A preference-based measure of health: the VR-6D derived from the veterans RAND 12-Item Health Survey

Alfredo J. Selim · William Rogers · Shirley X. Qian · John Brazier · Lewis E. Kazis

Accepted: 1 February 2011

Abstract
Purpose The Veterans RAND 12-Item Health Survey (VR-12) is currently the major endpoint used in the Medicare managed care outcomes measure in the Healthcare Effectiveness Data and Information Set (HEDIS®), referred to as the Health Outcomes Survey (HOS). The purpose of this study is to adapt the Brazier SF-6D utility measure to the VR-12 to generate a single utility index.
Methods We used the HOS cohorts 2 and 3 for SF-36 data and 9 for VR-12 data. We calculated SF-6D scores from the SF-36 using the algorithms developed by Brazier and colleagues. The values of the Brazier SF-6D were used to estimate utility scores from the VR-12 using a mapping approach based on a 2-stage mapping procedure, named as VR-6D.
Results The VR-6D derived from the VR-12 has similar distributional properties as the SF-6D. The change in VR-6D showed significant variations across disease groups with different levels of morbidity and mortality.
Conclusions This study produced a utility measure for the VR-12 that is comparable to the SF-6D and responsive to change. The VR-6D can be used in evaluations of health care plans and cost-effectiveness analysis to compare the health gains that health care interventions can achieve.

Keywords Utility measures · SF-6D · Cost-effectiveness analysis · Health-related quality of life

Abbreviations

SF-36	Short form 36-Item Health Survey
VR-36	Veterans RAND 36-Item Health Survey
MRE	Modified regression estimate
VR-12	Veterans RAND 12-Item Health Survey
HEDIS	Healthcare Effectiveness Data And Information Set
MA	Medicare advantage
HOS	Health Outcomes Survey
MRE	Modified regression estimation
QALYs	Quality adjusted life years

SF-36® and SF-12® are registered trademarks of the Medical Outcomes Trust.

A. J. Selim · S. X. Qian · L. E. Kazis (✉)
The Center for the Assessment of Pharmaceutical Practices (CAPP), Department of Health Policy and Management, Boston University School of Public Health, 715 Albany Street, T3 W, Boston, MA 02118, USA
e-mail: lek@bu.edu

A. J. Selim · W. Rogers · S. X. Qian · L. E. Kazis
The Center for Health Quality, Outcomes, and Economic Research (CHQOER), A Health Services Research and Development Field Program, VA Medical Center, Bedford, MA, USA

A. J. Selim
The Section of General Internal Medicine and Emergency Services, Boston VA Health Care System, West Roxbury, MA, USA

A. J. Selim
The Boston University School of Medicine, Boston, MA, USA

W. Rogers
The Health Institute, New England Medical Center, Boston, MA, USA

J. Brazier
The School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

Published online: 19 February 2011

Introduction

There is a growing movement toward the use of cost-effectiveness analysis to inform health care decisions [1]. The US Panel on Cost-Effectiveness in Health and Medicine proposed a single preference-based score be used in cost-effectiveness analysis [2]. A Blue Ribbon Panel recommended that the Centers for Medicare & Medicaid Services (CMS) consider adopting a utility assessment in the Health Outcomes Survey (HOS) for use in cost-effectiveness analysis [3]. The HOS is the first Medicare managed care outcomes measure in the Healthcare Effectiveness Data and Information Set (HEDIS®) [4]. The cost-effectiveness information can help Medicare to target its health care resources more efficiently [5].

A preference-based score, referred to as a "health utility," reflects value judgments on health states [6]. Generally, these measures evaluate subjective preferences for multi-dimensional health measurement instruments on a scale of 0 to 1 (or 100), where 0.00 represents death and 1.00 optimum health, with negative values for states worse than being dead [7]. A number of different utility measures have been developed for use in cost-effectiveness analysis [8–10]. Among them is the Brazier SF-6D derived from the SF-36 [11]. The SF-36 contains 36 "items" that address 8 domains of health status and which generate scales that yield scale values for each and summary scores for physical and mental health. The SF-36 was modified into a six-dimensional health state classification called the SF-6D to determine a patient's preferences for (or "utilities" of) different health states. The six dimensions are physical functioning, role limitations, social functioning, pain, mental health, and vitality. These six dimensions each have between two and six levels. A matrix of 18,000 possible SF-6D health states was defined by selecting one level from each dimension, and 249 of these states were selected and evaluated by 611 general population respondents in a standard gamble format.

One consideration with the SF-6D is its dependence on the SF-36, which has undergone a number of changes. First, the range of possible responses to two questions was expanded in order to reduce "floor effects" and to improve performance of the SF-36 in patient populations with poor health. This resulted in the Veterans RAND 36-Item Health Survey (VR-36) [12]. Later, both the SF-36 and the VR-36 were reduced from 36 to 12 items with minimal loss of information to form the SF-12 and the Veterans RAND 12-Item Health Survey (VR-12), respectively [13, 14]. These shorter questionnaires were found to perform well relative to their progenitors, but it is not clear whether they can be used to generate a preference-based index from the SF-6D.

The VR-12 is a widely used health survey in population-based studies such as the Medicare HOS [15]. There is currently an algorithm for converting the SF-12 into a shortened version of the SF-6D [16]. This could be adapted for use in developing a preference-based index from the VR-12, but for a number of reasons, this is not an optimal solution. The VR-12 differs from the SF-36 version 1 in the use of 5-point response choices for the role limitations due to physical problems and the role limitations due to emotional problems [17]. The SF-12 version of the SF-6D includes only 7 out of the 11 SF-36 items from the SF-36, and this significantly reduces the range of illness space it covers. While initial applications of the SF-12 version of SF-6D suggested this might not be important in impacting on the scores obtained, more recent evidence suggests that the scores are different and the SF-12 version of the SF-6D suffers from floor effects [18]. For this reason, a novel multi-stage mapping process was undertaken to map the VR-12 onto the SF-36 version of the SF-6D to provide accurate estimates.

The goal of this study is to use the Brazier SF-6D as the basis for developing a new utility measure, the VR-6D, which is based upon the VR-12. The specific objectives were as follows: (1) to develop an algorithm to compute utilities for the VR-12, (2) to examine the distributional properties of the utility scores from the VR-12, (3) to examine the concurrent validity of the utility scores from the VR-12 algorithm by using the SF-6D scores from the SF-36 as the standard, and (4) to evaluate the responsiveness to change of the utility scores from the VR-12 across 5 patient clinical condition groups (diabetes, hypertension, coronary artery disease (CAD), chronic obstructive lung disease (COPD)/asthma, and stroke).

Methods

Study population

We used VR-12 data collected by the Medicare HOS. The goal of the Medicare HOS program is to gather valid and reliable health status data in Medicare managed care for use in quality improvement activities, plan accountability, and public reporting. All managed care plans with Medicare Advantage (MA) contracts must participate. Since 1998 to the present, every year a random sample of Medicare beneficiaries, who were continuously enrolled for a six-month period, is drawn from each participating MA plans and surveyed every spring. Two years later, these same respondents are surveyed again.

The HOS Performance Measurement analysis is limited to beneficiaries who meet the following analytic criteria:

1. Were age 65 and older at the time of completing the baseline survey.

2. Had enough data to score baseline PCS or MCS scores using a previously validated Modified Regression Estimation (MRE) algorithm for missing data as well as adjustments for contextual issues such as mode of administration (phone versus mail-out) [19].
3. Were members of a health plan at baseline that remained in HOS at follow-up.

We used the Medicare HOS cohort 9 (April 2006–May 2008) since it provides a dataset with the VR-12 scores at baseline and at 2 years of follow-up (see Fig. 1). The sample size was 188,515 MA patients, including both the aged and the disabled. The response rate was 66.8%. Among the 188,515 beneficiaries, 103,661 met all 4-study criteria. By the time of the two-year follow-up, 7,820 died and 17,916 were not resurveyed because they were no longer enrolled in the plan in 2008 (voluntary disenrollment) or some plans no longer offered managed care to Medicare beneficiaries (involuntary disenrollment). There were 77,925 eligible to the HOS follow-up HOS and 62,875 completed surveys.

We also used the Medicare HOS cohorts 2 (March 1999–June 2001) and 3 (April 2000–May 2002) for the analysis as they provided SF-36 data at baseline and 2-year follow-up. Among the 301,184 beneficiaries in the Medicare HOS cohort 2, 194,374 completed the surveys, 124,835 were in the analytic sample, 11,946 died, and 68,742 responded to the follow-up HOS survey. Among the 298,883 beneficiaries in the HOS cohort 3, 208,655 completed the surveys, 122,317 were in the analytic sample, 13,099 died, and 59,578 responded to the follow-up HOS survey.

Health-related quality of life measures

The VR-12 is a modification of the RAND 36-Item Health Surveys 1.0 that was developed at the RAND Corporation as part of the Medical Outcomes Study [20]. It differs from the SF-12 in the use of 5-point response choices for the role limitations due to physical problems and the role limitations due to emotional problems [21]. This has resulted in a reduction to floor and ceiling effects of the scales with important gains to their distributional properties and increases to the reliability and validity of the assessments [22]. We summarized the 12 items into physical (PCS) and mental component (MCS) scales by applying a linear t-score transformation with a mean of 50 and a standard deviation of 10 based upon the 1990 US population norms and using the MRE algorithm for missing data and adjustments for contextual issues (phone versus mail-out administration) [23, 24].

The SF-6D is a classification for describing health states derived from a selection of SF-36 items [11]. It is composed of six multi-level dimensions of health, including physical functioning, role limitations due to physical and emotional problems, social functioning, pain, mental health, and vitality. Each dimension has two to six levels, and thus, 18,000 possible health states can be defined using the SF-6D descriptive system. The SF-6D measure scores on a 0.29 to 1.00 scale, with 1.00 indicating "full health."

Analytic plan

Building upon the SF-6D, we proposed a multi-step mapping approach for estimating a preference-based single

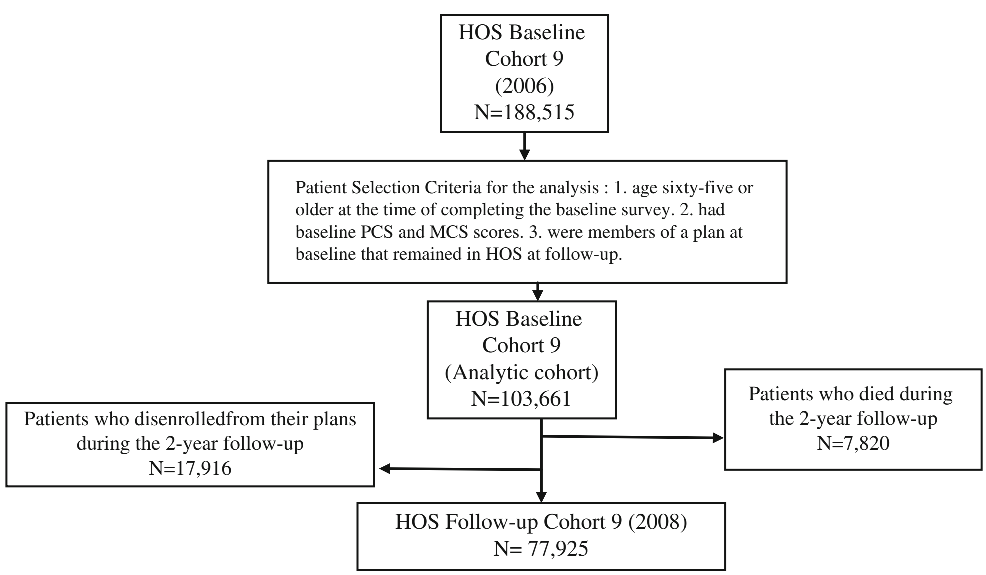


Fig. 1 Medicare Health Outcomes Survey (HOS) analytic cohort 9

index measure for health from the VR-12 (objective 1). The first step ("the linearization of the SF-6D") involved the calculation of the 8 scales of the SF-36 where the health descriptions were more complete, the regression of the SF-6D on the 8 scales of the SF-36, and the examination of the predicted values. The second step ("transformation of the predictive values") took in the predictive values to be calibrated to track the SF-6D more accurately than the linear model allowed. The third step ("adjustments for other methodological issues") included corrections for missing values and the VR-12 contextual effects.

There were four important methodological issues. First, the VR-12 differs from the SF-36 in the use of 5-point response choices for the role limitations due to physical problems and the role limitations due to emotional problems when the SF-36 version 1 uses dichotomous choices (Yes/No). Second, there are contextual differences in the order and presentation of the questions between the "non-native" VR-12 format (the 12 items embedded in a 36-item questionnaire) and the "native" VR-12 format (the 12 items embedded in a 12-item questionnaire). Third, the mode of administration (mail-out vs. phone) affects the VR-12 scores. The values of the mail-out VR-12 were systematically lower than those from the phone VR-12. For example, the pf02 ("moderate activities") score for a mailed VR-12 should have 1.68 subtracted from it, so if a person scores "50" (limited a little) then we would use 48.32. Fourth, one or more survey responses can be missing. To help in addressing such issues, we used the MRE [19].

The MRE algorithm was originally developed for missing data as well as adjustments for contextual issues such as mode of administration (phone versus mail-out). It was created as a linear function of the items (a regression estimate) modified to reduce the unwanted regression to the mean effects when scored with missing data. When all items are present, the MRE and the original measure are virtually identical, but when there are missing data, the MRE is able to estimate the scores with a reduced set of items. In principle, the MRE could be used to score a 12-item subset of a 36-item questionnaire (VR-36), referred to as the non-native VR-12. If the 12 items (each on a 0–100 scale) are arranged in a vector X, and the modified regression coefficients b are also made into a vector, the MRE takes the following form:

$$X_1b_1 + X_2b_2 + \cdots + X_{12}b_{12} + C$$

However, we could not simply apply the MRE computed in the larger 36-item to the "native" VR-12 given the contextual differences between the "non-native" and "native" VR-12 formats. In alternate forms and modes, the coefficients remain the same and the adjustments for contextual effects become a part of the constant term (C) in the MRE.

This approach was based on a previously validated method that is explained in our published paper on PCS and MCS norms [25].

Operationally, we used the following four steps (Fig. 2):

1. Linearization of the SF-6D: We computed the 8 scales of the SF-36 and then regressed the Brazier SF-6D from the SF-36 on those SF-36 scale values. The data consisted of 124,835 participants (analytic sample) from the Medicare HOS cohort 2. The following table shows the results of regressing SF-6D onto 8 scales (dimension scores) of the SF-36 based on a specification used previously by Ara and Brazier [26].

SF-36 scale	Coefficients
Physical functioning (PF)	0.082
Role physical (RP)	0.047
Bodily pain (BP)	0.127
General health (GH)	0.004
Vitality (VT)	0.072
Social functioning (SF)	0.091
Role emotional (RE)	0.019
Mental health (MH)	0.155
Constant	30.005
R^2	0.885

Using the above regression coefficients, we calculated the SF-6D from the VR-36. There are validated conversion formulas that allow for comparisons of VR-36 scores to those from the SF-36 [27]. We used VR-36 data from the 1999 Large Health Survey of Veteran Enrollees (LHSVE) [28].

2. Transformation of the predictive values: We examined the relationship between the predicted SF-6D values and the actual SF-6D scores and identified some curvature and established the low and high end of the scale (see Appendix 1). To overcome this problem, we fitted a smooth non-linear function with square and cubic adjustments at the lower end (0.42) and a spline at the higher end (0.86). The non-linear function T was smoothed from the actual item-dependent function because of the sufficiency of the SF-36 and Jensen's inequality [29]. The result was a transformed prediction that tracks the observed mean SF-6D. With variables defined as follows:

utility prediction (p) from step 1 and sf6dhat:
p2 = p^2;
p3 = p^3;

Step 1:
Linearization of the SF-6D

We regressed the Brazier's SF-6D linearly on the SF -36 scales and the coefficients were used to compute the SF-6D from the VR-36.

Step 2:
Transformation of the predictive values

We fitted a smooth non-linear function of the SF-6D with square and cubic adjustments at the lower end (.42) and a spline at the higher end (.86). With variables defined as follows:
utility prediction (p) from step 1 and sf6dhat:
p2=p^2;
p3=p^3;
p4=Max(p-86, 0);
p42=p4^2;
p5=Min(p-42, 0);
and a constraint in place that the maximum predicted value should yield a utility of 1, we calculated the transformation (T) coefficients.

Step 3:
Adjustments for missing data and VR-12 contextual effects

We used the MRE to regress the SF-6D from the VR-36 on the 12-item subset of the VR-36 (non-native VR-12). The MRE algorithm was adjusted for the (native) VR-12 contextual effects including context and mode of administration. This resulted in the utility predictions from the VR-12.

Step 4:
Calculation of the VR-6D

We used the T coefficients from step 2 and utility prediction (p) from step 3 were used in the following formula to estimate the VR-6D scores:
VR-6D = p*4.001629 + p2*-.0571654 + p3*.000337 + p4*1.895635 + p42*.1869775 + p5*-.0518862 + (-47.5366)

SF-36 = Short Form 36-item Health Survey
VR-36 = Veterans RAND 36-item Health Survey
MRE = Modified Regression Estimate
VR-12 = Veterans RAND 12-item Health Survey

Fig. 2 Development of the VR-6D

p4 = Max(p-86, 0);
p42 = p4^2;
p5 = Min(p-42, 0);
and a constraint in place that the maximum predicted value should yield a utility of 1, we calculated the transformation (T) coefficients.

3. Adjustments for missing data and VR-12 contextual effects: We used the MRE algorithm to regress predicted utilities for the VR-36 on the 12-item subset embedded in the VR-36 (non-native VR-12) to account for missing data. We utilized cases with complete data from the 1999 LHSVE to simulate each potential missing data pattern. We estimated a different regression model for each pattern of missing data for the 12 items, resulting in 4,095 sets of coefficients (after eliminating the situation in which all 12 items had missing data). Each set of these "original MRE coefficients" consisted of one coefficient for each non-missing item, plus a constant. We eliminated from the analysis those equations with R^2 less than 0.5 because they were deemed not to have sufficient information to estimate a coefficient pattern. The resulting equations were assembled into a dataset of regression coefficients with a row for each regression equation and 13 columns (one for each of the 12 items and the constant term). The derived coefficients can be used to estimate SF-6D scores from the "non-native" VR-12.

To account for the contextual differences in the order and presentation of the questions between the "non-native" and "native" VR-12 and the mode of administration (mail-out vs. phone), we made small adjustments in the constant term in the MRE. The following table shows the contextual effects on the 8 items that are in common between the VR-12 and the SF-36/SF-12.

Item	Mail-out	Phone	Mail-out*phone
General health	0.85	0.29	−0.03
Moderate activities	1.68	6.02	−2.50
Climb several stairs	2.68	5.80	−0.18
Pain interfere	−1.38	5.65	−0.42
Peaceful	5.97	5.54	−3.83
Energy	5.78	4.89	−2.39
Blue/sad	−2.66	1.33	−1.37
Social time	0.66	2.99	−1.38

Each item was mapped on a 0–100 point scale. The model included age, Supplemental Security Income (SSI), gender, race, education, health plans, and indicator variables for the VR-12. Building upon this table, we determined the proper adjustment for any subsets of these 8 VR-12 items.

4. Calculation of the VR-6D: We used the *T* coefficients calculated in step 2 and the utility prediction (p) from the MRE algorithm in step 3 to estimate the VR-6D scores:

$$\begin{aligned}\text{VR - 6D} = {} & \text{p} * 4.001 + \text{p2} * -.057 + \text{p3} * .001 + \text{p4} \\ & * 1.895 + \text{p42} * .186 + \text{p5} * -.0518 \\ & + (-47.536) \leftarrow\end{aligned}$$

To describe the distributional properties of the resulting VR-6D scores from the VR-12 (objective 2), we calculated the mean, median, mode, and range.

We compared the VR-6D scores from the VR-12 algorithm with the SF-6D scores from the SF-12 and SF-36 algorithms. The rationale for this objective 3 was to examine the concurrent validity of the utility scores from the VR-12 algorithm. The SF-6D scores served as the "standard." First, we calculated the baseline, the follow-up at 2 years, and the change in utility scores of persons who are alive (A). Second, we calculated the percentage that died at 2 years (D) and the change in utility scores of people who died (L). We assigned a value of 0 to the follow-up utility score for those that died. Third, we calculated the total change in utility scores (C) at 2 years that combines the utility scores of persons who are alive (A) with those who died (L). We used the following algorithm to calculate the total change: $C = D*L + (1 - D)*A$.

To evaluate the responsiveness of the VR-6D (objective 4), we examined the change in VR-6D across 5 patient groups from the Medicare HOS cohort 9. These groups were defined by using the following 5 self-reported conditions: diabetes, hypertension, coronary artery disease, stroke, and chronic obstructive lung disease/asthma. These conditions were selected because they are chronic conditions with different levels of morbidity and mortality. For this analysis of responsiveness to change, we used new cases. We defined "new cases" as those patients who reported the condition only at follow-up survey. Responsiveness was assessed by calculating the standardized response mean (SRM). We applied the following formula: SRM = mean change divided by the standard deviation of the change. The expectation was to find larger SRM values among conditions with a higher level of morbidity and mortality such as coronary artery disease, chronic obstructive lung disease/asthma, and stroke. We also evaluated the importance of mortality in the algorithm since mortality rates vary across cohorts with different conditions. For this analysis, we used established cases. We defined "established cases" as those patients who reported the condition only at the baseline survey. We combined the percentage of patients who died at 2 years (D), the change in VR-6D scores of people who died (L), and the change in VR-6 scores of those who are alive (A) to calculate the total change in VR-6D ($C = D*L + (1 - D)*A$).

Results

Table 1 shows the sociodemographic characteristics of the analytic sample from the Medicare HOS cohort 9 ($N = 103{,}661$). The study participants had a mean age of 75.4 years (SD ± 6) with representation of 40.8% men, 85.7% whites, 55.2% married, and 27.3% less than a high school education. Forty-three percent of the participants had income of less than $20,000, 72.3% owned their homes, and 6.7% were medicaid eligible. The average number of comorbid conditions was 2.8 (SD ± 2). Hypertension had the highest prevalence (64.2%) followed by arthritis of the hip (41.7%) and hand (36.9%), other heart conditions (22.6%), diabetes (22.3%), and sciatica (21.3%). A small proportion of the population (<3%) was under treatment for breast, prostate, lung, or colon cancer. The mean baseline PCS and MCS scores were 39.3 (SD ± 12) and 51.8 (SD ± 10), respectively. The PCS scores for our sample are substantially lower by close to one standard deviation below the norm of the general US population that has a mean score of 50. MCS is slightly higher by about 20% of one standard deviation above the norm for the general US population.

Table 1 Patient characteristics in the Medicare Health Outcome Survey (HOS) cohort 9

	$N = 103{,}661$
Age, years (Mean ± SD)	75.4 (±6)
65–74	48.2%
75–84	39.5%
85+	12.1%
Gender–male	40.8%
Race–whites	85.7%
African Americans	9.0%
Hispanics	1.7%
Others	3.6%
Married	55.2%
Education < 12 years	27.3%
Income < $20,000	43.1%
Home Owner	72.3%
Medicaid eligible	6.7%
Number of comorbidities (Mean ± SD)	2.8 (+2)
Comorbidity 0	10.0%
1	18.4%
2	20.5%
3	18.4%
>=4	32.7%
Hypertension	64.2%
Diabetes	22.3%
Angina/coronary artery disease	15.4%
Acute myocardial infarction	11.4%
Congestive heart failure	9.1%
Other heart conditions	22.6%
Stroke	9.1%
Asthma/COPD	14.0%
Gastrointestinal disorders	4.8%
Arthritis of hip or knee	41.7%
Arthritis of hand or wrist	36.9%
Sciatica	21.3%
Cancer other than skin cancer	15.5%
Treatment for colon/rectal cancer	1.1%
Treatment for lung cancer	0.6%
Treatment for breast cancer	2.1%
Treatment for prostate cancer	2.9%
Baseline Physical Health (Mean ± SD)[a]	39.3 (+12)
Baseline Mental Health (Mean ± SD)[a]	51.8 (+10)

[a] The scores were normed to 50.0 based upon a US population where higher scores denote better health

Table 2 gives distributional properties of the VR-6D scores using the Medicare HOS cohort 9. The mean VR-6D score was 0.694 (SD ± 0.129). The median and mode were 0.697 and 0.875, respectively. The VR-6D scores ranged from 0.264 to 1.094. Out of 103,654, only 15 subjects had scores slightly above 1. However, these scores are unlikely to affect the overall results. The rate of missing VR-6D values was only 0.01% ($N = 7$ patients) using the MRE algorithm.

Table 2 Distributional properties of the VR-6D scores (Medicare Health Outcomes Survey cohort 9)

N	103,654		
Mean	0.694	Std deviation	0.129
Median	0.697	Mode	0.875
Skewness	−0.174	Range (min–max)	0.264–1.094
Uncorrected SS	51629.41		
Coefficient Variation	18.54		
Variance	0.016		
Kurtosis	−0.505		
Corrected SS	1714.80		
Standard error mean	0.001		
Quantile	Estimate		
100% Max	1.094		
99%	0.954		
95%	0.880		
90%	0.860		
75% Q3	0.793		
50% Median	0.697		
25% Q1	0.600		
10%	0.525		
5%	0.481		
1%	0.394		
0% Min	0.264		
Missing	N	%	
Values	7	0.01	

Table 3 shows the SF-6D and VR-6D scores after applying the SF-36 and SF-12 algorithms to the Medicare HOS cohorts 2 and 3 and the VR-12 algorithm to the Medicare HOS cohort 9. The SF-36 and VR-12 algorithms generated similar baseline, follow-up, and change in utility scores. The SF-12 algorithm produced higher utility scores when compared with the other two algorithms. The total changes in utility scores (the combination of the change in utility scores of people who are alive and those who died) were a little larger than the mean change in utility scores of persons who are alive. The magnitude of the change decreased across the cohorts independently of the applied algorithm.

Table 4 provides the VR-6D scores by disease groups. New cases of diabetes and hypertension had the lowest change in VR-6D (−0.024 (SD ± 0.103) and −0.029 (SD ± 0.101), respectively). They were followed by new cases of coronary artery disease and COPD/asthma (−0.035 (SD ± 0.102) and −0.035 (SD ± 0.100), respectively). New cases of stroke had the highest VR-6D score change (−0.048 (SD ± 0.112)). The larger SRM values among conditions with high levels of morbidity and mortality such

Table 3 Utility scores after applying the SF-36, SF-12, and VR-12 algorithms

Health Outcomes Survey Cohort	Algorithm	Baseline Utility Scores			Follow-up Utility Scores			Change in Utility Scores of those who are alive			Change in Utility scores of those who died			Total Change in Utility Scores[a]
		N	Mean	STD	*N*	Mean	STD	*N*	Mean	STD	*N*	Mean	STD	Mean
Cohort 2	SF-6D SF-36 algorithm	93,804	0.706	0.135	52,443	0.700	0.132	49,162	−0.019	0.106	6,474	−0.610	0.140	−0.060
	SF-6D SF-12 algorithm	95,615	0.756	0.151	53,871	0.751	0.151	51,382	−0.022	0.124	6,548	−0.645	0.154	−0.065
Cohort 3	SF-6D SF-36 algorithm	112,477	0.703	0.136	55,601	0.698	0.131	51,915	−0.017	0.106	8,162	−0.606	0.139	−0.060
	SF-6D SF-12 algorithm	114,916	0.751	0.152	56,560	0.747	0.151	53,758	−0.019	0.122	8,299	−0.642	0.154	−0.065
Cohort 9	VR-6D VR-12 algorithm	103,654	0.694	0.129	63,030	0.693	0.126	63,030	−0.015	0.094	7,819	−0.599	0.128	−0.059

[a] Total change in utility scores combines the utility scores of persons who are alive with those who died at 2 years

as CAD, CODP/asthma, and stroke support the ability of the algorithm to detect change. Among established cases, the total changes in VR-6D, a combination of the VR-6D scores of patients who are alive with those who died, were the largest for COPD/asthma and stroke (−0.081 (SE ± 0.0021) and −0.091 (SE ± 0.0028), respectively) and the smallest for hypertension and diabetes (−0.058 (SE ± 0.0008) and −0.062 (SE ± 0.0015), respectively), indicating that morbidity and mortality contribute in an important way to the measure. As can be seen by the standard errors reported in Table 4, the changes were highly significant statistically.

Discussion

This study developed an algorithm for computing a preference-based single index, named the VR-6D, from the VR-12. The transformation coefficients have enhanced the VR-6D items making them comparable with the SF-6D derived from the SF-36. The resulting VR-6D scores are responsive to change. The VR-6D offers a suitable tool to measure and summarize the health of populations that can be used in cost-effectiveness analysis.

The VR-6D from the VR-12 has similar distributional properties as the SF-6D from the SF-36. The VR-6D preference-based measure can be regarded as a continuous outcome scored on a theoretical range from 0.00 to 1.00, with 1.00 indicating "full health". Estimating the VR-12 preference-based index avoids the problems identified with the SF-12 version of the SF-6D health state classification. The VR-6D scores from the VR-12 were similar to the SF-6D scores from the Brazier algorithm. This is evidence of concurrent validity. The Medicare HOS can calculate comparable utility scores by applying the appropriate utility algorithms based upon the SF-36 and the VR-12 across HOS cohorts to evaluate and compare health care plans.

The change in VR-6D scores showed significant variations across disease groups. This is evidence of differentiating responsiveness to change by specific condition subgroups. These findings are consistent with the literature [30]. While the differences might be small in absolute terms, they are entirely consistent with the types of changes you would expect in an observational sample compared to say the context of a placebo controlled trial, such as for example Harrison et al. [31]. They found that the mean SF-6D change in patients with early and more severe arthritis was −0.02 (SD ± 0.11) and −0.04 (SD ± 0.10), respectively. It is important to note that we included death in the utility metric of change. Mortality is an important component in the algorithm as mortality rates vary across cohorts. The value of being dead is

Table 4 The VR-6D scores by disease groups (Medicare Health Outcomes Survey cohort 9)

Disease	Sample Size	Baseline VR-6D Scores	Follow-up VR-6D Scores	2 years Mortality Rates	Change in VR-6D of those who died at 2 years	SD	Change in VR-6D of those who are alive at 2 years	SD	SRM	Total Change in VR-6D[a]	SE
New cases of hypertension	4,481	0.723	0.694	–	–	–	−0.029	0.101	−0.287	−0.029	0.0015
Established cases of hypertension	64,585	0.692	0.679	7.7%	−0.592	0.122	−0.013	0.093	−0.139	−0.058	0.0008
New cases of diabetes	2,818	0.692	0.668	–	–	–	−0.024	0.103	−0.233	−0.024	0.0019
Established cases of diabetes	22,391	0.669	0.656	8.8%	−0.578	0.119	−0.012	0.094	−0.127	−0.062	0.0015
New cases of CAD	2,944	0.672	0.637	–	–	–	−0.035	0.102	−0.343	−0.035	0.0018
Established cases of CAD	15,308	0.653	0.645	11.7%	−0.565	0.114	−0.008	0.091	−0.087	−0.073	0.0019
New cases of COPD/asthma	2,533	0.671	0.636	–	–	–	−0.035	0.100	−0.350	−0.035	0.0020
Established cases of COPD/asthma	13,987	0.650	0.638	12.5%	−0.562	0.112	−0.011	0.091	−0.120	−0.081	0.0021
New cases of stroke	1,698	0.669	0.621	–	–	–	−0.048	0.112	−0.428	−0.048	0.0027
Established cases of stroke	9,080	0.632	0.622	15.0%	−0.553	0.115	−0.010	0.096	−0.104	−0.091	0.0028

CAD Coronary artery disease, *COPD* Chronic obstructive pulmonary disease, *SD* Standard deviation, *SRM* Standardized response mean, *SE* Standard error

[a] Total change in VR-6D combines the scores of patients who are alive with those who died at 2 years

determined by the method used to obtain values for the VR-6D health states. The SG valuation task uses full health and dead as anchors and these are given values of one and zero, respectively. Given the axioms of expected utility theory, the values for the health states are on a cardinal scale with zero for the lower anchor of dead and so it is entirely legitimate to assign dead states zero in the calculation presented. However, some readers may only be interested in the properties of the VR-6D scale in those who stay alive, and this is a component of the overall measures of change given by the following equation $C = D*L + (1 - D)*A$; where C = total change in VR-6D scores, D = death, L = the change in VR-6D scores of people who died, and A = the change in VR-6D scores of persons who are alive.

A related question is whether the primary focus of measuring and reporting utility scores should be for public accountability of those managed care plans serving Medicare beneficiaries. The VR-6D offers several advantages. First, the VR-6D provides comparable utility scores to the SF-6D scores from the SF-36. This offers the Medicare HOS with the possibility of having comparable utility results across all Medicare HOS cohorts, given that the SF-36 was used in earlier Medicare HOS cohorts and the VR-12 in later ones. Second, the VR-6D has the potential to complement the current Medicare HOS outcome measures of change in physical (PCS) and mental (MCS) for use in plan accountability and public reporting. Third, the VR-6D allows the analyst to calculate quality adjusted life years (QALYs) from the VR-12 [32, 33]. Today, QALYs are used in most economic evaluations, and by many regulatory agencies, which have made cost-effectiveness analysis an integral part of their decision-making processes for purposes of resource allocation decisions [34].

The VR-6D preference-based index might suffer from a floor effect as the other SF-6D versions [35, 36]. This has been shown in comparison with the SF-6D from the SF-36 with the EQ-5D and HUI3 [37]. As Brazier et al. stated, this has been made slightly worse for the SF-12 index because the model overpredicts the poorer states. Using a Bayesian approach, the SF-6D values went down to 0.21 [38]. Future research may focus on the application of the Bayesian approach on the VR-6D.

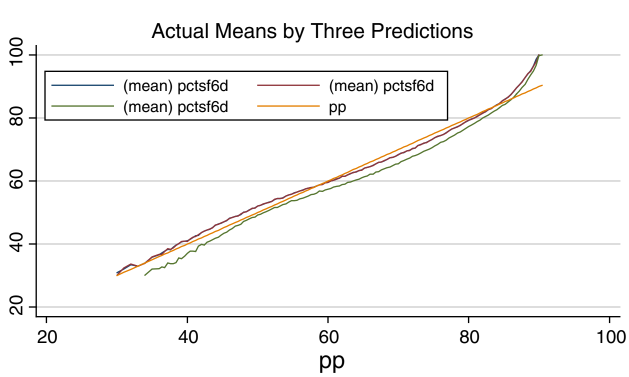
Ultimately, the source of values is a value judgment. Given the context of the Medicare system, it seems entirely appropriate to use the values of those affected, namely the Medicare population. This is not a patient sample per se (e.g. stroke patients valuing stroke states), but rather a general sample of Medicare population valuing general health states that is consistent with the recommendations of the Washington Panel. The values of the elderly Medicare populations are not necessarily higher, and indeed for the SF-6D health state, values were shown to follow an inverted U-shaped distribution with the minimum at 60–65 years of age and then values falling rapidly after 75 years of age [39]. Development of a set of preference-based weights using a representative sample from the Medicare population will provide a basis for future implementation of a utility metric in the HOS cohorts with potential for widespread application in Medicare Fee for Service.

In summary, this study produced a utility measure for the VR-12 that is comparable with the SF-6D from the SF-36 and is responsive to change. The VR-6D is germane to health care systems, both public and private, such as managed care organizations, which are increasingly adopting results from cost-effectiveness analysis as one of the measures to inform decisions on allocation of health care resources. We plan to make publicly available the algorithm to calculate the VR-6D through the website address http://www.va.gov/chqoer.

Acknowledgments The research in this article was supported in part by the Centers for Medicare and Medicaid Services and the National Committee for Quality Assurance (NCQA) under contract number HHSM-500-2009-00051C; the Center for the Assessment of Pharmaceutical Practices (CAPP) Boston University School of Public Health, Boston MA and the Center for Health Quality, Outcomes and Economic Research (CHQOER), Veterans Administration Medical Center, Bedford MA.

Appendix 1

In the figure below, the *Y*-axis is the actual value. The *X*-axis is the predicted value. The red curve is the regression of the SF-6D on 7 SF-36 scales (without General Health scale). The blue curve is the regression of the SF-6D on 8 SF-36 scales (with General Health). The dark green curve is the Ara and Brazier SF-6D [23]. The orange line is the line from (0,0) to (1,1). For each type of prediction, the predicted values were lumped into buckets 0.4% wide and the actual SF-6D values are averaged within them. The lines represent the non-linear regression lines organized by predicted values. The values by the original SF-6D algorithm and the regression method tracked each other in the mean utility up to about .8, and from there, the Brazier SF-6D goes up to 1 at roughly twice the rate of the linear version. The curvature in the values reflected the non-linearity of the SF-6D.



We estimated the transformation (T) coefficients using variables defined as follows utility prediction (p) estimated from step 1 and sf6dhat:

p2 = p^2;
p3 = p^3;
p4 = Max(p-86, 0);
p42 = p4^2;
p5 = Min(p-42, 0);

and a constraint in place that the maximum predicted value should yield a utility of 1.

sf6dhat	*T* Coef.	SE	*t*	*P* > |*t*| [95% Conf. Interval]
p	4.002	.022	180.06	0.000 3.958 4.045
p2	−.0571	.001	−161.00	0.000 −.0579 −.056
p3	.001	1.850	181.93	0.000 .000 .001
p4	1.895	.042	45.33	0.000 1.813 1.977
p42	.186	.016	11.73	0.000 .156 .218
p5	−.051	.005	−10.06	0.000 −.062 −.042
cons	−47.536	.453	−104.78	0.000 −48.425 −46.647

References

1. Eichler, H. G., Kong, S. X., Gerth, W. C., Mavros, P., & Jönsson, B. (2004). Use of cost-effectiveness analysis in health-care resource allocation decision-making: How are cost-effectiveness thresholds expected to emerge? *Value Health, 7*, 518–528.
2. Gold, M. R., Russell, L. B., & Weinstein, M. C. (Eds.). (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.
3. Kazis, L. (Chair of committee report) (in alphabetical order) Bierman, A., Diehr, P., Hays, R., Holzman, J., Iqbal, S.U., Kaplan, R., Lawrence, W., Rogers, W., Scholle, S.H., Selim, A., Shapiro, G., Smith, K. A White Paper report commissioned by CMS with Recommendations for the Medicare Health Outcomes Survey. Submitted to the Center for Medicare and Medicaid Services, Baltimore Maryland, October 26, 2007.
4. HEDIS®2003 Volume 6. February 2003. *Specifications for the medicare health outcomes survey*. Washington, DC: National Committee for Quality Assurance.
5. Neumann, P. J., Rosen, A. B., & Weinstein, M. C. (2005). Medicare and cost-effectiveness analysis. *New England Journal of Medicine, 353*, 1516–1522.
6. Weinstein, M. C., Siegel, J. E., Gold, M. R., Kamlet, M. S., & Russell, L. B. (1996). Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA, 276*, 1253–1258.

7. Coons, S. J., Rao, S., Keininger, D. L., et al. (2000). A comparative review of generic quality-of-life instruments. *Pharmacoeconomics, 17*(1), 13–35.
8. Andresen, E. M., Rothenberg, B. M., & Kaplan, R. M. (1998). Performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire among older adults. *Medical Care, 36*(9), 1349–1360.
9. Brooks, R., Rabin, R., & de Charro, F. (2003). *The measurement and valuation of health status using EQ-5D: A European Perspective*. Dordrecht, The Netherlands: Kluwer.
10. Feeny, D., Toorance, G., Furlong, W., et al. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care, 40*, 113–128.
11. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics, 21*, 271–292.
12. Hays, R. D., Sherbourne, C. D., & Mazel, R. M. (1993). The RAND 36-Item Health Survey 1.0. *Health Economics, 2*, 217–227.
13. Ware, J. E., Jr., Kosinski, M., & Keller, S. D. (1996). A 12 Item Short Form Health Survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*, 220–233.
14. Kazis, L. E., Miller, D. R., Skinner, K. M., et al. (2006). Applications of methodologies of the Veterans Health Study in the VA healthcare system: Conclusions and summary. *The Journal of Ambulatory Care Management, 29*, 182–188.
15. The Veterans RAND 12 Item Health Survey: What it is and how it is used. Available from: http://www.va.gov/chqoer/VR12.pdf. Accessed 9 August 2010.
16. Brazier, J. E., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care, 42*, 851–859.
17. Kazis, L. E., Ren, X. S., Lee, A., et al. (1999). Health status in VA patients: Results from the Veterans Health Study. *American Journal of Medical Quality, 14*, 28–38.
18. Hanmer, J. (2009). Predicting an SF-6D preference-based score using MCS and PCS Scores from the SF-36 or SF-12. *Value in Health, 12*, 958–966.
19. Imputing the Physical and Mental Summary Scores (PCS and MCS) for the MOS SF-36 and the Veterans SF-36 Health Survey in the presence of Missing Data. http://www.hosonline.org/surveys/hos/download/HOS_Veterans_36_Imputation.pdf. Accessed 9 June 2010.
20. Kazis, L. E., Selim, A., Rogers, W., Ren, X. S., Lee, A., & Miller, D. R. (2006). Dissemination of methods and results from the veterans health study: Final comments and implications for future monitoring strategies within and outside the veterans healthcare system. *The Journal of Ambulatory Care Management, 29*, 310–319.
21. Kazis, L. E., Miller, D. R., Skinner, K. M., et al. (2004). Patient-reported measures of health: The Veterans Health Study. *The Journal of Ambulatory Care Management, 27*, 70–83.
22. Kazis, L. E., Miller, D. R., Clark, J. A., Skinner, K. M., Lee, A., Ren, X. S., et al. (2004). Improving the response choices on the veterans SF-36 health survey role functioning scales: Results from the Veterans Health Study. *The Journal of Ambulatory Care Management, 27*(3), 263–280.
23. Ware, J. E., Kosinski, M., & Keller, S. K. (1994). *SF-36 physical and mental health summary scales: A users's manual*. Boston, MA: The Health Institute, New England Medical Center.
24. Ware, J. E., Kosinski, M., & Keller, S. D. (1995). *SF-12: How to score the SF-12 physical and mental health summary scores*. Boston, Massachusetts: The Health Institute, New England Medical Center.
25. Selim, A. J., Rogers, W., Fleishman, J. A., Qian, S. X., Fincke, B. G., Rothendler, J. A., et al. (2009). Updated U.S. population standard for the Veterans RAND 12-item Health Survey (VR-12). *Quality of Life Research, 18*, 43–52.
26. Ara, R., & Brazier, J. (2009). Predicting SF-6D preference-based utilities using the SF-36 health summary scores: Approximating health related utilities when patient level data is not available. *Value in Health, 12*, 346–353.
27. Kazis, L. E., Miller, D. R., Clark, J. A., et al. (2004). Improving the response choices on the veterans SF-36 health survey role functioning scales: Results from the Veterans Health Study. *The Journal of Ambulatory Care Management, 27*, 263–280.
28. Perlin, J., Kazis, L. E., Skinner, K., et al. (2000). Health status and outcomes of veterans: Physical and mental component summary scores veterans SF-36. 1999 large health survey of veteran enrollees, executive report. Office of Quality and Performance (10Q) and Center for Health Quality, Outcomes and Economic Research, HSR&D Field Program, Washington D.C. and Bedford Massachusetts.
29. Bullen, P. S., & Mitrinović, D. S., Vasić, P. M. "Means and their inequalities," Reidel (1988) pp. 27ff.
30. Walters, S. J., & Brazier, J. E. (2003). What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health and Quality of Life Outcomes, 1*, 4.
31. Harrison, M., Davies, L., Bansback, N. J., McCoy, M. J., Verstappe, S. M. S., Watson, K., Symmons, D. P. M. and the British Society for Rheumatology Biologics Register Control Centre Consortium (2009). Qual Life Res 18:1195–1205.
32. Pliskin, J. S., Shepard, D. S., & Weinstein, M. C. (1980). Utility functions for life years and health status. *Operations Research, 28*, 206–224.
33. Tsuchiya, A., & Dolan, P. (2005). The QALY model and individual preferences for health states and health profiles over time: A systematic review of the literature. *Medical Decision Making, 25*, 460–467.
34. Smith, M. D., Drummond, M., & Brixner, D. (2009). Moving the QALY forward: Rationale for change. *Value Health, 1*, S1–S4.
35. Ferreira, P. L., Ferreira, L. N., & Pereira, L. N. (2008). How consistent are health utility values? *Quality of Life Research, 17*, 1031–1042.
36. Seymour, J., McNamee, P., Scott, A., & Tinelli, M. (2010). Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. *Health Economics, 19*, 683–696.
37. Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics, 13*, 873–884.
38. Kharroubi, S. A., Brazier, J. E., Roberts, J., & O'Hagan, A. (2007). Modelling SF-6D health state preference data using a nonparametric Bayesian method. *Journal of Health Economics, 26*, 597–612.
39. Kharroubi, S. A., Brazier, J. E., & O'Hagan, A. (2007). Modelling covariates for the SF-6D standard gamble health state preference data using a nonparametric Bayesian method. *Social Science and Medicine, 64*, 1242–1252.