

NBER WORKING PAPER SERIES

A PROPOSED METHOD FOR MONITORING U.S. POPULATION HEALTH:
LINKING SYMPTOMS, IMPAIRMENTS, AND HEALTH RATINGS

Susan T. Stewart
Rebecca M. Woodward
Allison B. Rosen
David M. Cutler

Working Paper 11358
<http://www.nber.org/papers/w11358>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2005
Revised January 2007

© 2005 by Susan T. Stewart, Rebecca M. Woodward, Allison B. Rosen, and David M. Cutler. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Proposed Method for Monitoring U.S. Population Health: Linking Symptoms, Impairments,
and Health Ratings

Susan T. Stewart, Rebecca M. Woodward, Allison B. Rosen, and David M. Cutler

NBER Working Paper No. 11358

May 2005, Revised January 2007

JEL No. I10,I12

ABSTRACT

We propose a method of quantifying non-fatal health on a 0-1 QALY scale that details the impact of specific symptoms and impairments and is not based on ratings of counterfactual scenarios. Measures of general health status are regressed on health impairments and symptoms in different domains, using ordered probit and ordinary least squares regression. This yields estimates of their effects analogous to disutility weights, and accounts for complex non-additive relationships. Health measures used include self-rated health status on a 5-point scale, EuroQol 5D (EQ-5D) scores, and ratings of current health using a 0-100 rating scale and a time-tradeoff. Data are from the nationally representative Medical Expenditure Panel Survey (MEPS) year 2002 (N=34,615), with validation in an independent sample from MEPS 2000 (N=21,067) and among 1420 adults age 45-89 in the Beaver Dam Health Outcomes Study. Decrement weights for symptoms and impairments are used to derive estimates of overall health-related quality of life, laying the groundwork for a detailed national summary measure of health.

Susan T. Stewart
Harvard Interfaculty Program
for Health Systems Improvement
c/o NBER
1050 Massachusetts Ave
Cambridge MA 02134
sstewart@nber.org

Allison B. Rosen
Departments of Internal Medicine
and Health Management and Policy
University of Michigan
300 North Ingalls
Suite 7E10
Ann Arbor, MI 48109
abrosen@umich.edu

Rebecca M. Woodward
Harvard Interfaculty Program
for Health Systems Improvement
c/o NBER
1050 Massachusetts Ave.
Cambridge, MA 02145
rebecca_woodward@harvard.edu

David M. Cutler
Department of Economics
Harvard University
1875 Cambridge Street
Cambridge, MA 02138
and NBER
dcutler@harvard.edu

Population health measurement is of interest for many purposes, from the evaluation of changes in health benefit structures to the valuation of medical treatments. To accurately measure population health, it is necessary to measure health-related quality of life (HRQOL) in addition to mortality. This paper proposes a method of health measurement that utilizes the broad range of reported symptoms and impairments available in national data. The method produces a summary health score on a 0-1 scale that can be combined with mortality to yield a more comprehensive picture of what is happening to the nation's health, while retaining the rich detail on symptoms and impairments needed to allow tracking of the impact of these over time. It accounts for co-relationships between impairments and uses regression methods that do not require individuals to rate specific health problems.

A number of instruments have been developed to measure general health status and HRQOL. However, no gold standard has been agreed upon (Gold, Siegel et al. 1996). Existing measures vary along a number of dimensions, three of which we examine via the method proposed in this paper: the breadth of domains covered, the method used to weight multiple problems within the domains, and the approach used to aggregate item/domain scores into an overall summary measure.

In terms of the breadth of coverage, current measures span the continuum from single-question ratings of general health to detailed questionnaires covering multiple domains with multiple items. Within the category of single question ratings of general health, utility-based ratings of overall current health status such as standard gamble (SG) and time-tradeoff (TTO) have been recommended as summary measures that incorporate morbidity and mortality on a single 0-1 scale. Another popular brief measure is an overall self-rated health status question (SRHS), which asks respondents to describe their health in one of 4 or 5 categories ranging from

‘poor’ to ‘excellent’. While not derived from expected utility theory, SRHS has been shown to be an accurate predictor of subsequent changes in functional status and of mortality and is related to a number of specific aspects of health (Idler and Kasl 1995; Wilcox, Kasl et al. 1996; Cutler and Richardson 1997; Idler and Benyamini 1997; Idler, Russell et al. 2000; DeSalvo, Bloser et al. 2006). However, both preference-based and psychometric ratings of general health lack detail regarding specific symptoms and impairments that may contribute to overall health. An understanding of the impact on quality of life of these symptoms and impairments (many of which are measured on an ongoing basis by our national health surveillance surveys) is quite useful for assessing changes in health over time. Our first goal is to examine the impact of symptoms and impairments from different domains on overall health assessment.

Detailed existing health questionnaires measure an individual’s functioning in specific health domains, with one or more questions used to measure each domain, and a scoring algorithm developed to form domain scores and possibly an overall summary score. There is substantial variation across general health instruments in the domains assessed and in the items that are implicitly or explicitly covered by these domains, as illustrated in Table 1. Psychometric ‘profile measures’ yield a set of domain scores that are not necessarily designed to be combined to form a single score. The most popular of these measures is the SF-36 (Ware, Kosinski et al. 2000) (and the shorter SF-12 and SF-8 versions). In contrast, ‘preference-based’ measures assign utility weights to items or health states, and the weights associated with an individual’s reported problems are combined to yield an overall summary measure of health on a 0-1 scale (by convention, 0 denotes death and 1 denotes perfect health). Such measures include the Health Utilities Index (HUI Mark I, II and III, developed in Canada, (Horsman, Furlong et al. 2003), the Quality of Well-Being Scale (QWB, interviewer and self-administered versions, developed in

Southern California, (Kaplan, Bush et al. 1976; McDowell and Newell 1996), the EQ-5D (“EuroQol”, developed in Europe, (Brooks, Rabin et al. 2003), the 15-D (Finland, (Sintonen 2001), the ‘Assessment of Quality of Life’ (AQOL, Australia, (Hawthorne, Richardson et al. 1999) and the SF-6D (Brazier, Roberts et al. 2002), an econometric transformation of SF-36 data that yields a score on a 0-1 scale based on SG valuations of a sub-set of 11 SF-36 questions.

Most instruments include items measuring physical and social function, and do not include symptoms unless they are severe enough to affect function. Examples of items typically excluded are headache, rash, sleep problems, weight problems, cognitive problems, sexual problems, and specific physical limitations such as lifting or use of fingers. Exclusion of such items can result in ceiling effects (Fryback, Dasbach et al. 1993; Essink-Bot, Krabbe et al. 1997; Jenkinson, Gray et al. 1997; Andresen, Rothenberg et al. 1998; Kaplan, Ganiats et al. 1998; Johnson and Pickard 2000); a large portion of the general population is sufficiently healthy that it does not report decrements in function. However, these people do experience a range of health problems that can reduce HRQOL. Failure to identify and value the sources of such HRQOL reductions precludes tracking improvements in health if these problems are alleviated.

Determining the weights to assign to different impairments is a second longstanding issue. Disutility weights for preference-based measures are typically derived from studies in which samples of community members are asked to rate a number of hypothetical health scenarios using a valuation method: SG, TTO, or a rating scale (RS) such as a thermometer from 0 to 100. The health states are presented as ‘counterfactual scenarios’, in which a respondent is asked to imagine having a particular health state, but may or may not have actually experienced the health problem. Typically those who have experienced a condition rate it less severely than do those who have not (Sackett and Torrance 1978; Epstein, Hall et al. 1989; Slevin, Stubbs et

al. 1990; Ubel, Loewenstein et al. 2003). A number of explanations have been put forth to explain this phenomenon. People with the condition may adapt to it over time, they may make a mental shift and rate themselves in comparison to other sick people (or to how healthy they could expect to be with their health problem), and/or those without the condition may focus too narrowly on the potential negative life impact of the problem (a focusing illusion)(Kahneman and Tversky 2000; Ubel, Loewenstein et al. 2003). In this circumstance, there is controversy regarding whose values more accurately reflect the true health of a person in a given state, with the standard recommendation being to use individual weights for personal treatment decisions, and “community preferences” for cost-effectiveness and policy analysis (Gold, Siegel et al. 1996).

However, there is an extensive literature in psychology suggesting that ratings of counterfactual health states may not be accurate—that people overestimate the negative impact of a future hypothetical event on their happiness (Gilbert, Pinel et al. 1998; Wilson, Wheatley et al. 2000). People have been found to overpredict the extent and duration of their negative emotional reactions to a range of events, from the loss of one’s favored sports team or political candidate, to the failure to achieve tenure in a professorial position, the end of a close romantic relationship, losing a child in a car accident, being diagnosed with cancer, becoming paralysed, or being sent to a concentration camp (Gilbert, Pinel et al. 1998), (Wilson, Wheatley et al. 2000). This ‘durability bias’ calls into question the ability of individuals to make accurate judgments about the impact of hypothetical health problems. Indeed, in one study, UK residents valuing 42 of the EQ-5D health states rated 38% of the states as worse than death (Dolan 1997). These

findings are clearly inconsistent with the reality that most people experiencing these health states do not commit suicide or profess a wish to be dead (Dolan and Kahneman working paper).¹

We propose an alternate method of deriving weights for health states that does not require the direct rating of counterfactual health situations by either people with a condition or people without it, but rather asks people to rate only their current general health and to report their current symptoms and impairments. The effect of the specific symptoms and impairments is then estimated based on their relationship to this general health rating. If there is consistency in underlying response scales across people with similar health problems, this method will yield an accurate estimate of the effect of the problem.

Finally, existing instruments incorporate different approaches to forming a summary health score. Questionnaires in which all items contribute equally to the final score implicitly assign equal weight to all problems. Some scales such as the SF-36 use factor analysis to assign different weights to items based on their relative frequency and correlations with other items (Ware, Kosinski et al. 2000). However factor analysis can inappropriately substitute variation in the prevalence of health problems for variation in the impact of these health problems on health, and can thus exclude items with important effects on health if they have a low correlation with others or occur infrequently (Kaplan, Bush et al. 1976). Among preference-based scales, some use an additive model (i.e. 15-D, SF-6D), assuming scores on one item or domain are unaffected by scores on others. Others use a multiplicative model (HUI, AQOL), or a combination of both (QWB, EQ5D). We allow for a flexible functional form that can accommodate additive or non-additive relationships between pairs of symptoms and impairments within and across domains.

¹ Population valuations of EQ5D health states using TTO have yielded worse weights in the U.K. than in the U.S., particularly for states characterized by extreme problems (Johnson et al., 2005), which may also serve as evidence that such ratings are problematic.

The paper is structured as follows. The first section describes the methodology that we use for measuring health. The second section discusses the data. The third section presents results, and the fourth section offers discussion and conclusion.

I. Methodology for Measuring Health

Our analysis proceeds in two parts: the development of weights for symptoms and impairments in different domains, and the calculation of summary measures. The first part of our analysis is to estimate the HRQOL decrement of specific health problems, by relating ratings of general overall health to the presence of specific impairments and symptoms. We estimate item weights using regression analysis. To illustrate the methodology, we consider two symptoms or impairments, SI_1 and SI_2 , with responses for each individual of SI_{1i} and SI_{2i} . Our equations are of the form:

$$\text{Health}_i = b_1 \cdot SI_{1i} + b_2 \cdot SI_{2i} + e_i \quad (1)$$

where Health_i is a general rating of current health, SI_i are specific symptoms and impairments and b 's are the estimates of the effects of these symptoms or impairments on health. As we discuss below, we performed these analyses using several different measures of general health as dependent variables. We use three data sources, all of which include a 5-point rating of general health worded: "In general, would you say your health is: excellent, very good, good, fair, or poor?" (self-rated health status, SRHS). For comparison we also replicated analyses using utility-based measures available in each sample. Despite our concerns about measures based on people's ratings of counterfactuals, this comparison is important in light of the policy

recommendation favouring preference-based measures for cost-effectiveness and policy analysis (Gold, Siegel et al. 1996), and the widespread use of these measures. The Medical Expenditure Panel Survey (MEPS) administered two utility-based metrics: an EQ5D and a rating of current health on a scale from 0 (worst possible) to 100 (best possible). The Beaver Dam Health Outcome Study (BDHOS) includes a time tradeoff (TTO) rating of current health.

An ordered probit model was used to predict SRHS, due to the ordinal nature of the question. To predict EQ5D, RS, and TTO ratings, OLS regression was used, since each is a linear outcome measure. Some researchers have used a two-part model for the TTO and EQ5D truncated distribution (i.e. Sullivan et al., 2005), but its use has been found to make little difference with the Beaver Dam data.²

There are often two or more questions about different aspects of health within a particular domain. For example, our pain domain in Beaver Dam includes three questions regarding pain/discomfort (described in appendix 2). Past summary measures have varied in their treatment of these multiple symptoms, for example assigning the weight associated with the worst symptom (0.349 for sexual organ pain in the QWB (Kaplan and Anderson 1996)), or weighting overall health states that include one of 3 levels on a general pain question (EQ5D, -0.17 for moderate pain/discomfort as the only symptom (Shaw, Johnson et al. 2005)) or one of 5 categories based on combined levels of pain and activity limitation (HUI II, -0.25 for frequent pain with occasional disruption of normal activities (Torrance, Feeny et al. 1996)). To test for the appropriate relationships between items, we include interaction terms between all pairs of items in the MEPS samples.³ Due to the smaller sample size in Beaver Dam, we include interactions

² Personal communication, D. Fryback.

³ We do not consider higher order interactions because of sample size issues.

between only some pairs of items within domains (20 interactions out of a possible 266).

Algebraically, suppose we are considering one domain that includes two items (SI_1 and SI_2). We relate overall health ratings to each of the items and to the interaction between the items:

$$\text{Health}_i = b_1 \cdot SI_{1i} + b_2 \cdot SI_{2i} + b_{12} \cdot SI_{1i} \cdot SI_{2i} + e_i \quad (2)$$

Several common assumptions are special cases of this framework. In the case of a worst symptom indicator, we would find (assuming SI_1 is worst) $b_1 < b_2 = -b_{12}$: the first symptom has an adverse effect on health, and the second symptom has an adverse effect only when the person does not have the first symptom. Alternative special cases are independent effects ($b_{12} = 0$), more-than-additive effects ($b_{12} < 0$), and decreasing marginal impairment ($b_{12} > 0$).

Our regression equation (1) includes all symptoms and impairments, but no sociodemographic variables. We chose not to control for such factors as age, gender, race, and socioeconomic status with the belief that these affect health primarily via specific symptoms and impairments caused by acute and chronic conditions. If we had controlled for sociodemographic variables, their coefficients would also reflect the effects of symptom and impairment variables that were related to these factors but were not adequately accounted for by our models.⁴

Because of the large number of possible interactions, there is a possibility of overfitting the data. To address this, we tested out-of-sample predictions. Predicted SRHS scores were calculated in MEPS 2000, based on results of regressions in MEPS 2002. Two models were fit: a more complete model including all possible two-way interactions between symptoms and impairments and a smaller model including terms for only those interactions with an N of 500 or

⁴ Still, alternative analyses controlling for age, sex and their interactions yielded similar results.

more (38 of the 66 possible interactions). A likelihood ratio test showed that the explanatory power was significantly greater with the 28 additional interactions (difference in log likelihood = 27.2, $p < 0.01$). Further, the mean-squared error in out-of-sample predictions was slightly lower for the more complete model (2.00 vs. 2.01). As a result, we use the model with more interactions. That said, the correlation between predicted health ratings for the two models was very high: 0.998. Thus in practical terms, the issue of overfitting is not particularly important.

In the case of preference-based measures, the health metric is on a 0 to 1 scale, so that the coefficients are directly interpretable as the utility reduction from the relevant symptom. This is advantageous. However the disadvantages of the TTO (and the EQ5D, whose scoring is based on TTO ratings) were noted above, especially the possibility that people are poor at performing counterfactuals about health states.

The SRHS and RS analyses avoid this by asking people only about their current health state. Because the SRHS scale is not 0 to 1, however, it needs to be transformed onto a 0 to 1 scale in order to estimate a quality measure that can be combined with mortality in population health assessment. To do this, we consider the best and worst cases. Mean predicted scores from the regressions were calculated first assuming that everyone reported the item (the worst case) and then assuming that no one reported the item (the best case). The difference between these mean predicted scores captures the broadest possible impact of having the symptom or impairment in light of the other conditions that people have, and was used to normalize SRHS-based weights.

Because the SRHS scale does not have the top and bottom anchors of a utility measure (perfect health, and death or ‘worst possible health’), we did not assume that these best and worst states were equal to 1 and zero on a utility scale. To estimate how far the best and

worst states were from these end points, we examined the scores for the best and worst case scenarios using weights derived from the rating scale and from TTO. SRHS-based item weights were then normalized to fall within this range, as described in the results section below.

The advantages of using the self-reported scale are that it is easily and frequently asked, it has minimal floor or ceiling effects, and it does not require people to answer counterfactual questions about their health. One limitation is that it assumes that all people consider their responses on the same scale—e.g. that excellent and good health mean the same thing to all people. If particular groups systematically report their health as better or worse than other groups, and particular health problems are over or under-represented among these groups, then the effects of these problems on HRQOL may be mis-estimated. Use of vignettes—descriptions of people whose health the individual rates—has recently been explored as a way to control for interpersonal differences in scale interpretation, and would be natural to examine if such vignettes were part of the data set (Salomon, Tandon et al. 2004). None of our surveys ask vignette questions, however.⁵

In addition, there is a question about whether true health is captured in a linear manner by the SRHS question. Equation (1) might be correct for true health, but it might be that people report SRHS as $SRHS = g(\text{Health})$. For example, if people are hesitant to use the bottom end of the scale, but make fine gradations at the top end, the impact of a given change in true health on self-reported health will be greater at the top of the scale than the bottom. We test for this in our empirical work by relating utility-based measures of health to predicted health using the self-reported scale and non-linear transformations of that (effectively recovering $g^{-1}(\cdot)$). We generally accept linearity or a very close approximation, as shown in the results section.

⁵ However, the use of vignettes could also introduce the problem of bias in rating counterfactuals, discussed above.

Once the item weights have been determined, we can form summary measures of health for each person, by applying the item weights to the symptoms and impairments that s/he reports. We compare the item weights derived from SRHS and from utility-based measures in each survey, as well as mean scores on all summary measures.

Although we focus in this paper on health decrements for specific symptoms and impairments, our results can be used to form disease-based disutilities. To do this, one would relate responses about specific symptoms and impairments to diseases the person has, and use that to form disease-specific disutilities. For data and space reasons, we do not pursue that in this paper.

II. Data

We use three sources of data in our analysis. The primary source is the 2002 wave of the Medical Expenditure Panel Survey (MEPS; N=39,165), a nationally representative survey of the U.S. civilian non-institutionalised population. In order to test the robustness of our findings and test out-of-sample predictions, we also replicate our analyses in the 2000 wave of MEPS, which contains an independent sample of 25,096 people. In both years, MEPS asked a variety of summary health questions: a five point general rating of health (SRHS), a rating of current health on a 0 to 100 scale, and the EQ5D, a summary measure based on questions about moderate and severe impairment in five domains—mobility, self-care, usual activities, pain, and anxiety/depression--which we score using U.S. community preferences measured with TTO weights (Shaw, Johnson et al. 2005). The EQ5D questions and rating scale are shown in Appendix 3. The SRHS was asked of all respondents, whereas the EQ5D and RS were part of a

paper and pencil questionnaire administered only to those 18 and over. A proxy was asked to answer health questions for those unable to respond. The sample of those with data on SRHS and all impairments and symptoms was 34,274 in MEPS 2002 and 20,767 in 2000. Statistical methods designed for survey data and sampling weights were used to account for the MEPS sampling factors including stratification, clustering, oversampling of minorities, and nonresponse.

Our third data source is the Beaver Dam Health Outcomes Study (BDHOS; (Fryback, Dasbach et al. 1993)), a community-based study of health status and quality of life. Though this survey is smaller, it enabled us to consider the broadest array of symptoms and impairments, due to its inclusion of the QWB, which was developed to cover the full array of health problems that affect people's daily lives (Kaplan, Bush et al. 1976; McDowell and Newell 1996).⁶ For the BDHOS, a random sample of 1431 residents age 45 and up was drawn from a larger sample developed for the Beaver Dam Eye Study. Interviews were performed face-to-face, lasted approximately one hour, and included questions about current health status, chronic medical conditions, sensory problems, current medications, past surgeries, and life stresses. Participants were excluded if they were cognitively unable to answer interview questions or were institutionalised in an acute or chronic care facility. We use data from 1422 respondents with complete data on the QWB, interviewed between 1/1/1991 and 8/14/1992. The age range of the sample was 45 to 89, with a mean age of 64 (SD =10.8), and the sample was almost entirely Caucasian (Fryback, Dasbach et al. 1993). BDHOS includes a TTO rating of current health as well as the five point summary rating of health.

⁶ The QWB includes a symptom list in addition to questions in mobility, social, and physical function. For our study, information on all of the symptoms from the symptom list reported by each respondent was added to the data set from original paper interview sheets; original BDHOS data included only the item from the symptom list reported by the respondent as most bothersome, as was required for QWB scoring.

In both MEPS and the BDHOS we examined all the health questions and grouped them into relatively standard domains, based on our review of the health measurement literature (Kaplan, Ganiats et al. 1998; Hawthorne, Richardson et al. 1999, Kaplan, 1998 #46; Ware, Kosinski et al. 2000; Sintonen 2001; Brooks, Rabin et al. 2003; Horsman, Furlong et al. 2003, 2003 #31). These include social activity, physical activity, pain, mental health, vitality, sensory, and others. Impairments and symptoms within each domain in each survey are listed in Table 3. Levels of severity were available for some items in MEPS (primary activity limitation, walking, pain, vitality), allowing the effects of these levels to be examined separately.

III. Results

Responses to Summary Measures

The distributions of the four summary measures of health used in our analyses are shown in Figure 1 and summary statistics are shown in Table 2. The mean EQ5D score in both MEPS surveys was 0.87. The mean rating scale was 0.81 in MEPS 2002 (0.79 in 2000). The correlation between the EQ5D and RS is reasonably high (0.66). The somewhat lower mean health using the rating scale could reflect several factors. The EQ5D is based on questions about 5 domains, whereas the RS asks about ‘current overall health’, which would lead an individual to consider a broader array of factors affecting health. Also, utilities obtained by RS are typically lower than those obtained using methods that incorporate risk or uncertainty (Torrance, Feeny et al. 2001). A much lower percentage had perfect RS scores than EQ5D scores (0.08 vs.0.48). There has been some evidence of end-aversion towards the top end of the RS(Torrance, Feeny et al. 2001), however it is natural to have fewer people end up in the top category when there are more categories on which to make a finer gradation of one's health.

The mean TTO in Beaver Dam was 0.86. Beaver Dam is an older population (45+ versus 18+ in MEPS). For comparison, for people age 45 and over in MEPS 2002, the mean EQ5D score was similar at .85, and the mean RS rating was lower, at .79.

None of the summary measures were normally distributed; all were skewed toward good health, as would be expected in a general community population. However, the EQ5D and TTO distributions were truncated, with 48 and 59 percent of people reporting 1. This is typical in a general population; for example, in MEPS 2002, 44% of the population age 18 and over reported having none of the symptoms and impairments that we considered. However, the fact that responses were not heaped at ‘excellent’ on the 5-point scale indicates that this measure is more sensitive to factors that affect people’s health. On the EQ5D, scores of 1 represent those who do not suffer from any problems in any of the 5 domains queried. This suggests that the EQ5D does not ask about a sufficient number of domains for community-based health assessment. On the TTO, the responses of 100 constitute a mix of respondents who felt they were in perfect health and those who disliked their current health state but were unwilling to reduce their life expectancy by even 0.5% to live in excellent health. While 0.5 percent seems small, it actually encompasses a large range of response. If a year in good health is worth \$100,000, a conventional estimate, in dollar terms, a health change of 0.5 percent is worth perhaps \$500. People might not be willing to pay \$500 annually to fix a number of significant health complaints, however such problems can nevertheless have significant impacts on HRQOL and be alleviated by medical care.

Comparisons of the distributions of SRHS and EQ5D, RS, and TTO measures are shown in Table 3. As the first panel shows, only 37 percent of those with perfect EQ5D scores rated their health as excellent on the 5-point question, again reflecting health problems not captured by

the EQ5D. EQ5D scores are rarely below 0.75 (12 percent of cases). But only half of those in this bottom group report fair or poor health. A higher portion (63 percent) of those with perfect RS ratings had excellent SHRS scores. Comparison of SRHS and TTO ratings indicates that only 20 percent of those with TTO ratings of 100 rated their health as excellent on the 5-point question. On all three measures, a small percentage (4%) of those with low utility-based scores (<0.75) rated their health as excellent. These contradictory scores may reflect the cognitive challenges inherent in utility elicitation. Indeed, the validity of utilities may be subject to the numeracy of those surveyed (Woloshin, Schwartz et al. 2001).

Regression Analyses

Tables 1 and 2 of Appendix 1 show the proportion of the sample reporting each item and combination of items within domains. These tables also report coefficients from ordered probit models predicting SRHS and OLS regression models predicting EQ5D, RS, and TTO.⁷ All of the main effect coefficients are negative (symptoms predict worse health), as expected, with a few exceptions in Beaver Dam. Almost all of the symptoms and impairments are statistically significantly related to the health measures in MEPS and, in Beaver Dam, 40 percent of them are statistically significantly related to SRHS and 20 percent significantly predict TTO. However, because many of the variables are correlated, the statistical significance of any single variable is less important than the relative magnitude of their effects.

Many of the interaction terms are positive, indicating that a combination of multiple impairments is not as bad as a simple additive model would suggest, although instances of

⁷ Regression results from MEPS 2000 were similar to those from MEPS 2002 and are not shown, but are available upon request. Pain and self-care were excluded from regressions on EQ5D because they are included nearly directly in the EQ-5D.

negative coefficients for interaction terms are also found. In a few cases in MEPS and several cases in Beaver Dam, positive interaction effects are larger than one or both of the associated main effects, such that having two problems rather than one appeared to improve health ratings. Although this is not likely the case—and most of these positive estimates include zero in the confidence interval—we did not attempt to constrain large interaction terms or main effects, reasoning that they were part of a complex set of results, and it was unclear which terms (if any) were overly large or small. Overall, our results reject a theory of worst symptoms dominating or of additivity across comorbid impairments.

As noted above, self-reported health scores were normalized using the difference between the hypothetical scenarios in which everyone had every symptom/impairment and in which no one had any of the symptoms/impairments. We equated the best self-reported health state to 0.90, the predicted score for those with none of the impairments using the Rating Scale in MEPS 2002. The worst state was equated to 0.17, the worst score based on the decrement weights for all impairments from the Rating Scale in MEPS 2002.

Predicted Summary Health

The predicted measure of health based on self-reported general health status is shown in Table 2. Mean predicted health was 0.79 in both the MEPS 2002 and 2000.⁸ close to the mean rating scale rating (0.81) and broadly similar to the mean EQ5D score (0.87). Predicted SRHS-based health from MEPS 2002 is plotted in Figure 2. There is a truncation at the highest possible score, 0.90, that would be expected in a healthy population. As we discuss below, this suggests

⁸ This is the mean summary score among those age 18+ in MEPS, for comparison with RS and EQ, which were only asked of adults. The mean SRHS-based summary score for all ages was 0.81.

that the set of symptoms and impairments in MEPS is not rich enough to capture some health impairments.

In BDHOS, mean predicted health based on SRHS was 0.75, well below the 0.86 TTO-based mean (Table 2).⁹ For comparison to the age range in the BDHOS sample, the mean SRHS-based score among those age 45 in MEPS 2002 and over is the same (0.73) and the mean rating scale among those 45 and older is .79.

In addition to mean health, we care about the correlation across measures. Table 2 shows that the correlation between predicted health using the 5-point scale and the EQ5D is reasonably high (0.82). The correlation is lower, but still high, for the rating scale (0.70). Health using the TTO is very different from health derived using the self-assessment, however; the correlation between predicted SRHS and TTO in Beaver Dam was only 0.41. The lower correlation with TTO may reflect respondents' difficulty in understanding the task, and/or their unwillingness to trade off years of life for health problems below a certain threshold of seriousness.¹⁰

A key issue for our analysis is whether the relation between predicted self-reported health and the utility measures of health is linear. To test this, we regress the utility-based measures of health on our predicted SRHS-based health score and on this score squared. The fitted values from this regression along with the regression equations are shown in Figure 3. Linearity implies that the squared term should be insignificant. That is the case for TTO in Beaver Dam and the Rating Scale in MEPS 2002. The EQ5D in MEPS 2002 is positively related to the squared prediction from the self-assessment model. This is not surprising given the large sample size in

⁹ An alternate Beaver Dam model including only the items available in MEPS yielded a somewhat higher mean score of 0.77, equivalent to the mean summary score in MEPS.

¹⁰ Correlations between predicted SRHS and *predicted* summary scores derived from the regression models (not reported in Table) are quite high (0.93 and above). This reflects the reasonably common impact of different impairments on health using regression methods.

MEPS. However, as Figure 3 shows, although statistically significant, this relationship does not appear to be substantively divergent from linear. In light of these results, we use the linear assumption in our analysis.

Weights for Impairments and Symptoms

Decrement weights based on SRHS probit regressions are shown in Table 4. Domains with the largest decrements in the MEPS 2002 and 2000 samples were social activity, pain, and vitality, with the largest decrements for inability to perform major role activity (-0.12/-0.15), severe pain (-0.19/-0.21), and having little or no energy (-0.15/-0.13). In the physical activity domain, limitations in activities of daily living had a similarly negative effect (-0.10/-0.11), whereas other impairments had a milder impact, on par with or less severe than sensory impairments. While differences between weights across the two MEPS samples are apparent for several impairments, very few symptoms and impairments had effects that differed significantly across the two MEPS surveys; coefficients are statistically different when accounting for everything else in the model only for depressive symptoms using SRHS (decrement greater by 0.03 in 2002, $p=0.007$) and moderate vitality loss using RS (decrement greater by 0.02 in 2002, $p=0.04$).¹¹

Decrements were generally similar in self-reported health status regressions using Beaver Dam data. Inability to perform major role activity had the worst impact (-0.12), with pain and vitality yielding somewhat smaller disutilities (-0.06, -0.05). Anxiety and some physical limitations yielded more severe weights in the Beaver Dam sample. This may be partially explained by differences in item definitions and the non-representative BD population. The older

¹¹ A few interaction effects also differed significantly across waves in SRHS, EQ5D, and RS regressions (results not shown).

age range in Beaver Dam does not appear to explain differences from MEPS, since analyses in MEPS that were restricted to those ages 45 and over yielded decrement weights similar to those from the full MEPS samples.

Table 4 shows that the sum of the weights for the worst case scenario (having the worst level of all symptoms and impairments) is negative (worse than death). This was not impossible, but rare; only 0.2% of the MEPS 2002 sample had scores worse than death based on the item weights.¹²

Decrement weights based on TTO, EQ5D, and RS are shown in Table 5, and domain-specific scores derived from these measures and SRHS are shown in Figure 4. The pattern of decrements is similar across methods, with weights derived from utility-based measures somewhat less severe than those derived from the self-rated health measure. This is likely due to the lumping of utility-based scores at higher values. The largest differences are for social activity limitations, where the utility-based methods yielded smaller decrements.

In the Beaver Dam data, where a larger array of health problems are examined, several problems, such as sensory impairments, depressive symptoms, gastrointestinal illness, sexual problems, and problems with weight or appearance, had little independent effect using either rating method. A few problems had small positive weights using TTO in Beaver Dam, however the confidence interval for each of these items includes zero, and all are negative predictors of health ratings in the bivariate case. This highlights the fact that our weights for symptoms and problems are to be considered as a group and not individually.

¹² Analytically, this occurs because of the non-linear process of forming weights from the ordered probit regression coefficients.

The baseline decrement among those with none of our symptoms and impairments (the penultimate line of Table 5) ranges across models from -0.03 to -0.12 , indicating that even the lengthy set of symptoms and impairments we have fails to fully capture the problems that affect people's health ratings. The bottom end of the scale ranged from -0.74 to -0.95 using utility-based measures.

Figure 4 compares mean scores on the summary measures derived from the different health rating methods among those with problems in each domain. Those with cognitive problems had the worst summary scores in the MEPS 2002 sample, followed by those with social and physical limitations. Those with pain or vitality problems had somewhat higher mean summary health scores, similar to those with mental health symptoms or sensory impairments. Given the relatively less severe decrement weights associated with mental health symptoms and sensory problems, however, this similarity in scores likely reflects comorbidities among those with pain or vitality problems. In the Beaver Dam sample, those with social activity limitations had the lowest mean scores, and scores were less severe among those with pain. Cognitive impairment did not affect scores as much as in MEPS, where a more severe definition of impairment was used and proxy respondents were used for the cognitively impaired. The SRHS-based measure was more sensitive to the impairments and symptoms we measured than the RS and EQ5D-based measures.

IV. Discussion

This paper lays out a method for estimating the health-related decrements in quality of life associated with different symptoms and impairments, and for estimating a summary measure of health based on this information. Our method was developed to address three issues of

concern in the measurement of general health: the limited range of impairments and symptoms typically employed in summary health measures, the use of counterfactual scenarios to obtain condition weights, and inconsistent accounting for co-relationships between health problems.

Economists are naturally attracted to utility-based measures of health, and the time tradeoff is commonly used. Time tradeoff questions are sometimes asked directly, and have been used as a basis for scoring of other measures such as the EQ5D. Despite this theoretical benefit, we encounter several difficulties with this measure. Most notably, there is substantial clustering of responses at perfect health. In Beaver Dam, which included a time tradeoff rating of current health, 59 percent of people were unwilling to trade off any length of life for higher quality of life. On the EQ5D in MEPS 2002, 48 percent of people had scores equal to perfect health.

But health is not perfect for those people. Only one fifth of those in Beaver Dam expressing no willingness to trade off length for quality of life report their health to be excellent; 31% report only 'good' health and a few (4%) even report fair or poor health. Similarly, of those reporting no impairments on the EQ5D, 20 percent report only 'good' health and 3 percent report themselves in fair or poor health. In the case of the EQ5D, this likely reflects the fact that there are only 5 domains in the survey, resulting in insufficient coverage of symptoms and impairments. The most common problems among those who had perfect EQ5D scores but did not have the top score on our SRSR measure in MEPS 2002 were vitality loss (45%), sensory problems (15-16%), and depressive/anxiety symptoms (13-15%). In the case of the TTO we suspect that the reason for the clustering at perfect health is that people are unwilling to trade off years for quality of life when the question is posed directly. This is understandable, but it does not mean that quality of life is unimportant to people. People are very willing to grade their health less than perfect when asked about it directly using a SRHS question.

Utility-based measures that incorporate risk and uncertainty via standard gambles or time tradeoffs are not designed to yield decrements for relatively minor health complaints. However, particularly in a healthy population, it is important to try to consider the full range of problems that can affect people—and that can potentially be alleviated through medical care. SRHS and rating scale methods are inherently more disposed to reflect the effects of such problems.

Time tradeoff and similar questions involving counterfactuals also suffer from possible biases. People are asked to compare their current health state to a health state that they are not in – perhaps a health state they have never been in.¹³ While descriptions of that health state can be provided, other evidence suggests that experience matters: people with a health problem very often think their health is better than people without that problem guess it would be. (Sackett and Torrance 1978; Epstein, Hall et al. 1989; Slevin, Stubbs et al. 1990; Ubel, Loewenstein et al. 2003) Concerns about such biases underlie Dolan and Kahneman’s argument that economists should develop measures focused more directly on experienced utility rather than decision utility (Dolan and Kahneman working paper).

For these reasons, our recommendation is for summary measures of health that ask people only about their current health state, without the counterfactual of perfect health or the use of weights based on ratings of counterfactual health states. There are two such measures that we consider: a rating scale from 0 to 100 (worst to best possible health), and a 5-point self-assessment of general health. The rating scale is naturally on a 0 to 1 scale (when simply divided by 100). We present a method for transforming the self-assessed general health question onto that scale. Empirically, these two scales yield relatively similar measures of health. The mean rating scale score is 0.81, compared to 0.77 using the self-assessed method; the correlation

¹³ In the case of a TTO rating of current overall health, the comparison state is perfect health.

between health measured the two ways is 0.70. The domain-specific health impairments tend to be larger with SRHS than with the RS. The SRHS-based measure was more sensitive to the impairments and symptoms in MEPS than the RS-based measure, the latter being slightly more affected by other unknown factors.

The tradeoff in not asking about counterfactuals is that it must be assumed that different people use the same response scale. If some people consider ‘very good’ health to be different from others, and those groups differ in symptom propensity, our measure will be biased. This could be tested using vignette questions that ask people to rate the health of others, although such questions are not on our surveys (and are subject to the same counterfactual issues noted above).

As important as the summary measure of health is the attribution of health problems to particular symptoms and impairments. There is frequently a desire to examine how health would change if a particular impairment were eliminated, or if medical care improved to control the symptoms of a particular disease. Many health metrics currently employed are very restrictive. They ask about only a few domains (e.g., 5 domains in the EQ5D), they assume particular relationships between different symptoms within a domain (e.g., the worst symptom in QWB), and they add up across domains in ways that may not be optimal (e.g., use of factor analysis in SF36).

We develop a methodology to surmount these difficulties, relating the summary measure of health flexibly to symptom responses and their interactions. Our results show the importance of allowing for flexible functional forms. Having a combination of health problems is generally not as bad as having each one individually, but is worse than having only one of the problems. We also demonstrate the robustness of our findings in two independent, nationally representative

samples. The similarity of weights across samples is encouraging and supports the stability of our model.

Considering the various symptoms and impairments, the ones with the greatest effect on health in MEPS 2002 are severe pain (-0.20), having little or no energy (-0.16), inability to perform major role activity (-0.12), and limitations in activities of daily living (-0.11). Additional impairments that we found to have important impacts on health were physical activity limitations (up to -0.09 in BD), respiratory symptoms (-0.05 in BD), moderate pain (-0.11 in MEPS 2002), and anxiety (-0.08 in BD). Our findings are consistent with scattered evidence found in previous studies of factors predicting SRHS, such as being on work-disability pay and perception of one's own physical performance (Fylkesnes and Forde 1991; Fylkesnes and Forde 1992; Leinonen, Heikkinen et al. 1999), persistent pain disorder (Gureje, Simon et al. 2001), and cognitive capacity (Leinonen, Heikkinen et al. 1999). The great importance of these factors suggests the possibility that targeting interventions to alleviate some particular impairments in a few domains (or modifying environments so that they are less problematic) could have a significant effect on health across a wide range of the population.

The decrement weights for some of our impairments may appear small relative to the impact that one would expect the problem to have on HRQOL. However, it is important to keep in mind that the impact of a particular condition such as depression will occur through many impairments and symptoms. We include a broad array of interrelated symptoms and impairments, which reduces the direct effect of each one individually, but provides a more comprehensive picture of the overall impact of different problems on HRQOL.

Even our rich set of symptoms and impairments is likely not enough. Only 44% of people with no symptoms or impairments reported excellent health, and only 15% reported 100 on the

rating scale. Beaver Dam asked about more symptoms and impairments, and some of these do matter.

Of the additional items in Beaver Dam that were not available in MEPS,¹⁴ difficulty with driving/public transport had the largest decrement at -0.09. Overall, the inclusion of these additional items did not dramatically lower mean summary scores; an alternate model in Beaver Dam using only those items available in MEPS yielded a mean score that was only slightly higher (0.03). However, a smaller portion (12%) had the highest possible score on the SRHS-based measure in Beaver Dam. In comparison, 28% of those age 45+ in MEPS 2002 had the highest possible score, and 28% had the highest possible score in Beaver Dam when only using those items available in MEPS. This indicates that the consideration of additional symptoms and impairments is important.

We do not know, of course, the exact additional symptoms and impairments that matter to people but were not asked of those in our samples. Possibilities include symptoms caused by particular diseases or their treatments (such as shortness of breath in emphysema, coughing blood/hair loss from chemotherapy, or broken bones in osteoporosis). In addition, mild anxiety and the blues are important symptoms to ask about. A slightly lower portion of the MEPS 2002 sample had perfect EQ5D scores (49%) than had the top score on our SRHS-based summary measure (55%), largely due to a subset of respondents who reported anxiety/depression on the EQ5D but were not captured by the cutoffs on the anxiety and depressive SF-12 questions. This is partly due to a difference in the time window specified by the different questions: the SF-12

¹⁴ Limping or using a cane/crutches/walker (prevalence = 0.13), not driving or requires help with public transport (0.10), rash (0.07), problems with urinary/bowel/sex organs (0.08), trouble sleeping (0.14), speech problem (0.05), gastrointestinal illness (0.06), respiratory illness (0.19), problems with sexual interest/performance (0.07), weight/appearance problem (0.48), missing limbs/extremities (0.04), and headache (0.20)

asks about symptoms in the past 4 weeks, whereas the EQ5D asks only about ‘today’ (the day the question is answered).

Although we focus in this paper on health decrements for specific symptoms and impairments, our results can be used to form disease-specific utilities. The strategy is to relate the symptoms and impairments to the presence of specific diseases reported by the individual. This will provide insight into the mechanisms by which specific diseases affect health and how this might change over time.¹⁵

Limitations

Our study has some limitations that are worth noting. We discussed above the issues associated with possible response differences across individuals. In addition, there are some data issues that are worth noting. While the MEPS data are nationally representative, the Beaver Dam data are not, and its relatively small sample size may underlie the few small positive main effects and the anomalous large positive interaction terms for some combinations of problems. The possibility of omitted symptoms and impairments means that we cannot fully predict or explain changes in health related quality of life due to these omitted variables. To overcome this limitation, our national surveys may benefit from the addition of questions on a more complete set of symptoms and impairments that affect quality of life.

¹⁵ Previously, chronic condition-specific HRQOL has been estimated by calculating the mean or median score on a preference-based instrument among people with the health condition (i.e. Gold et al., 2002) or by regressing scores from a preference-based instrument on chronic conditions (Cutler and Richardson, 1997; Sullivan et al, 2005). However, this approach does not illustrate the relative contributions of specific symptoms and impairments to the overall HRQOL decrement for a specific chronic condition. Nor does it allow researchers to be able to tell how alleviation of one symptom of a disease would affect health.

Conclusion

Despite the large expenditure on medical care in the U.S. as a portion of our GDP, there has not been a systematic effort to measure the overall impact of national health care spending on U.S. population health. While existing instruments are used to monitor population health in many countries, no single measure has been adopted for the U.S.¹⁶ Our goal in this project is to lay the groundwork for a national health-related quality of life scale, as a main output in a set of National Health Accounts (Abraham and Mackie 2005).

Various national surveys ask about symptoms, impairments, and/or chronic diseases on an ongoing basis. Our method or a variation of it can be used to estimate a detailed summary measure of population health based on reported symptoms and impairments, or on chronic conditions. We develop in this paper a set of weights for 2002, with comparison to 2000 and an earlier community sample. While these weights did not significantly differ between 2000 and 2002, they could well change over longer time periods due to factors that affect what it is like to live with specific symptoms. Thus, we would recommend that these weights be re-estimated in a nationally representative sample every few years, perhaps once a decade. The use of vignettes to identify and adjust for differences in the way people value health would also be valuable. Finally, combining quality of life with projected length of life among those with different conditions would enable a more comprehensive measure of population health.

¹⁶ Measures used by agencies of the U.S. government include a 14-item measure (http://www.cdc.gov/hrqol/hrqol14_measure.htm#1), and a 'Health and Activity Limitation Index' also called 'Years of Healthy Life' (HALex/YHL) (Erickson, 1998) based on age, ADLs and 5-point SRHS.

Acknowledgements

This work was supported by National Institute on Aging (NIA) research grants: P01 AG17625, P30 AG12810 and R01 AG015911. We appreciate the support of the Harvard Interfaculty Program for Health Systems Improvement and the Lasker foundation. We thank Dennis Fryback for his generous provision of the Beaver Dam data, including the additional data entry required for our project, performed by Katie A. Gonring. We are also grateful to Dennis Fryback, Emmett Keeler, Willard Manning and Milton Weinstein for their comments on earlier versions of the manuscript and the methodology employed.

Table 1: Domains Covered by Existing General Health Instruments

Domain*	SF-36 (Ware, Kosinski et al. 2000)	HUI II/III (Horsman, Furlong et al. 2003)	EQ-5D (Brooks, Rabin et al. 2003)	15-D (Sintonen 2001)	AQoL (Hawthorne, Richardson et al. 1999)	QWB (Kaplan, Ganiats et al. 1998)
Social function	√		√	√	√	√
Physical Function	√	√	√	√		√
Pain	√	√		√	√	√
Mental Health	√	√	√	√	√	√
Vitality	√			√		√
Sensory		√		√	√	√
Cognition		√		√		√
Health Regimen					◇	√
Sexual function		√		√		√
Illness					◇	√
Appearance						√
General health	√		◇			◇

*This table was constructed based on the domains that appeared to be covered by each instrument, and does not necessarily reflect the domains as described by instrument developers. For example, the Quality of Well-being scale technically has a symptom list and three domains, however there are additional areas covered by the symptom list that may be seen as domains, such as mental health.

◇part of questionnaire but not included in final scoring.

Table 2: Summary Measures of Overall Health

MEPS 2002			
	EQ5D	Rating Scale	SRHS*
Mean	0.87	0.81	0.79
Correlation coefficient			
EQ5D	1.00		
Rating scale	0.66	1.00	
SRHS*	0.82	0.70	1.00
BDHOS			
	TTO	SRHS*	
Mean	0.86	0.75	
Correlation coefficient			
TTO	1.00		
SRHS*	0.41	1.00	

* Predicted health based on model.

Table 3: Distribution of Self-Rated Health Status Responses by Level of Response on Utility-Based Measures

EQ5D MEPS 2002 (N= 23790)	SRHS				Proportion
	Poor/Fair	Good	Very Good	Excellent	
< 75 (n= 3095)	0.55	0.29	0.12	0.04	1.0
75-99 (n = 9597)	0.13	0.35	0.36	0.15	1.0
100 (n = 11098)	0.03	0.20	0.41	0.37	1.0
Rating Scale MEPS 2002 (N= 22763)					
< 75 (n= 5950)	0.42	0.38	0.16	0.04	1.0
75-99 (n = 14857)	0.04	0.26	0.44	0.26	1.0
100 (n = 1956)	0.02	0.10	0.24	0.63	1.0
TTO Beaver Dam (N=1388)					
< 75 (n= 251)	0.30	0.44	0.22	0.04	1.0
25-99 (n = 318)	0.14	0.47	0.32	0.08	1.0
100 (n = 819)	0.04	0.31	0.45	0.20	1.0

Table 4: Decrement Weights for Symptoms and Impairments Derived From Regressions of Self-Rated General Health on Symptoms and Impairments in MEPS and Beaver Dam

Domains and symptoms/impairments	MEPS 2002	MEPS 2000	Beaver Dam
Social Activity^a			
Limited or performed no major role activity			-0.12
Limited in major role activity	-0.08	-0.08	
Completely unable to perform major role activity	-0.12	-0.15	
Limited in other role activity	-0.06	-0.10	-0.05
Physical Activity^b			
Self-care limitations	-0.10	-0.11	-0.05
Bed/chair bound			-0.06
Lifting/bending difficulty	-0.02	-0.04	-0.08
Limp/cane/crutches/walker			-0.03
Limited walking or other physical limitation			-0.05
Moderate walking limitation	-0.03	-0.04	
Severe walking limitation	-0.09	-0.07	
Do not drive / limited in public transport			-0.09
Pain^c			
Moderate pain	-0.10	-0.10	-0.06
Severe pain	-0.19	-0.21	
Rash (burning or itching rash on large area)			-0.04
Urinary/bowel/sex organs			-0.03
Mental Health^d			
Depressive symptoms	-0.07	-0.04	-0.02
Anxiety symptoms	-0.04	-0.04	-0.08
Trouble sleeping			-0.04
Vitality^e			
General tiredness/weakness/weight loss			-0.05
Have a lot of energy only some of the time	-0.09	-0.08	
Have a lot of energy a little or none of the time	-0.15	-0.13	
Senses^f			
Vision impairment	-0.06	-0.04	-0.03
Hearing impairment	-0.04	-0.05	-0.02
Speech problem			-0.04
Cognitive impairment^g			
	-0.08	-0.06	-0.06

Continued

Miscellaneous^h

Gastrointestinal illness			-0.01
Respiratory illness			-0.06
Problems with sexual interest / performance			-0.00
Weight/appearance			-0.02
Limbs/extremities			-0.03
Headache			-0.04
Baseline decrement with no impairments or symptoms	-	-	-
	0.08**	0.08**	0.08**
<u>Sum of weights for all impairments and symptoms*</u>	<u>-1.03</u>	<u>-1.05</u>	<u>-1.15</u>

*Includes baseline decrement. For impairments and symptoms with weights for two levels of severity, this sum includes only the weight for the worst level

**Average of decrements for RS in two MEPS waves and of TTO

Table 5: Decrement Weights for Symptoms and Impairments Derived From Regressions of Preference-Based General Health Ratings on Symptoms and Impairments in MEPS and Beaver Dam

Domains and symptoms/impairments	MEPS 2002		MEPS 2000		Beaver Dam
	EQ5D	RS	EQ5D	RS	TTO
Social Activity^a					
Limited or performed no major role activity					-0.14
Limited in major role activity	-0.04	-0.04	-0.06	-0.06	
Completely unable to perform major role activity	-0.03	-0.07	-0.10	-0.11	
Limited in other role activity	-0.04	-0.03	-0.02	-0.07	-0.01
Physical Activity^b					
Self-care limitations		-0.13		-0.07	0.01
Bed/chair bound					-0.09
Lifting/bending difficulty	-0.07	-0.00	-0.08	-0.03	-0.02
Limp/cane/crutches/walker					-0.02
Limited walking or other physical limitation					-0.08
Moderate walking limitation	-0.02	0.00	-0.03	-0.01	
Severe walking limitation	-0.05	-0.03	-0.07	-0.04	
Do not drive / limited in public transport					-0.08
Pain^c					
Moderate pain		-0.06		-0.06	
Severe pain		-0.17		-0.19	
Rash (burning or itching rash on large area)					-0.03
Urinary/bowel/sex organs					-0.04
Mental Health^d					
Depressive symptoms	-0.07	-0.05	-0.06	-0.03	-0.02
Anxiety symptoms	-0.06	-0.03	-0.07	-0.04	-0.04
Trouble sleeping					0.01
Vitality^e					
General tiredness/weakness/weight loss					-0.01
Have a lot of energy only some of the time	-0.07	-0.07	-0.06	-0.05	
Have a lot of energy a little or none of the time	-0.13	-0.13	-0.13	-0.12	

Continued

Senses^f

Vision impairment	-0.03	-0.03	-0.04	-0.02	-0.05
Hearing impairment	-0.03	-0.02	-0.04	-0.03	-0.01
Speech problem					-0.02

Cognitive impairment^g

Miscellaneous^h

Gastrointestinal illness					0.01
Respiratory illness					-0.04
Problems with sexual interest / performance					0.01
Weight/appearance					0.00
Limbs/extremities					-0.00
Headache					-0.01
Baseline decrement with no impairments or symptoms	-0.06	-0.10	-0.07	-0.12	-0.03
<u>Sum of weights for all impairments and symptoms*</u>	<u>-0.63</u>	<u>-0.83</u>	<u>-0.74</u>	<u>-0.95</u>	<u>-0.76</u>

*Includes baseline decrement. For impairments and symptoms with weights for two levels of severity, this sum includes only the weight for the worst level.

Figure 1: Percent distribution of Original Summary Health Measures

Figure 1a: SRHS, all three surveys

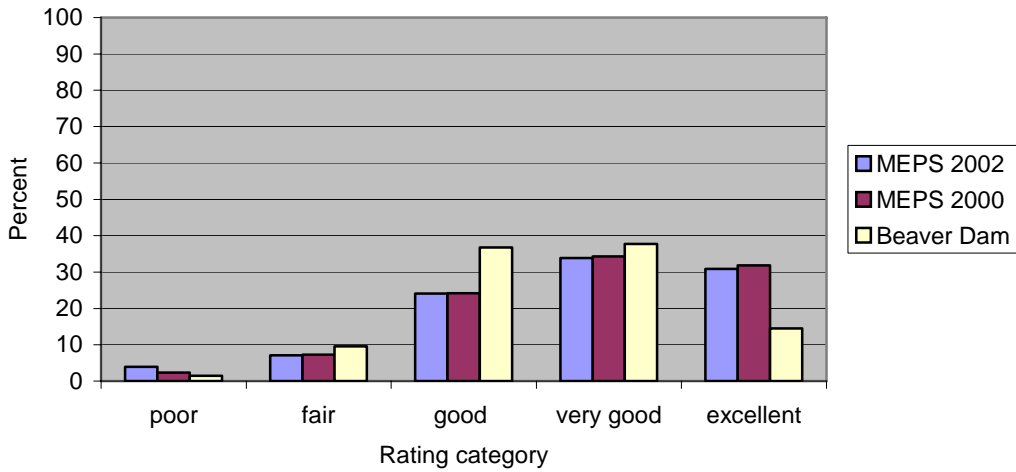


Figure 1b: EQ5D, MEPS 2002 (N = 11098)

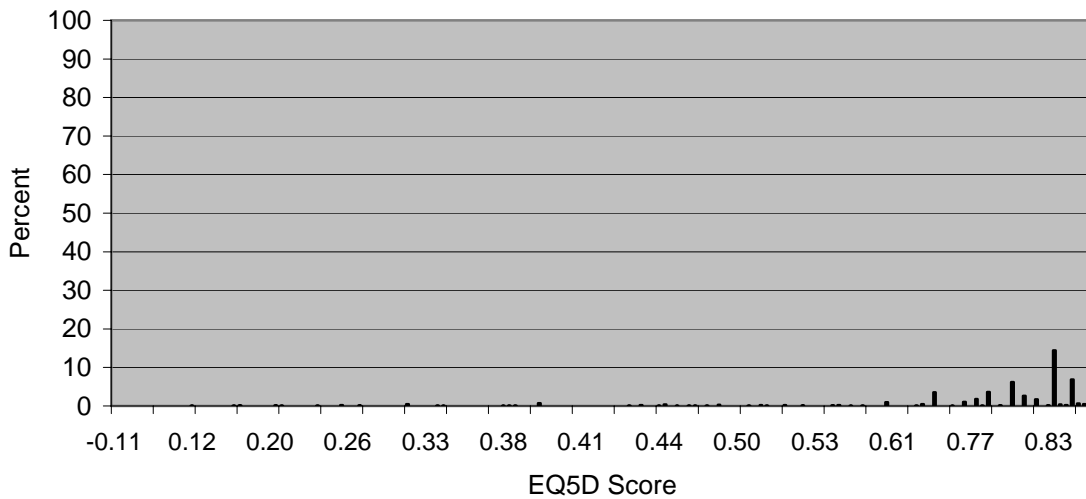


Figure 1c: Rating Scale, MEPS 2002 (N=22763)

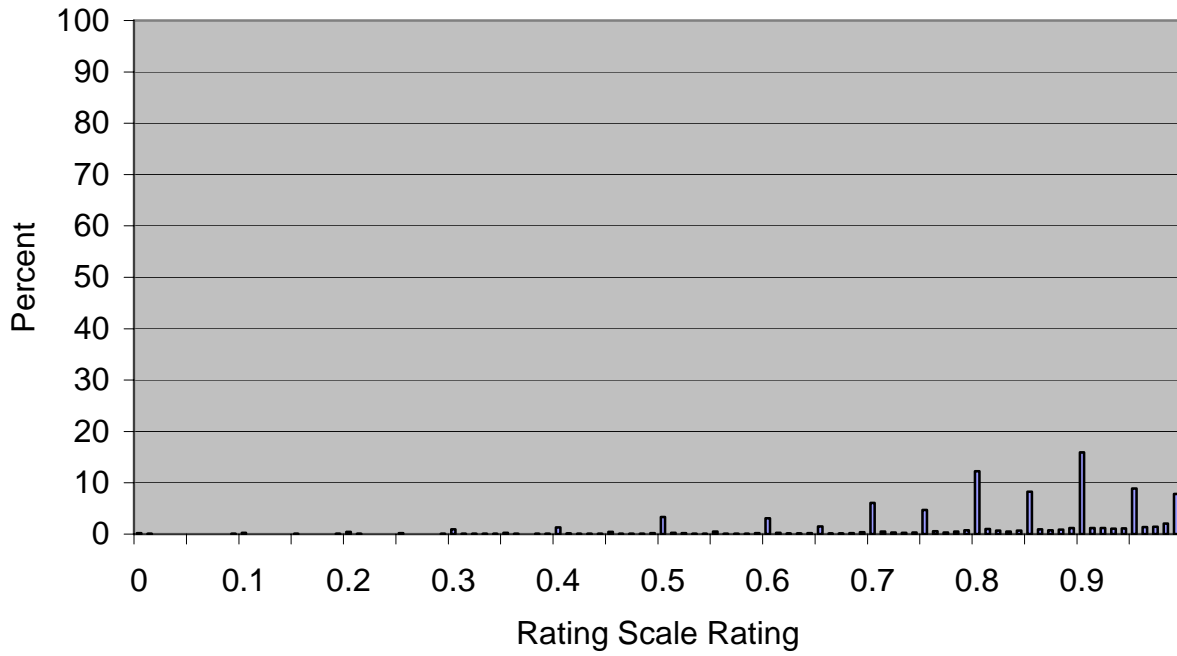


Figure 1d: Time-Tradeoff Ratings of current health Beaver Dam (N = 1380)

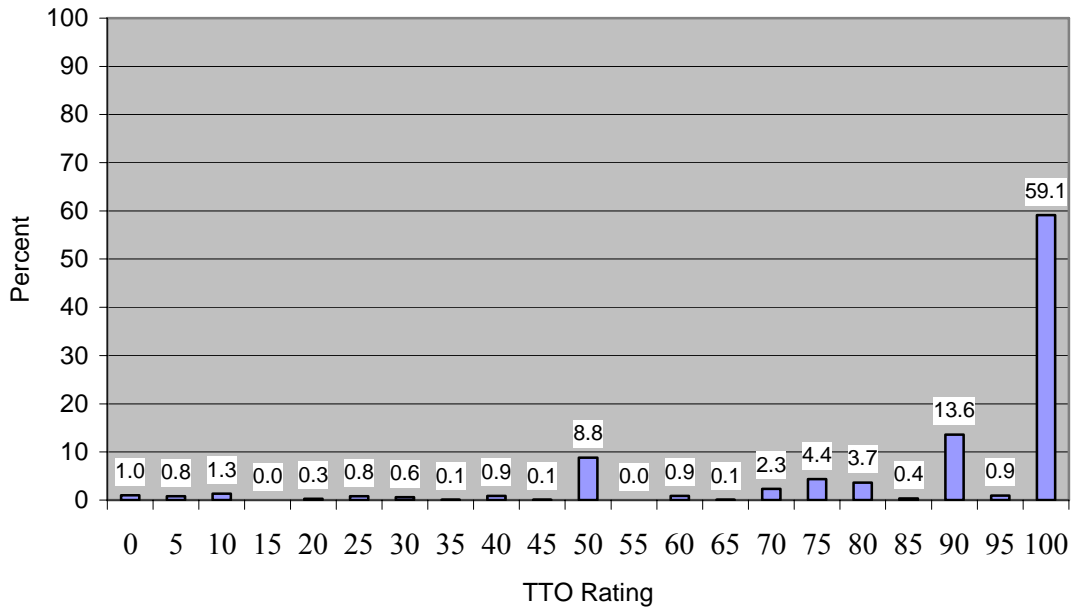


Figure 2: Distribution of Predicted SHRS-based Score

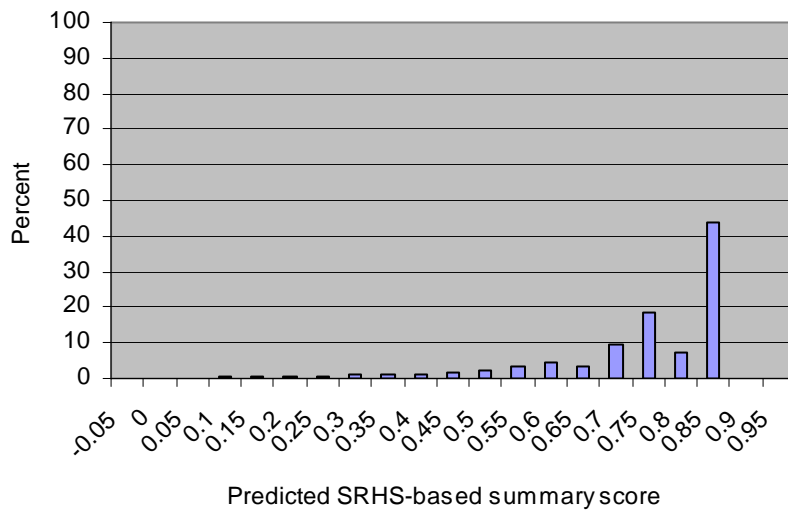
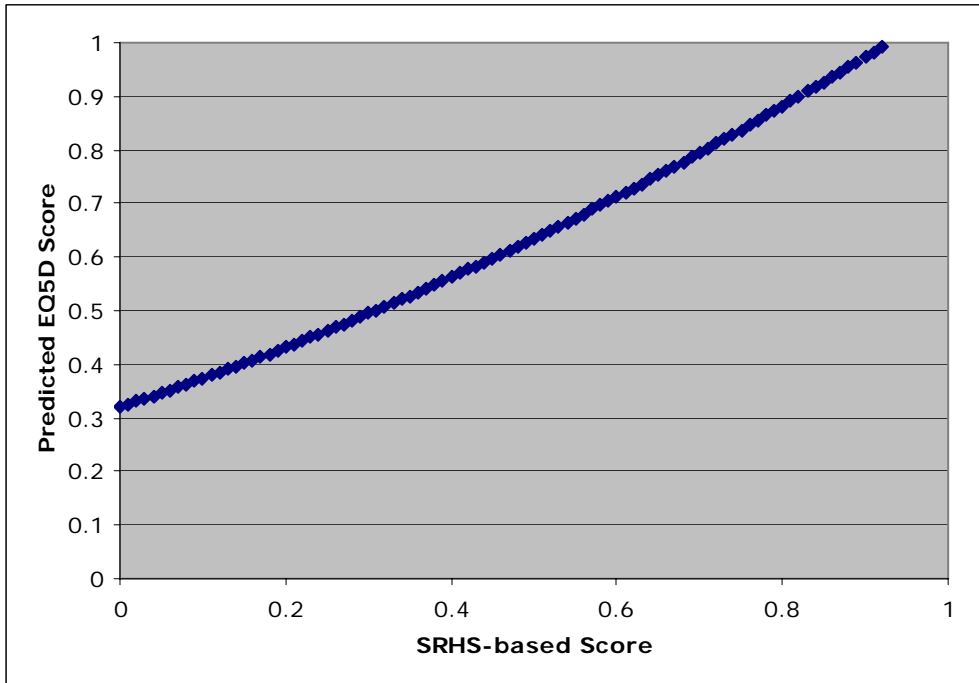


Figure 3: Predicted Utility-Based Scores Corresponding to Hypothetical Values of Predicted Self-Rated Health-based Scores

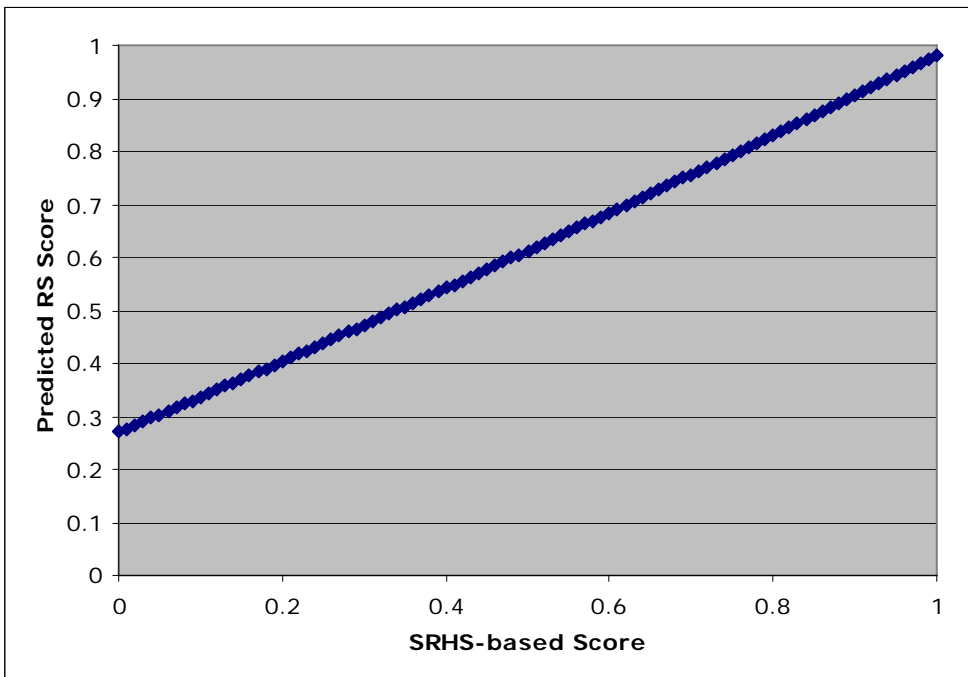
Figure 3a: Squared SRH score predicting EQ5D score, MEPS 2002



$$\text{EQ5D} = 0.32 + 0.51 * \text{SRHS} + 0.24 * \text{SRHS}^2$$

(0.01) (0.03) (0.02)

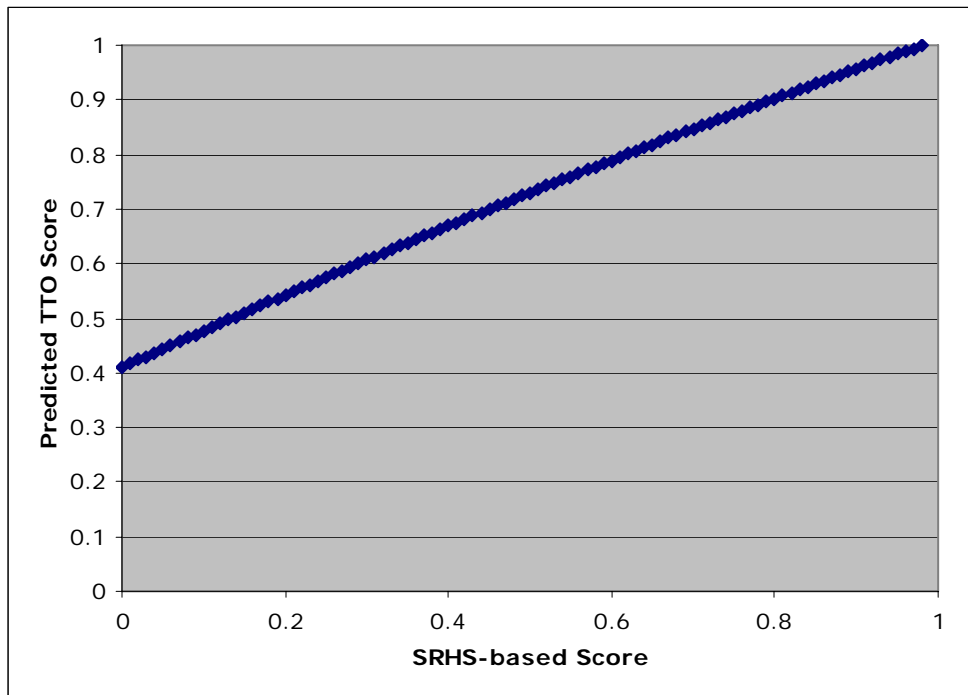
Figure 3b: Squared SRH score predicting RS rating, MEPS 2002



$$\text{RS} = 0.27 + 0.66 * \text{SRHS} + 0.05 * \text{SRHS}^2$$

(0.01) (0.05) (0.03)

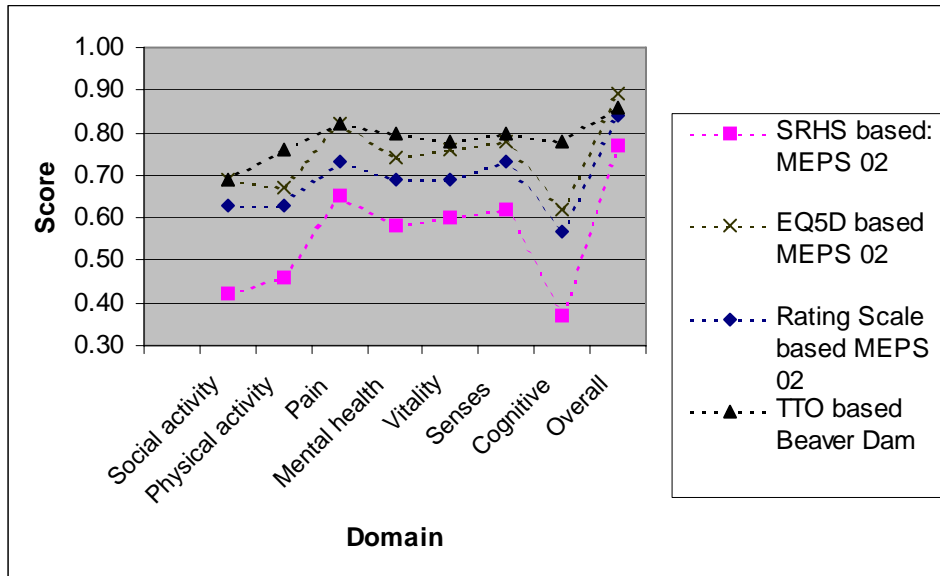
Figure 3c: Squared SRH score predicting TTO rating, Beaver Dam 1991/92



$$\text{TTO} = 0.41 + 0.68 * \text{SRHS} - 0.08 * \text{SRHS}^2$$

(0.06) (0.19) (0.15)

Figure 4: Predicted Summary Scores by Domain



Note: mean TTO scores are not shown for those with miscellaneous impairments that did not fall into above domains and were not available in MEPS. This explains why the TTO does not yield the highest overall mean score.

References

- Abraham, K. G. and C. Mackie, Eds. (2005). Beyond the Market: Designing Nonmarket Accounts for the United States. Washington, DC, National Academies Press.
- Andresen, E. M., B. M. Rothenberg, et al. (1998). "Performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire among older adults." Med Care **36**(9): 1349-60.
- Brazier, J., J. Roberts, et al. (2002). "The estimation of a preference-based measure of health from the SF-36." J Health Econ **21**(2): 271-92.
- Brooks, R., R. E. Rabin, et al., Eds. (2003). The measurement and valuation of health status using EQ-5D: a European perspective. Dordrecht, Kluwer Academic Publishers.
- Cutler, D. and E. Richardson (1997). "Measuring the health of the U.S. population." Bookings Papers on Economic Activity, Microeconomics: 217-272.
- DeSalvo, K. B., N. Bloser, et al. (2006). "Mortality prediction with a single general self-rated health question. A meta-analysis." J Gen Intern Med **21**(3): 267-75.
- Dolan, P. (1997). "Modeling valuations for EuroQol health states." Med Care **35**(11): 1095-108.
- Dolan, P. and D. Kahneman (working paper). "Interpretations of utility and their implications for the valuation of health."
- Epstein, A. M., J. A. Hall, et al. (1989). "Using proxies to evaluate quality of life. Can they provide valid information about patients' health status and satisfaction with medical care?" Med Care **27**(3 Suppl): S91-8.
- Essink-Bot, M. L., P. F. Krabbe, et al. (1997). "An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument." Med Care **35**(5): 522-37.
- Fryback, D. G., E. J. Dasbach, et al. (1993). "The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors." Med Decis Making **13**(2): 89-102.
- Fylkesnes, K. and O. H. Forde (1991). "The Tromso Study: predictors of self-evaluated health--has society adopted the expanded health concept?" Soc Sci Med **32**(2): 141-6.
- Fylkesnes, K. and O. H. Forde (1992). "Determinants and dimensions involved in self-evaluation of health." Soc Sci Med **35**(3): 271-9.
- Gilbert, D. T., E. C. Pinel, et al. (1998). "Immune neglect: a source of durability bias in affective forecasting." J Pers Soc Psychol **75**(3): 617-38.
- Gold, M., J. Siegel, et al. (1996). Cost-Effectiveness in Health and Medicine. New York, Oxford University Press.
- Gureje, O., G. E. Simon, et al. (2001). "A cross-national study of the course of persistent pain in primary care." Pain **92**(1-2): 195-200.

- Hawthorne, G., J. Richardson, et al. (1999). "The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life." Qual Life Res **8**(3): 209-24.
- Horsman, J., W. Furlong, et al. (2003). "The Health Utilities Index (HUI(R)): concepts, measurement properties and applications." Health Qual Life Outcomes **1**(1): 54.
- Idler, E. L. and Y. Benyamini (1997). "Self-rated health and mortality: a review of twenty-seven community studies." J Health Soc Behav **38**(1): 21-37.
- Idler, E. L. and S. V. Kasl (1995). "Self-ratings of health: do they also predict change in functional ability?" J Gerontol B Psychol Sci Soc Sci **50**(6): S344-53.
- Idler, E. L., L. B. Russell, et al. (2000). "Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study, 1992. First National Health and Nutrition Examination Survey." Am J Epidemiol **152**(9): 874-83.
- Jenkinson, C., A. Gray, et al. (1997). "Evaluation of index and profile measures of health status in a randomized controlled trial. Comparison of the Medical Outcomes Study 36-Item Short Form Health Survey, EuroQol, and disease specific measures." Med Care **35**(11): 1109-18.
- Johnson, J. A. and A. S. Pickard (2000). "Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada." Med Care **38**(1): 115-21.
- Kahneman, D. and A. Tversky, Eds. (2000). Choices, Values and Frames, Cambridge University Press.
- Kaplan, R. and J. Anderson (1996). The general health policy model: an integrated approach. Quality of life and pharmacoeconomics in clinical trials. S. B. Philadelphia, Lippincott-Raven: 309-322.
- Kaplan, R. M., J. W. Bush, et al. (1976). "Health status: types of validity and the index of well-being." Health Serv Res **11**(4): 478-507.
- Kaplan, R. M., T. G. Ganiats, et al. (1998). "The Quality of Well-Being Scale: critical similarities and differences with SF-36." Int J Qual Health Care **10**(6): 509-20.
- Leinonen, R., E. Heikkinen, et al. (1999). "A path analysis model of self-rated health among older people." Aging (Milano) **11**(4): 209-20.
- McDowell, I. and C. Newell (1996). Measuring health : a guide to rating scales and questionnaires. New York, Oxford University Press.
- Sackett, D. L. and G. W. Torrance (1978). "The utility of different health states as perceived by the general public." J Chronic Dis **31**(11): 697-704.
- Salomon, J. A., A. Tandon, et al. (2004). "Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes." Bmj **328**(7434): 258.
- Shaw, J., J. Johnson, et al. (2005). "US valuation of the EQ-5D health states: development and testing of the D1 model." Medical Care **43**(3): 203-220.
- Sintonen, H. (2001). "The 15D instrument of health-related quality of life: properties and applications." Ann Med **33**(5): 328-36.

- Slevin, M. L., L. Stubbs, et al. (1990). "Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses, and general public." Bmj **300**(6737): 1458-60.
- Torrance, G. W., D. Feeny, et al. (2001). "Visual analog scales: do they have a role in the measurement of preferences for health states?" Med Decis Making **21**(4): 329-34.
- Torrance, G. W., D. H. Feeny, et al. (1996). "Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2." Med Care **34**(7): 702-22.
- Ubel, P. A., G. Loewenstein, et al. (2003). "Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public." Qual Life Res **12**(6): 599-607.
- Ware, J. J., M. Kosinski, et al. (2000). Sf-36 Health survey: Manual & Interpretation Guide. Lincoln, RI, QualityMetric Incorporated.
- Wilcox, V. L., S. V. Kasl, et al. (1996). "Self-rated health and physical disability in elderly survivors of a major medical event." J Gerontol B Psychol Sci Soc Sci **51**(2): S96-104.
- Wilson, T. D., T. Wheatley, et al. (2000). "Focalism: a source of durability bias in affective forecasting." J Pers Soc Psychol **78**(5): 821-36.
- Woloshin, S., L. M. Schwartz, et al. (2001). "Assessing values for health: Numeracy matters." Med Decis Making **21**(5): 382.

Appendix 1: Regression results and proportion reporting symptoms and impairments

Table 1: MEPS 2002: Proportion Reporting Items in Each Domain and Coefficients from Probit Regressions on Self-Rated Health Status and OLS Regressions on EQ5D Scores and Rating Scale Ratings

Domains and symptoms/impairments	Proportion reporting*	Probit coefficient for SRHS	P value	OLS coefficient for EQ5D	P value	OLS coefficient for Rating Scale	P value
Social Activity	0.08						
Limited major role activity	0.04	-0.41	<.001	-0.04	<.001	-0.04	0.00
Unable to perform major role activity	0.04	-0.61	<.001	-0.08	<.001	-0.06	<.001
Limited in other role activity	0.05	-0.40	<.001	-0.04	0.00	-0.03	0.02
Physical Activity	0.09						
Self-care limitations	0.01	-0.49	0.00				
Bending/Lifting/Standing/Reaching/Fingers	0.08	-0.20	0.05	-0.07	<.001	-0.12	<.001
Moderate walking limitation	0.07	-0.12	0.10	-0.02	0.05	0.00	0.68
Severe walking limitation	0.06	-0.40	<.001	-0.04	<.001	-0.03	0.00
Pain	0.28						
Moderate pain	0.31	-0.62	<.001			-0.06	<.001
Severe pain	0.03	-1.22	<.001			-0.18	<.001
Mental Health ^d	0.12						
Depressive symptoms	0.08	-0.42	<.001	-0.07	<.001	-0.05	<.001
Anxiety symptoms	0.08	-0.26	<.001	-0.06	<.001	-0.03	<.001
Vitality	0.22						
Have a lot of energy a little/none of time	0.14	-0.60	<.001	-0.08	<.001	-0.08	<.001
Have a lot of energy only some of time	0.08	-0.94	<.001	-0.14	<.001	-0.14	<.001
Senses	0.10						
Vision impairment	0.05	-0.35	<.001	-0.03	<.001	-0.03	<.001
Hearing impairment	0.06	-0.26	<.001	-0.03	<.001	-0.02	0.00
Cognitive impairment	0.03	-0.49	<.001	-0.05	<.001	-0.03	0.00
Interactions							
Primary & secondary activity limitations	0.04	0.12	0.17	0.02	0.14	0.00	0.87
Primary activity & self-care limitations	0.01	0.28	0.09			0.03	0.11
Primary activity & walking limitations	0.04	0.12	0.31	0.00	0.86	-0.01	0.56
Primary activity & bending/lifting limitations	0.05	0.24	0.04	0.00	0.73	0.04	0.00
Primary activity limitation & depressive symptoms	0.02	-0.19	0.05	0.00	0.85	0.04	0.00
Primary activity limitation & anxiety symptoms	0.02	0.02	0.85	-0.05	<.001	0.01	0.50
Primary activity limitation & vision impairment	0.01	-0.15	0.13	0.03	0.02	-0.02	0.11
Primary activity limitation & hearing impairment	0.01	0.09	0.35	0.02	0.05	0.00	0.98
Primary activity limitation & cognitive impairment	0.02	0.22	0.03	0.04	0.00	-0.01	0.44
Primary activity limitation & pain	0.05	-0.18	0.04			-0.03	0.01
Primary activity limitation & vitality loss	0.05	0.04	0.63	-0.02	0.01	0.00	0.75
Secondary activity & self-care limitations	0.01	-0.26	0.04			-0.03	0.06
Secondary activity & walking limitations	0.03	-0.16	0.27	-0.02	0.13	-0.02	0.30

Domains and symptoms/impairments	Proportion reporting*	Probit coefficient for SRHS	P value	OLS coefficient for EQ5D	P value	OLS coefficient for Rating Scale	P value
Secondary activity & bending/lifting limitations	0.03	0.04	0.77	0.01	0.49	0.01	0.49
Secondary activity limitation & depressive symptoms	0.01	0.15	0.15	-0.02	0.08	-0.02	0.17
Secondary activity limitation & anxiety symptoms	0.01	0.02	0.86	-0.01	0.43	-0.01	0.46
Secondary activity limitation & vision impairment	0.01	0.10	0.32	0.00	0.95	0.02	0.11
Secondary activity limitation & hearing impairment	0.01	-0.24	0.02	0.01	0.57	-0.02	0.13
Secondary activity limitation & cognitive impairment	0.01	0.13	0.19	-0.03	0.00	-0.01	0.62
Secondary activity limitation & pain	0.03	0.25	0.02			0.00	0.99
Secondary activity limitation & vitality loss	0.03	-0.02	0.83	0.02	0.16	0.03	0.04
Self-care & walking limitations	0.01	0.12	0.69			0.00	0.91
Self-care & bending/lifting limitations	0.01	0.20	0.52			0.11	0.00
Self-care limitations & depressive symptoms	0.005	-0.24	0.09			-0.04	0.02
Self-care limitations & anxiety symptoms	0.005	-0.07	0.62			0.01	0.41
Self-care limitations & vision impairment	0.004	0.34	0.01			0.02	0.33
Self-care limitations & hearing impairment	0.004	-0.22	0.10			-0.01	0.41
Self-care limitations & cognitive impairment	0.01	0.12	0.31			0.02	0.20
Self-care limitations & pain	0.01	-0.44	0.01			-0.01	0.69
Self-care limitations & pain	0.01	0.27	0.11			-0.05	0.01
Walking & bending/lifting limitations	0.06	-0.21	0.03	-0.01	0.25	-0.03	0.01
Walking limitations & depressive symptoms	0.02	0.30	0.05	0.00	0.85	0.01	0.48
Walking limitations & anxiety symptoms	0.02	-0.14	0.33	-0.04	0.02	-0.01	0.61
Walking limitations & vision impairment	0.01	-0.02	0.86	-0.01	0.42	-0.03	0.11
Walking limitations & hearing impairment	0.02	0.15	0.23	0.00	0.86	0.00	0.85
Walking limitations & cognitive impairment	0.02	-0.09	0.57	-0.01	0.58	0.05	0.01
Walking limitations & pain	0.06	0.13	0.21			0.00	0.69
Walking limitations & vitality loss	0.05	-0.07	0.47	0.00	0.98	0.01	0.35
Bending/lifting limitations & depressive symptoms	0.02	-0.05	0.74	0.01	0.77	-0.02	0.22
Bending/lifting limitations & anxiety symptoms	0.02	-0.12	0.38	0.02	0.30	0.00	0.88
Bending/lifting limitations & vision impairment	0.02	0.12	0.35	0.02	0.13	0.02	0.15
Bending/lifting limitations & hearing impairment	0.02	0.16	0.18	0.01	0.43	0.02	0.10
Bending/lifting limitations & cognitive impairment	0.02	0.08	0.61	0.01	0.65	-0.02	0.38
Bending/lifting limitations & Bendpain	0.07	0.05	0.58			0.00	0.90
Bending/lifting limitations & vitality loss	0.06	0.09	0.29	0.03	0.01	-0.02	0.10
Depressive & anxiety symptoms	0.03	0.00	0.99	-0.04	<.001	-0.02	0.02
Depressive symptoms & vision impairment	0.01	-0.09	0.31	0.02	0.08	0.01	0.17
Depressive symptoms & hearing impairment	0.01	0.00	0.96	0.01	0.55	0.02	0.06
Depressive symptoms & cognitive impairment	0.01	0.08	0.42	-0.02	0.06	-0.03	0.01
Depressive symptoms & pain	0.05	0.04	0.46			-0.02	0.00
Depressive symptoms & vitality loss	0.05	0.20	0.00	0.00	0.60	0.02	0.00
Anxiety symptoms & vision impairment	0.01	-0.05	0.54	-0.02	0.08	-0.04	0.00
Anxiety symptoms & hearing impairment	0.01	-0.04	0.68	0.01	0.36	0.00	0.87
Anxiety symptoms & cognitive impairment	0.01	0.04	0.70	0.02	0.14	-0.03	0.03
Anxiety symptoms & pain	0.05	-0.04	0.44			-0.02	0.02
Anxiety symptoms & vitality loss	0.05	0.31	<.001	0.03	<.001	0.02	0.01
Vision & hearing impairments	0.01	0.06	0.43	-0.01	0.44	0.01	0.21
Vision & cognitive impairments	0.01	0.09	0.34	-0.01	0.35	0.04	0.00
Vision impairment & pain	0.03	0.19	0.00			0.02	0.04
Vision impairment & vitality loss	0.03	-0.07	0.26	-0.01	0.08	-0.01	0.25
Hearing & cognitive impairments	0.01	0.25	0.01	0.02	0.08	0.01	0.22
Hearing impairment and pain	0.04	0.06	0.28			0.00	0.96

Domains and symptoms/impairments	Proportion reporting*	Probit coefficient for SRHS	P value	OLS coefficient for EQ5D	P value	OLS coefficient for Rating Scale	P value
Hearing impairment & vitality loss	0.03	-0.07	0.23	0.00	0.72	-0.01	0.04
Cognitive impairment & pain	0.03	0.20	0.06			0.00	0.89
Cognitive impairment & vitality loss	0.03	-0.15	0.13	-0.01	0.40	-0.01	0.55
Pain & vitality loss	0.15	0.07	0.05			-0.02	<.001
Break points:							
Cut 1		-0.12		--		--	
Cut 2		0.92		--		--	
Cut 3		2.10		--		--	
Intercept		3.23		0.94		.90	
Summary statistics							
N		34,752		23,497		22,335	
R ² / Ln (likelihood)		12847.80		0.474		0.515	

Table 2: Beaver Dam data: Proportion Reporting Items in Each Domain and Coefficients from OLS Regressions on Time-Tradeoff Ratings and Probit Regressions on Self-Rated Health Status

Domains and symptoms/impairments	Proportion reporting*	Probit coefficient for SRHS	P value	OLS coefficient for TTO	P value
Social Activity	0.17				
Limited or performed no major role activity	0.09	-0.59	<.0001	-0.15	<.0001
Limited in other role activity	0.11	-0.28	0.04	-0.03	0.27
Interaction term: limitations in major and other	0.04	0.18	0.44	0.17	0.00
Physical Activity	0.35				
1+ ADL limitations (dress/feed/toilet/bath)	0.03	-0.24	0.25	0.01	0.89
Most or all of day in chair, couch or bed	0.07	-0.16	0.50	-0.09	0.06
Trouble lifting/stooping/bending over/stairs	0.20	-0.42	<.0001	-0.01	0.71
Limp/cane/crutches/walker	0.13	-0.25	0.11	-0.04	0.23
Limited walking or other physical limitation	0.18	-0.32	0.01	-0.07	0.00
Do not drive and/or limited in public transport	0.10	-0.56	0.00	-0.09	0.01
Interaction terms:					
Chair or bed all day and lift/stoop/bend/stairs	0.05	-0.55	0.08	0.02	0.70
Chair or bed all day and limp/walking aid	0.04	0.61	0.04	0.05	0.44
Chair or bed all day and limited walking	0.05	-0.52	0.08	-0.08	0.18
Lift/stoop/bend/stairs and limp/walking aid	0.07	-0.13	0.55	0.01	0.91
Lift/stoop/bend/stairs and limited walking	0.10	0.45	0.02	-0.06	0.11
Limp/walking aid and limited walking	0.08	0.26	0.24	0.07	0.13
Drive/public transport and chair or bed all day	0.04	0.29	0.31	0.04	0.44
Drive/public transport and lift/stoop/bend	0.06	0.18	0.48	0.02	0.71
Drive/public transport and limp/walking aid	0.05	0.33	0.22	-0.01	0.88
Drive/public transport and limited walking	0.06	0.06	0.81	-0.00	0.98
Pain	0.54				
Pain ^a	0.49	-0.26	0.00	-0.02	0.19
Rash (burning or itching rash on large area)	0.07	0.01	0.95	-0.01	0.80
Urinary/bowel/sex organ	0.08	-0.08	0.69	0.01	0.86
Interaction terms:					
Pain and urinary/bowel/sex organ	0.05	-0.11	0.64	-0.09	0.08
Pain and rash	0.04	-0.43	0.08	-0.03	0.49
Mental Health	0.22				
Depressive symptoms ^b	0.06	-0.14	0.48	-0.02	0.64
Excessive worry or anxiety	0.10	-0.44	0.00	-0.05	0.05
Trouble sleeping	0.14	-0.20	0.05	-0.01	0.70
Interaction terms:					
Depressive symptoms and anxiety	0.03	0.54	0.05	0.01	0.83
Depressive symptoms and trouble sleeping	0.02	-0.00	1.00	-0.00	0.95
Anxiety and trouble sleeping	0.04	0.24	0.30	0.13	0.01
Vitality (general tiredness/weakness/weight loss)	0.19	-0.23	0.01	-0.01	0.70
Senses	0.96				
Vision problem ^c	0.07	-0.16	0.20	-0.05	0.04
Hearing problem ^d	0.25	-0.10	0.15	-0.01	0.52
Speech problem ^e	0.05	-0.17	0.19	-0.02	0.38

Domains and symptoms/impairments	Proportion reporting*	Probit coefficient for SRHS	P value	OLS coefficient for TTO	P value
Miscellaneous	0.95				
Gastrointestinal illness ^f	0.06	-0.25	0.09	0.01	0.81
Respiratory illness ^g	0.19	-0.36	<.0001	-0.04	0.02
Interaction: Gastrointestinal and respiratory	0.02	0.99	0.00	-0.02	0.77
Problems with sexual interest or performance	0.07	-0.00	0.97	0.00	0.85
Weight/appearance ^h	0.48	-0.11	0.07	0.00	0.77
Limbs/extremities ⁱ	0.04	-0.16	0.32	-0.00	0.94
Headache ^j	0.20	-0.19	0.01	-0.00	0.75
Trouble learning/remembering/thinking clearly	0.14	-0.29	0.00	-0.03	0.06
Break points:					
Cut 1		1.41	<.0001	--	--
Cut 2		1.51	<.0001	--	--
Cut 3		1.30	<.0001	--	--
Intercept		3.71	0.18	0.94	<.0001
Summary statistics					
N		1422		1380	
R ² / Ln (likelihood)		1593.9		0.213	

*includes anyone reporting one or more items in the domain; ^aPain, stiffness, weakness, numbness or other discomfort in chest, stomach, side, neck, back, hips, joints of hands/feet/arms or legs; ^bSpells of feeling upset, being depressed, or of crying; ^cScore below 80 on a 100 point vision scale; ^dScore below 80 on a 100 point hearing scale ^etrouble talking, such as lisp, stuttering, hoarseness, or inability to speak; ^fsick or upset stomach, vomiting or loose bowel movements, with or without fever, chills, or aching all over; ^gcough, wheezing, or shortness of breath with or without fever, chills, or aching all over. ^hScore below 90 on a 100-point hearing scale; ⁱoverweight or underweight for age and height; or skin defect of face, body, arms or legs, such as scars, pimples, warts, bruises, or changes in color; ^jany combination of one or more hands, feet, arms, or legs either missing, deformed (crooked), paralyzed (unable to move) or broken—includes wearing artificial limbs or braces; ^kHeadache, or dizziness, or ringing in ears, or spells or feeling hot, or nervous, or shaky.

Appendix 2: Symptom/Impairment Question wording

Primary role activity:

MEPS: limited in ability or completely unable to work at a job, do housework, or go to school because of an impairment or a physical or mental health problem

BD: workers: reasons related in any way to your health that you did not work at all (on day prior to interview), or were/would have been limited in any way in the amount or kind of work done, such as using special aids, not doing certain tasks, taking special rest periods, or working only part of the day?

Homemakers: reasons related in any way to your health you are not now working (more) hours for pay/did no housework/were/would have been limited in any way in the amount or kind of housework, such as not lifting small children, not cooking, washing or ironing, not doing heavy cleaning, or taking special rest periods?

Retired: reasons related in any way to your health you are not now working (more) hours for pay/did no kind of work activities at all (on day prior to interview), including not only a job, but also such activities as shopping, cooking, cleaning, or working in or around the house, yard, or garden/ or would have been limited in any way in the amount or kind of work done, such as using special working aids, not doing certain tasks or strenuous work, taking special rest periods, or only working part of the day?

Other role activity:

MEPS: besides the limitations we just talked about, limited in participating in social, recreational, or family activities because of an impairment or a physical or mental health problem

BD: This card lists a wide range of activities people might engage in outside of work/housework/work around the house/schoolwork/play activities: (examples only, as usual for age): going shopping, handling personal business, and so on; taking part in hobbies, games, play, recreational activities, and so on; visiting or meeting with friends, relatives, and so on; taking part in church, Sunday school, religious activities, and so on; taking part in community work, going to meetings, and so on; attending movies, ballgames, other entertainment, and so on. Limited in any way related to your health in doing any of these types of activities (on day prior to interview).

Self-care:

MEPS: receive help or supervision with personal care such as bathing, dressing, or getting around the house (confirmed for people under age 13 that this is due to an impairment or physical or mental health problem).

BD: Did not or have more than the usual help for most persons your age in dressing, feeding yourself, using the toilet, or taking a bath (on day prior to interview).

Bed/chair bound (BD): spent most or all of day prior to interview in any type of chair or couch or in bed

Lifting/bending:

MEPS: difficulty bending or stooping, lifting 10 pounds, reaching over head, standing 20 minutes, using fingers to grasp

BD: had trouble or did not try to lift, stooping, bend over, or use stairs or inclines due to health related reasons (on day prior to interview).

Walking:

Limp/cane/crutches/walker (BD): limp, or use a cane, crutches, or walker (on day prior to interview).

Limited walking or other physical limitation (BD): have any other physical limitation or try not to walk as far or fast as most persons your age are able.

MEPS:

Moderate walking limitation: unable or some/a lot of difficulty walking a mile, some difficulty walking 3 blocks.

Severe walking limitation: unable or a lot of difficulty walking 3 blocks, unable or some/a lot of difficulty walking up 10 steps.

Drive/public transport (BD): Did not drive, did not use public transportation, or needed/would have used more help from someone else than usual for your age to do so, for reasons related in any way to your health (on day prior to interview).

Pain:

MEPS: Have moderate pain or discomfort/have extreme pain or discomfort (EQ5D).

BD: 1) Pain, stiffness, weakness, numbness or other discomfort in chest, stomach, side, neck, back, hips, joints of hands/feet/arms or legs (on day prior to interview). 2) Pain, bleeding, itching, or discharge(drainage) from sexual organs (excluding normal monthly menstrual bleeding). 3) Burning or itching rash on large areas of face, body, arms, or legs.

Depressive symptoms:

MEPS: Felt downhearted and blue a good bit of the time, most of the time, or all of the time during the past 4 weeks (vs. some, a little, or none of the time) (SF-12).

BD: Spells of feeling upset, being depressed, or of crying (on day prior to interview)

Anxiety symptoms:

MEPS: Felt calm and peaceful a little or none of the time during the past 4 weeks (vs. some of the time, a good bit of the time, most of the time, all of the time.) (SF-12)

BD: Excessive worry or anxiety (on day prior to interview).

Sleep (BD): Trouble sleeping (on day prior to interview).

Vitality:

MEPS: Had a lot of energy some of the time/a little or none of the time during the past 4 weeks (vs. a good bit of the time, most of the time, all of the time.) (SF-12)

BD: General tiredness, weakness, or weight loss (on day prior to interview).

Vision:

MEPS: difficulty seeing (with glasses or contacts, if used): some difficulty seeing or can not read ordinary newspaper print or can not recognize familiar people standing two or three feet away or blind

BD: score of less than 80 on a 100-point Vision Functional Assessment.

Hearing:

MEPS: difficulty hearing (with hearing aid, if used): some difficulty hearing or can not hear some or most things people say or deaf

BD: score of less than 80 on a 100-point Hearing Functional Assessment.

Speech (BD): trouble talking, such as lisp, stuttering, hoarseness, or inability to speak (on day prior to interview).

Cognition:

MEPS: Experience confusion or memory loss such that it interferes with daily activities; have problems making decisions to the point that it interferes with daily activities; require supervision for own safety.

BD: Trouble learning/remembering/thinking clearly (on day prior to interview)

Miscellaneous (BD; on day prior to interview)

Sick or upset stomach, vomiting or loose bowel movements, with or without fever, chills, or aching all over

Cough, wheezing, or shortness of breath with or without fever, chills, or aching all over.

Problems with sexual interest or performance.

Overweight or underweight for age and height; or skin defect of face, body, arms or legs, such as scars, pimples, warts, bruises, or changes in color.

Any combination of one or more hands, feet, arms, or legs either missing, deformed (crooked), paralyzed (unable to move) or broken—includes wearing artificial limbs or braces.

Headache, or dizziness, or ringing in ears, or spells or feeling hot, or nervous, or shaky

Appendix 3: EQ5D question wording

As shown below, the EQ-5D contains five questions asking about the extent of problems in mobility, self-care, daily activities, pain, and anxiety/depression. Each question has three possible responses: no problem, mild problem, or severe problem. The combination of responses to these questions comprises a 'health state'. Scores are assigned to each of the 243 possible states based on previous research. The absence of problems--the most highly valued state—is equal to 1, death is equal to zero, and other states are assigned a score based on community ratings of a portion of the states (with extrapolation to other states), meant to represent the average preference of one state over another (REF).

The rating scale is a sixth, separate question on the EQ5D, not included in EQ5D scoring, and shown below.

Health Questionnaire
(English version for the US)

Note: this copy of the EQ5D for viewing only; any use for research or clinical purposes will require approval from the developers (contact the EuroQol Executive Office at userinformationservice@euroqol.org)

By placing a checkmark in one box in each group below, please indicate which statements best describe your own health state today.

Mobility

I have no problems in walking about

I have some problems in walking about

I am confined to bed

Self-Care

I have no problems with self-care

I have some problems washing or dressing myself

I am unable to wash or dress myself

Usual Activities (*e.g. work, study, housework, family or leisure activities*)

I have no problems with performing my usual activities

I have some problems with performing my usual activities

I am unable to perform my usual activities

Pain/Discomfort

I have no pain or discomfort

I have moderate pain or discomfort

I have extreme pain or discomfort

Anxiety/Depression

I am not anxious or depressed

I am moderately anxious or depressed

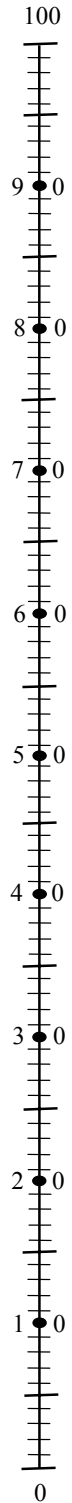
I am extremely anxious or depressed

To help people say how good or bad a health state is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked 100 and the worst state you can imagine is marked 0.

We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your health state is today.

**Your own
health state
today**

Best
imaginable
health state



Worst
imaginable
health state