U.S. NORMS FOR SIX GENERIC HEALTH-RELATED QUALITY-OF-LIFE INDEXES FROM THE NATIONAL HEALTH MEASUREMENT STUDY

Dennis G. Fryback, PhD, **Nancy Cross Dunham, PhD**, **Mari Palta, PhD**, **Janel Hanmer, BS**, **Jennifer Buechner, BS**, **Dasha Cherepanov, BS**, and **Shani Herrington, MS**
University of Wisconsin-Madison

Ron D. Hays, PhD,
University of California, Los Angeles, and RAND, Santa Monica, CA

Robert M. Kaplan, PhD,
University of California, Los Angeles

Theodore G. Ganiats, MD,
University of California, San Diego

David Feeny, PhD, and
Health Utilities, Inc., and Kaiser Permanente Northwest

Paul Kind, MPhil
University of York

Abstract

Background—A number of indexes measuring self-reported generic health-related quality-of-life (HRQoL) using preference-weighted scoring are used widely in population surveys and clinical studies in the U.S.

Objective—To obtain age-by-gender norms for older adults on 6 generic HRQoL indexes in a cross-sectional U.S. population survey and compare age-related trends in HRQoL.

Methods—The EQ-5D, HUI2, HUI3, SF-36v2™ (used to compute SF-6D), QWB-SA and HALex were administered via telephone interview to each respondent in a national survey sample of 3,844 non-institutionalized adults aged 35-89. Persons aged 65 to 89 and telephone exchanges with high percentages of African Americans were over-sampled. Age-by-gender means were computed using sampling and post-stratification weights to adjust results to the U.S. adult population.

Results—The six indexes exhibit similar patterns of age-related HRQoL by gender; however, means differ significantly across indexes. Females report slightly lower HRQoL than do males across all age groups. HRQoL appears somewhat higher for persons age 65-74 compared to people in the next younger age decade, as measured by all indexes.

Conclusions—Six HRQoL measures show similar but not identical trends in population norms for older U.S. adults. Results reported here provide reference values for 6 self-reported HRQoL indexes.

Keywords

Health-related quality-of-life; Health Status; EQ-5D; SF-6D; QWB-SA; SF-6D; Health Utilities Index; HUI2; HUI3; SF-36; Population survey; Adults; Aging; Patient-Reported Outcomes; Health Outcomes measures; Comparative studies

INTRODUCTION

The past 35 years has seen significant progress in methodologies for health assessment using self-reported, preference-based summary measures of health-related quality-of-life (HRQoL). Advances include a growing number of available indexes and the use of generic health outcome measures in clinical and cost-effectiveness studies creating a legacy of published data [1-5]. More recently, interpretation of the measures has been improved by their inclusion in large health surveys thereby creating population norms against which to evaluate results obtained in clinical studies. These surveys also provide points for population tracking and comparison.

Generic HRQoL indexes score health using standardized weighting representing community preferences for health states on a scale anchored by 0 (dead) and 1 (full health). Representing community preferences is important for cost-effectiveness analysis in health and medicine [11]. There are currently six such indexes used in the United States: the EuroQol EQ-5D (EQ-5D) [12]; the Health Utilities Index Mark 2 (HUI2) and Mark 3 (HUI3) [13,14]; the SF-6D [15,16]; the Quality of Well-being scale self-administered form (QWB-SA) [17]; and the Health and Activities Limitations index (HALex) [18,19]. Scores from these indexes purport to represent the same evaluation of a given level of health, so we could expect them to be similar if administered in the same population. The indexes, however, emphasize overlapping but different health concepts, operationalize these concepts differently, refer to different time frames, and use scoring algorithms derived with different elicitation methods from different populations.

Indeed, within the few surveys which have collected more than one of these health indexes, distributions of summary scores differ by index and by scoring system used. Comparing distributions across the few published studies presenting US population normative data is also limited by differences in sample, mode of administration, and year of the survey, but again differences are seen across indexes [6,7,9,20]. Luo, *et al.*, [6] report gender-age-specific U.S. means for EQ-5D, HUI2, and HUI3 assessed during a household survey of U.S. adults from which the U.S. scoring system for EQ-5D was derived (US Valuation of the EQ-5D, "USVEQ"); they show lower HRQoL in older cohorts for both genders, with females reporting slightly lower HRQoL than males. Hanmer, *et al.*, [7] used Medical Expenditure Panel Survey (MEPS) data to estimate gender-age-specific means for EQ-5D and SF-6D (the latter computed from SF-12 data), and report results similar to Luo, *et al.* with USVEQ data. The two analyses report means which differ from each other, but with overlapping confidence intervals for EQ-5D means in comparable strata. In USVEQ data, means for the three indexes differed when compared in the same strata, with EQ-5D and HUI2 both significantly higher than HUI3, and EQ-5D slightly higher than HUI2 [6]. Hanmer, *et al.*, [20] suggest that mode of administration (*i.e.*, face-to-face interview, telephone, or self-administered) may lead to differences in means. Noyes, *et al.*, [21] show that relatively small differences in HRQoL scores associated with different scoring systems (U.K. *vs.* U.S.) for the EQ-5D can lead to different conclusions in cost-effectiveness analysis of medical interventions. Thus, it is important that we know better where differences between indexes exist and the relative magnitudes of differences.

The purpose of the National Health Measurement Study (NHMS) was to add to this growing base of knowledge about the different indexes in two ways. First, we add QWB-SA and HALex and SF-6D (based on SF-36) to the list of indexes for which U.S. non-institutionalized adult norms are available. Second, we administer all six indexes using the same mode to the same individuals in a population-based survey, minimizing effects of mode of administration and sample when interpreting differences.

Background and terminology

These indexes are used around the globe to measure and summarize the health of populations. The EuroQol EQ-5D (EQ-5D) [12], is very widely used and now has a U.S. scoring system [22] and U.S. population-based data [6,23]. The Health Utilities Index with its two distinct summary scores, HUI2 and HUI3, is used in many clinical trials [13,14]; population-based norms are available in Canada and the U.S. from several different population-based surveys [6,24]. The oldest of preference-based measures is the Quality of Well-being Scale (QWB) which was originally proposed for health policy analysis [25] and which now has a self-administered form, QWB-SA [17]. Though less used than the EQ-5D or HUI2/3, the QWB-SA is unique among this class of indexes because it makes use of an extensive set of self-reported symptoms to describe health. The SF-36, perhaps the most widely used generic health status measure in the world, was developed for the Medical Outcomes Study [26]; version 2.0, the SF-36v2™, was used in this study [27]. A preference-scored summary index, the SF-6D, has been developed specifically to use the SF-36v2™ questionnaire [15]. The Health and Activities Limitations Index (HALex) was constructed *ad hoc* to proxy as a preference-based summary index of health, retrospectively using data from the U.S. National Health Interview Survey and the U.S. Behavioral Risk Factor Surveillance System to evaluate U.S. health goals under Healthy People 2000 and Healthy People 2010 [18,19]. The HALex has also been used to evaluate population health in Japan [28].

Each measure involves a generic multidimensional health state classification, or descriptive system, employing multiple health domains to classify health broadly (*i.e.*, these are formulated in generic, not disease- or organ-specific, terms), and a standardized weighting (or "scoring") system derived from a community preference valuation of health states. The health state classification system is most commonly a set of health domains, or "attributes" or "dimensions," (such as pain) which have pre-defined levels (*e.g.*, "none", "moderate", "severe"). Levels range from fully healthy state to a very unhealthy state in each domain. The measures are self-reported; a person's answers to a standardized questionnaire are used in a prescribed manner to specify the level of each domain in the index's descriptive system with which to associate the respondent. We will refer to each specific combination of questionnaire, health state classification system, and standardized scoring system as an "index" or "measure" synonymously, and numbers assigned to individuals as "scores."

METHODS

We used a random digit dialed (RDD) telephone interview of a sample of adults aged 35 to 89 years, designed to represent the older half of the U.S. population in 2005-06 (2005 median age was 36.4 years, http://factfinder.census.gov) from the continental U.S. The upper age cutoff was selected because the non-institutionalized portion of the population over age 89 is sufficiently small (~ 1% of population), and the prevalence of dementia sufficiently high, that means other than telephone interview would likely be needed to assess HRQoL adequately in this segment of the population. RDD surveys use random telephone numbers from non-cellular exchanges to sample households. By design, to allow greater power for analyses of two important subgroups, we over-sampled black Americans and persons aged 65 to 89 using standard survey sampling methods. Blacks were over-sampled by approximately a factor of 2

by increasing the probability of calling in geographically scattered phone exchanges with known higher proportions of black households. Older individuals were over-sampled by approximately a factor of 3 as described below. Because all sampling was done using pre-determined probabilities, accurate sample weights for each respondent could be computed in order to adjust estimates of means to the U.S. population aged 35-89 using standard survey statistical methods.

HRQoL indexes

EQ-5D—Permission to administer EQ-5D was given without charge by the EuroQol Group (http://www.euroqol.org). The EQ-5D questions refer to "your health today." The EQ-5D descriptive system uses 5 domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each with 3 response options (no problems, moderate problems, severe problems), defining a total of 243 unique health states [29]. For this study, we applied the scoring algorithm derived for the U.S. general population. This scoring algorithm was derived from time tradeoff assessments of EQ-5D health states made by a population sample of some 4,000 U.S. adults in face-to-face household interviews [22].

HUI2, HUI3—License to use the proprietary HUI2/3 English-language questionnaire and mapping algorithm with a 1 week recall period was purchased from Health Utilities, Inc. (http://www.healthutilities.com/). A condition of the license is that users not reveal the content of the questions or the mapping algorithm. Respondents are asked to consider "your level of ability or disability during the past week." Scoring algorithms for both HUI2 and HUI3 were derived from standard gamble assessments made by adults in community samples in Hamilton, Ontario, and employ multiplicative multi-attribute utility functions. The algorithms map data from the same 40-item interviewer-administered questionnaire to each of the HUI2 and HUI3 classification systems. The HUI2 defines health status on 6 attributes (sensation, mobility, emotion, cognition, self-care and pain—we excluded an optional fertility dimension as is usual in the literature). Each attribute is divided into 4 or 5 levels, resulting in 8,000 unique health states [14]. The HUI3 defines health on 8 attributes (vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain), each having 5 or 6 levels, and jointly describing 972,000 unique health states [13].

Both HUI2 and HUI3 scoring functions have health states scored less than 0 (dead). HUI2 scores range from -0.03 to 1.0; HUI3 scores range from -0.36 to 1.0.

QWB-SA—Permission to use the QWB-SA was obtained free of charge from the University of California, San Diego, Health Services Research Center, La Jolla, CA (http://medicine.ucsd.edu/fpm/hoap/index.html). Usually the QWB-SA is self-administered using a two-sided optical scan form. We adapted the QWB-SA for this study to be administered by computer-assisted telephone interview. The QWB-SA assesses health over the past 3 days. The QWB-SA combines 3 domains of functioning (mobility, physical activity, social activity) with a lengthy list of symptoms and health problems, each assigned a weight, using an algorithm that yields a single summary score [17,30] based on presence or absence of activities and symptoms on each of the past 3 days. The final QWB-SA score is the average of the 3 single-day scores. The original QWB algorithm was developed using visual analog scale (VAS) ratings of health state descriptions by a community sample of adults located in the San Diego, CA, area. The QWB-SA algorithm is conceptually similar to that of the original QWB, but was derived from ratings by a convenience sample of people in family medicine clinics around San Diego; VAS scales were used to rate domain levels and some case descriptions formed from special combinations of domains in a multi-attribute utility elicitation process. Excluding dead (0.00), the minimum possible QWB-SA score is 0.09 and the maximum is 1.0.

SF-6D—License to administer the SF-36v2™ was purchased from its vendor (http://www.sf-36.org/). SF-36v2™ refers to several time frames. One question asks for self-rated health "in general." Some questions ask how much one's health "now limits" doing certain activities. Other questions refer to the "past four weeks." The SF-6D is computed from a subset of 11 of the 36 questions in the proprietary questionnaire. While SF-36v2™ yields a health profile summary using 8 domains, the SF-6D has reduced this to 6 domains (physical function, role limitation, social function, pain, mental health, and vitality), each comprised of 5 to 6 levels, and jointly defining about 18,000 health states [15]. Scoring was derived from standard gamble assessments by a population sample from the U.K. The SF-6D scoring algorithm is distributed by the SF-36v2™ vendor. We separately coded a SAS algorithm and verified its output scores with both the developer and vendor, leading to clarification and minor update to the algorithm distributed by the vendor (personal communication, Prof. J. Brazier, Dr. J. Bjorner, March-April, 2007). The scoring algorithm produces scores ranging from 0.30 to 1.0.

HALex—No permission is needed to use the HALex. The HALex is the only summary index available for the U.S. National Health Interview Survey, and it is used to track years of healthy life in Healthy People 2000 and 2010 [18]. HALex questions refer to "your health in general." It consists of two domains, 6 levels of activity limitation (ranging from "no limitations" to "unable to perform activities of daily living") and 5 levels of self-reported health ("excellent", "very good", "good", "poor", "fair"), jointly defining 30 health states. This is the only one of the six indexes to use self-rated health to describe health states. For the self-rated health domain we used question 1 from the SF-36v2™. For the activity domain we used the questions from Appendix 1 of Erickson [18], adapted for computer-assisted telephone administration.

The scoring algorithm was developed *ad hoc* without actual preference survey data using correspondence analysis to the Health Utilities Index Mark I. The worst of the 30 health states is scored 0.10 and the best scored 1.0.

Survey Administration

Trained interviewers at the University of Wisconsin Survey Center conducted the interviews from June 2005 through August 2006 using computer assisted telephone interview (CATI) software. For the 40% of households where street addresses were available from reverse directories, advance letters were sent explaining the purpose of the study and including $2 cash as pre-incentive to increase survey participation. When a household was reached by telephone, the interviewer conducted a brief screening interview eliciting an enumeration of adults in the household and their ages. CATI software assigned persons to three age ranges, 35-44, 45-64, and 65-89, and sampled among populated age ranges with pre-set probabilities favoring the oldest age group. If there were more than one adult in the household in the selected age group, the eligible respondent was selected using the Troldahl-Carter-Bryant technique [31]. The selected respondent was contacted and asked to participate in the telephone interview. Respondents were provided explanations of the purposes and content of the study, guaranteed anonymity, and offered a $25 incentive for completion of the entire survey.

Respondents were told that a number of health measures were being tested and that some questions would sound redundant. The first question of the interview was the categorical self-rating of health question asking respondents to rate their overall health as excellent, very good, good, fair, or poor. Following this the EQ-5D, HUI, SF-36v2™, and QWB-SA questionnaires were administered in an order randomized across respondents by the CATI software. Following these four questionnaires, the HALex questions were administered. Later sections of the interview elicited a wide range of other information including sociodemographic and financial variables. When a respondent could not complete the interview in one session (n=734) a call-

back was arranged and the interview picked up where it left off. Interviewers encouraged respondents not to break within a questionnaire for an index, and call-backs were generally within a few days.

Analyses

Regression analyses were completed using the SURVEYREG and SURVEYMEANS procedures of SAS version 9.0 software (The SAS Institute, Cary, NC). Sampling weights were computed as the inverse sampling probability for each participant based on the sampling scheme and then post-stratified to the U.S. Census 2000 population by age (35-44, 45-64, 65-89), gender (male, female), and race (black, white, other). Combined survey and post-stratification weights were trimmed within age-decade-, gender-, race-subgroups so that no one observation constituted more than 5% of its subgroup survey weight. These trimmed final weights were applied to adjust results from the sample distribution, where blacks and older people were over-sampled, to the U.S. adult population.

RESULTS

Sample and response rate

We sampled 29,844 telephone numbers deemed potentially in scope (*i.e.*, working, residential, non-fax/data lines, etc.). Although telephone numbers were called a minimum of 10 times before being abandoned as unreachable, 15,450 (54%) of these could not be verified as in scope because of non-contact (*e.g.*, phone never answered, immediate hang-ups or phone problems). Of 14,394 identified households contacted, 2,738 informants broke off the call before initial screening questions to determine household/respondent eligibility could be completed. Screening was completed in 11,656 households and 6,822 were determined to have at least one eligible respondent and 1 person from each of these households was invited to participate. Of those invited, 4,334 agreed to begin the interview and 3,853 completed. Nine of these were found retrospectively to be ineligible. The final sample was N=3,844 respondents.

There are different approaches to calculate response rates depending on how many households with potentially eligible respondents are assumed to be in the unscreened telephone numbers [32]. We used two recommended methods to compute response rates. The simple estimate is the ratio of completed interviews (3,844) to identified eligible households (6,822), 56.3%. However, we may assume there were unidentified eligible respondents in the 2,738 households for which a screening interviewed was not completed. A second estimate takes the proportion eligible among screened households (6,822/11,656 = 0.585) to estimate the proportion eligible among identified households not screened. Thus, 1,602 eligible respondents may have resided in the 2,738 unscreened households. Under this assumption the response rate is 45.6% (3,844/ (6822+1602)). The advance letter with $2 pre-incentive increased response rates by 5% compared to households not getting these (data not shown).

Table 1 shows demographic characteristics of the raw sample, the survey-weighted sample, and U.S. Census 2000 for comparison. We applied post-stratification weights based on gender, age, and race; the remaining characteristics show our sample to be somewhat higher income and better educated than the population in general (both the opposite of the USVEQ sample bias [6]). For results below, "weighted" analyses adjust the result to the U.S. population using the survey weights; "unweighted" analyses do not use the survey and post-stratification weights, and thus describe the raw sample.

Average administration times, percentage of cases for which scores could not be computed due to missing data, and distributional characteristics of the indexes in the unweighted sample data are in Table 2. Though this was a sample of non-institutionalized adults, essentially the full

range of each index's values was observed. EQ-5D had substantial ceiling effect (36% scoring the highest value), QWB-SA and SF-6D had no or minimal ceiling effect, and the remaining indexes showed modest ceiling effects. Weighted Pearson correlations among pairs of indexes vary from 0.60 to 0.71, except the 0.89 correlation between the HUI2 and HUI3, which use the same underlying questionnaire (all correlations $p<.001$).

Table 3 presents our main result: estimated population means by gender and age for each of the six HRQoL indexes. These estimates were computed using weighted regressions within each age-by-gender stratum and the standard errors presented also reflect the population weighting. In weighted analyses of variance within each age-by-gender stratum there is significantly more variation among index means than would be expected if the indexes scored health the same (all p-values for the effect of index $< .001$). The EQ-5D means were the highest, along with HUI2 which were slightly lower in each stratum. HUI3, SF-6D, and HALex formed a mid-range cluster of means and the QWB-SA means were always lowest, differing from the other means by 0.1 or more. All indexes show the same general relationship with age. All indexes, except HALex computed for females, show a leveling or increase in HRQoL in the 65-74 year old group compared to the 55-64 year old group; this "bump" in the indexes in aggregate is statistically significant ($p=0.04$). Females reported slightly lower HRQoL across all age groups than did men. Although the associations between age and index scores appear similar across indexes, the weighted age trend-by-index interaction is significant ($p<0.05$) in these data, indicating the measures somewhat differently exhibit the age trend. Of the 6 indexes, HALex tends to exhibit the largest difference in mean score between youngest and oldest groups. Figure 1 shows the means in Table 3 as graphs to emphasize the similarity of trends in results using the different indexes; confidence intervals are suppressed for visual clarity but can be derived from the standard errors in Table 3.

DISCUSSION

Generic, self-reported, preference-based HRQoL instruments provide a picture of health complementary to the health indicators and mortality rates which are often used to summarize aspects of population health. The NHMS administered six widely used measures in a national telephone survey of older U.S. adults. Although all the indexes are nominally scored with anchors of 1 for full health and 0 for dead, there is evident difference among the indexes in mean scores observed for age-by-gender groups. Reasons for these differences may include varying degrees of ceiling effect in the non-institutionalized population and varying sensitivity to different aspects of health. The QWB-SA produced by far the lowest scores. Perhaps this is due to this index's valuation method as it is the only one of those administered which is based on a visual analog scale estimation of utilities, the scores for which tend to be lower than time tradeoff or standard gamble scores; or it may be due to the formulation of this index in terms of symptoms, a distinctly different approach from the other descriptive systems.

Our means for EQ-5D are similar to those reported previously for younger ages in two data sets, USVEQ and MEPS [6,7]; however in the decade around age 60 the means observed in our study begin to diverge, appearing significantly higher than those in the other data sets, averaging .02 to .07 higher in males, and .03 to .12 higher in females, these differences increasing with age. Our means for HUI3 are very similar at all ages to another U.S. population telephone survey, the Joint Canada US Survey of Health [20], but diverge from the HUI3 estimates in the USVEQ household survey [6]. Our results are similar to all previous surveys in finding slightly lower HRQoL for females than males regardless of index or survey mode [20].

The differences we observed present a quandary to those who rely on exact numerical values of these indexes for cost-effectiveness analysis (CEA), and at a minimum imply that scores

from different indexes should never be mixed in one CEA. In spite of these absolute numerical differences, it is striking how similar the general pattern of results is across instruments in Table 3/Figure 1, raising the possibility that they all tell the same general story in population tracking in spite of their structural and etiologic differences.

There appears to be either a slight "dip" in HRQoL at ages 55-64 or a slight "bump" up in HRQoL in the 65-74 age group, compared to a linear, downward trend associated with age group. Are people in the older cohort really healthier than in the younger group? People aged 55-64 during our survey were born in 1941-50; those aged 65-74 were born in 1931-40. A recent National Bureau of Economic Standards [33] used an *ad hoc* index and U.S. Health and Retirement Study data to analyze self-reported health by persons in birth cohorts from 1936-41, 1942-47, and 1948-53 at age 51-56. That report concluded those in the youngest cohort ("baby boomers") self-reported worse health at age 51-56 than did those in the oldest cohort 10-12 years earlier when they were the same age. It is difficult to assess whether this finding reflects a cohort-associated reporting bias or is a true reflection of worse health. Our cross-sectional survey data compares HRQoL from approximately these birth cohorts, but collected concurrently, and appear to support a present-day difference between them.

Most surveys, and telephone surveys in particular, are limited and NHMS is no exception. Perhaps telephone surveys tend to have higher response rates from more educated, wealthier people (as seen comparing our weighted sample to the U.S. Census 2000 data in Table 1), and face-to-face household surveys may have the opposite bias (Table 1 in [6]), resulting in our "seeing" healthier older people compared to face-to-face surveys. Increasing use of cell phones, excluded in household RDD sampling, apparently has little effect on population health estimates in the time our survey was completed (especially among the age groups we sampled) [34]. However non-response bias, apparent in the differences between our sample distributions of education and household income, may have affected our results.

In spite of limitations, our results contribute important added observations to previously reported population means in the U.S. by better informing users of self-reported HRQoL indexes in population surveys and in clinical investigations and by providing critical comparative data by gender and age, and a better sense of variability of results among carefully conducted surveys.

Acknowledgements

REFERENCES

1. Patrick DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health status index. Health Serv Res 1973;8:228–245. [PubMed: 4761617]

2. Ware JE Jr. Brook RH, Davies AR, et al. Choosing measures of health status for individuals in general populations. Am J Public Health 1981;71:620–5. [PubMed: 7235100]

3. Fryback DG, Dasbach EJ, Klein R, et al. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. Med Decis Making 1993;13:89–102. [PubMed: 8483408]

4. Gold, MR. Cost-effectiveness in health and medicine. Oxford University Press; New York: 1996.

5. Gold MR, Stevenson D, Fryback DG. HALYS AND QALYS AND DALYS, OH MY: Similarities and Differences in Summary Measures of Population Health. Annu Rev Public Health 2002;23:115–34. [PubMed: 11910057]

6. Luo N, Johnson JA, Shaw JW, et al. Self-Reported Health Status of the General Adult U.S. Population as Assessed by the EQ-5D and Health Utilities Index. Med Care 2005;43:1078–1086. [PubMed: 16224300]

7. Hanmer J, Lawrence WF, Anderson JP, et al. Report of Nationally Representative Values for the Noninstitutionalized US Adult Population for 7 Health-Related Quality-of-Life Scores. Med Decis Making 2006;26:391–400. [PubMed: 16855127]

8. Revicki D. National Health Preference Data: A Useful Resource for Health Services Research. Med Decis Making 2006;26:310–312. [PubMed: 16855120]

9. Sullivan P, Ghushchyan V. Preference-Based EQ-5D Index Scores for Chronic Conditions in the United States. Med Decis Making 2006;26:410–420. [PubMed: 16855129]

10. Lasser KE, Himmelstein DU, Woolhandler S. Access to care, health status, and health disparities in the United States and Canada: results of a cross-national population-based survey. Am J Public Health 2006;96:1300–7. [PubMed: 16735628]

11. Gold, MR.; Patrick, DL.; Torrance, GW., et al. Identifying and valuing outcomes. In: Gold, MRSJ.; Russell, LB.; Weinstein, MC., editors. Cost-effectiveness in health and medicine. Oxford University Press; New York: 1996. 1996

12. Brooks, R.; Rabin, R.; de Charro, F. The Measurement and Valuation of Health Status using EQ-5D: A European Perspective. Kluwer Academic Publishers; Dordrecht, The Netherlands: 2003.

13. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care 2002;40:113–28. [PubMed: 11802084]

14. Feeny, D.; Torrance, G.; Furlong, W. Health Utilities Index. In: Spilker, B., editor. Quality of Life and Pharmacoeconomics in Clinical Trials. Lippincott-Raven Press; Philadelphia, PA: 1996. 1996

15. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271–92. [PubMed: 11939242]

16. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004;42:851–9. [PubMed: 15319610]

17. Andresen EM, Rothenberg BM, Kaplan RM. Performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire among older adults. Med Care 1998;36:1349–60. [PubMed: 9749658]

18. Erickson P. Evaluation of a population-based measure of quality of life: the Health and Activity Limitation Index (HALex). Qual Life Res 1998;7:101–14. [PubMed: 9523491]

19. Molla, MT.; Wagener, DK.; Madans, JH. Healthy People Statistical Notes, no. 21. National Center for Health Statistics; Hyattsville, Maryland: 2001. Summary measures of population health: Methods for calculating healthy life expectancy.

20. Hanmer J, Fryback DG, Hays RD. Mode of administration is important in U.S. national estimates of health related quality of life. Med Care. 2007 in revision

21. Noyes K, Dick AW, Holloway RG. The Implications of Using US-Specific EQ-5D Preference Weights for Cost-Effectiveness Evaluation. Med Decis Making. 2007

22. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: Development and testing of the D1 valuation model. Med Care 2005;43:203–220. [PubMed: 15725977]

23. Sullivan P, Lawrence W, Ghushchyan V. A National Catalog of Preference-Based Scores for Chronic Conditions in the United States. Med Care 2005;43:736–749. [PubMed: 15970790]

24. Sanmartin, C.; Ng, E.; Blackwell, DL., et al. Joint Canada/United States Survey of Health: Findings and Public-use Microdata File. [accessed Nov. 11, 2006]. Statistics Canada, Updated June 2, 2004. Catalog No. M0022XIE. http://www.statcan.ca/bsolc/english/bsolc?catno=82M0022X

25. Kaplan RM, Anderson JP. A general health policy model: update and applications. Health Serv Res 1988;23:203–35. [PubMed: 3384669]

26. Ware JE Jr. Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83. [PubMed: 1593914]

27. Ware JE Jr. SF-36 health survey update. Spine 2000;25:3130–9. [PubMed: 11124729]

28. Asada Y, Ohkusa Y. Analysis of health-related quality of life (HRQOL), its distribution, and its distribution by income in Japan, 1989 and 1998. Soc Sci Med 2004;59:1423–1433. [PubMed: 15246171]

29. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;33:337–43. [PubMed: 11491192]

30. Kaplan RM, Sieber WJ, Ganiats TG. The Quality of Well-being Scale: Comparison of the interviewer-administered version with a self-administered questionnaire. Psychol & Health 1997;12:783–791.

31. Gaziano C. Comparative analysis of within-household respondent selection techniques. Public Opinion Quarterly 2005;69:124–157.

32. Schwartz L, Woloshin S, Fowler F Jr. et al. Enthusiasm for Cancer Screening in the United States. JAMA 2004;291:71–8. [PubMed: 14709578]

33. Soldo, BJ.; Mitchell, OS.; Tfaily, R., et al. Cross-cohort differences in health on the verge of retirement. National Bureau of Economic Research; Cambridge, MA: Dec. 2006 Working paper 12762

34. Blumberg SJ, Luke JV, Cynamon ML. Telephone coverage and health survey estimates: evaluating the need for concern about wireless substitution. Am J Public Health 2006;96:926–31. [PubMed: 16571707]

FIGURE 1.
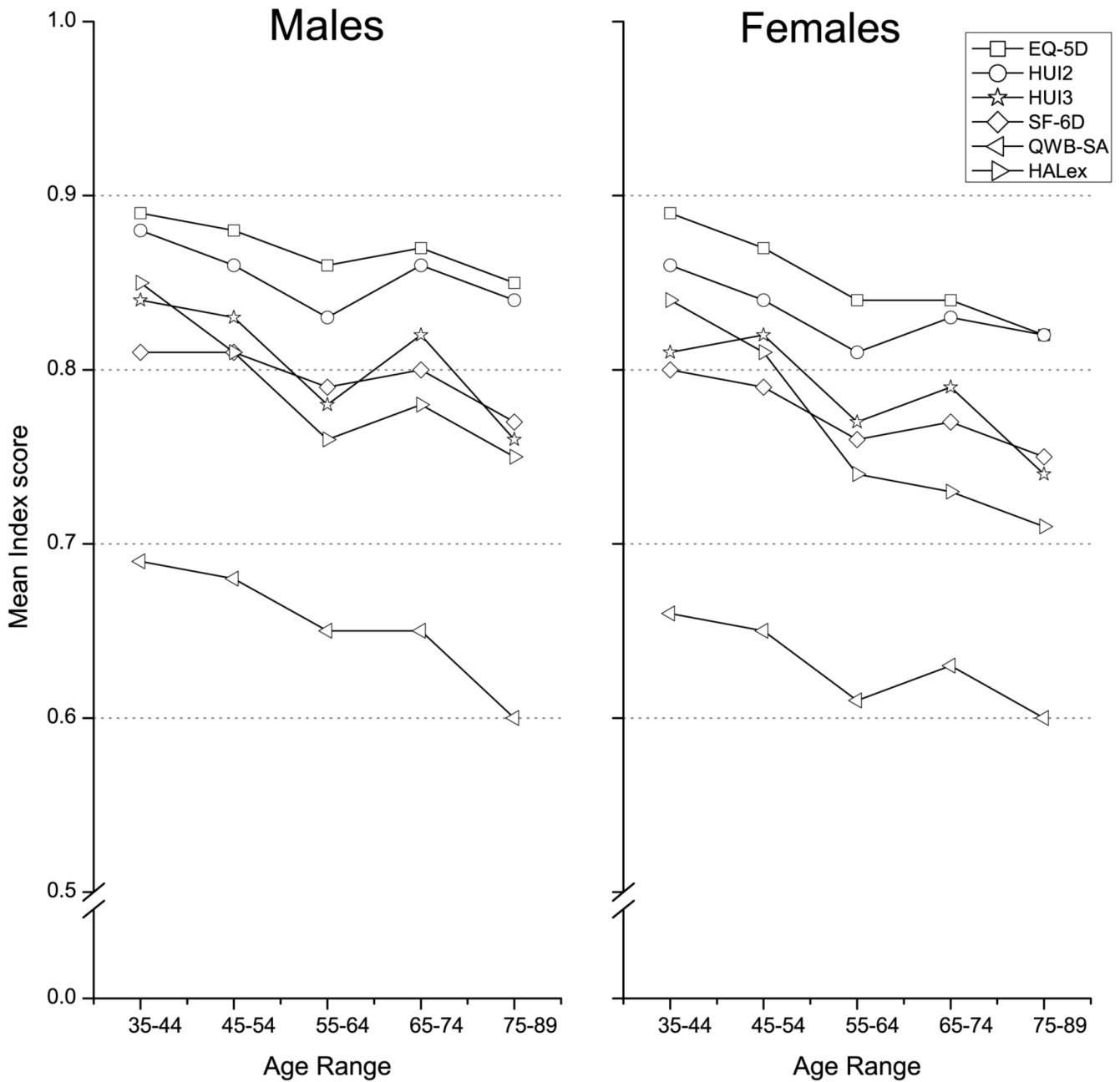Means estimating non-institutionalized US population health-related quality-of-life (HRQoL) by age are shown for males and females using 6 standardized indexes: EQ-5D, HUI2, HUI3, QWB-SA, SF-6D, and HALex. Cases were stratified by age range and gender and means computed within each stratum using survey-weighted regression of index scores on age centered to ages 40, 50, 60, 70, and 80 years as appropriate to the stratrum.

Table 1

Sample characteristics unweighted, weighted, and compared to U.S. Census 2000

| | | Sample | | |
| | | unweighted | weighted | U.S. Census 2000 [*] |
Item	N	%	%	%
Gender				(ages 35-89)
Male	1,641	42.7	47.1	47.4
Female	2,203	57.3	52.9	52.6
Age				
35-44	642	16.7	31.6	32.1
45-54	826	21.5	23.8	26.8
55-64	684	17.8	19.9	17.3
65-74	965	25.1	14.2	13.1
75-89	727	18.9	10.6	10.8
Race				(ages 35-89)
White	2,562	66.7	81.2	80.9
Black	1,086	28.3	10.3	10.5
Other Races	178	4.6	7.6	8.6
(Missing)	18	0.5	0.7	
Household Income				(householder aged 35+)
LT 20,000	781	20.3	10.4	21.5
20,000 - 34,999	699	18.2	14.6	18.3
35,000 - 74,999	1176	30.6	34.2	35.3
75,000 +	861	22.4	34.9	24.9
(Missing)	327	8.5	5.8	
Education (highest level)				(ages 25+)
< High School	464	12.1	8.2	19.6
High School	1159	30.2	28.2	28.6
Some post-High School	856	22.3	22.1	27.3
4-yr College degree or				
higher	1341	34.9	40.9	24.4
(Missing)	24	0.6	0.6	
Census Region				(ages 35-89)
Northeast	578	15.0	16.8	20.1
Midwest	1016	26.4	29.3	23.2
South	1884	49.0	35.0	35.8
West (excludes AK, HI)	366	9.5	18.9	20.9
Total	3844	100.0	100.0	100.0

[*] Census distributions from U.S. Census 2000; age range varies for different items as indicated. For household income, the respondent was not necessarily the householder making it difficult to match ages. Percentages may not total 100.0 due to rounding.

TABLE 2

Characteristics of the Six Indexes

Item	EQ-5D	HUI2[§]	HUI3	QWB-SA	SF-6D	HALex
Mean time to administer by telephone in minutes* (5%–tile, 95%-tile)	1.9 min. (1.4, 4.5)	3.4 min (2.1, 5.6)		11.1 min (7.7,17.5)	7.9 min (5.3, 12.3)	1.0 min (0.5, 1.8)
N missing (%)	32 (0.8)	277 (7.2)	286 (7.0)	86 (2.2)	105 (2.7)	3 (0.1)
Minimum value given not dead** (theoretical)	-0.11	-0.02[§]	-0.36	0.093	0.30	0.10
Minimum observed value (n)	-0.11 (n=1)	-0.01 (n=2)	-0.34 (n=2)	0.09 (n=1)	0.30 (n=12)	0.10 (n=54)
Lower quartile	0.800	0.770	0.670	0.540	0.667	0.630
Median	0.827	0.880	0.880	0.635	0.791	0.840
Upper Quartile	1.0	0.950	0.970	0.990	0.887	0.920
N scoring 1.0 (%)	1393 (36.2)	416 (10.8)	439 (11.4)	0 (0.0)	166 (4.3)	480 (12.5)
Weighted Correlations:[§§]						
HUI2	0.71	--	--	--	--	--
HUI3	0.68	0.89	--	--	--	--
QWB-SA	0.64	0.66	0.66	--	--	--
SF-6D	0.70	0.71	0.69	0.65	--	--
HALex	0.64	0.65	0.65	0.60	0.65	--

Note: Results other than correlations in this table are not survey sample weighted.

*
HUI2 and HUI3 are scored from the same set of questions; the SF-6D time is for the entire SF-36 and not just the 11 SF-6D questions; the HALex time excludes the self-rated health question from SF-36.

**
This was calculated by inserting answers for each question on the questionnaires which minimize the summary score. For all indexes, the health state "dead" is scored 0.0.

§
HUI2 computed without the optional fertility domain, minimum is -0.02.

§§
All correlations significant, $p < .001$.

TABLE 3

Mean HRQoL Index Scores of U.S. Adults Ages 35-89 by Gender and Age

Index:	Male		Female		Total	
age range	Mean	(s.e.)	Mean	(s.e.)	Mean	(s.e.)
EQ-5D:						
35-44	0.89	(0.01)	0.89	(0.01)	0.89	(0.01)
45-54	0.88	(0.01)	0.87	(0.01)	0.87	(0.01)
55-64	0.86	(0.01)	0.84	(0.01)	0.85	(0.01)
65-74	0.87	(0.01)	0.84	(0.01)	0.86	(0.01)
75-89	0.85	(0.01)	0.82	(0.01)	0.84	(0.01)
HUI2:						
35-44	0.88	(0.01)	0.86	(0.01)	0.87	(0.01)
45-54	0.86	(0.01)	0.84	(0.01)	0.85	(0.01)
55-64	0.83	(0.01)	0.81	(0.02)	0.82	(0.01)
65-74	0.86	(0.01)	0.83	(0.01)	0.85	(0.01)
75-89	0.84	(0.01)	0.82	(0.01)	0.83	(0.01)
HUI3:						
35-44	0.84	(0.02)	0.81	(0.02)	0.83	(0.01)
45-54	0.83	(0.01)	0.82	(0.01)	0.83	(0.01)
55-64	0.78	(0.02)	0.77	(0.02)	0.77	(0.02)
65-74	0.82	(0.02)	0.79	(0.02)	0.80	(0.01)
75-89	0.76	(0.02)	0.74	(0.02)	0.75	(0.01)
SF-6D:						
35-44	0.81	(0.01)	0.80	(0.01)	0.80	(0.01)
45-54	0.81	(0.01)	0.79	(0.01)	0.80	(0.01)
55-64	0.79	(0.01)	0.76	(0.01)	0.78	(0.01)
65-74	0.80	(0.01)	0.77	(0.01)	0.78	(0.01)
75-89	0.77	(0.01)	0.75	(0.01)	0.76	(0.01)
QWB-SA:						
35-44	0.69	(0.01)	0.66	(0.01)	0.67	(0.01)
45-54	0.68	(0.01)	0.65	(0.01)	0.66	(0.01)
55-64	0.65	(0.01)	0.61	(0.02)	0.63	(0.01)

Index:	Male		Female		Total	
age range	Mean	(s.e.)	Mean	(s.e.)	Mean	(s.e.)
65-74	0.65	(0.01)	0.63	(0.01)	0.64	(0.01)
75-89	0.60	(0.01)	0.60	(0.01)	0.60	(0.01)
HALex:						
35-44	0.85	(0.01)	0.84	(0.01)	0.85	(0.01)
45-54	0.81	(0.01)	0.81	(0.01)	0.81	(0.01)
55-64	0.76	(0.02)	0.74	(0.03)	0.75	(0.02)
65-74	0.78	(0.01)	0.73	(0.02)	0.76	(0.01)
75-89	0.75	(0.02)	0.71	(0.02)	0.73	(0.01)

Note: This table was computed first by stratifying cases into gender by age-range strata and then, within each stratum, using survey sample-weighted regression of index scores on ages centered at 40, 50, 60, 70, or 80 years as appropriate to the stratum. The intercept term in each regression then estimates the stratum mean index score for the U.S. population in that stratum at the time of the survey. The standard error of the regression intercept indicates precision of the estimate accounting for survey weights.