practical application of this algorithm may be useful to end-users. First, in most cases, only the MCS-12 and PCS-12 scores are available, and values for sociodemographic variables are missing. In these cases, it is recommended that the CLAD MPE model with the derived constant

(-0.01067) be used. This method uses the following prediction equation: EQ-5D = 0.057867 + 0.010367·PCS-12 + 0.00822·MCS-12 - 0.000034·PCS-12·MCS-12 - 0.01067. When predicting EQ-5D index scores, it is important to note that predicted scores >1.0 are possible. Hence, analysts should truncate any predicted score greater than 1 at 1.0, as was done in the empirical comparison. It is important that end-users incorporate appropriate sensitivity analyses of point estimates mapped from SF-12 scores. More detailed information on the appropriate use of sensitivity analysis is provided elsewhere.³³ One can use the standard error or 95% confidence intervals (CIs) of the mean predicted EQ-5D score to derive an estimate of the uncertainty in the point estimates. The mean, standard error, and 95% CI for the CLAD MPE-predicted EQ-5D score in the full MEPS sample were 0.8912, .0010636, and 0.8891-0.8933, respectively. However, an equation for calculating an adjusted standard error for mapping is provided by Franks and others:¹²

$$\sqrt{\text{Var}/((N-1)\cdot(R^2))}$$

where var is the sample variance of the individual predictions in a sample of size N , and $R^2 = 0.41$.

Researchers can calculate EQ-5D index scores and adjusted standard errors by entering PCS-12 and MCS-12 scores and sample standard deviation at the journal's Web site, mdm.sagepub.com.

Although it would have been feasible to develop a model based on all 12 items of the SF-12, it may not be practical for end-users (many of whom will only have data on the PCS-12 and MCS-12 scales). In addition, the model specification becomes unwieldy and difficult for applied use, as discussed by Franks and others.¹²

The current research compares the empirical performance of different analytic methods to estimate EQ-5D index scores in an independent, large, nationally representative validation set, which has implications for the broader econometric analysis of preference-based HRQL scores that tend to be clustered at 1.0. In this nationally representative data set, EQ-5D index scores were found to have a high degree of censoring (46%), with errors that were not normally distributed and exhibited heteroskedasticity. Theoretically, these properties suggest that CLAD is the only unbiased estimator of EQ-5D index scores.²⁵ However, given the statistical properties of EQ-5D index scores in this data set, it is surprising how well OLS performed and how poorly Tobit performed. The mean prediction error with OLS was not significantly higher than with CLAD and was lower than Tobit. The 95% CI of the prediction errors overlapped for all 3 analytic methods. Future research is needed to examine the most appropriate theoretical approach to estimating preference-based HRQL scores with these statistical properties, as well as the relative empirical performance of analytic methods.

The current research is not without limitations. There is a significant amount of variance that is unexplained by the SF-12 and sociodemographic variables in this analysis. This uncertainty should be incorporated in probabilistic sensitivity analyses in cost-effectiveness analyses using these estimates. In addition, the method of reducing predicted scores above 1.0 may improve the empirical performance of all 3 methods but likely improves the performance of CLAD and Tobit more than OLS. Unlike previous mapping models, the current model does not include squared PCS-12 and MCS-12 terms due to the lack of apparent nonlinearity and potential for multicollinearity. The range of predicted EQ-5D scores for the full CLAD model ranges from 0.33 to 1.04. This range covers approximately 98% of actual EQ-5D scores in MEPS. The range for the simplified

CLAD model (CLAD MPE) was .42 to 1.02, which covers >96% of all actual EQ-5D scores. If using the minimum and maximum values of actual PCS-12 and MCS-12 scores, the CLAD MPE model-predicted EQ-5D scores range from .20 to >1.0. This range covers >99.9% of actual EQ-5D scores in MEPS and demonstrates the potential range of the model given actual SF-12 scores in MEPS. Nonetheless, it should be noted that the mapping algorithms provided in this research may not provide the full range of individual EQ-5D scores possible at the lower bounds.

Although the construct validity, reliability, and responsiveness of the EQ-5D have been documented extensively in both general and specific disease populations, the EQ-5D is not without limitations. The parsimony in the 5 items and 3 levels of the EQ-5D questionnaire translates into ease of administration and efficiency of preference assessment but also results in a potential lack of discrimination. As discussed, 46% of the nationally representative MEPS sample had an EQ-5D score of 1.0. In addition, from the full health state of 1.0, the minimum possible decrement in the EQ-5D index score is 0.14. (However, it should be noted that using the mapping algorithm provided in this article will provide predicted scores within this range.) Taking these limitations into account, there is no consensus that other theoretically based MAHSCS are superior to the EQ-5D.³⁴⁻³⁷ Each has its own advantages and limitations. The EQ-5D is currently the only instrument with a preference scoring algorithm based on the general US population that can be mapped from the SF-12 in a nationally representative data set.

The generalizability of the estimates provided in this research is limited to the MEPS sample of individuals with valid EQ-5D index, MCS-12, and PCS-12 scores. Individuals with complete responses for all 3 scores are likely to have slightly better health than others. MEPS is also restricted to the noninstitutionalized population. Nonetheless, it is rare to find both instruments in a nationally representative population, and the sample used in this research is likely as generalizable to the nation as is available.

In a recent examination of cost-utility analyses, 77% did not incorporate community-based preferences, and 33% used arbitrary expert or author judgment.³⁸ Although not perfect, the mapping algorithm provided in this research may provide analysts an avenue to obtain preference-based HRQL scores appropriate for reference-case cost-effectiveness analyses in the United States when only SF-12 scores are available.

REFERENCES

1. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.
2. Drummond MF. *Methods for the Economic Evaluation of Health Care Programmes*. 2nd ed. Oxford, UK: Oxford University Press; 1997.
3. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health*. 2000;21:587–611.
4. Von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. 2nd ed. Princeton, NJ: Princeton University Press; 1947.
5. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res*. 1972;7:118–33.
6. Froberg DG, Kane RL. Methodology for measuring health-state preferences: II. Scaling methods. *J Clin Epidemiol*. 1989;42:459–71.
7. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. *Health Utilities Index. Pharmacoeconomics*. 1995;7:490–502.
8. The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16:199–208.
9. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37:53–72.
10. Kaplan RM, Anderson JP. A general health policy model: update and applications. *Health Serv Res*. 1988;23:203–35.
11. Ware JJ, Kosinski M, Keller S. *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales*. 3rd ed. Lincoln, RI: Quality Metric; 1998.
12. Franks P, Lubetkin EI, Gold MR, Tancredi DJ, Jia H. Mapping the SF-12 to the EuroQol EQ-5D Index in a national US sample. *Med Decis Making*. 2004;24:247–54.
13. Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D preference scores from the SF-12 Health Survey in a nationally representative sample. *Med Decis Making*. 2004;24:160–9.
14. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35:1095–1108.
15. Johnson JA, Luo N, Shaw JW, Kind P, Coons SJ. Valuations of EQ-5D health states: are the United States and United Kingdom different? *Med Care*. 2005;43:221–8.
16. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43:203–20.
17. Agency for Healthcare Research and Quality (AHRQ). Puf Documentation Files. Agency for Healthcare Research and Quality (AHRQ). Available at: <http://www.meps.ahrq.gov/PUFFiles/H52/H52doc.htm>. Accessed 10 November 2003.
18. US Health Care Financing Administration. *The International Classification of Diseases, 9th Revision, Clinical Modification: ICD-9-CM*. 2nd ed. Washington, DC: US Department of Health and Human Services, Public Health Service; 1980.
19. Elixhauser A, Andrews RM, Fox S. *Clinical Classifications for Health Policy Research: Discharge Statistics by Principal Diagnosis and Procedure*. AHCPH Publication No. 93-0043. Rockville, MD: Agency for Health Policy and Research; 1993.
20. McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires*. 2nd ed. New York: Oxford University Press; 1996.
21. The EuroQol Group. Available at: <http://www.euroqol.org/download/ref.pdf>. Accessed 11 November 2003.
22. Sullivan PW, Lawrence WF Jr, Ghushchyan V. A national catalogue of preference-based scores for chronic conditions in the US *Med Care*. 2005;43:736–49.
23. Sullivan PW, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. *Med Decis Making*. 2006;26:410–20.
24. Austin PC. Bayesian extensions of the Tobit model for analyzing measures of health status. *Med Decis Making*. 2002;22:152–62.
25. Green WH. *Econometric Analysis*. 4th ed. Upper Saddle River, NJ: Prentice Hall; 2000.
26. Maddala GS. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, UK: Cambridge University Press; 1999.
27. Powell J. Least absolute deviations estimation for the censored regression model. *J Econometrics*. 1984;25:303–25.
28. Clarke P, Gray A, Holman R. Estimating utility values for health states of type 2 diabetic patients using the EQ-5D (UKPDS 62). *Med Decis Making*. 2002;22:340–9.
29. Austin PC. A comparison of methods for analyzing health-related quality-of-life measures. *Value Health*. 2002;5:329–37.
30. Petersen D, Waldman D. The treatment of heteroscedasticity in the limited dependent variable model. Mimeo. November 1981.
31. <http://www.stata.com/stb/stb58>
32. Agency for Healthcare Research and Quality. *Computing Standard Errors for MEPS Estimates*. Agency for Healthcare Research and Quality. Accessed September 2003.
33. Briggs AH. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics*. 2000;17:479–500.
34. Gold MR, Franks P, McCoy KI, Fryback DG. Toward consistency in cost-utility analyses: using national measures to create condition-specific values. *Med Care*. 1998;36:778–92.
35. Macran S, Weatherly H, Kind P. Measuring population health: a comparison of three generic health status measures. *Med Care*. 2003;41:218–31.
36. Luo N, Chew LH, Fong KY, et al. A comparison of the EuroQol-5D and the Health Utilities Index Mark 3 in patients with rheumatic disease. *J Rheumatol*. 2003;30:2268–74.
37. Bosch JL, Hunink MG. Comparison of the Health Utilities Index Mark 3 (HUI3) and the EuroQol EQ-5D in patients treated for intermittent claudication. *Qual Life Res*. 2000;9:591–601.
38. Bell CM, Chapman RH, Stone PW, Sandberg EA, Neumann PJ. An off-the-shelf help list: a comprehensive catalog of preference scores from published cost-utility analyses. *Med Decis Making*. 2001;21:288–94.