NBER WORKING PAPER SERIES

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

Susan T. Stewart
Rebecca M. Woodward
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

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

Susan T. Stewart, Rebecca M. Woodward, and David M. Cutler

NBER Working Paper No. 11358
May 2005
JEL No. I10, I12

ABSTRACT

We propose a method of quantifying non-fatal health that details the mechanisms through which chronic conditions affect health. Self-rated health status and time-tradeoff ratings of current health are regressed on impairments and symptoms from the Quality of Well-Being Scale, using OLS regression and ordered probit. This yields estimates of their effects analogous to disutility weights but not based on counterfactual scenarios, and accounts for complex non-additive relationships. Data are from 1420 adults age 45-89 in the Beaver Dam Health Outcomes Study. Chronic condition weights and summary measures of health are derived, laying the groundwork for a detailed national summary measure of health.

Susan T. Stewart
Harvard University
sstewart@nber.org

Rebecca M. Woodward
Harvard University
rebecca_woodward@harvard.edu

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

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 is based on a broad range of reported symptoms and impairments and quantifies their relationship to underlying chronic conditions. An understanding of how specific conditions affect health can highlight problems for which successful interventions have the greatest potential return on expenditures.

A number of instruments have been developed to measure general health status and health-related quality of life. However, no gold standard has been agreed upon to date (Gold et al., 1996). 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 problems in these domains, and the approach used to aggregate them into a summary score. Our goal was to build upon existing measures and propose methods for dealing with our concerns in these areas.

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. 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 endorse one of 4 or 5 categories between ‘poor’ and ‘excellent’ that describes their health. 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 (Cutler

and Richardson, 1997; Idler and Benyamini, 1997; Idler and Kasl, 1995; Idler et al., 2000; Wilcox et al., 1996). However, both utility-based and likert scale ratings of general health lack detail regarding specific symptoms and impairments that contribute to the overall assessment. This detail is essential in evaluating the mechanisms of health change. Our first goal is to examine the role of symptoms and impairments from different domains in overall health assessment.

Detailed 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 and items that are implicitly or explicitly covered, as illustrated in Table 1. '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 et al., 2000) (and shorter SF-12 and SF-8 versions). In contrast, 'preference-based' measures assign utility weights to items or domains, 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. Such measures include the Health Utilities Index (HUI Mark I, II and III, developed in Canada; Horsman et al., 2003), the Quality of Well-Being Scale (QWB, interviewer and self-administered versions, developed in Southern California; Kaplan et al., 1976; McDowell and Newell, 1996), the EQ-5D ("EuroQol", developed in Europe; Brooks et al., 2003), the 15-D (Finland; Sintonen, 2001), the 'Assessment of Quality of Life' (AQOL, Australia; Hawthorne et al., 1999) and the SF-6D (Brazier 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, or having to follow a health regimen. Exclusion of such items can result in ceiling effects (Andresen et al., 1998; Essink-Bot et al., 1997; Fryback et al., 1993; Jenkinson et al., 1997; Johnson and Pickard, 2000; Kaplan et al., 1998); 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. Failing to account for the sources of such HRQOL reductions does not allow measurement of improvement in health if these problems are alleviated.

Determining the extent to which different problems reduce HRQOL 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. It is common that less severe ratings are found among those who have experienced the condition being rated. A number of explanations have been put forth to explain this phenomenon, including patient vantage point, adaptation to health problems,

a focusing illusion, and response shift (Ubel et al., 2003). 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 et al., 1996).

There is also an extensive literature in psychology that suggests that community ratings of health states many not be accurate—that people overestimate the negative impact of a future hypothetical event on their happiness (Gilbert et al., 1998; Wilson et al., 2000). This 'durability bias' puts 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, 2005). 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 health. If there is consistency in ratings across people with the same health problem, this method will yield an accurate estimate of the effect of the problem. If particular groups in which specific health problems are over or under-represented systematically report their health as better or worse than others, then the effects these problems may be mis-estimated.

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 et al., 2000). However factor analysis can inappropriately substitute variation in frequency for variation in social preference, and can thus exclude items with important effects on health if they have a low correlation with others or occur infrequently (Kaplan 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, EQ-5D). We allow for a flexible functional form that can accommodate additive or non-additive relationships between pairs of symptoms and impairments within domains.

Methods

Data

A source of data with the richness to enable empirical examination of these three issues is the Beaver Dam Health Outcomes Study (BDHOS; Fryback et al., 1993), a community-based longitudinal study of health status and quality of life in the U.S. state of Wisconsin. 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 institutionalized in an acute or chronic care facility. The current study uses 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 et al., 1993).

This survey includes multiple measures of health, including the QWB, which is unique in its inclusion of a broad list of symptoms and impairments in addition to more commonly asked questions on mobility, social activity, and physical function, and was developed with the intention of covering the full array of health problems that affect people's daily lives (Kaplan et al., 1976; McDowell and Newell, 1996). 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 (Kaplan and Anderson, 1996).

The BDHOS also includes a 5-point general self-rated health question, the SF-36 scale, a TTO rating of current health, and over 30 self-reported chronic health conditions. This range of health data enabled us to assess relationships between particular domains of impairment and ratings of overall health, empirically examine and account for non-additive relationships between domain items, and derive weights for chronic health conditions and summary measures of population health based on profiles of symptoms and impairments.

To examine the importance of different domains of health, we re-assigned QWB items to seven domains based on our review of the health measurement literature (Brooks et al., 2003; Hawthorne et al., 1999, Kaplan, 1998; Horsman et al., 2003; Kaplan et al., 1998; Sintonen, 2001;

Ware et al., 2000): social activity, physical activity, pain, mental health, vitality, and sensory. We also included a 'miscellaneous' category, containing ten items that did not fit into these domains. Impairments and symptoms within each domain are listed in the first column of Table 5. (The "other symptoms" item under miscellaneous includes a range of problems that respondents experienced but felt were not included in the main questionnaire.)

Analyses

Our analysis proceeds in three parts: the development of weights for symptoms/impairments and domains, the estimation of health decrements for specific chronic conditions, and the calculation of summary measures. Each part of the analysis is done using both a TTO rating of current health and a 5-category self-rated health question (SRHS). The TTO has the advantage of incorporating elements of expected utility theory, which is considered to be the gold standard in the decision-making field (although TTO ratings are not technically considered utilities). SRHS, in contrast, is commonly asked, and does not require extrapolation to any other health state (whereas the TTO rating of current health requires respondents to imagine the absence of their health problems). For the TTO, respondents were asked to imagine that there was a medical treatment that could cure their health problems, and were reminded by the interviewer which health problems they had reported. They were told that the treatment would on average reduce their life expectancy by a certain percentage, beginning with 10%. Respondents were presented with tradeoffs until they reached a point of indifference. Those unwilling to trade 0.5% of their life expectancy were assigned scores of 100, and those willing to trade for as little as one day in perfect health were given scores of zero. The SRHS question was

worded: “in general, would you say your health is: excellent, very good, good, fair, or poor”, with excellent giving the highest score.

Estimating weights for domains, symptoms and impairments

The first part of our analysis is to estimate the HRQOL decrement of specific health problems, by relating TTO and SHRS ratings of general overall health to specific impairments and symptoms. We estimate domain and item weights using regression analysis. To illustrate the methodology, we consider two domains, D_1 and D_2, with responses for each individual of D_{1i} and D_{2i}. Our equations are of the form:

$$\text{Health}_i = \beta_1 \cdot D_{1i} + \beta_2 \cdot D_{2i} + \varepsilon_I \qquad (1)$$

where D_i are the functioning in domains, $\beta_{\cdot s}$ are the estimates of the effects of these domains on health, and Health_i is a general rating of current health using TTO or SRHS.

To predict TTO ratings, OLS regression was used. Some researchers have used a two-part model for the TTO truncated distribution, but its use has been found to make little difference with these data (personal communication, D. Fryback). An ordered probit was used to predict SRHS, due to the ordinal nature of the question. Including all of the domains in the same equation gives us an independent effect of each domain on health, for example, the extent to which pain adversely affects health independent of the physical limitations with which it is often associated. In other cases, we desire a full effect of a domain on health, including all the correlated symptoms and problems. To determine this, we estimate equation 1 including only the

items from a particular domain in the equation. We term these two measures the 'independent' and 'full' impact of each domain on health.

Within domains, there are often several questions about different aspects of health. For example, our pain domain includes questions regarding pain, rash, and problems with urinary, bowel, and sexual organs. Past summary measures have varied in their treatment of these multiple symptoms, for example assigning the weight associated with 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 (EQ-5D, -0.17 for moderate pain/discomfort as the only symptom (Shaw 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 et al., 1996)). In order to test for the appropriate relationships between items, we include interaction terms between pairs of items within domains when the number of respondents reporting both problems is sufficiently large (at least 15). Algebraically, suppose we are considering one domain (D_1) which includes two items (D_{1a} and D_{1b}). We relate overall health ratings to each of the items and to the interaction between the items:

$$Health_i = (\beta_{1a} \cdot D_{1ai} + \beta_{1b} \cdot D_{1bi} + \beta_{1ab} \cdot D_{1ai} \cdot D_{1bi}) + \varepsilon_i \quad (2)$$

Several common assumptions are special cases of this framework. In the case of a worst symptom indicator, we would find (assuming D_{1ai} is worst) $\beta_{1a} < \beta_{1b} = -\beta_{1ab}$: 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 ($\beta_{1ab} = 0$), more-than-additive effects ($\beta_{1ab} < 0$), and effects of decreasing marginal impairment ($\beta_{1ab} > 0$).

This methodology might be applied to different domain scores as well. For example, if D_1 and D_2 are two domains (each with a set of specific items and their interactions), we could estimate models of the form:

$$Health_i = (\beta_1 \cdot D_{1i} + \beta_2 \cdot D_{2i} + \beta_{12} \cdot D_{1i} \cdot D_{2i}) + \varepsilon_I \quad (3)$$

where β_{12} shows the evaluation of the combination of the two domains. In practice, these estimates were not sufficiently stable due to the limited number of respondents with certain combinations of problems. Thus, only within-domain interaction terms were included in our final models, and across domains, weights were combined additively in our scoring. Future analyses with larger data samples will be useful in examining interactions across domains.

Our regression equation (1) includes all domains and their component symptoms and impairments, but no demographic variables. We chose not to control for demographics with the belief that these do not fundamentally affect health. If we had controlled for them, 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. (Although alternative analyses controlling for age, sex and their interactions yielded similar results.)

In order to calculate health decrement weights for each symptom and impairment, we used a best case / worst case hypothetical scenario. Mean predicted scores from the TTO and SRHS 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. We used a similar methodology for each domain as a whole, calculating the effect of the domain by setting people to have all and then none of the particular items in the domain, and taking the difference between these mean predicted scores.

In the case of TTO as the dependent measure, scores are on a 0 to 1 scale. This, along with the utility-based reference of the question, is advantageous. The disadvantage of the TTO is the possibility that people are poor at performing counterfactuals about perfect health status. The SRHS analysis avoids this by asking people only about their current health state.

Because SRHS is not on a 0-1 QALY scale anchored at death, two issues need to be addressed in our SRHS approach. First, this methodology requires that 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 others, and particular health problems are over or under-represented among these groups, then the effects 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 et al., 2004).

The second issue is translating predicted SHRS scores from the 1-5 scale into a 0-1 QALY metric. To develop a QALY metric, we first estimated the maximum range of the SRHS scale using predicted scores from hypothetical scenarios in which everyone had every problem, assumed to be the worst health state, and in which no one had any had any problems, assumed to be the best health state. Because our worst SRHS state was not equal to death, we did not equate the score for this state to 0 on the QALY scale. To estimate the lowest level of health in our sample with reference to death, we examined the mean TTO rating among those with the lowest 10% of ratings on the TTO. All predicted SRHS scores were then mapped to the 0-1 QALY scale so that the value for our hypothetical worst state was equal to that lowest endorsed level on the TTO, and our best SRHS state was equal to 1. Formally:

$$Y = 1+(X-x_1)/(x_1-x_0) \cdot (1-Z) \quad (4)$$

where Y is the SRHS score normalized to a 0-1QALY metric, X is the predicted SRHS score, x_0 is the mean predicted score for the worst SHRS scenario, x_1 is the mean predicted score for the best scenario, and Z is the mean rating among those with the lowest 10% of ratings on the TTO.

SRHS-based item weights were then calculated by taking the difference in means on these re-scaled SRHS predicted scores among those with and without each symptom and impairment.

Estimating Weights for Chronic Conditions

The second part of our methodology is to estimate the health decrements of thirty specific chronic conditions, based on the relationship of these conditions to domain-specific impairments and symptoms. Typically, chronic condition-specific HRQOL has been estimated by calculating the mean score on a preference-based instrument among people with the health condition (i.e. Gold et al., 2002; Gold and Muennig, 2002). However, this approach does not illustrate the relative contributions of specific symptoms and impairments to the overall HRQOL decrement for a specific chronic condition. Thus, researchers are not able to tell how alleviation of one symptom of a disease would affect health. Nor does this methodology control for other chronic conditions that a person may have.

To estimate disease-specific HRQOL decrements, we pursue a method similar to the one used to calculate our item and domain weights. We first use probit analysis to regress each impairment and symptom on chronic conditions, estimating the probability of the impairment occurring among those with each condition. We then multiply these probabilities by our decrement weights for symptoms and impairments, and sum across symptoms and impairments to form an overall decrement weight for each condition. Once again, there are two possible metrics of chronic disease QOL. The first is the independent contribution of that condition to all impairments and symptoms, controlling for other possible conditions the person may have. One might want this, for example, to learn about the impact of diabetes on health separate from any comorbid renal or cardiovascular conditions. Alternatively, one might want to know about the full impact of diabetes, accounting for all the channels through which it affects health. We

determine independent and full effects by regressing each symptom/impairment on the full list of conditions as well as on each condition on its own, effectively looking at the proportion of those with each chronic condition who report each impairment/symptom.

Estimating a Summary Measure of Health

The third and final step in our methodology is to form two summary measures of health for each person, based on TTO and SHRS weights. The summary health scores can be generated in one of two ways: using data on impairments and symptoms, or data on chronic diseases. The mean in each case will be very similar, with slight differences across people for health states involved in interactions. We use the method based on chronic conditions, to show the application of our metric to data typically collected in other surveys. We use probabilities from the probit analyses estimating the independent effect of each chronic condition on each impairment and symptom. Based on the chronic conditions reported by a respondent, we estimate the overall probability that the respondent experienced each symptom and impairment (capping the probability of experiencing a symptom or impairment at 1). We then multiply these probabilities by our weights for these symptoms/impairments, developed as described above. Finally, we sum across symptoms and impairments to obtain a summary score. We compare our summary measures of health to two other commonly used measures: the QWB and summary scores from the SF-36 (physical and mental component scores, and the SF6D).

Results

The distributions of TTO and SRHS ratings of current health are shown in Figures 1 and 2. Neither variable was normally distributed; both were skewed toward good health (mean TTO

was 0.86), as would be expected in a general community population. However, the TTO distribution (Figure 1) was severely truncated, with 59 percent of responses at 100. These responses appear to 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.

Comparison of SRHS and TTO ratings in Table 2 indicates that only 65 percent of those with TTO ratings of 100 rated their health as excellent on the 5-point question, and 33 percent of those with low TTO ratings (<25) rated their health as excellent. Indeed, the correlation between the TTO measure of health and self-rated health status is only .0.33 ($p < .0001$). This suggests that factors unrelated to health were affecting one or both of the two scores. Such factors possibly include attitudes regarding length vs. quality of life, and perhaps misunderstanding of the questions. But other interpretations are also possible.

As a first examination of the relationship between health ratings and health problems, Table 3 illustrates the distribution of SRHS and TTO scores among those with one or more impairments in each domain. A much higher proportion of those with problems in each domain reported their health as excellent using TTO than using SRHS, including individuals with substantial impairments.

The last two columns of Table 3 show the number and proportion of the sample reporting one or more symptoms/impairments in each domain. The most common problems were in the sensory and miscellaneous domains: wearing corrective lenses, prescribed medication or diet,

problems with hearing/ears/nose/mouth (including crooked teeth), and problems with weight or appearance. The portion reporting items in the pain domain was also high, at 54 percent. Table 1 of the Appendix gives more detail regarding the proportion of the sample reporting each item and combination of items within domains.

Table 1 of the Appendix also reports coefficients from OLS regression model predicting TTO and the ordered probit model predicting SRHS. Virtually all of the main effect coefficients are negative (symptoms predict worse health), as expected. Over half of the items are statistically significantly related to TTO, and 25 percent significantly predict SRHS. 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 diseases is not as bad as a simple additive model would suggest, although instances of negative coefficients for interaction terms are also found. In several cases, 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 any additivity across comorbid impairments.

The mean predicted SRHS score was 0.41 for the hypothetical scenario in which everyone had every symptom/impairment and 4.18 for the scenario in which no one had any of

the symptoms/impairments. The mean TTO rating among those with the lowest 10% of ratings on the TTO was 0.30. These values were used to normalize SRHS scores to a 0-1 QALY metric using formula (4) above.

Independent and full decrement weights for each domain are shown in Table 4, with each domain including any respondents who reported one or more of its symptoms/impairments. When SRHS is used as a dependent measure, the physical activity, social activity, pain, and 'miscellaneous' domains have the largest decrement weights. Decrement weights for the specific items in each domain are shown in Table 5. Within the physical functioning domain, the single biggest coefficients are for limitations in driving, and troubles with basic mobility such as lifting and bending. The miscellaneous symptom with the largest effect on health is having to take prescribed medications or follow a prescribed diet. Independent of symptoms and problems queried, people who take prescribed medications report themselves in worse health than people who do not. This result may seem counterintuitive given that medications are intended to improve health. However if this item did not have a negative weight, then curing a person of a medicated ailment could not show up as an improvement in health. The negative result may also reflect the burden of regimen, and health problems associated with taking medications that are not completely captured by the health questions asked.

For several domains, independent effects are roughly half as large as full effects (Table 4), indicating significant shared variance between that domain and others. Indeed, the zero weight for mental health as an independent effect likely reflects that these problems are unlikely to occur alone without affecting other domains. It also reflects the fact that the worst case

scenario assigning these interactions to everyone includes interaction terms for impairment pairs that few people have. For the sensory domain, the independent effect is only 25 percent as large as the full effect. This reflects the fact that although a large portion of the sample reported sensory problems (including the need for glasses), many of these individuals had problems in other domains that were more important to health rating than sensory problems alone. In contrast, for the pain domain the independent and full effects are closer in value, indicating that pain is a key factor in influencing self-rated health regardless of other health problems.

When TTO is used as the dependent measure, the pattern of HRQOL decrements across domains and items is similar to when SRHS is used, with most TTO weights somewhat less severe, likely due to the truncated TTO distribution. There is more apparent shared variance between domains using TTO; independent effects for most domains are quite small compared to the full effects. Indeed, the weight for mental health comes out positive when controlling for other domains.

The single biggest decrement to health in the TTO models comes from physical activity limitations, with many specific items contributing to this. The pain and miscellaneous domains have decrement weights from the TTO analysis that are particularly less severe than those from the SRHS analysis. People who are in pain or take medications report themselves in worse health but are not willing to give up years of life to alleviate those symptoms.

Table 5 shows that many other problems, such as sensory impairments, speech problems, sexual problems, and problems with weight or appearance, had little independent effect using

either rating method. A few problems have zero or small positive weights (that round up to 0.01), 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 independent weights for symptoms and problems are to be considered as a group and not individually.

The original QWB weights/scores most comparable to each of our domains and items are shown in the last columns of Tables 4 and 5. For the social and physical activity domains, QWB weights were less severe than both our independent and full weights. For the other domains, which were derived from the QWB symptom list, QWB weights were more comparable to our full weights. However, to illustrate the effects of considering all symptoms rather than only the one reported as most bothersome, we also compare our independent weights. Our independent domain weights (Table 4) were much less severe than QWB weights for the worst symptom in each domain, and our independent item weights (Table 5) were on average over 6 times lower than QWB weights for these items.

Table 6 illustrates our method of obtaining full decrement weights for particular chronic conditions based on SHRS, for the examples of MI, mood disorder, and arthritis. The first column lists our decrement weights derived for each item using SRHS as a dependent measure (taken from Table 5). We then report for each condition the probability of experiencing each symptom/impairment. This is multiplied by our SRHS weight for that item to yield a decrement weight for each item among those with that chronic condition. For example, the symptoms with the largest impact on health ratings among those with heart attack are work limitations, pain, and

having to be on medications or a prescribed diet. The sum of these decrements is the overall disutility weight for the chronic condition.

This method of obtaining disutility weights was replicated for each of the chronic conditions in the BDHOS data. To summarize the symptom profiles for all 30 chronic conditions examined (not shown): pain and medication/diet were top contributors to the health decrement for most chronic conditions, and approximately half of the chronic conditions had among the largest contributions from work limitations, trouble with lifting/stooping/bending or stairs, not driving or using public transport independently, and ear/nose/mouth or throat problems. Not surprisingly, coughing/shortness of breath was a major symptom among those with respiratory conditions (asthma, emphysema, bronchitis), and anxiety was prominent among those with an anxiety or mood disorder. Vitality was a key symptom only among those with mood disorder.

The full and partial disutility weights for each chronic condition are shown in Table 7. When TTO-based weights are used, decrements for conditions are less severe than when SRHS-based weights are used, again reflecting the positively skewed TTO distribution. A respondent with none of the chronic conditions has an SRHS disutility of -0.11. Compared to this, full decrement weights based on SHRS ranged from -0.23 for sinus problems to -0.46 for congestive heart failure. Other conditions with large HRQOL decrements include diabetes (-0.41 for Type I and –0.32 for Type II), MI (-0.40), and mood disorders (-0.39). This pattern generally accords with expectations. Independent weights are on average roughly half as severe as full weights for each chronic condition, reflecting correlations between chronic conditions. Our weights are compared to weights previously published for these chronic conditions in the last three columns

of Table 7. For most conditions our full weights are similar to and our independent weights are less severe than mean QWB or HALex scores among those with the chronic condition (Gold et al., 1998) and weights by Cutler and Richardson (Cutler and Richardson, 1997) based on these chronic conditions as predictors of SRHS.

The overall mean summary health score using our method based on reported diseases is 0.77 using SRHS-derived weights (5-95 percent range = 0.66-0.89) and 0.88 using TTO-derived weights (5 to 95 percent range = 0.81-0.95), with a correlation of 0.99 between the two. The high correlation reflects the reasonably common impact of different conditions on health measured both ways. By contrast, the mean QWB score is 0.72, and the mean SF6D is 0.84. Our measures are thus in a range similar to these existing metrics. Figure 3 compares mean scores on our measures, the QWB, and the SF6D among those with problems in each domain. Our TTO-based summary measure yield the highest scores across domains, whereas our summary measure using SRHS-based weights shows decrements more severe than the SF-6D and less severe than the QWB across all domains. Despite the similarity in distribution of scores across domains, correlations between measures at the individual level are only moderate. Both of our summary measure scores correlate 0.55 with the QWB, 0.49 with the SF-6D, 0.58 with the physical component score of the SF-36, and 0.20 with its mental component score. The QWB is similarly related to the SF-6D (r=0.52) but more closely related to the SF-36 physical component score (r=0.72) and less closely related to the mental component score (r=0.12) than our measures. The relatively low correlation between our measures and other extant measures shows the importance of relaxing functional form assumptions in our methodology.

Discussion

This paper lays out a method for estimating the HRQOL decrements associated with different chronic conditions and their associated symptoms and impairments, and estimating a summary measure of health based on this information. In addition to providing detail regarding the effects of chronic conditions, our method was developed to address three issues of concern in the measurement of general health: the limited range of impairments and symptoms typically measured, the use of counterfactual scenarios to obtain condition weights, and inconsistent accounting for co-relationships between health problems.

Our results show important differences between TTO and SRHS-based measures of health. The SRHS-based measure yields greater levels of average impairment than does the TTO-based measure. Average health is 0.77 using SRHS weights and 0.88 using TTO weights. The mean SF6D score is in between these, and the mean QWB score is more severe. Further, the correlations between scores from our method and these other measures were only moderate (r= 0.49 to 0.55). We find some of the TTO results anomalous. The very large share of people who appear unwilling to trade off any reduction in length of life for quality suggests that people may—even when reporting severe impairment—find the concept of hastening death too difficult to consider. If this is the case, TTO values of quality of life may be systematically too high.

Methodologically, our approach to health assessment has several advantages over existing measures. In basing our weights on global assessments of current health, we avoid a range of biases that have been found to affect people's ratings of experienced and counterfactual states (Gilbert et al., 1998; Ubel et al., 2003; Wilson et al., 2000). 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, 2005), and our measure is more consistent than others with this recommendation. We also relax the assumptions about comorbidities, which is important substantively. Results for interaction terms reveal complex relationships between different pairs of comorbidities, in some cases indicating complementarity between conditions and in others, a synergistic negative effect of having both conditions. Thus, the scoring rules used in many other health metrics seem overly constraining.

Further, expanding the range of health problems queried adds important information. No single questionnaire besides the QWB includes all of the domains/items that we found to have the greatest impact on health. Indeed, among the items we find to be important but that are not routinely accounted for are: cognitive problems (with a decrement of -0.05 based on SRHS, reported by 14 percent of our sample), low vitality (-0.04, reported by 19 percent), limitations in major role activity (-0.10, reported by 17 percent), and use of medications or a prescribed diet (-0.08, reported by 78 percent).

Additional impairments that we found to have important impacts on health were physical activity limitations (up to –0.09), respiratory symptoms (-0.05) pain (–0.05), and anxiety (–0.07). 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, 1992; Leinonen et al., 1999), persistent pain disorder (Gureje et al., 2001), cognitive capacity (Leinonen et al., 1999), and taking prescription medications (Hogan et al., 2003; Mansson et al., 2002). The great importance of these factors

suggests the possibility that targeting interventions to alleviate some particular impairments or promoting adaptive behavior in a few domains could have a significant effect on health across a wide range of the population.

Chronic conditions with the worst impact on health ratings in our study included heart and respiratory conditions, mood disorders, chronic pain, sleep disorders, ulcers, and Type I diabetes. Most of our full chronic condition weights were similar to previously reported mean HRQOL scores by chronic condition. However, an important advantage of our method is that it illustrates the effects of chronic conditions on specific symptoms and areas of function. Most of these chronic conditions were associated with a greater likelihood of limitations in primary role activity, trouble with lifting/bending/stairs, pain, and the need to take medications or follow a prescribed diet. Over half of these chronic conditions caused limitations in driving or using public transportation. Other impairments and symptoms had increased prevalence only among those with specific chronic conditions.

The relatively large negative effect of prescription medications or diets on general health ratings has interesting implications. It may reflect aspects of health that are not fully captured in other questions, such as side effects of medications and problems not relieved by medication. However, even in absence of symptoms, people who take medications are regularly reminded of the underlying health problems that require this treatment, and likely take these into account when rating their health. While this may be seen as a labeling effect arising from the burden of regimen, it can be argued that eliminating underlying causes would indeed render a person fundamentally healthier than medicating symptoms. This may change over time with a shifting

societal perception of medications and diets as wellness tools to maintain optimal health, rather than as negative indicators that one is not healthy. Our data is from 1991, and it would be interesting to see if the medication item would have less impact on contemporary ratings now that medications are increasingly used as tools in the treatment, control, and prevention of health conditions.

Limitations

The community sample that was available to illustrate our method is not nationally representative, and its relatively small size resulted in small samples of respondents with certain conditions and combinations of symptoms. This may underlie the few small positive main effects and the anomalous large positive interaction terms for some combinations of problems. A larger sample will better elucidate the complex relationships between symptoms and impairments within and across domains. It is also possible that respondents with multiple symptoms experienced some response shift in their use of the SHRS scale. It would have been ideal if vignettes had been incorporated into the original data collection to account and adjust for this potential bias (Salomon et al., 2004). Finally, while the QWB is quite comprehensive in its coverage of symptoms and impairments, a few of the symptom categories (such as the ones we labeled ‘appearance’, and ‘ear/nose/throat’) are quite broad, and would be best separated into unique questions in order to capture more specifically the factors underlying differences in health ratings.

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 some countries, no single measure has been adopted for the U.S.. 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. 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 (Cutler, 2005).

Various surveys ask about chronic disease 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 chronic conditions, or on combinations of symptoms and conditions. We develop in this paper a set of weights for 1991, however these are based on a community sample. Also, these weights could well change over time, due to alterations in disease manifestation and 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.

Acknowledgements

This work was supported by National Institute on Aging (NIA) research grants: P01 AG17625, P30 AG012810 and R01 AG015911.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 Allison Rosen, Willard Manning, and Dennis Fryback for their comments on the manuscript.

Table 1: Domains Covered by Existing General Health Instruments

Domain*	SF-36 (Ware et al., 2000)	HUI II/III (Horsman et al., 2003)	EQ-5D (Brooks et al., 2003)	15-D (Sintonen, 2001)	AQoL (Hawthorne et al., 1999)	QWB (Kaplan et al., 1998)
Physical Function	√	√	√	√		√
Social function	√		√	√	√	√
Cognition		√		√		√
Mental Health	√	√	√	√	√	√
Pain	√	√		√	√	√
Vitality	√			√		√
Sensory		√		√	√	√
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: Distribution of Self-Rated Health Status Reponses by Level of Response on Time-Tradeoff (n = 1380)

	SRHS			Proportion
TTO	Poor/ Fair	Good	Very Good/ Excellent	
< 25 (n= 48)	0.29	0.37	0.33	1.0
25-99 (n = 516)	0.20	0.46	0.34	1.0
100 (n = 816)	0.04	0.31	0.65	1.0

Table 3: Distribution of Self-Rated Health Status and Time-Tradeoff Ratings Among Those Reporting One or More Symptoms or Impairments in Each Domain; Number and Proportion Reporting Symptoms or Impairments in Each Domain

	TTO			SRHS			Sample	
Domain	< 25	25-99	100	Poor/ Fair	Good/ Very Good	Excellent	N	Proportion
Social activity	0.10	0.62	0.29	0.36	0.61	0.03	229	0.17
Vitality	0.06	0.53	0.40	0.27	0.68	0.05	267	0.19
Physical activity	0.06	0.56	0.37	0.25	0.71	0.04	500	0.35
Mental health	0.04	0.50	0.55	0.23	0.70	0.07	319	0.22
Pain	0.04	0.47	0.49	0.16	0.76	0.08	766	0.54
Miscellaneous	0.04	0.38	0.58	0.12	0.75	0.13	1356	0.95
Senses	0.04	0.38	0.58	0.11	0.75	0.14	1371	0.96
Overall	0.03	0.37	0.59	0.06	0.77	0.17	1420	1.00

Table 4: Independent and Full Decrement Weights for Entire Domains

	Effect with TTO as measure of health		Effect with SRHS as measure of health		Decrement from Quality of Well-Being scale
Domain*	Independent	Full	Independent	Full	
Miscellaneous	-0.10	-0.47	-0.22	-0.46	-0.34
Physical activity	-0.25	-0.34	-0.18	-0.35	-0.106
Social activity	-0.02	-0.23	-0.13	-0.30	-0.061
Pain	-0.14	-0.26	-0.16	-0.27	-0.349
Senses	-0.05	-0.19	-0.05	-0.22	-0.23
Mental health	0.05	-0.08	0.00	-0.16	-0.257
Vitality	-0.01	-0.09	-0.04	-0.13	-0.259

*domains include anyone reporting one or more items in the domain

in the QWB scoring, this weight applies only if the person is also limited in work or leisure activities

Table 5: Decrement Weights for Symptoms and Impairments by Domain,* Derived From Regressions of Time-Tradeoff Ratings and Self-Rated Health Status on these Items; Comparison to Original QWB Decrement Weights

Domains and symptoms/impairments from Quality of Well-being Scale	Weights using TTO	Weights using SRHS	Original QWB weights
Social Activity			
Limited or performed no major role activity	-0.13	-0.10	-0.061
Limited in other role activity	-0.01	-0.04	-0.061
Physical Activity			
1+ ADL limitations (dress/feed/toilet/bath)	0.01	-0.04	-0.106♦
Most or all of day in chair, couch or bed	-0.09	-0.04	-0.077
Trouble lifting/stooping/bending over/stairs	-0.02	-0.07	-0.060
Limp/cane/crutches/walker	-0.02	-0.02	-0.060
Limited walking or other physical limitation	-0.08	-0.04	-0.060
Do not drive and/or limited in public transport	-0.07	-0.08	-0.062†
Pain			
Pain[a]	-0.03	-0.05	-0.299
Rash (burning or itching rash on large area)	-0.03	-0.03	-0.240
Urinary/bowel/	-0.04		-0.292
sex organ		-0.03	-0.349
Mental Health			
Depressive symptoms[b]	-0.02	-0.02	-0.257
Excessive worry or anxiety	-0.03	-0.07	-0.257
Trouble sleeping	0.01	-0.03	-0.257
Vitality (general tiredness/weakness/weight loss)	-0.02	-0.04	-0.259
Senses			
Eye pain or trouble seeing after correction[c]	-0.02	-0.02	-0.230
Glasses or contact lenses	-0.01	-0.01	-0.257
Speech problem[d]	-0.02	-0.02	-0.237
Miscellaneous			
Gastrointestinal illness[e]	0.01	-0.01	-0.290
Respiratory illness[f]	-0.04	-0.05	-0.257^
Problems with sexual interest or performance	0.01	0.00	-0.257
Ear/nose/teeth[g]	0.01	-0.03	-0.170
Weight/appearance[h]	0.00	-0.01	-0.186
Medication or prescribed diet	-0.05	-0.08	-0.144
Limbs/extremities[i]	0.00	-0.02	-0.333
Headache[j]	-0.00	-0.03	-0.244
Trouble learning/remembering/thinking clearly	-0.03	-0.05	-0.340
Other symptoms[k]	-0.04	0.01	--

*Item weights within each domain do not sum to domain weights from Table 4 due to interaction effects;

♦this weight applies only if the person is also limited in work or leisure activities; †worst score on QWB mobility scale is –0.09 if in the hospital because of health reasons; our sample includes only a few people who were in the hospital, and we included these as unable to drive/take bus unaided; [a]Pain, stiffness, weakness, numbness or other discomfort in chest, stomach, side, neck, back, hips, joints of hands/feet/arms or legs; [b]Spells of feeling upset, being depressed, or of crying; [c]pain or discomfort in one or both eyes (such as burning or itching) or any trouble seeing after correction; [d]trouble talking, such as lisp, stuttering, hoarseness, or inability to speak; [e]sick or upset stomach, vomiting or loose bowel movements, with or without fever, chills, or aching all over; [f]cough, wheezing, or shortness of breath with or without fever, chills, or aching all over. [g]Pain in ear, tooth, jaw, throat, lips, tongue; missing or crooked permanent teeth—includes wearing bridges or false teeth; stuffy, runny nose; any trouble hearing—includes wearing a hearing aid; [h]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; [i]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; [j]Headache, or dizziness, or ringing in ears, or spells or feeling hot, or nervous, or shaky; [k]symptoms or problems that people felt were not captured on symptom list.

^0.257 was the average QWB weight and is used in QWB scoring as a weight for items that were not given weights in community rating studies.

Table 6: Illustration for Selected Chronic Conditions of the Method Used to Obtain Decrement Weights: Full Effect of Each Chronic Condition using Weights Based on Self-Rated Health Status

Domains and items from Quality of Well-being Scale	Weight using SRHS (from Table 5)	Myocardial Infarction		Mood Disorder		Arthritis	
		Prob-ability	Probability x weight	Prob-ability	Probability x weight	Prob-ability	Probability x weight
Social Activity							
Limited/performed no major role activity	-0.10	0.40	-0.04	0.28	-0.03	0.13	-0.01
Limited in other role activity	-0.04	0.25	-0.01	0.22	-0.01	0.16	-0.01
Physical Activity							
1+ ADL limitations(dress/feed/toilet/bath)	-0.04	0.10	0.00	0.04	0.00	0.05	0.00
Most or all of day in chair, couch or bed	-0.04	0.40	-0.02	0.21	-0.01	0.10	0.00
Trouble lifting/stooping/bending over/stairs	-0.07	0.55	-0.04	0.43	-0.03	0.32	-0.02
Limp/cane/crutches/walker	-0.02	0.25	-0.01	0.12	0.00	0.21	0.00
Limited walking or other physical limitation	-0.04	0.45	-0.02	0.31	-0.01	0.29	-0.01
Do not drive and/or limited in public transport	-0.08	0.40	-0.03	0.19	-0.02	0.15	-0.01
Pain							
Pain[a]	-0.05	0.75	-0.04	0.70	-0.04	0.70	-0.04
Rash (burning/itching rash on large area)	-0.03	0.05	0.00	0.13	0.00	0.09	0.00
Urinary/bowel/ sex organ	-0.03	0.10	0.00	0.18	-0.01	0.10	0.00
Mental Health							
Depressive symptoms[b]	-0.02	0.15	0.00	0.30	-0.01	0.08	0.00
Excessive worry or anxiety	-0.07	0.10	-0.01	0.34	-0.02	0.14	-0.01
Trouble sleeping	-0.03	0.10	0.00	0.25	-0.01	0.18	-0.01
			0.00		0.00		0.00
Vitality (tiredness/weakness/weight loss)	-0.04	0.35	-0.02	0.57	-0.02	0.24	-0.01
Senses							
Eye pain or trouble seeing after correction[c]	-0.02	0.20	0.00	0.33	-0.01	0.26	-0.01
Glasses or contact lenses	-0.01	1.00	-0.01	0.97	-0.01	0.97	-0.01
Speech problem[d]	-0.02	0.00	0.00	0.15	0.00	0.07	0.00
Miscellaneous							
Gastrointestinal illness[e]	-0.01	0.10	0.00	0.09	0.00	0.07	0.00
Respiratory illness[f]	-0.05	0.25	-0.01	0.33	-0.02	0.21	-0.01
Problems with sexual interest or performance	0.00	0.15	0.00	0.04	0.00	0.08	0.00
Ear/nose/teeth[g]	-0.03	0.85	-0.03	0.81	-0.03	0.76	-0.02
Weight/appearance[h]	-0.01	0.25	0.00	0.69	-0.01	0.54	-0.01
Medication or prescribed diet	-0.08	1.00	-0.08	0.91	-0.07	0.85	-0.07
Limbs/extremities[i]	-0.02	0.10	0.00	0.04	0.00	0.05	0.00
Headache[j]	-0.03	0.30	-0.01	0.42	-0.01	0.22	-0.01
Trouble learning/remembering/thinking	-0.05	0.30	-0.02	0.22	-0.01	0.18	-0.01
Other symptoms[k]	0.01	0.10	0.00	0.09	0.00	0.09	0.00
Total decrement for chronic condition*			-0.40		-0.39		-0.29

[a]Pain, stiffness, weakness, numbness or other discomfort in chest, stomach, side, neck, back, hips, joints of hands/feet/arms or legs; [b]Spells of feeling upset, being depressed, or of crying; [c]pain or discomfort in one or both eyes (such as burning or itching) or any trouble seeing after correction; [d]trouble talking, such as lisp, stuttering, hoarseness, or inability to speak; [e]sick or upset stomach, vomiting or loose bowel movements, with or without fever, chills, or aching all over; [f]cough, wheezing, or shortness of breath with or without fever, chills, or aching all over. [g]Pain in ear, tooth, jaw, throat, lips, tongue; missing or crooked permanent teeth—includes wearing bridges or false teeth; stuffy, runny nose; any trouble hearing—includes wearing a hearing aid; [h]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; [i]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; [j]Headache, or dizziness, or ringing in ears, or spells or feeling hot, or nervous, or shaky; [k]symptoms or problems that people felt were not captured on symptom list.

*same as the weights from Table 7, column 7.

Table 7: Independent and Full Decrement Weights for Chronic Conditions; Comparison to Weights from Other Sources

Chronic Condition	Prevalence in Beaver Dam		TTO derived decrements		SRHS derived decrements		QWB score	NHIS HALex score**	Cutler & Richardson 1997
	N	Per-cent	Indep-endent	Full	Indep-endent	Full			
Cancer	62	4.4	-0.07	-0.18	-0.15	-0.31	-0.31	--	--
Musculoskeletal									
Arthritis	644	45.4	-0.07	-0.17	-0.13	-0.29	-0.32	*-0.31*	-0.21
Gout	58	4.1	-0.07	-0.18	-0.14	-0.31	-0.26	-0.35	--
Back pain (severe)	263	18.5	-0.07	-0.19	-0.15	-0.32	-0.33	-0.25	--
Neck pain (severe)	105	7.4	-0.07	-0.19	-0.13	-0.33	-0.33	--	--
Endocrine									
Thyroid trouble/goiter	93	6.5	-0.08	-0.17	-0.15	-0.30	-0.31	*-0.30*	--
Diabetes (Type II)	89	6.3	-0.10	-0.20	-0.19	-0.32	-0.31	--	
Diabetes (Type I)	37	2.6	-0.13	-0.28	-0.23	-0.41	-0.35	--	
Either diabetes	126	8.9		-0.22		-0.35	-0.32	*-0.40*	-0.34
Circulatory									
Hypertension	515	36.3	-0.07	-0.16	-0.14	-0.27	-0.29	*-0.30*	-0.14
Stroke	14	1.0	-0.10	-0.22	-0.17	-0.35	-0.34	*-0.51*	-0.26
Angina	68	4.8	-0.04	-0.24	-0.08	-0.37	-0.35	*-0.43*	--
CHF	29	2.0	-0.10	-0.32	-0.18	-0.46	-0.39	*-0.47*	--
MI	19	1.3	-0.13	-0.29	-0.21	-0.40	-0.37	--	--
All heart (CHF/angina/MI)	92	6.5		-0.25		-0.38	-0.35	--	-0.20
Respiratory									
Asthma	46	3.2	-0.09	-0.19	-0.16	-0.30	-0.32	*-0.36*	-0.26
Bronchitis	49	3.5	-0.08	-0.21	-0.16	-0.35	-0.33	-0.26	-0.14
Emphysema	41	2.9	-0.09	-0.25	-0.16	-0.39	-0.36	*-0.51*	--
Sinus problem	106	7.5	-0.08	-0.13	-0.15	-0.23	-0.28	*-0.27*	-0.07
Allergy	30	2.1	-0.09	-0.15	-0.18	-0.27	-0.27	--	--
Digestive									-0.24
Diverticulosis/Colitis/Crohns	52	3.7	-0.06	-0.15	-0.13	-0.27	-0.30	*-0.36*	--
Ulcer (gastric, peptic, duod.)	82	5.8	-0.07	-0.19	-0.14	-0.33	-0.33	*-0.38*	--
Hiatal hernia	48	3.4	-0.09	-0.18	-0.17	-0.30	-0.32	*-0.34*	--
Mental Health									
Mood disorder	67	4.7	-0.09	-0.24	-0.17	-0.39	-0.35	--	--
Anxiety disorder	56	3.9	-0.07	-0.22	-0.13	-0.37	-0.32	--	--
Sleep disorder	145	10.2	-0.08	-0.21	-0.16	-0.35	-0.33	--	--
Sensory/Nerveous									
Cataracts	337	23.7	-0.06	-0.18	-0.13	-0.30	-0.31	*-0.32*	--
Glaucoma	70	4.9	-0.06	-0.19	-0.12	-0.30	-0.32	*-0.33*	-0.03
Macular degeneration	46	3.2	-0.09	-0.23	-0.15	-0.35	-0.35	--	--
Vision loss*	339	23.9	-0.10	-0.19	-0.18	-0.30	-0.31	-0.35	-0.07
Hearing loss*	633	44.6	-0.09	-0.16	-0.18	-0.27	-0.29	*-0.26*	-0.07
Migraines	79	5.6	-0.06	-0.13	-0.13	-0.26	-0.28	-0.32	--
Pain (chronic)	599	42.2	-0.10	-0.17	-0.20	-0.29	-0.32		
None of above conditions	130	9.0	-0.05	-0.05	-0.11	-0.11	-0.19	--	--

*score on visual or hearing function scale > 90
**based on 1987-92 NHIS age 18+, weights in italics were age-adjusted to the mean age of the Beaver Dam sample (Gold et al. 1992)

Figure 1: Percent distribution of Time-Tradeoff Ratings of current health (N = 1380)

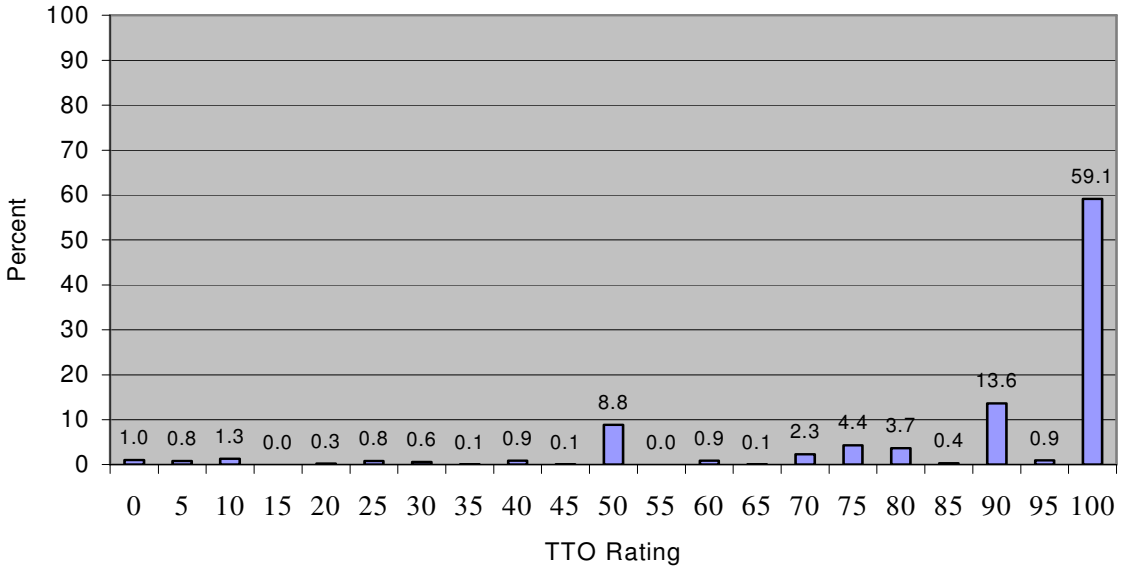
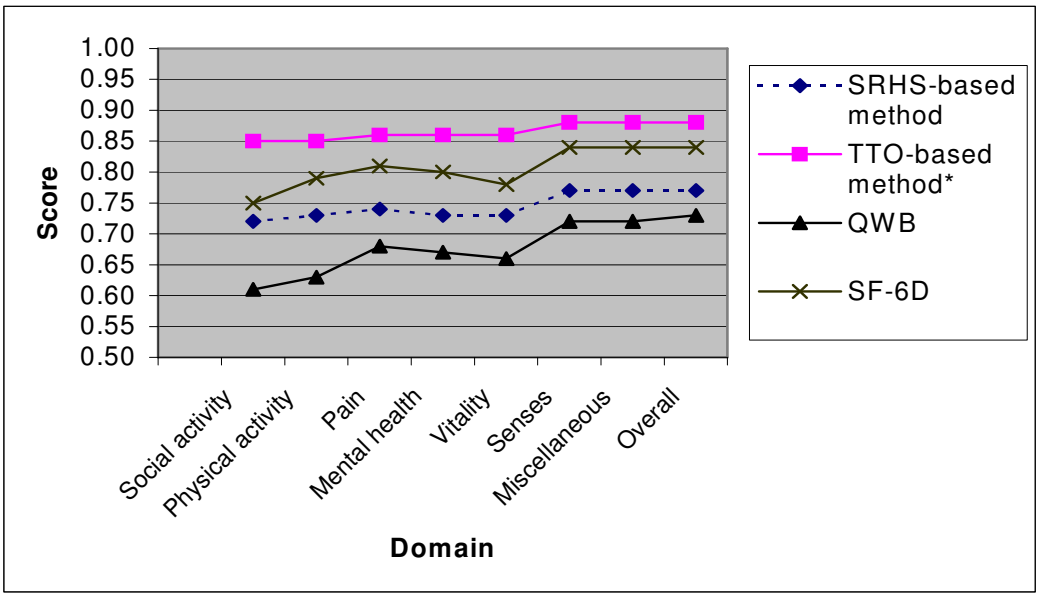


Figure 2: Distribution of responses to 5-point Self-Rated Health question (N = 1422)

Figure 3: Mean scores among those with problems in each domain and overall on our TTO and SRHS-based measures, the SF-6D, and the QWB



*N=1380 vs 1422 for other measures

References

Andersen, E. M., Rothenberg, B. M. and Kaplan, R. M., 1998. Performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire among older adults. Medical Care, 36(9), 1349-1360.

Brazier, J., Roberts, J. and Deverill, M., 2002. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics, 21(2), 271-292.

Brooks, R., Rabin, R.E. and de Charro, F. (Eds.), 2003. The measurement and valuation of health status using EQ-5D: a European perspective. Kluwer Academic Publishers, Dordrecht.

Cutler, D.M., 2005, Health. In: KG Abraham and C Mackie (Eds.), Designing Nonmarket Accounts for the United States. National Academies, Washington, DC, pp. 117-140.

Cutler, D.M. and Richardson, E., 1997. Measuring the health of the U.S. population. Bookings Papers on Economic Activity, Microeconomics, 217-272.

Dolan, P., 1997. Modeling valuations for EuroQol health states. Medical Care, 35(11), 1095-1108.

Dolan, P., 2005. Interpretations of utility and their implications for the valuation of health. Presentation at the Harvard/MIT/BU Health Economics Seminar.

Dolan, P. and Kahneman, D. (2005) Interpretations of utility and their implications for the valuation of health. Working paper, cited with permission.

Erickson, P., 1998. Evaluation of a population-based measure of quality of life: the Health and Activity Limitation Index (HALex). Quality of Life Research, 7(2), 101-114.

Essink-Bot, M. L., Krabbe, P.F., Bonsel, G.J. and Aaronson, N.K., 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. Medical Care, 35(5), 522-537.

Fryback, D.G., Dasbach, E.J., Klein, R., Klein, B.E., Dorn, N. et al., 1993. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. Medical Decision Making, 13(2), 89-102.

Fylkesnes, K. and Forde, O.H., 1991. The Tromso Study: predictors of self-evaluated health--has society adopted the expanded health concept? Social Science and Medicine, 32(2), 141-146.

Fylkesnes, K. and Forde, O. H., 1992. Determinants and dimensions involved in self-evaluation of health. Social Science and Medicine, 35(3), 271-279.

Gilbert, D.T., Pinel, E.C., Wilson, T.D., Blumberg, S.J. and Wheatley, T.P., 1998. Immune neglect: a source of durability bias in affective forecasting. Journal of Personality and Social Psychology, 75(3), 617-638.

Gold, M. R. and Muennig, P., 2002. Measure-dependent variation in burden of disease estimates: implications for policy. Medical Care, 40(3), 260-266.

Gold, M.R., Franks, P., McCoy, K.I. and Fryback, D.G., 1998. Toward consistency in cost-utility analyses: using national measures to create condition-specific values. Medical Care, 36(6), 778-792.

Gold, M.R., Siegel, J.E., Russell, L.B. and Weinstein, M.C., 1996, Cost-Effectiveness in Health and Medicine. Oxford University Press, New York.

Gureje, O., Simon, G. E. and Von Korff, M., 2001. A cross-national study of the course of persistent pain in primary care. Pain, 92(1-2), 195-200.

Hawthorne, G., Richardson, J. and Osborne, R., 1999. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health-related quality of life. Quality of Life Research, 8(3), 209-224.

Hogan, D.B., Maxwell, C.J., Fung, T.S. and Ebly, E.M., 2003. Prevalence and potential consequences of benzodiazepine use in senior citizens: results from the Canadian Study of Health and Aging. Can J Clin Pharmacol, 10(2), 72-77.

Holland, R., Smith, R. D., Harvey, I., Swift, L. and Lenaghan, E., 2004. Assessing quality of life in the elderly: a direct comparison of the EQ-5D and AQoL. Health Economics, 13(8), 793-805.

Horsman, J., Furlong, W., Feeny, D. and Torrance, G., 2003. The Health Utilities Index (HUI(R)): concepts, measurement properties and applications. Health and Quality of Life Outcomes, 1(1), 54.

Idler, E.L. and Benyamini, Y., 1997. Self-rated health and mortality: a review of twenty-seven community studies. Journal of Health and Social Behavior, 38(1), 21-37.

Idler, E.L. and Kasl, S. V., 1995. Self-ratings of health: do they also predict change in functional ability? Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 50(6), S344-353.

Idler, E.L., Russell, L.B. and Davis, D., 2000. Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study, 1992. First National Health and Nutrition Examination Survey. American Journal of Epidemiology, 152(9), 874-883.

Jenkinson, C., Gray, A., Doll, H., Lawrence, K., Keoghane, S. 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. Medical Care, 35(11), 1109-1118.

Johnson, J.A. and Pickard, A.S., 2000. Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. Medical Care, 38(1), 115-121.

Kaplan, R.M., Bush, J. W. and Berry, C.C., 1976. Health status: types of validity and the index of well-being. Health Services Research, 11(4), 478-507.

Kaplan, R.M., Ganiats, T.G., Sieber, W.J. and Anderson, J.P., 1998. The Quality of Well-Being Scale: critical similarities and differences with SF-36. International Journal for Quality in Health Care, 10(6), 509-520.

Kaplan, R.M. and Anderson, J.P., 1996, The general health policy model: an integrated approach. In: Spilker B (Ed.), Quality of life and pharmacoeconomics in clinical trials. Lippincott-Raven, Philadelphia, pp. 309-322.

Leinonen, R., Heikkinen, E. and Jylha, M., 1999. A path analysis model of self-rated health among older people. Aging, 11(4), 209-220.

Mansson, N.O., Merlo, J. and Ostergren, P.O., 2002. Is there an interaction between self-rated health and medication with analgesics and hypnotics in the prediction of disability pension? Scandinavian Journal of Public Health, 30(4), 267-273.

McDowell, I. and Newell, C., 1996, Measuring health : a guide to rating scales and questionnaires. Oxford University Press, New York.

Salomon, J.A., Tandon, A. and Murray, C.J., 2004. Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. BMJ, 328(7434), 258.

Shaw, J.W., Johnson, J.A. and Coons, S.J., 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. Annals of Medicine, 33(5), 328-336.

Torrance, G.W., Feeny, D.H., Furlong, W.J., Barr, R.D., Zhang, Y. et al., 1996. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. Medical Care, 34(7), 702-722.

Ubel, P.A., Loewenstein, G. and Jepson, C., 2003. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Quality of Life Research, 12(6), 599-607.

Ware, J.E. Jr., Kosinski, M. and Gandek, B., 2000, Sf-36 Health survey: Manual & Interpretation Guide. QualityMetric Incorporated, Lincoln, RI.

Wilcox, V.L., Kasl, S.V. and Idler, E.L., 1996. Self-rated health and physical disability in elderly survivors of a major medical event. Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 51(2), S96-104.

Wilson, T.D., Wheatley, T., Meyers, J.M., Gilbert, D.T. and Axsom, D., 2000. Focalism: a source of durability bias in affective forecasting. Journal of Personality and Social Psychology, 78(5), 821-836.

Appendix

Table 1: Proportion Reporting Items in Each Domain and Coefficients from OLS Regressions on Time-Tradeoff Ratings (N=1380) and Probit Regressions on Self-Rated Health Status (N= 1422)

Domains and symptoms/impairments from the Quality of Well-being Scale	Proportion reporting*	OLS coefficient for TTO	P value	Probit coefficient for SRHS	P value
Social Activity	0.17				
Limited or performed no major role activity	0.09	-0.15	<.0001	-0.56	<.0001
Limited in other role activity	0.11	-0.03	0.28	-0.24	0.07
Interaction term: limitations in major and other	0.04	0.16	0.00	0.09	0.70
Physical Activity	0.35				
1+ ADL limitations (dress/feed/toilet/bath)	0.03	0.01	0.88	-0.24	0.25
Most or all of day in chair, couch or bed	0.07	-0.09	0.05	-0.13	0.61
Trouble lifting/stooping/bending over/stairs	0.20	-0.01	0.67	-0.44	<.0001
Limp/cane/crutches/walker	0.13	-0.03	0.25	-0.21	0.19
Limited walking or other physical limitation	0.18	-0.07	0.01	-0.29	0.02
Do not drive and/or limited in public transport	0.10	-0.08	0.01	-0.53	0.00
Interaction terms:					
Chair or bed all day and lift/stoop/bend/stairs	0.05	0.02	0.76	-0.61	0.05
Chair or bed all day and limp/walking aid	0.04	0.05	0.43	0.62	0.04
Chair or bed all day and limited walking	0.05	-0.08	0.19	-0.54	0.08
Lift/stoop/bend/stairs and limp/walking aid	0.07	0.00	0.99	-0.16	0.47
Lift/stoop/bend/stairs and limited walking	0.10	-0.06	0.14	0.51	0.01
Limp/walking aid and limited walking	0.08	0.07	0.12	0.23	0.31
Drive/public transport and chair or bed all day	0.05	0.04	0.45	0.32	0.25
Drive/public transport and lift/stoop/bend	0.07	0.02	0.63	0.26	0.31
Drive/public transport and limp/walking aid	0.05	-0.01	0.86	0.27	0.32
Drive/public transport and limited walking	0.07	-0.02	0.70	-0.02	0.94
Pain	0.54				
Pain[a]	0.49	-0.02	0.17	-0.25	0.00
Rash (burning or itching rash on large area)	0.07	-0.01	0.74	0.01	0.97
Urinary/bowel/sex organ	0.08	0.00	0.98	-0.15	0.47
Interaction terms:					
Pain and urinary/bowel/sex organ	0.05	-0.08	0.09	-0.06	0.79
Pain and rash	0.04	-0.03	0.54	-0.40	0.11
Mental Health	0.22				
Depressive symptoms[b]	0.06	-0.02	0.53	-0.15	0.45
Excessive worry or anxiety	0.10	-0.05	0.06	-0.45	0.00
Trouble sleeping	0.14	0.00	0.85	-0.20	0.05
Interaction terms:					
Depressive symptoms and anxiety	0.03	0.02	0.71	0.59	0.03
Depressive symptoms and trouble sleeping	0.02	-0.01	0.92	-0.03	0.92
Anxiety and trouble sleeping	0.03	0.12	0.01	0.24	0.30

Domains and symptoms/impairments from the Quality of Well-being Scale	Proportion reporting*	OLS coefficient for TTO	P value	Probit coefficient for SRHS	P value
Vitality (general tiredness/weakness/weight loss)	0.19	-0.01	0.63	-0.24	0.01
Senses	0.96				
Eye pain or trouble seeing after correction[c]	0.20	-0.02	0.21	-0.12	0.13
Glasses or contact lenses	0.95	-0.01	0.69	-0.04	0.76
Speech problem[d]	0.05	-0.02	0.43	-0.13	0.34
Miscellaneous	0.95				
Gastrointestinal illness[e]	0.06	0.01	0.67	-0.22	0.12
Respiratory illness[f]	0.19	-0.04	0.02	-0.33	<.0001
Interaction: Gastrointestinal and respiratory	0.01	-0.01	0.86	0.99	0.00
Problems with sexual interest or performance	0.07	0.01	0.76	0.02	0.88
Ear/nose/teeth[g]	0.71	0.01	0.56	-0.17	0.01
Weight/appearance[h]	0.48	0.00	0.68	-0.07	0.27
Medication or prescribed diet	0.78	-0.05	0.00	-0.44	<.0001
Limbs/extremities[i]	0.04	0.00	0.88	-0.12	0.46
Headache[j]	0.20	0.00	0.81	-0.17	0.03
Trouble learning/remembering/thinking clearly	0.14	-0.03	0.07	-0.29	0.00
Other symptoms[k]	0.07	-0.04	0.11	0.05	0.64
Break points:					
Cut 1	--	--		1.43	<.0001
Cut 2	--	--		1.53	<.0001
Cut 3	--	--		1.33	<.0001
Intercept	0.98	<.0001		4.18	0.18
Summary statistics					
N	1380			1422	
R^2/ Ln (likelihood)	0.222			1571.9	

*includes anyone reporting one or more items in the domain; [a]Pain, stiffness, weakness, numbness or other discomfort in chest, stomach, side, neck, back, hips, joints of hands/feet/arms or legs; [b]Spells of feeling upset, being depressed, or of crying; [c]pain or discomfort in one or both eyes (such as burning or itching) or any trouble seeing after correction; [d]trouble talking, such as lisp, stuttering, hoarseness, or inability to speak; [e]sick or upset stomach, vomiting or loose bowel movements, with or without fever, chills, or aching all over; [f]cough, wheezing, or shortness of breath with or without fever, chills, or aching all over. [g]Pain in ear, tooth, jaw, throat, lips, tongue; missing or crooked permanent teeth—includes wearing bridges or false teeth; stuffy, runny nose; any trouble hearing—includes wearing a hearing aid; [h]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; [i]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; [j]Headache, or dizziness, or ringing in ears, or spells or feeling hot, or nervous, or shaky; [k]symptoms or problems that people felt were not captured on symptom list.