Relative Disutilities of 47 Risk Factors and Conditions Assessed with Seven Preference-Based Health Status Measures in a National U.S. Sample

Toward Consistency in Cost-Effectiveness Analyses

Peter Franks, MD,* Janel Hanmer,† and Dennis G. Fryback, PhD†

Background: Preference-based health measures yield summary scores that are compatible with cost-effectiveness analyses. There is limited comparative information, however, about how different measures weight health conditions in the U.S. population.

Methods: We examined data from 11,421 adults in the 2000 Medical Expenditure Panel Survey, a nationally representative sample of the U.S. general population, using information on sociodemographics (age, gender, race/ethnicity, income, and education), health status (EQ-5D, EQ-VAS, and SF-12), 4 risk factors (smoking, overweight, obesity, and lacking health insurance), and 43 conditions. From the EQ-5D, we derived summary scores using U.K. [EQ(UK)] and U.S. weights. From the SF-12 we derived SF-6D, and regression-predicted EQ-5D (U.S. and U.K. weights) and Health Utility Index scores. Each of the 7 preference measures was regressed on each of the 47 problems (risk factors and conditions) to determine the disutility associated with the problem, adjusting for socio-demographics.

Results: The adjusted disutilities averaged across the 47 problems for the 7 preference measures ranged from 0.059 for the SF-6D to 0.104 for the EQ(UK). Correlations between each of the measures of the adjusted disutilities ranged from 0.85–1.0. Standardization, using linear regression, attenuated between measure differences in disutilities.

Conclusions: Absolute incremental cost-effectiveness analyses of a given problem would likely vary depending on the measure used, whereas the relative ordering of incremental cost-effectiveness analyses of a series of problems would likely be similar regardless of the measure chosen, as long as the same measure is used in each series of analyses. Absolute consistency across measures may be enhanced by standardization.

Key Words: health status, population health, socioeconomic status, chronic disease

Cost-effectiveness analyses (CEAs) can be used to inform policy by putting the health outcomes of disparate interventions on the same scale thereby making them comparable to weigh against costs. The U.S. Panel on Cost-Effectiveness in Health and Medicine recommended that these analyses use quality adjusted life years (QALYs) as the health outcome. QALYs are calculated by multiplying the duration of health conditions by their associated health-related quality of life (HRQL). QALY-compatible HRQL measures use preference-based interval level scores anchored at 0.0 (death) and 1.0 (perfect health). In a CEA, the incremental costs of an intervention are compared with the incremental gain in QALYs from use of the intervention, not the total number of QALYs accrued after the intervention. There are many off-the-shelf indexes of HRQL that could be used for CEAs, but there is concern that different indexes may lead to different estimates of QALYs gained from any given intervention, limiting the potential for CEA to contribute to policy.

Concern about the exchangeability of different health evaluations precedes most of the HRQL measures currently in use; in 1984 Read et al² reported that the absolute scores for health states varied widely when using time-trade off, standard gamble, and visual analog scales, the methods used to score current indexes. More recent research has shown that HRQL estimates by different indexes are not exchangeable based on absolute scores obtained from patients with certain health conditions.⁴⁻⁶ Some authors have concluded that these differences would result in noncomparable estimates in economic evaluation.⁸⁻¹¹ Technically, these concerns are misdirected for CEA, however, because CEA requires only that incremental QALY gains from an intervention are exchangeable, not the absolute score for a given health state or condition.

Concerns about index exchangeability have been raised in research examining changes in patient health status.⁷⁻¹⁴ These studies report that different indexes result in
different HRQL changes and thus different incremental QALY gains. Although most authors warn against exchanging indexes (ie, using HRQL weights from different indexes for different health conditions within 1 comparison study) because of the potential to affect CEA outcomes, 1 CEA study found these differences to be irrelevant. In general, the studies warning against the exchangeability of indexes focus on the absolute magnitude of change without considering measurement error. Typically, these studies also have focused only on a single condition.

In this report, we examine 3 questions. First, are estimates of disutilities associated with given conditions different depending on the HRQL measure used and, if so, how different? We define adjusted disutility as the difference in the average utility reported by persons with a given problem compared with those without the problem, adjusted for sociodemographics, but not for comorbidities. Second, are the relative ordering or ranking of conditions by their associated disutility similar when examined by different HRQL measures? Third, can the absolute disutilities associated with a given condition captured by each HRQL measure be made more similar through standardization? We analyze data from the 2000 Medical Expenditure Panel Survey (MEPS), a nationally representative sample of the U.S. noninstitutionalized adult civilian population. The self-administered questionnaire included in MEPS contained 3 of the off-the-shelf measures of HRQL: the EuroQol EQ-5D (EQ-5D), the EuroQol Visual Analog Scale (EQ VAS), and the SF-12 version 1. From these 3 measures, we derived 7 different HRQL summary scores: the EQ-5D with U.S. weights, the EQ-5D with U.K. weights, the EQ VAS, the SF-6D, a predicted Health Utilities Index Mark 3, and a predicted EQ-5D with both U.K. and U.S. weights. We present the relationships among the adjusted disutilities associated with 47 risk factors and conditions assessed by each HRQL measure to explore the issue of exchangeability in CEAs intended for resource allocation.

METHODS

We used the 2000 Medical Expenditure Panel Survey (MEPS), a nationally representative survey of the U.S. noninstitutionalized civilian population conducted by the Agency for Healthcare Research and Quality (AHRQ). MEPS has a panel design, following individual respondents for 2 years. The interviewer administered household component of MEPS includes sociodemographic and insurance information enabling classification of respondents as privately insured or uninsured. The self-administered questionnaire, in English and Spanish, contains the EQ-5D Index and VAS, and the SF-12 version 1. The questionnaire also includes height and weight (allowing calculation of body mass index [BMI] and, therefore, classification of persons as normal weight [BMI 20–25 kg/m²], overweight [BMI >25–30], and obese [BMI >30]), smoking status (current smoker or not), and the self-reported presence or absence of the following 8 common conditions: hypertension, coronary heart disease, angina, myocardial infarction, diabetes, stroke, asthma, and emphysema. In addition, for each health care encounter, respondents provide information on the condition associated with that encounter. AHRQ categorizes these conditions into Clinical Classification Categories (CCC). We considered all chronic CCCs reported by at least 100 respondents (35 CCCs met these criteria, including at least one from each organ system, and 3 chronic psychiatric conditions. The full list is available from the authors). Further details of the survey are available at www.meps.ahrq.gov.

Data were obtained from 15,438 adults (>18 years of age). Completion rates were 97% for the EQ-5D Index, 86% for the SF-12, and 85% for the EQ VAS. This analysis focuses on the 11,421 (75%) persons providing complete responses to all 3 measures. Persons who provided complete HRQL responses (compared with those without complete responses) tended to be younger, more educated, have higher incomes, and were less likely to be minorities (results available from authors).

The EQ-5D Index includes 5 questions that classify health on 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The respondent has a choice of 3 levels for each dimension: no problems, some problems, or extreme problems/unable to. These data may be converted into a single summary score (EQ-5D) by applying scores from a valuation set. We used 2 valuation sets, one derived from a U.K. sample [herein called the EQ(UK)], and one derived using a U.S. sample [EQ(US)]. In addition, the EQ-5D captures a self-rating of health by the VAS, anchored at 100 (best imaginable health) and 0 (worst imaginable health), we rescaled VAS scores from 1 to 0. Both the EQ-5D Index and EQ VAS exhibit acceptable reliability and validity.

The SF-12 was constructed as an abridged version of the SF-36 to reflect the physical and mental component summaries of the parent scale. The SF-12 measures physical functioning, physical or emotional role limitations, pain, general health, vitality, social functioning, and mental health. The responses are combined to generate physical component summary (PCS-12) and mental component summary (MCS-12) scales. The SF-12 has been used increasingly in cross-sectional and longitudinal health studies because of its low respondent burden; its ability to reproduce the SF-36 summary scales; and its reported reliability, validity, and responsiveness.

We derived 4 preference-based summary scores from the SF-12. First, we derived SF-6D scores using the algorithm developed by Brazier and Roberts. The SF-6D uses 7 questions from the SF-12 to construct 6-dimensional health states that were evaluated with the standard gamble technique. Regression analysis was used to model the preferences assigned to SF-6D health states through the standard gamble technique. The derived preference-based score exhibits reasonable performance characteristics. Second, we used the MCS-12 and PCS-12 scores to derive a predicted EQ-5D using the U.K. weights [pEQ(UK)] by applying the regression mapping algorithm reported by Franks et al; this predicted measure explains 62% of the variance in EQ(UK) scores in this sample, and produces results very similar to those reported previously. Next, we derived a predicted EQ-5D using the U.S. weights [pEQ(US)] using an identical regression mapping approach to that reported.
by Franks et al\textsuperscript{26}; this predicted measure (details available from the authors) produced results consistent with those previously observed for the pEQ(UK), and again the predicted measure explains 62\% of the variance in the EQ(US) scores. The equation for EQ(US) = 0.88082 + 0.00920 * (PCS12 – 49.8) + 0.00583 * (MCS12 – 51.4) – 0.00007 * (PCS12 – 49.8) * (MCS12 – 51.4) – 0.00004 * (PCS12 – 49.8)\textsuperscript{2} – 0.00009 * (MCS12 – 51.4)\textsuperscript{2}, where PCS12 and MCS12 are the respondent’s scores.

Finally, we derived a predicted Health Utility Index Mark 3 (HUI-3)\textsuperscript{29} score (pHUI) using a regression algorithm based on the MCS-12 and PCS-12.\textsuperscript{27} Two algorithms were considered,\textsuperscript{27,30} they both explain about 50\% of the variance in HUI-3 scores, and in this sample were highly correlated; one of them,\textsuperscript{30} however, yielded utilities greater than 1, and thus was not used.\textsuperscript{30}

Data on sociodemographic characteristics were obtained by in-person interviews with 1 household member providing information for each household resident. We characterized race/ethnicity as white, African-American, Latino, or other. Income was characterized as a percentage of the federal poverty level (<100\%, 100–124\%, 125–199\%, 200–399\%, and ≥400\%). We categorized education as years of schooling completed (<12, 12, 13–15, and ≥16).

Analyses

Data were analyzed using STATA (Version 8.2, StataCorp, College Station, TX) and adjusted for the complex survey design of MEPS. Reported results incorporate both design and sampling weights, yielding nationally representative sample of U.S. persons,\textsuperscript{17} though our approach could use any of the summary scores as the standard. To derive standardization parameters we regressed the disutility found with the EQ(US) onto that found with each of the other HRQL measures in turn:

\[
\text{EQ(US)} = \text{Intercept} + \beta_1 \cdot \text{HRQL}_x + \beta_2 \cdot (\text{HRQL}_x - \text{mean}[\text{HRQL}_x])^2
\]

Where EQ(US) is the adjusted disutility measured by the EQ(US) for a given problem, HRQL\textsubscript{x} is the adjusted disutility measured by one of the HRQL measures. We include a squared term to allow for nonlinear relationships between the 2 HRQL measures, that is, the relationship between any 2 measures may vary by the level of disutility. To reduce collinearity, the squared term is included as the square of the deviation of the HRQL disutility from its mean.

RESULTS

The person-level mean summary scores among the 11,421 adults providing complete information on all 3 HRQL measures (EQ-5D, VAS, SF-12) are shown in Table 1, together with the correlations among these 7 scores. The mean summary scores were highest for the EQ(US) (0.87) and lowest for the pHUI (0.79) indices. The correlations between each of the scores varied from a low of 0.62 between the VAS and both the EQ(US) and EQ(UK), to 0.99 between the EQ(UK) and EQ(US). The correlations between each of the SF-12 derived instruments all exceeded 0.9.

The adjusted disutilities (and 95\% CIs) as measured by the 7 preference measures are illustrated in Figure 1 for the 4 risk factors and the 8 conditions addressed in the self-administered questionnaire (selected for illustration pur-

| Table 1. Sample Means and Correlations Among 7 Preference Measures (n = 11,421) |
|---------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                               | Means (Range) | EQ(UK)         | EQ(US)         | VAS            | SF-6D          | pHUI           | pEQ(UK)        |
| EQ(UK)                         | 0.832 (−0.594, 1) | 0.99          |                |                |                |                |                |
| EQ(US)                         | 0.871 (−0.109, 1) | 0.62          | 0.62           |                |                |                |                |
| VAS                            | 0.797 (0, 1)     | 0.72          | 0.72           | 0.63           |                |                |                |
| SF-6D                          | 0.824 (0.345, 1) | 0.72          | 0.72           | 0.63           |                |                |                |
| pHUI                           | 0.788 (−0.265, 0.932) | 0.77          | 0.76           | 0.69           | 0.91           |                |                |
| pEQ(UK)                        | 0.831 (0.012, 0.997) | 0.79          | 0.78           | 0.71           | 0.91           | 0.97           | 1.00           |
| pEQ(UK)                        | 0.871 (0.252, 0.984) | 0.79          | 0.78           | 0.72           | 0.91           | 0.97           | 1.00           |

\( \text{VAS} \) was rescaled 0 (death)-1 (perfect health), from the original 0–100.

© 2006 Lippincott Williams & Wilkins

Copyright © Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
poses). In comparing the charts within the figure, 4 key findings emerge visually. First, the absolute point estimate of disutility for a given problem is different for most of the HRQL measures for most problems. For example, the disutility for angina is 0.13 using the EQ(UK) and 0.08 using the SF6. Second, for each problem, the CIs for each point estimate overlap considerably, suggesting that, even in this large sample, there is uncertainty whether the observed differences by problem across HRQL measures represent true differences in disutility. Third, the average disutility within each problem, no matter which summary measure is used, appears more similar than do disutilities between problems. The disutilities for risk factors appear less than those for any of the health conditions. Within the set of risk factors, being obese appears to have a greater impact than being overweight. Among conditions, the disutilities associated with emphysema and stroke appeared more marked than those for hypertension and asthma. Fourth, the pattern of impact measured by each summary score compared with other summary scores is quite consistent across the different problems. Typically, the EQ(UK) and the pHUI and pEQ(UK) suggest the largest disutilities, while the SF-6D and EQ(US) suggest the smallest.

The relationships among the 7 preference measures for all 47 problems are summarized in Table 2, which shows that the mean problem disutility (averaged across all 47 problems, separately for each preference measure) varies between 0.059 for the SF-6D to 0.100 for the EQ(UK). The EQ(UK) showed the greatest range of disutilities (0.000–0.202) whereas the SF-6D showed the smallest range (0.008–0.128). The Spearman correlations between each of the HRQL measures were all 0.85 or greater. The VAS exhibited the lowest correlations with the 2 EQ-5D measures; other correlations were all 0.91 or greater.

The results of the regression standardizations are shown in Table 3, together with an illustration of how to use the results to translate an obtained disutility into EQ(US) disutility units. The regressions of the EQ(US) on the EQ(UK) and on the VAS did not yield a statistically significant squared term. Those for the other regression analyses had statistically significant squared terms, but these terms were small in size, suggesting little evidence overall of nonlinear relationships between the EQ(US) and other HRQL measures. Regression analyses examining the relationships among the other HRQL also revealed little evidence of nonlinear relationships (results not shown).

![Figure 1](https://example.com/figure1.png)

**FIGURE 1.** Each graph shows the parameter estimate and 95% CI for the disutility associated with each problem on 7 preference scores adjusted for sociodemographics. The y-axis is size of disutility in HRQL units.
Table 4 illustrates the impact of regression standardization on the estimates of disutilities using the different HRQL measures compared with the EQ(US). The EQ(US) shows the greatest mean absolute difference from the EQ(US) in disutility estimates, followed by the pEQ(UK). After regression standardization, the absolute differences are reduced, though by very little for the VAS. Differences between the disutilities observed with the EQ(US) and VAS look similar before and after regression standardization. The standardization does reduce the number of problems for which the difference in disutility between the EQ(US) and the other HRQL measures exceeds 0.03. Before standardization, the number ranges from 24 (for the EQ[UK]) to 6 (for the SF-6D); whereas after standardization the number ranges from 4 (for the VAS) to 0 (for the EQ[UK]).

**DISCUSSION**

This analysis confirmed that different HRQL measures produce different absolute estimates of disutilities for the same problem, though with considerable overlap in the confidence intervals for each measure. The relative impact of 47 different self-reported conditions and risk factors was measured similarly by the 7 different index scoring systems; the Spearman correlations of conditions by instrument were all 0.85 or greater. Finally, standardizing the scores using linear regression increased the similarity of the absolute disutility estimates among the measures.

Our findings suggest that any one of the HRQL index systems examined would likely yield a similar rank-order of disutilities associated with the diseases and risk factors examined to any of the other measures and that it may be possible to compare absolute values of disutilities across these index systems after standardization. Without standardization, however, comparing the absolute values of CEAs derived using different index systems would likely produce inconsistent results.

These results suggest some steps toward consistency in CEAs. Herein, we focus on the cautions to this statement and the limitations of the study. Questions about the internal validity of the findings center largely on our analytic choices. First, we did not adjust for comorbidity, to obtain “pure” estimates of the disutilities examined. Our intent was not to produce a catalog of weights but to explore the relative and absolute comparability of disutilities captured by different HRQL scores. In our view, there is no satisfactory way to make the adjustment. There is no complete list of comorbidities available that would allow adjustment for all possible conditions. One possibility is to adjust for the number of comorbid conditions, but such an approach assumes each condition affects HRQL equally; clearly, this is not the case. Further, it is uncertain whether such adjustment is appropriate given the complex causal interactions among various risk factors and morbidities. For example, persons without health insurance and obese persons have lower HRQL—the reasons

| TABLE 2. Correlations Among 7 Preference Measure Estimates of Adjusted Mean Problem Disutilities (n = 47) |
|---------------------------------------------------------------|---------------------------------------------------------------|
| **Mean Adjusted Disutility (Range)** | **Correlations** |
| EQ(US) | 0.104 (0.000, 0.202) | EQ(UK) | 1.00 |
| VAS | 0.072 (0.000, 0.144) | EQ(US) | 0.85 |
| pHUI | 0.075 (0.008, 0.164) | VAS | 0.92 |
| pEQ(UK) | 0.092 (0.004, 0.196) | pHUI | 0.92 |
| SF-6D | 0.059 (0.008, 0.128) | SF-6D | 0.94 |
| pEQ(US) | 0.067 (0.007, 0.145) | pEQ(UK) | 0.97 |

| TABLE 3. Regression Standardization Analyses: EQ(US) Disutility Regressed on Disutility Captured by Other HRQL Measures |
|---------------------------------------------------------------|---------------------------------------------------------------|
| **Main Effect** | **Squared Deviation Effect** | **Intercept** |
| PE | 95% CI | P | PE | 95% CI | P | PE | 95% CI | P | R² |
| EQ(US) | 0.688 | 0.676–0.701 | 0.000 | — | — | — | — | — | — | 1.00 |
| VAS | 0.772 | 0.625–0.920 | 0.000 | — | — | — | — | — | — | 0.014 |
| pHUI | 0.668 | 0.574–0.762 | 0.000 | — | — | — | — | — | — | 0.015 |
| pEQ(UK) | 0.676 | 0.584–0.769 | 0.000 | — | — | — | — | — | — | 0.013 |
| SF-6D | 1.055 | 0.912–1.198 | 0.000 | — | — | — | — | — | — | 0.015 |
| pEQ(US) | 0.961 | 0.828–1.095 | 0.000 | — | — | — | — | — | — | 0.013 |

Each row represents the results of a regression (n = 47) of the adjusted disutility captured by the EQ(US) regressed on the disutility captured by one of the other HRQL measures and the squared deviation of the disutility from the mean observed with that measure. R² indicates the variance explained by the regression equation. Nonsignificant parameter estimates are not shown.

For example, supposing a user obtains a mean disutility of 0.05 using the SF-6D. Given that the mean disutility for the SF6 is 0.059 (from Table 2), then expressed in EQ(US) units the disutility is given by 1.055 * 0.05 – 5.024 * (0.05 – 0.059) * (0.05 – 0.05) + 0.015 – 0.068.

PE indicates parameter estimate.
presumably being because both risk factors are associated with more frequent and more severe morbidity, so that adjusting for that comorbidity would be over-adjusting. The same may be true for many of the conditions as well. For example, many prevalent cardio-vascular conditions are interconnected in causally complex ways. Consider 4 of the 8 conditions in the self-administered questionnaire—hypertension, angina, myocardial infarction, and congestive heart failure. Treating, curing or preventing any one of them would likely affect the risks and morbidity associated with any of the others. For example, beta-blockers will treat or prevent all 4 conditions—thus improving the comorbidities along with the index condition. So, again, for many CEA purposes, excluding the contributions of comorbidities may be over-adjusting. However, we conducted some analyses using adjustment for the 8 conditions in the self-administered questionnaire; these results (available from the authors) were quite consistent (though with smaller adjusted disutilities) with those presented.

Second, ordinary least squares (OLS) regression is not an entirely satisfactory method for analyzing HRQL scores that are skewed and have an upper bound (producing heteroskedasticity problems and possibly biased estimates). This may be a particular problem for the EQ-5D because 47% of respondents have scores at 1, as do 33% of those with at least 1 condition. However, alternatives, such as using censored quantile regression (CLAD) are also not completely satisfactory. For example, using censoring implies scores theoretically extend above 1, which is not consistent with the assumptions of preference measures. Quantile regression has the advantage of producing less biased estimates of the median when data are skewed. However, most CEAs typically use mean disutilities, so that advantages over OLS regression are less clear. Our analytic methods used robust standard errors to adjust for the complex cluster survey design; such robust standard errors also mitigate heteroskedasticity problems. We also conducted some analyses comparing our findings to those using CLAD regression for the EQ-5D, with reassuring findings (results available from the authors).

Our findings raise key issues for further exploration. First, how close should estimates be to be exchangeable? Despite concerns about the nonexchangeability of values across indexes because of absolute differences, there is little discussion of the differences that would be acceptable. Such a discussion should properly incorporate the error in estimates. Currently, the issue of a minimum clinically important difference (MCID) is debatable. However, a MCID of 0.03 has been suggested for the SF-6D, the HUI, and the EQ(UK). Based on this MCID, the mean absolute differences in disutilities observed after regression standardization (see Table 4) may be acceptable.

The VAS conformed least well after standardization, reflecting its lower correlation with the EQ-5D. The VAS differs from the other measures in representing personal preferences versus community preferences. The U.S. Panel on Cost-Effectiveness in Health and Medicine recommended the use of community preferences in reference case CEAs. It is interesting that the VAS values are similar to the others systems for most conditions and risk factors. While most literature suggests that personal preferences of those with chronic illnesses show evidence of adaptation (ie, are higher than those that the general population would assign to those conditions) this phenomenon is not seen in all studies. It is possible that these disparate findings reflect that adaptation varies by condition. Such heterogeneity in adaptation may explain the lower correlations we observed of the VAS with the other measures. The lower correlations may also reflect heterogeneity in the extent to which the EQ-5D captures the disutility associated with different conditions.

In considering the generalizability of the findings, the following considerations emerge. We used the 2000 MEPS dataset, a nationally representative sample of the U.S. noninstitutionalized civilian population. We also used all the available conditions and risk factors of sufficient prevalence in the dataset. Although the spectrum of conditions examined in this study is broader than previous studies, it would be useful to explore the issue with more conditions. It is also important to explore the performance of the standardization equations for the same conditions in a new sample (for example, the 2002 MEPS). Of note, none of the disutilities examined were greater than 0.209. We are uncertain how the standardization would perform for conditions with larger utilities. This is a particular concern because the squared terms will tend to attenuate the size of a standardized disutility at higher levels of disutility. Given the wide confidence

### TABLE 4. Mean Absolute Differences in Disutility Scores Between the EQ(UK) and the Other HRQL Scores Before and After Regression Standardization

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unstandardized</th>
<th>Regression Standardized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean, Median, Range</td>
<td>n &gt; 0.03*</td>
</tr>
<tr>
<td>EQ(UK)</td>
<td>0.032, (0.031, 0.005–0.061)</td>
<td>24</td>
</tr>
<tr>
<td>VAS</td>
<td>0.0016, (0.012, 0.000–0.057)</td>
<td>7</td>
</tr>
<tr>
<td>pHU</td>
<td>0.0016, (0.020, 0.000–0.057)</td>
<td>14</td>
</tr>
<tr>
<td>pEQ(UK)</td>
<td>0.0024, (0.019, 0.000–0.091)</td>
<td>13</td>
</tr>
<tr>
<td>SF-6D</td>
<td>0.0015, (0.011, 0.000–0.059)</td>
<td>6</td>
</tr>
<tr>
<td>EQ(UK)</td>
<td>0.0012, (0.008, 0.000–0.045)</td>
<td>3</td>
</tr>
</tbody>
</table>

*n > 0.03 indicates the number of (of 47) cases in which disutility difference between the EQ(UK) and the other HRQL measure exceeded 0.03.
intervals around the squared terms, it is quite possible that a different function would apply at those higher levels of disutility.

We were limited to the HRQL measures included in the survey and the 7 possible scoring algorithms for these measures. It would be useful to explore the performance of more intact HRQL indexes. It is, perhaps, not surprising that the 2 EQ-5D measures and the 4 values derived from the SF-12 responses produce highly correlated findings.

We included only those subjects completing all 3 preference measures included in the survey, which was 75% of the total sample. Persons who provided complete responses tended to be younger, more educated, have higher incomes, and were less likely to be minorities; therefore, our study may have limited generalizability. Analyses not shown here found that correlations among the measures were lower when all available data were used. This change in correlations illustrates the importance of using identical samples when comparing different measures. For this study, using the sample with complete information was a priority because we wanted to assess the relative disutilities of different problems when using different summary scores. The sample used in this study is likely to be the most nationally representative available to explore the issues addressed in this paper.

All of the index systems evaluated in this study are used to measure QALYs. Other methods exist to measure the burden of disease, including disability-adjusted life years and years of healthy life. This study does not address the exchangeability of HRQL estimates among the QALY, disability-adjusted life years, and years of healthy life frameworks. Gold and Muennig found different rank orders of 5 diseases among these 3 frameworks, though they used different population samples to examine each framework and did not consider the confidence intervals around the estimates. Further research is needed to examine the exchangeability of these 3 approaches to measuring HRQL.

In conclusion, our data illustrate that 7 different health preference scores will rank 47 conditions and risk factors from a large population sample in a similar order despite differences in the absolute mean scores within each condition or risk factor. We have also shown that standardization will produce more similar disutilities. Thus, standardization of disutility estimates from different preference measures (at least those examined in this study) may make them more comparable for use in cost-effectiveness analyses.

REFERENCES


34. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Qual Life Outcomes*. 2003;1:4.

