

Comparison of Trends in US Health-Related Quality of Life Over the 2000s Using the SF-6D, HALex, EQ-5D, and EQ-5D Visual Analog Scale Versus a Broader Set of Symptoms and Impairments

Susan T. Stewart, PhD,* David M. Cutler, PhD,*†‡ and Allison B. Rosen, MD, ScD*§

Background: A number of instruments have been developed to measure health-related quality of life (HRQoL), differing in the health domains covered and their scoring. Although few such measures have been consistently included in US national health surveys over time, the surveys have included data on a broad range of symptoms and impairments, which enables the tracking of population health trends.

Objectives: To compare trends in HRQoL as measured using existing instruments versus using a broader range of symptoms and impairments collected in multiple years of nationally representative data.

Data and Measures: Data were from the 2000–2010 Medical Expenditure Panel Survey, which is nationally representative of the noninstitutionalized US population. Level of and trends in HRQoL derived from a broad range of survey symptoms and impairments (SSI) was compared with HRQoL from the SF-6D, the HALex, and, between 2000 and 2003, the EuroQol-5D (EQ-5D) and EQ-5D Visual Analog Scale.

Results: Trends in HRQoL were similar using different measures. The SSI scores correlated 0.66–0.80 with scores from other measures and mean SSI scores were between those of other measures. Scores from all HRQoL measures declined similarly with increasing age and with the presence of comorbid conditions.

Conclusions: Measuring HRQoL using a broader range of symptoms and impairments than those in a single instrument yields population health trends similar to those from other measures while making maximum use of existing data and providing rich detail on the factors underlying change.

Key Words: health measurement, health trends, health status, health-related quality of life, health monitoring, population health, quality of life

(*Med Care* 2014;00: 000–000)

A number of nationally representative health surveys are conducted in the United States with the goal of monitoring population health over time. Particularly valuable for tracking overall health are measures that consider a range of symptoms and impairments and weigh their severity and relative importance, combining them to yield a single overall score on a 0–1 scale. This score can then be combined with life expectancy to form a measure of quality-adjusted life expectancy (QALE). Some of the best known of these measures include the EuroQol-5D (EQ-5D),¹ the Short-Form 6D (SF-6D),² and the Health Utilities Index (HUI).³ Such measures have seldom been included in multiple years of national health surveys,⁴ in part due to the additional survey space required to ask the questions specific to a single instrument, and in part because no gold standard measure has been identified; each has strengths and weaknesses and different measures are best suited for different purposes.

However, the breadth of symptoms and impairments asked about in current national surveys is actually greater than that covered by any one instrument. For example, surveys often include questions about specific sensory, cognitive, and physical abilities (such as walking and bending) that are not included in specific instruments. By calculating disutilities associated with this broader set of symptoms and impairments,⁵ we have made use of the wide range of existing data to develop health-related quality of life (HRQoL) scores and track health over a longer time period and in a more comprehensive population than possible using an existing measure.⁶ Although this approach has the strength of enabling detailed tracking and disaggregation of HRQoL change, it is important to examine how the HRQoL scores formed in this manner compare to those from existing measures, at a point in time and over time. This paper compares the level of and trends in HRQoL scores derived from the broad range of questions in national health data from 2000 to 2010 to HRQoL scores from several other measures that can be calculated in the same national dataset:

From the *National Bureau of Economic Research; †Department of Economics, Harvard University; ‡Harvard University Interfaculty Program for Health Systems Improvement, Cambridge; and §Department of Quantitative Health Sciences and Meyer's Primary Care Institute, University of Massachusetts Medical School, Worcester, MA.

Supported by National Institute on Aging research grant P01AG031098.

The authors declare no conflict of interest.

Reprints: Susan T. Stewart, PhD, National Bureau of Economic Research, 1050 Massachusetts Ave, Cambridge, MA 02138. E-mail: sstewart@nber.org.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Website, www.lww-medicalcare.com.

Copyright © 2014 by Lippincott Williams & Wilkins
ISSN: 0025-7079/14/000-000

the SF-6D, the Health and Activity Limitations Index (HA-Lex),⁷ and, for a shorter time period, the EQ-5D and EQ-5D Visual Analog Scale (VAS).

METHODS

The Medical Expenditure Panel Survey (MEPS) is a household-based medical expenditure survey sponsored by the Agency for Healthcare Research and Quality (AHRQ) that is nationally representative of the noninstitutionalized civilian US population in each survey year.⁸ In the year 2000, 2 existing health measurement instruments were added to MEPS: the SF-12, which continues to be included in the survey, and the EQ-5D, which was included through 2003. The SF-6D can be calculated using a subset of 7 questions from the SF-12 questionnaire that cover physical and emotional role function, mobility, vitality, depression, and pain.² The EQ-5D asks broad questions in 5 domains of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, with respondents indicating either no difficulty, moderate difficulty, or extreme difficulty in each domain. An additional measure from the EQ-5D is an overall rating of general health on a scale from 0 (equivalent to the worst imaginable health state) to 100 (the best health imaginable), which is administered as a visual thermometer. We divide these scores by 100 to transform them to a 1–0 scale, and refer to this as the VAS in our comparisons.

Another HRQoL measure that can be estimated using data collected in MEPS is the HALex,⁷ which was developed as part of the Healthy People 2000 initiative and is scored based on the presence of impairments in self-care (Activities of Daily Living, or ADLs) and a self-rating of overall health on a 5-point scale, from excellent to poor. The original HALex was developed for the National Health Interview Survey, which includes a question about being “limited in any way in any activities because of any impairment or health problem.” We used an alternate HALex scoring that was developed for the MEPS survey, which does not contain this question.⁹

The SF-6D, EQ-5D, and HALex measures use different scoring/weighting algorithms to yield scores on a 0–1 scale. The SF-6D and EQ-5D use community-elicited preference weights for each health state. That is, individuals were asked to rate the potential impact of different combinations of health problems on HRQoL, yielding disutility weights for all of the instrument’s possible combinations of problems. Although the original weights for EQ-5D scoring were measured in the UK, we used preference weights measured in the US population.¹⁰ HALex weights for each combination of self-rated health and ADL function were imputed by its developers using mathematical techniques and reference to preference weights developed for the HUI instrument.⁷

As a result of the inclusion of the EQ-5D and SF-12 as well as many other specific questions on health and functioning, MEPS is the national survey that has included the broadest array of symptoms and impairments over time. Table 1 groups symptoms and impairments into 7 broad domains: problems with primary activity (working, keeping house, school) or social activity; physical activity limitations (self-care, performance of tasks for routine needs,

and specific movement difficulties); pain; mental health symptoms; low energy; sensory problems (vision and hearing impairment); and cognitive impairment.

To estimate the impact of this broad range of symptoms and impairments on QOL, we used a previously published method to develop disutility weights for each problem by relating them to the EQ-5D VAS in MEPS 2002.⁶ We regressed the VAS rating on the set of symptoms and impairments and their interactions, and then calculated disutility weights for each symptom and impairment that were independent of the others and reflected the effects of interactions. To do this we calculated mean predicted scores, first assuming that everyone reported the item (worst case) and then assuming no one reported the item (best case). The difference between these mean predicted scores yielded a weight that captured the broadest possible impact of having the symptom or impairment in light of the other symptoms and impairments that people have. These weights were then used to calculate HRQoL scores based on an individual’s reported symptoms and impairments.⁶ A decrement equal to the intercept of the regression was added to each score to represent the average utility loss not accounted for by this set of symptoms and impairments. We refer to this as the Survey-based Symptom/Impairment (SSI) method of deriving HRQoL scores.

The wording of the relevant questions in MEPS remained the same between 2000 and 2010 with the exception of those that were part of the SF-12, which changed slightly in 2003 due to the adoption the SF-12v2 questionnaire. Specifically, in version 2, questions about the frequency of depressive feelings, anxious feelings, and energy omitted the response category “good bit of the time,” thus dropping from 6 response categories to 5. In addition, role performance questions were expanded from having only 2 response categories (yes/no) in v1 to having 5 categories in v2 (all, most, some, a little, or none of the time). To adjust for the change in depressive and energy questions, the SF-6Dv1 scoring was revised to randomly assign half of those who responded “a good bit of the time” to the adjacent categories on either side (“most of the time” and “some of the time”). To adjust for the change in role performance questions, the SF-6D v2 scoring assigned all categories to “yes” except “none of the time” [Program for calculation of SF-6D scores, personal communication from Janel Hanmer. An alternate item mapping for the SF-6D using version 1 of the SF-12 is described in the online Appendix (Supplemental Digital Content 1, <http://links.lww.com/MLR/A757>)].

For the SSI, we used role performance questions in the MEPS survey that were not part of the SF-12. For depressive, anxious, and tired symptoms, we examined different methods of handling the question changes as described in the online Appendix (Supplemental Digital Content 1, <http://links.lww.com/MLR/A757>). Not adjusting for the changes had virtually the same effect as holding constant the change in these symptoms from 2002 to 2003, because the effects of the question changes acted in different directions for different symptoms.

Sample sizes for each year are shown in the Appendix (Supplemental Digital Content 1, <http://links.lww.com/MLR/A757>). We restrict our comparison of trends to the same sample over time, comparing SF-6D, SSI, and HALex only among those nonmissing on all measures. We did not

TABLE 1. Symptom/Impairment Question Coverage Across Instruments

	Primary Activity	Physical Activity	Mental	Sensory	Pain	Cognitive	Energy	Self-rated Health
MEPS 2000–2010	Major Social	Self-care Routine needs Bending Lifting Walking Standing Reaching Dexterity	Depressive Anxious	Vision Hearing				—
SF-6D	Major Social	Moderate activities	Depressive	—		—		—
EQ-5D	Major/social Combined	Self-care/routine needs Combined	Depressive/anxious Combined	—		—	—	—
HALex	—	Self-care Routine needs	—	—	—	—	—	—

EQ-5D indicates EuroQol-5D; HALex, Health and Activity Limitations Index; MEPS, Medical Expenditure Panel Survey; SF-6D, Short-Form 6D.

condition these trends on nonmissing response to VAS and EQ-5D, as they were only available for the first 4 years. To examine convergent validity, we tested correlations between measures, plotted them against one another, tested time trends in each measure, and compared their rank order over time. To plot SSI scores against those from other measures, we divided each other measure into categories based on the distribution of scores. The categories for these plots were as close to quartiles as possible while keeping those with the same score grouped together. Time trends were tested by regressing a continuous MEPS year variable on each measure in pooled data, controlling for 10-year age group, sex, race (black, with white and other in referent group), and sex interacted with race.

To assess construct validity, we also examined scores on each HRQoL measure by age and by the presence of none, 1, or ≥ 2 of the following self-reported conditions: diabetes, emphysema, stroke or transient ischemic attack, coronary heart disease, myocardial infarction, or any other kind of heart condition/disease (other than coronary heart disease/myocardial infarction/angina).

Analyses were performed using SAS (version 9 for Unix; The SAS Institute, Cary, NC) and Stata (version 12.1 for Unix; StataCorp., College Station, TX) using statistical methods designed for survey data to account for sampling factors including stratification, clustering, oversampling of minorities, and nonresponse to the overall survey and the mail-in health questionnaire.

RESULTS

Figure 1 shows trends in HRQoL using each of the measures, adjusted by 10-year age group to the 2010 population. We show the data points without confidence intervals because the SEs were sufficiently small to be unnoticeable (ranging from 0.001 to 0.002). Scores increased significantly from 2000 to 2010 using the SSI ($P=0.016$) and SF-6D ($P<0.0001$), whereas HALex was unchanged ($P=0.7480$). The notable drop in SF-6D scores in 2003 was due to the recoding of responses to role model questions in SF-6D scoring to try to match the binary question used in

prior years, which resulted in a higher prevalence of role performance problems in 2003 and beyond. Separately testing trends from before and after the SF-12 questionnaire change, from 2000 to 2003, there was no significant change in mean SSI ($P=0.661$) or SF-6D ($P=0.368$) scores, HALex scores declined ($P=0.030$), and there was an increase in mean EQ-5D ($P=0.0004$) and VAS ($P<0.0001$) scores.

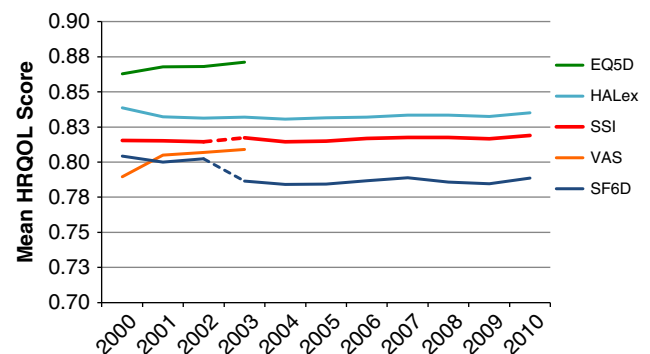


FIGURE 1. Comparison of trends in health-related quality of life using different measures, overall population, aged 18 years and over, MEPS survey years 2000–2010, age-adjusted to the 2010 population by 10-year age group. EQ-5D and VAS trends are truncated at 2003 because they were discontinued from the MEPS in 2004. Trends are shown for those with complete data on HALex, SSI, and SF-6D. Those missing on EQ-5D and VAS from 2000 to 2003 are not omitted; on the measures however, trends are almost identical when these additional respondents are omitted (not shown). The notable drop in SF-6D scores in 2003 was at least partially due to the adoption of the new SF-12 questionnaire in that year and the reassignment of responses to role model questions to try to match to the dichotomous response categories in the earlier SF-12 version. Survey Based Symptom/Impairment scores did not use the role performance questions. They did use other SF-12 questions that changed in 2003 (depressive and anxious symptoms and energy). Adjustments are discussed in the Appendix (Supplemental Digital Content 1, <http://links.lww.com/MLR/A757>).

TABLE 2. Correlations Between Measures in Pooled 2000–2002 MEPS Data

	Survey-based Symptom/ Impairment (SSI)	Short-Form 6D (SF-6D)	Health and Activity Limitations Index (HALex)	EuroQoL-5D (EQ-5D)
SF-6D	0.80			
HALex	0.66	0.54		
EQ-5D	0.75	0.71	0.59	
VAS	0.69	0.64	0.63	0.64

Correlations between all pairs of measures were significant ($P < 0.0001$).

MEPS indicates Medical Expenditure Panel Survey; VAS, Visual Analog Scale from EQ-5D.

From 2003 to 2010, SSI increased significantly ($P = 0.023$), but not the SF-6D ($P = 0.312$) or HALex ($P = 0.313$).

Differences between the measures at a point in time were larger than the differences in trends over time. Compared with SSI scores, the mean EQ-5D and HALex scores were significantly higher and the mean SF-6D and VAS scores were significantly lower (P for all comparisons < 0.0001). Two factors contributing to the slightly lower scores with the SF-6D than with the other measures were the additional disutility assigned to those with higher levels of any given problem, and the disutility assigned to milder levels of depressive symptoms (feeling down a little of the time), and low energy (had a lot of energy most of the time but not all of the time). Despite these differences, mean scores on all measures remained within 0.07 of one another. Table 2 shows correlations between pairs of measures in pooled data from 2000 to 2002, the years that included the EQ-5D/VAS and were before the switch in SF-12 versions. All correlations were significant ($P < 0.0001$) and they ranged from 0.54 to 0.80; correlations between the SSI and other measures were the highest.

Figure 2 depicts relationships between the SSI and the other measures in pooled data from 2000 to 2002, with each other measure divided into categories based on the distribution of scores. As shown in box plots, mean SSI scores increased with each ascending category of SF-6D, EQ-5D, and VAS scores.

Regarding the distribution of scores on each measure, SSI scores ranged from 0.1 to 0.9, with 27% of respondents reporting none of the problems in 2000 and thus having the maximum possible score of 0.92 (peaking at 31% in 2003 and 2010). In comparison, SF-6D scores ranged from 0.3 to 1; only 6% had no disutility/perfect scores in 2000, rising to 8% in 2010. The EQ-5D and HALex, which allow scores below 0 (when a health state is considered worse than death), had the broadest range of scores (-0.14 to 0.95). On the EQ-5D, 46% had the top score (1) in 2000 (rising to 49% in 2003), indicating a ceiling effect. On the HALex, 27% had the top score (0.95) in 2000, falling slightly to 24% in 2010. Mean self-rated overall health using the VAS ranged from 0 to 1, with 7.3% reporting a score of 1 in 2000, rising to 7.8% in 2003.

SSI scores were lowered by the presence of additional impairments not asked about in other HRQOL instruments, including specific physical abilities and cognitive impairment. At least one of the broader set of symptoms and impairments was reported by 16% of those with perfect SF-6D scores, 33% of those with perfect VAS ratings, 48% of those

with perfect EQ-5D scores, and 50% of those with the top HALex score. Conversely, among those with the top SSI score, 79% had a disutility on the SF-6D, 57% on the HALex, 13% on the EQ-5D, and 82% on the VAS. For EQ-5D, a few people with the top SSI score reported imperfect mobility (15%) or trouble with usual activity (7%) but the majority (93%) reported some pain, showing that our use of the SF-6D pain item (pain affects work) is different than just asking about pain separate from work.

Scores on all measures declined similarly with age, as shown in Figure 3A. The VAS dropped slightly more steeply with age than other measures, yielding the lowest scores from middle age onward. Measures also showed similar trends across increasing numbers of self-reported diseases, as shown in Figure 3B. The HALex and VAS measures showed the steepest drop with increasing numbers of diseases.

DISCUSSION

Although there were some differences across the 5 measures in the level of HRQoL, the trends in HRQoL over time were similar regardless of the measure used. Comparison of HRQOL levels across measures were consistent with previous studies finding similar rank ordering of instrument scores across diseases¹¹ and similar distributions of scores on different measures by age.¹² Our estimate of HRQoL based on the broader range of survey-based symptom and impairment questions yielded scores in between those of the other measures, made maximal use of available data, and provided the same information about trends in HRQOL as the other measures.

Our results are consistent with past findings of correlation between different HRQoL measures. The correlation among measures, particularly those based on the presence of specific impairments, supports their convergent validity. However, the measures clearly differed in their manner of capturing health, as demonstrated by the only partial overlap between top scores on different measures. Given the differences in questions and weighing/scoring across measures, this is not surprising: researchers have cautioned against assuming equivalence between scores from different questionnaires.¹³ Indeed, for the calculation of cost-effectiveness ratios, different measures can yield widely divergent results.¹⁴ However, although these differences across measures in the levels of HRQoL at a point in time are important, this is not the case for tracking population health trends over time, where we find similar results across measures.

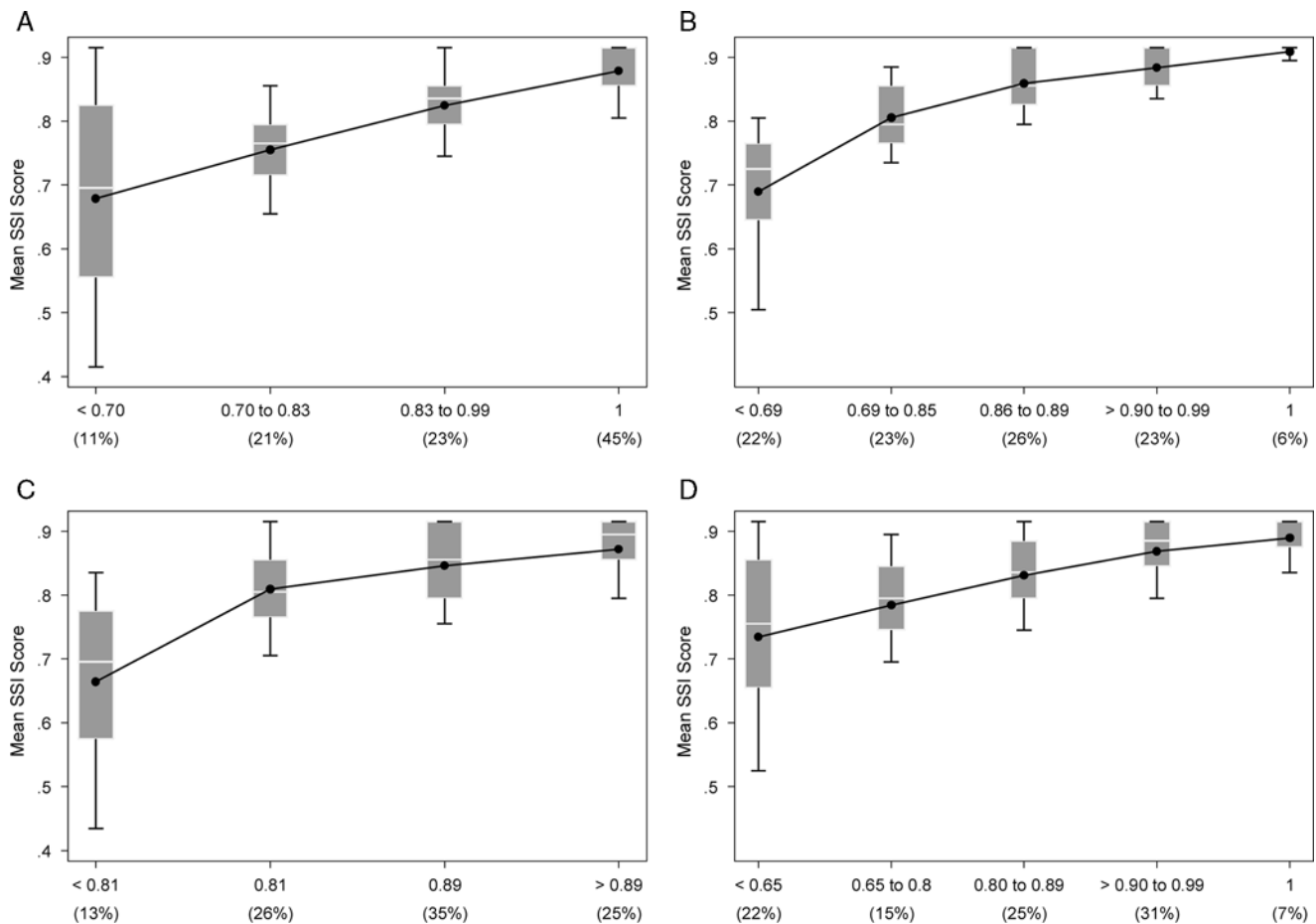


FIGURE 2. Relationships between SSI and other measures in pooled MEPS data from 2000 to 2002. A, Mean SSI scores by categories of EQ-5D scores (and percentage of sample in category). B, Mean SSI scores by categories of SF-6D scores (and percentage of sample in category). C, Mean SSI scores by categories of HALex scores (and percentage of sample in category). D, Mean SSI scores by categories of VAS scores (and percentage of sample in category). Line connects mean SSI scores among those in each category of the other measure. Box plots depict range of SSI scores from 25th to 75th percentile, and 10% and 90% confidence intervals. White line within each bar indicates median SSI score. Percentages in categories may not sum to 100 due to rounding.

In terms of the point-in-time comparisons of the level of HRQoL, our results are consistent with prior studies finding that the EQ-5D exhibits a ceiling effect in healthier populations.^{15–17} Scores based on the broader range of symptoms and impairments fell between other measures in terms of the proportion with top scores, but had a lower worst possible score than the SF-6D (0.1 vs. 0.4). The EQ-5D is based on a small number of broad health questions, and thus does not capture decrements to health that may occur due to more specific symptoms and impairments. The VAS health rating may capture these decrements, but the specific problems contributing to VAS scores are unmeasured. An instrument with very broad coverage, the Quality of Well-Being Scale (QWB),¹⁸ has been recommended for avoiding ceiling effects and providing extensive detail on health in a broad swath of the population¹³; however, it is lengthy to administer and its questions are not a part of MEPS, hence QWB scores could not be calculated for the current study.

For tracking health trends over time, it seems that the scores derived from the broad range of questions already asked in national data capture the largest level of detail possible without adding unnecessarily to survey length. In support of their construct validity, all measures yielded lower mean scores with increased age and among those with increased burden of self-reported diseases. The finding that VAS ratings and HALex scores dropped more than other measures in successive age groups and among those with more diseases suggests that self-rated health (which is also part of the HALex scoring) may reflect an element of poor health or frailty not captured by other measures.^{19,20} However, general health ratings lack detail regarding the specific symptoms and impairments that may contribute to overall health—detail which is important in enabling understanding of the specific factors driving trends in population health over time. Global self-ratings of health have also shown discrepant trends across different nationally representative US health

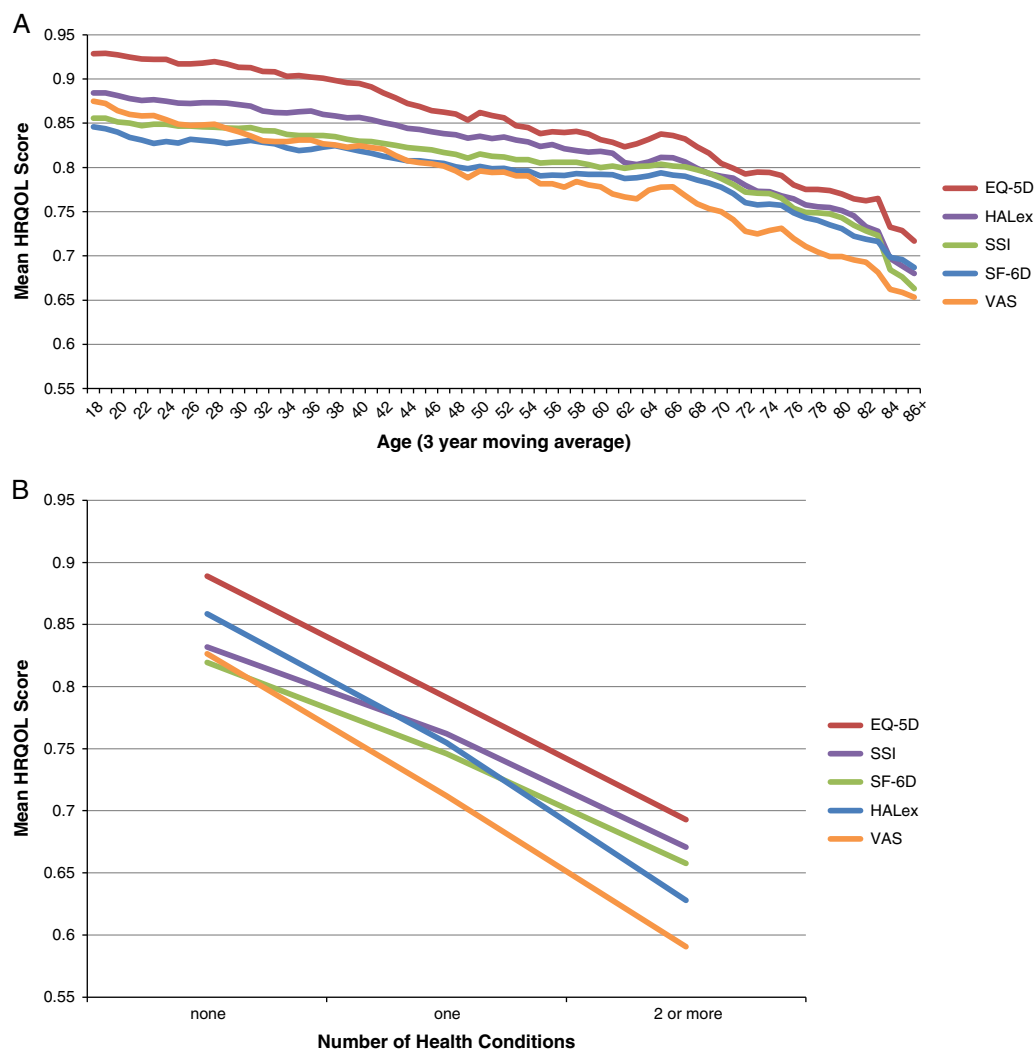


FIGURE 3. Comparison of trends in health-related quality of life using different measures in pooled 2000–2002 data. A, Comparison of age trends across measures. To smooth trends, means by age are depicted using a 3-year moving average centered around the indicated year. B, Comparison scores across measures among those with none, 1, or 2+ self-reported health conditions: coronary heart disease, myocardial infarction, other heart disease, stroke, diabetes, and emphysema.

surveys, over the same time period suggesting that they may be unsuitable for population health tracking and reinforcing recommendations to use more detailed measures.²¹

Although an examination of the underlying factors contributing to the overall HRQOL trend is beyond the scope of this paper, we have examined this in detail for the SSI measure in a prior paper.⁶ From 2000 to 2008, we found that HRQOL remained unchanged among nonelderly adults, and increased slightly among those over age 65. This increase was driven primarily by improved energy, as well as increased ability to perform one's primary activity, and decreased pain, ADL limitations, and depressive symptoms. Comparing HRQoL change by sex and race, HRQoL improved among black and white respondents of both sexes, with dips in the mid 2000s for all groups, and some variation across groups. A HRQoL rise among white males in 2001 was driven by small reductions in severe depressive symptoms and low energy.

There are some limitations to our method. The SSI measure does not use traditional utility measures such as Standard Gamble or Time-Tradeoff, which require respondents to rate health scenarios directly by indicating what they would give up in order to live without the health problem. Thus, our SSI scores do not reflect a true health utility in decision analytic terms. However, our scores are derived from the VAS, which is often used to rate health states and has been defended as a choice-based health instrument with advantages over others.²² Another limitation of our method, which is shared with utility-based measures, is our assumption that the effects of impairments and symptoms on health remain the same over time. Although we have found stability in the effects of symptoms and impairments on overall health in different years of MEPS⁵ (2000 and 2002), these effects may be expected to change over longer time periods, as population norms and the built environment change. The

occasional reinclusion of the VAS scale in a national health survey would enable reassessment of these relationships over time. However, the effects of symptoms and impairments on health are much more stable than the effects of diseases on health over time. For example, while improved diabetes management in a population can dramatically affect the proportion of diabetics who develop impairments (such as problems with walking or vision), the effect of a walking or vision impairment on overall health would not be expected to change dramatically. Finally, our analyses exclude those living in institutions and children. Although in other work we have included the institutionalized using additional data sources,⁶ the current comparison was restricted to the MEPS survey, which is a sample of the noninstitutionalized only.

CONCLUSIONS

Calculating HRQoL using a broad range of symptoms and impairments already collected in existing national data yields population health trends similar to those seen with other well-known measures and has several important advantages. It maximizes the use of existing data and avoids the ceiling effects frequently seen with the EQ5D. It allows trends in HRQoL to be disaggregated into the specific symptoms and impairments with the greatest (and smallest) contribution to health change over time.⁶ Finally, HRQoL scores can be calculated based on the full set of symptoms and impairments available in any national dataset,²³ and do not require the consistent inclusion of a particular instrument in the same dataset over time. Given the competition for space in national surveys and the absence of a gold-standard health measure for inclusion in national surveys, this is an important advantage.

REFERENCES

- Brooks R, Rabin RE, de Charro F. *The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective*. Dordrecht: Kluwer Academic Publishers; 2003.
- Brazier JE, Roberts J. Estimating a preference-based index from the SF-12. *Med Care*. 2004;42:851–859.
- Horsman J, Furlong W, Feeny D, et al. The Health Utilities Index (HUI(R)): concepts, measurement properties and applications. *Health Qual Life Outcomes*. 2003;1:54.
- Jia H, Zack MM, Thompson WW. State quality-adjusted life expectancy for US adults from 1993 to 2008. *Qual Life Res*. 2011;20:853–863.
- Stewart ST, Woodward RM, Rosen AB, et al. The impact of symptoms and impairments on overall health in US national health data. *Med Care*. 2008;46:954–962.
- Stewart ST, Cutler DM, Rosen AB. US trends in quality-adjusted life expectancy from 1987 to 2008: combining national surveys to more broadly track the health of the nation. *Am J Public Health*. 2013; 103:e78–e87.
- Erickson P. Evaluation of a population-based measure of quality of life: the Health and Activity Limitation Index (HALex). *Qual Life Res*. 1998;7:101–114.
- Agency for Healthcare Research and Quality (AHRQ) Medical Expenditure Panel Survey Background. Available at: http://www.meps-ahrq.gov/mepsweb/about_meps/survey_back.jsp. Accessed August 31, 2012.
- Lawrence WF, Yabroff KR, Fleishman JA. Measuring health utilities in national data sets: a new approach to valuing the HALex. *J Gen Intern Med*. 2003;18(suppl 1):164.
- Shaw J, Johnson J, Coons S. US valuation of the EQ-5D health states: development and testing of the D1 model. *Med Care*. 2005;43: 203–220.
- Franks P, Hanmer J, Fryback DG. Relative disutilities of 47 risk factors and conditions assessed with seven preference-based health status measures in a national US sample: toward consistency in cost-effectiveness analyses. *Med Care*. 2006;44:478–485.
- Fryback DG, Dunham NC, Palta M, et al. US norms for six generic health-related quality-of-life indexes from the National Health Measurement Study. *Med Care*. 2007;45:1162–1170.
- Kopeck JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. *J Clin Epidemiol*. 2003;56:317–325.
- Davis JC, Liu-Ambrose T, Khan KM, et al. SF-6D and EQ-5D result in widely divergent incremental cost-effectiveness ratios in a clinical trial of older women: implications for health policy decisions. *Osteoporos Int*. 2012;23:1849–1857.
- Bharmal M, Thomas J III. Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. *Value Health*. 2006;9:262–271.
- Brazier J, Roberts J, Tsuchiya A, et al. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ*. 2004;13:873–884.
- Seymour J, McNamee P, Scott A, et al. Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. *Health Econ*. 2010;19:683–696.
- Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. *Health Serv Res*. 1976;11:478–507.
- Idler EL, Kasl SV. Self-ratings of health: do they also predict change in functional ability? *J Gerontol B Psychol Sci Soc Sci*. 1995;50: S344–S353.
- Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav*. 1997;38:21–37.
- Salomon JA, Nordhagen S, Oza S, et al. Are Americans feeling less healthy? The puzzle of trends in self-rated health. *Am J Epidemiol*. 2009;170:343–351.
- Parkin D, Devlin N. Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Econ*. 2006;15:653–664.
- Stewart ST, Woodward RM, Cutler DM. *A Proposed Method for Monitoring US Population Health: Linking Symptoms, Impairments, Chronic Conditions, and Health Ratings*. NBER Working Paper No. 11358. Cambridge: National Bureau of Economic Research; 2005.