ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 497−506

Journal of Clinical Epidemiology

# Five preference-based indexes in cataract and heart failure patients were not equally responsive to change

Robert M. Kaplan[a,*], Steven Tally[b,c], Ron D. Hays[d], David Feeny[e], Theodore G. Ganiats[b,c], Mari Palta[f], Dennis G. Fryback[f]

[a]*Department of Health Services, University of California, Los Angeles, CA, USA*
[b]*Department of Family and Preventive Medicine, University of California, San Diego, CA, USA*
[c]*Health Services Research Center, University of California, San Diego, CA, USA*
[d]*Department of Medicine, University of California, Los Angeles, CA, USA*
[e]*Center for Health Research, Kaiser Permanente Northwest, Portland, OR, USA*
[f]*Department of Population Health Sciences, University of Wisconsin-Madison*

Accepted 22 April 2010

## Abstract

**Objective:** To compare the responsiveness to clinical change of five widely used preference-based health-related quality-of-life indexes in two longitudinal cohorts.

**Study Design and Setting:** Five generic instruments were simultaneously administered to 376 adults undergoing cataract surgery and 160 adults in heart failure management programs. Patients were assessed at baseline and reevaluated after 1 and 6 months. The measures were the Short Form (SF)-6D (based on responses scored from SF-36v2), Self-Administered Quality of Well-being Scale (QWB-SA), the EuroQol-5D developed by the EuroQol Group, the Health Utilities Indexes Mark 2 (HUI2) and Mark 3 (HUI3). Cataract patients completed the National Eye Institute Visual Functioning Questionnaire-25, and heart failure patients completed the Minnesota Living with Heart Failure Questionnaire. Responsiveness was estimated by the standardized response mean.

**Results:** For cataract patients, mean changes between baseline and 1-month follow-up for the generic indices ranged from 0.00 (SF-6D) to 0.052 (HUI3) and were statistically significant for all indexes except the SF-6D. For heart failure patients, only the SF-6D showed significant change from baseline to 1 month, whereas only the QWB-SA change was significant between 1 and 6 months.

**Conclusions:** Preference-based methods for measuring health outcomes are not equally responsive to change. 

*Keywords:* Quality of life; Measurement; Responsiveness; Cost−utility analysis; Quality-adjusted life years (QALY); Preference measure

## 1. Responsiveness of five preference-based indexes in cataract and heart failure patients

Estimates of quality-adjusted life years (QALYs) are required for several purposes, including population monitoring and cost-effectiveness analysis. For example, in the health objectives for the United States, one of two overarching goals is to increase the number of healthy years of life. Unfortunately, there has been no way to address this objective, because we do not have consensus on how to measure this construct. Similarly, there is increasing demand for cost-effectiveness evaluations in medicine and health care. Yet, comparisons of alternative health care investments are limited, because the measures of health outcome used for these analyses are not standardized. In this article, we compare alternative methods for estimating health outcome. The measures we compare are known as preference-based measures of health-related quality of life

Supported by Grant P01AG020679 from the National Institute on Aging. Drs. Kaplan and Hays were also provided support by National Institutes of Health (NIH) grant 1 P01 AG020679-01A2, University of California, Los Angeles (UCLA) Claude D. Pepper Older Americans Independence Center, NIH/National Institute of Aging (NIA) 5P30AG028748, and Center for Disease Control and Prevention (CDC) Grant U48 DP000056-04. Dr. Hays also received support from the UCLA Resource Center for Minority Aging Research/Center for Health Improvement in Minority Elderly (P30AG021684) and the UCLA/Drew Project EXPORT (P20MD000148 and P20MD000182). David Feeny has a proprietary interest in Health Utilities Incorporated, Dundas, Ontario, Canada. HUInc. distributes copyrighted Health Utilities Index (HUI) materials and provides methodological advice on the use of HUI. The authors also appreciate the help of Barbara Brody, MPH, and Denise Herman, MD, from University of California, San Diego (UCSD); Nancy Sweitzer, MD, PhD, and Neal Barney, MD, from UW; Greg Fonerow, MD, and John Bartlett, MD, from UCLA, for their collaboration on subject acquisition.

* Corresponding author. Department of Health Services, School of Public Health, University of California, CHS 31-293-C, Los Angeles, CA 90095, USA. Tel.: +1-310-825-7652; fax: +1-310-825-3317.

*E-mail address*: rmkaplan@ucla.edu (R.M. Kaplan).

0895-4356/$ - see front matter 
doi: 10.1016/j.jclinepi.2010.04.010

**What is new?**

- In two patient populations, this is the first head-to-head comparison of responsiveness to change for the five most common methods for estimating the quality-of-life component of quality-adjusted life years (QALYs).
- The measures are the Short Form-6D, the EuroQol-5D, the Self-Administered Quality of Well-being Scale, and two versions of the Health Utilities Indexes (HUI2 and HUI3).
- The results suggest that preference-based methods for measuring health outcomes are not equally responsive to change.
- Although most of the measures captured change after cataract surgery, the standardized response means varied.
- The results suggest that ''league tables,'' which are commonly used for policy analysis and resource allocation, may be misleading, because they assume that QALY change is interchangeable across assessment methods.

(HRQOL). These methods are required for cost–utility analysis and for population indicators, such as of QALYs and years of healthy life. The measures are a hybrid of two assessments. First, individuals are placed into observable levels of health status, typically on the basis of questionnaire responses. Then, these health states are weighted by level preference or utility on a continuum ranging from 0.0 for death to 1.0 for optimum health. The utility or preference weights can be provided by those who occupy the health states or by groups of external judges. The hybrid measures are used to adjust life expectancy for quality of life.

Investigators have multiple options when selecting preference-based measures for outcome studies. Measures are of little value if they are not responsive to the effects of health care interventions. In this article, we evaluate the responsiveness to change for the five most widely used preference-based HRQOL instruments. The measures are the Short Form-6D (SF-6D), the EuroQol (EQ)-5D, the Self-Administered Quality of Well-being Scale (QWB-SA), and two versions of the Health Utilities Index (HUI2 and HUI3). Furthermore, these generic measures were compared with disease-targeted measures: (1) for cataract patients, the National Eye Institute Visual Functioning Questionnaire-25 (NEI VFQ-25); and (2) for heart failure patients, the Minnesota Living with Heart Failure Questionnaire (MLHFQ).

The two patient populations were (1) those soon to undergo cataract extraction surgery with lens replacement and (2) patients newly referred to congestive heart failure clinics. The disease groups were selected, because they represent common health problems with different etiologies and expected HRQOL changes after treatment. Vision impairment affects people of all ages with the primary concentration in the elderly. For cataract surgery, significant sudden and noteworthy change after intervention is expected. Heart failure is a significant health problem that affects the cardiovascular system and is particularly common in older adults. Improvements after treatment are often small and may be transitory.

## 2. Methods

### *2.1. Subjects*

Subjects in both components of the study had to be at least 35 years old, able to give competent consent, able to hear and understand verbal instructions in English, and have sufficient vision and ability in reading and writing English to complete the questionnaires. For the vision impairment component of the study, patients were excluded if they were undergoing simultaneous glaucoma, corneal, or vitroretinal procedures. Patients with traumatic cataract and with visual impairment so severe that they were unable to read a large print version of the self-administered questionnaire were also excluded.

Heart failure patients were included if there was evidence of the presence of heart failure for at least 3 months, defined as a left ventricular ejection fraction less than 40%. Patients were excluded if their New York Heart Association classification was class IV, if they had a recent myocardial infarction (less than 6 months), if they had an unstable angina, or if they had a coronary artery bypass graft within the last 3 months. Patients were also excluded if they were on a heart transplant list or if they experienced symptomatic or sustained ventricular tachycardia during the previous 3 months that was not controlled by medical therapy or a defibrillator.

Participants were recruited from clinical sites at three academic medical centers: The University of California, Los Angeles (UCLA); the University of California, San Diego (UCSD); and the University of Wisconsin. In addition, some participants in the cataract component were recruited from the University of Southern California.

At enrollment, patients were given the measurement packet (and a self-addressed, stamped return envelope) to take home, complete, and return by mail, within 7 days, to the project's data collection coordinator, the UCSD Health Services Research Center (HSRC). HSRC mailed out the same measurement packet at the 1- and 6-month follow-ups along with a postage-paid self-addressed return envelope to each study participant who returned the baseline questionnaires [1].

### 2.2. Measures

The measures administered at baseline and at 1 and 6 months are described briefly in the following sections.

#### 2.2.1. Short Form-6D

Perhaps, the most commonly used outcome measure in the world today is the Medical Outcome Study SF-36 Health Survey. The SF-36 grew out of work by the RAND Corporation in the Medical Outcomes Study (MOS) [2]. The SF-36v2 includes eight health concepts: physical functioning, role limitations because of physical health problems, bodily pain, general health perceptions, vitality, social functioning, role limitations because of emotional problems, and mental health [3]. The SF-36v2 can be either administered by a trained interviewer or self-administered. There is substantial evidence supporting the reliability and validity of the SF-36v2 [4–6].

Our study focuses on preference-based outcome measures. Although the SF-36v2 is not a preference-based measure, Brazier et al. obtained independent utility ratings of 249 health states composed of SF-36 components. They used these ratings to estimate health state evaluations for 18,000 states that could be derived from a subset of the SF-36v2 items [7]. The method is known as the SF-6D.

#### 2.2.2. EuroQol-5D

The EQ-5D was developed by a collaborative group from Western Europe known as the EuroQol group [8]. The EQ-5D questions refer to "your health today." The EQ-5D descriptive system uses five domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). For each domain, the respondent is asked to describe his or her health on that day using three response options (no problems, moderate problems, severe problems). The five domains combined with the three response options yields 5 [2] or 243 unique health states. Adding perfect health and death gives 245 possible states [9]. Although the EQ-5D was originally validated in Europe, a scoring algorithm derived for the U.S. general population is available, and it was applied in this study. This scoring algorithm was derived from time trade-off assessments of EQ-5D health states made by a population sample of some 4,000 U.S. adults in face-to-face household interviews [10].

The EQ-5D is now used in a substantial number of clinical and population studies. [11,12] Although the EQ-5D is easy to use and comprehensive, there have been some concerns about ceiling effects. Substantial numbers of people obtain the highest possible score. However, we did not anticipate this problem in the current study, as all participants were recruited because they had serious medical conditions. Information on the EQ-5D is available at http://www.euroqol.org.

#### 2.2.3. Self-Administered Quality of Well-being Scale

The QWB-SA assesses self-reported functioning using a series of questions designed to record limitations over the previous 3 days, within three separate domains (Mobility, Physical Activity, and Social Activity). In addition, the QWB-SA includes a series of questions that ask about the presence or absence of different symptom/problem complexes. The four domain scores are combined into a total score that provides a numerical point-in-time expression of well-being that ranges from 0 for dead to 1.0 for asymptomatic optimum functioning. The original QWB obtained preference ratings of 856 people from the general population. The QWB-SA used convenience samples to model preference for case descriptions, and the models were shown to be highly correlated with the population ratings in the original QWB general-population preferences elicitation study [13].

The self-administered QWB-SA has been shown to be highly correlated with the interviewer-administered QWB and to retain the psychometric properties [13]. Extensive evaluations of reliability and validity have been published [14–19]. Access to the measure and details about its development are available at http://qwbsa.ucsd.edu.

#### 2.2.4. Health Utilities Index

The HUI is a family of health status and preference-based HRQOL measures [20,21]. Each member of the family includes a health status classification system, a preference-based multiattribute utility function, data collection questionnaires, and algorithms for deriving HUI variables from questionnaire responses. HUI focuses on capacity rather than performance. This study used the HUI Mark 2 (HUI2) and Mark 3 (HUI3). HUI2 consists of six dimensions of health status: sensation (vision, hearing, speech); mobility; emotion; cognition; self-care; and pain [21]. HUI3 includes eight dimensions of health status: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain and discomfort, with five or six levels per attribute. Multiplicative multiattribute utility functions based on community preferences have been estimated for HUI2 [22] and HUI3 [23]. The utility function was derived to represent preference for attributes and interaction among the attributes. Evidence supporting construct validity (including responsiveness) of the HUI has been published [24–27]. Reference information on the HUI is available at http://www.fhs.mcmaster.ca/hug/ and http://www.healthutilities.com/

### 2.3. Disease-targeted measures

#### 2.3.1. National Eye Institute Visual Functioning Questionnaire-25

Participants in the Cataract study were also evaluated using the NEI 25-item VFQ-25. The VFQ-25 was developed by RAND and UCLA with support from the NEI. The VFQ-25 assesses self-reported vision-related functioning

and well-being. There is extensive support for the reliability and validity of the VFQ-25 [28,29].

#### 2.3.2. Minnesota Living with Heart Failure Questionnaire

The MLHFQ assesses the extent to which heart failure affects daily life. The 21 MLFHQ items can be completed in 5–10 minutes. The content covers the most frequent and important ways heart failure affects daily functioning. The MLFHQ yields an overall score and two subscores: physical and emotional. Support for reliability and validity of the MLHQ is provided at www.mlhfq.org [30,31].

### 2.4. Statistical analysis

Number and percentage of cataract and congestive heart failured (CHF) study participants in age, race, education, and gender categories are given. We also provide the number of patients with data at each time point for each HRQOL measure. We estimate the change in HRQOL scores as the differences between 1-month and baseline scores and between 6- and 1-month scores. Statistical significance of the change was assessed by paired *t*-tests. Linear trend was modeled with time points equally spaced, by mixed models with random intercept and slope. The standardized response mean (SRM), defined as the mean change divided by the standard deviation of change, was used as the indicator of responsiveness. Pearson correlations among the five generic indexes and the respective disease-targeted measures are presented for cataract and heart failure patients separately.

## 3. Results

A total of 536 adults participated in the study. Among these, 376 were recruited, because they had cataract disease, and 160 had been diagnosed with heart failure. Demographic characteristics of the two groups are summarized in Table 1. Most of the patients were whites (87% for cataract and 79% for heart failure). The cataract patients tended to be females (59%), with most being 65 years or older. The heart failure patients tended to be males (67%) and somewhat younger; 78% were aged less than 65 years.

The number of subjects participating in each follow-up is shown in Table 2. Figure 1 shows the distributions on all generic measures at baseline for cataract and for heart failure patients. There was strong negative skew for the EQ-5D, the HUI2, and the HUI3. The distributions for the QWB-SA and SF-36 were nearly normal.

Results for the cataract patients are summarized in Table 3. Differences between the baseline and 1-month follow-up are shown in the table. At 1 month, differences were statistically significant for all indexes except the SF-6D. The mean difference in scores range from −0.005 (for the SF-6D, SRM = −0.05) to 0.052 (for the HUI3, SRM = 0.25).

Table 1
Demographic characteristics of the samples

| Characteristic | Cataract patients, *n* (%) | Heart failure patients, *n* (%) |
|---|---|---|
| Age (years) | | |
| 35–44 | 5 (1) | 24 (15) |
| 45–64 | 115 (31) | 101 (63) |
| 65–91 | 256 (68) | 35 (22) |
| Race | | |
| White | 328 (87) | 126 (79) |
| Black | 12 (3) | 19 (12) |
| Asian | 19 (5) | 5 (3) |
| Other | (1) | 2 (1) |
| Missing | 13 (3) | 8 (5) |
| Education | | |
| Less than high school | 21 (6) | 20 (13) |
| High school graduate | 60 (16) | 45 (28) |
| Some college | 78 (21) | 47 (29) |
| 2-yr Associate degree | 27 (7) | 12 (8) |
| 4-yr College graduation | 90 (24) | 16 (10) |
| Master's degree | 57 (15) | 9 (6) |
| Doctorate or professional | 34 (9) | 6 (4) |
| Missing | 9 (2) | 5 (3) |
| Female | 222 (59) | 52 (33) |

Hence, the absolute differences that would be used to calculate QALYs were quite different across measures. For example, if we assume these differences last for 10 years, the EQ-5D difference of 0.017 units would produce a difference of 0.17 QALYs (undiscounted) or one QALY for every six patients, whereas the HUI3 would produce 0.52 QALYS or one QALY for less than every two patients.

The largest SRMs were observed for change between baseline and 1-month follow-up for the VFQ-25 (SRM = 0.77) and the HUI3 (SRM = 0.25). The SRMs

Table 2
Number of patients in each group at each time point

| Index | Baseline | 1mo | 6mo |
|---|---|---|---|
| Cataract group total | 376 | 315 | 302 |
| VFQ-25 | 361 | 309 | 293 |
| SF-6D | 351 | 298 | 286 |
| QWB-SA | 376 | 315 | 302 |
| EQ-5D | 369 | 308 | 288 |
| HUI2 | 352 | 306 | 290 |
| HUI3 | 355 | 304 | 289 |
| CHF group total | 160 | 138 | 110 |
| MLHFQ | 160 | 138 | 110 |
| SF-6D | 152 | 133 | 107 |
| QWB-SA | 160 | 138 | 110 |
| EQ-5D | 155 | 136 | 110 |
| HUI2 | 152 | 133 | 109 |
| HUI3 | 151 | 133 | 109 |

*Abbreviations:* VFQ-25, Visual Functioning Questionnaire-2; SF-6D, Short Form-6D; QWB-SA, Self-Administered Quality of Well-being Scale; EQ-5D, EuroQol-5D; HUI2, Health Utilities Index Mark 2; HUI3, Health Utilities Index Mark 3; MLHFQ, Minnesota Living with Heart Failure Questionnaire.

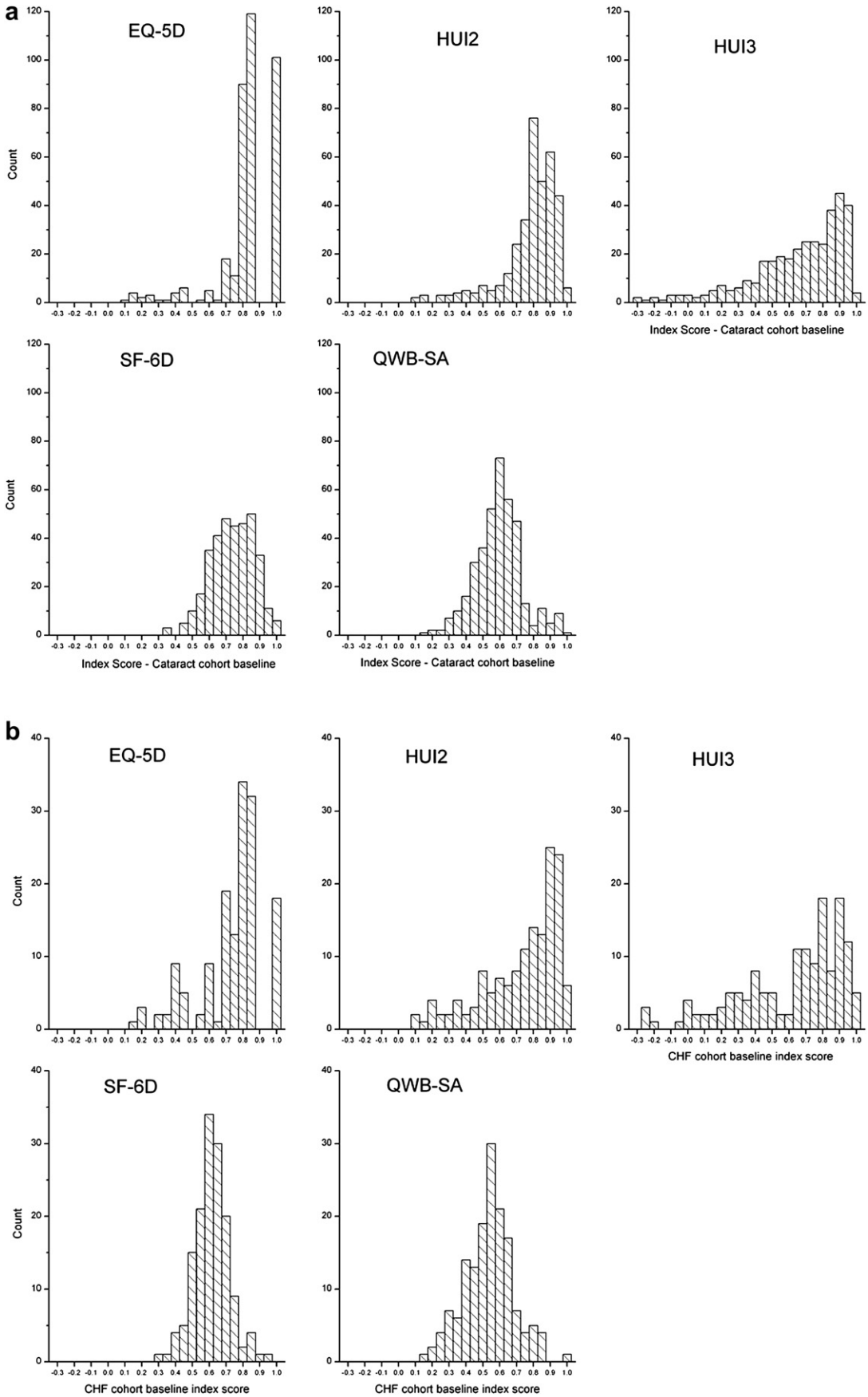


Fig. 1. Distributions of baseline scores on 5 indexes at baseline for cataract (a) and heart failure patients (b). EQ-5D, EuroQol-5D; HUI2, Health Utilities Index Mark 2; HUI3, Health Utilities Index Mark 3; SF-6D, Short Form-6D; QWB-SA, Self-Administered Quality of Well-being Scale.

Table 3
Changes in cataract patients between baseline and 1 month and between 1 and 6 months by index

| Index | *N* | Mean difference | SD | *t*-Value | SRM |
|---|---|---|---|---|---|
| Difference (SD): 1month−baseline | | | | | |
| VFQ-25 | 297 | 8.74 | 11.32 | 13.30 | 0.77 |
| SF-6D | 284 | −0.005 | 0.09 | −0.87 | −0.05 |
| QWB-SA | 315 | 0.018 | 0.13 | 2.41 | 0.14 |
| EQ-5D | 303 | 0.017 | 0.12 | 2.53 | 0.15 |
| HUI2 | 286 | 0.030 | 0.14 | 3.75 | 0.22 |
| HUI3 | 286 | 0.052 | 0.21 | 4.21 | 0.25 |
| Difference (SD): 6month−1month | | | | | |
| VFQ-25 | 257 | 1.92 | 7.40 | 4.16 | 0.260 |
| SF-6D | 240 | 0.007 | 0.10 | 1.04 | 0.067 |
| QWB-SA | 269 | 0.004 | 0.12 | 0.59 | 0.036 |
| EQ-5D | 250 | −0.027 | 0.14 | −3.10 | −0.190 |
| HUI2 | 251 | 0.003 | 0.15 | 0.35 | 0.022 |
| HUI3 | 249 | 0.008 | 0.20 | 0.64 | 0.040 |

*Abbreviations:* SD, standard deviation; SRM, standardized response mean.

for the other measures were smaller, and as noted earlier, the differences on the SF-6D were not statistically significant.

The lower portion of Table 3 shows changes in the cataract patients between 1 and 6 months. The analysis suggests that after 1 month, HRQOL scores remain stable for all indexes, although there is a significant reduction of −0.027 for the EQ-5D (SRM = −0.19). The SRMs for the other generic measures were all less than 0.10. Considering the three time points (baseline, 1 month, 6 months), there was a significant linear trend for improved quality of life for the QWB-SA ($t = 3.85$, $P < 0.0001$) HUI2 ($t = 3.31$, $P < 0.001$) and HUI3 ($t = 4.58$, $P < 0.0001$). There was also a strong linear trend for the VFQ-25 ($t = 18.31$, $P < 0.0001$). Trends for the SF-6D ($t = 0.78$, $P = 0.43$) and EQ-5D ($t = -1.15$, $P = 0.25$) were nonsignificant.

Table 4
Changes in heart failure patients between baseline and 1 month and between 1 and 6 months by index

| Index | *N* | Mean difference | SD | *t*-Value | SRM |
|---|---|---|---|---|---|
| Difference (SD): 1month−baseline | | | | | |
| MLHFQ | 138 | −3.90 | 21.51 | −2.13 | −0.18 |
| SF-6D | 126 | 0.022 | 0.087 | 2.91 | 0.26 |
| QWB-SA | 138 | 0.007 | 0.12 | 0.65 | 0.06 |
| EQ-5D | 132 | 0.005 | 0.14 | 0.38 | 0.03 |
| HUI2 | 127 | 0.009 | 0.16 | 0.66 | 0.06 |
| HUI3 | 126 | 0.012 | 0.20 | 0.69 | 0.06 |
| Difference (SD): 6month−1month | | | | | |
| MLHFQ | 107 | −4.72 | 17.80 | −2.74 | −0.26 |
| SF-6D | 101 | 0.014 | 0.086 | 1.66 | 0.17 |
| QWB-SA | 107 | 0.031 | 0.12 | 2.60 | 0.25 |
| EQ-5D | 105 | −0.000 | 0.16 | −0.03 | −0.00 |
| HUI2 | 102 | 0.003 | 0.15 | 0.22 | 0.02 |
| HUI3 | 102 | 0.020 | 0.20 | 1.00 | 0.10 |

MLHFQ is scored so that a higher score is worse health-related quality of life.

Results for the heart failure group are shown in Table 4. In contrast to the cataract analysis, only the SF-6D changed significantly between baseline and 1 month (Table 4). After 1 month, the heart failure patients remained stable on all measures, except the QWB-SA, which suggested some continued improvement (Table 4). The disease-targeted MLHFQ was no more responsive to change (SRM = −0.26) than was the generic QWB-SA (SRM = 0.287). The difference in signs for change between these measures occurs, because the low scores on the MLHFQ indicate better health, whereas low scores on the QWB-SA suggest poor health. Changes on the other measures were not statistically significant.

Changes over time for the QWB-SA and the HUI3 are shown in Fig. 2. Treatment for heart failure is expected to produce gradual gains [32], whereas cataract surgery is expected to produce change shortly after treatment [33]. Figure 2 summarizes the mean scores on each index at baseline, 1 month, and 6 months. For reference, the top line on the figure gives the estimated score on this index for the general population, matched to the average age of the participants in this study (65.5 years). This estimate comes from the National Health Measurement Study [34]. The figures show that cataract patients gain most of their improvement by 1 month (first two graphs in Fig. 2a). In contrast, heart failure patients continue to improve up to the 6-month evaluation on the QWB-SA and MLHFQ, although the magnitude of improvement tends to be very modest (Fig. 2b). Similar patterns were seen for the other measures (data not shown). Considering the three time points (baseline, 1 month, and 6 months), there was a significant linear trend for improved quality of life for the QWB-SA ($t = 2.84$, $P < 0.005$) and SF-6D ($t = 4.39$, $P < 0.0001$). The linear trend for the MLHFQ was comparable in strength to most of the generic measures ($t = -5.10$, $P < 0.0001$). The trend was nonsignificant for the EQ-5D ($t = 0.67$, $P = 0.50$), HUI2 ($t = 0.70$, $P = 0.48$), and the HUI3 ($t = 1.40$, $P = 0.16$).

Although statistically significant, the SRMs for the cataract study were all modest. They ranged from −0.05 for the SF-6D to 0.25 for the HUI3. All were lower than the moderate SRM for the VFQ (0.77). For the heart failure study, most of the change occurred between 1 and 6 months. The 1- to 6-month SRM observed with the disease-targeted MLHFQ was no larger (SRM = −0.26) than that for the generic QWB-SA (SRM = 0.25). SRMs for the other indexes were not noteworthy.

Results reported earlier indicate that the estimates of the amount of change and the degree of responsiveness vary across the measures. This raises the question about the degree of association among the measures. Correlations between baseline scores for the generic and disease-targeted measures are shown in Table 5 for the cataract patients and Table 6 for the heart failure patients. Table 5 shows that the generic measures are highly intercorrelated among cataract patients (*r*'s are 0.53 or higher) and that each measure is substantially correlated with the VFQ-25. Table 6 offers

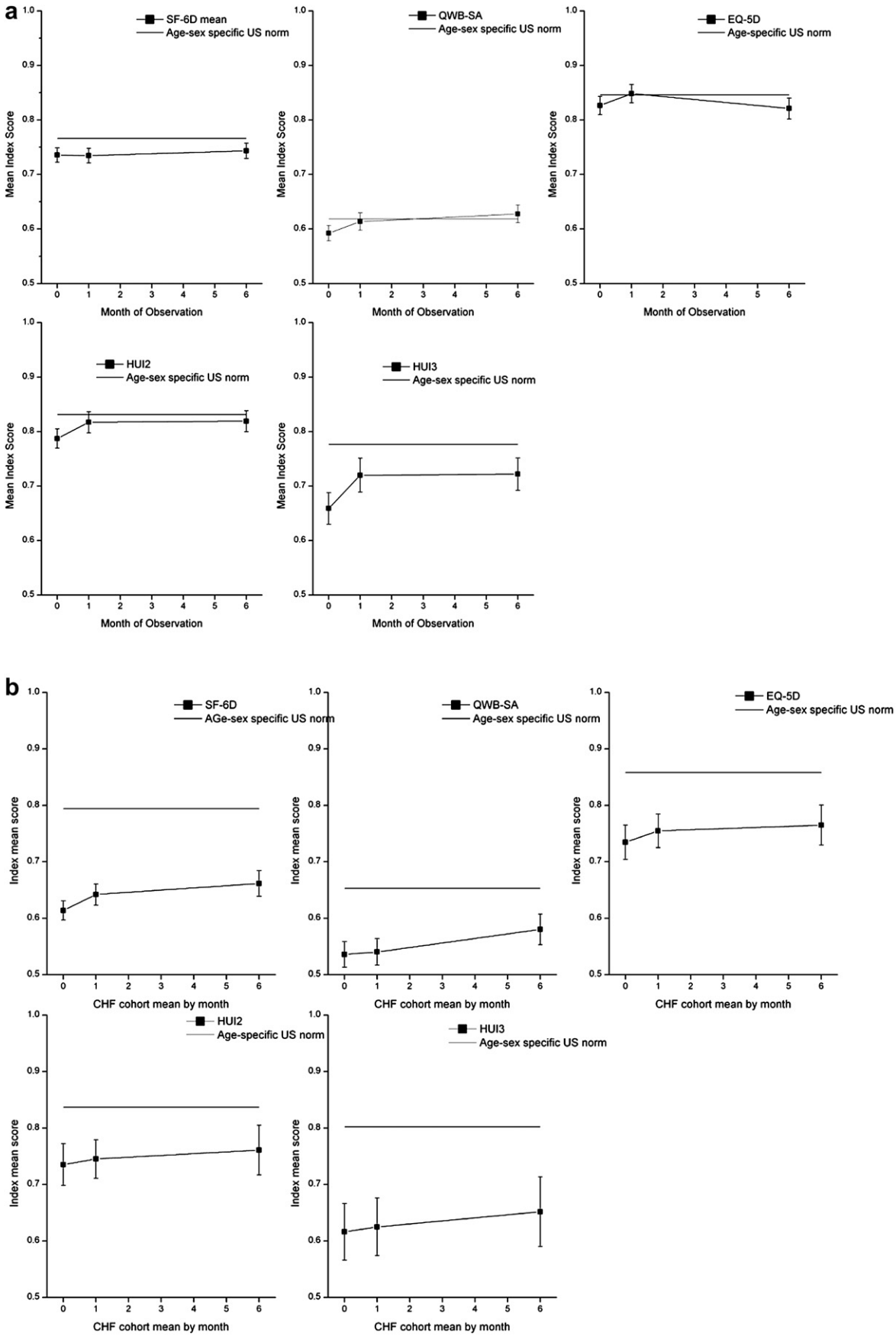


Fig. 2. QWB-SA and HUI3 over time in National Health Measurement Study (NHMS) for cataract patients (a) and heart failure patients (b). (a) Cataract cohort index means by month of observation; (b) CHF cohort means by month of observation. Vertical bars indicate 95% confidence intervals.

Table 5
Correlations between generic measures and the VFQ for cataract patients at baseline

| | VFQ-25 | QWB-SA | HUI2 | HUI3 | EQ-5D | SF-6D |
|---|---|---|---|---|---|---|
| VFQ-25 | 1.00 | 0.49 | 0.52 | 0.58 | 0.50 | 0.54 |
| QWB-SA | | 1.00 | 0.53 | 0.60 | 0.55 | 0.54 |
| HUI2 | | | 1.00 | 0.88 | 0.77 | 0.67 |
| HUI3 | | | | 1.00 | 0.70* | 0.67 |
| EQ-5D | | | | | 1.00 | 0.68 |
| SF-6D | | | | | | 1.00 |

All correlations were statistically significant: $P < 0.01$.

a similar story for the heart failure patients; the generic measures tended to be substantially correlated with the MLHFQ. Even though the measures have noteworthy associations, they produce different estimates of change and differ in responsiveness.

## 4. Discussion

At least five preference-based measures of HRQOL can be used for cost–utility analysis. Our analysis suggests that the five measures are not equally responsive to change after cataract surgery or medical management of heart failure. Among the measures we considered, the SF-6D tended to be an outlier. It did not appear to capture the same change as the other measures. This might be expected, because the SF-6D was derived from a different measurement tradition than the other measures. The SF-6D is built upon responses to the SF-36v2 questionnaire. Clearly, there are substantial similarities among the measurement systems; each has a health status classification system, questionnaires, algorithms to derive health status vectors from questionnaire responses, and algorithms for generating preference-based overall scores. However, the HUI2, HUI3, EQ-5D, and the QWB-SA were developed with the intention of developing a preference-based scoring function to provide overall summary scores on the conventional 0 = dead to 1.0 = perfect health scale. Both HUI measures and the EQ-5D allow for scores lower than 0.0. The original intent with the SF-36 was to generate eight domain scores. Later, the two summaries, physical and mental, were added. Much later, the algorithm for providing preference-based scores was added.

The HUI2, HUI3, EQ-5D, and QWB-SA were developed with the intention of creating a health status classification system. The plan for the measures included the development of a multiattribute utility function. This planning affected choices about which dimensions of health status to include and the relationship among those dimensions. The plan for the SF-36 was more focused on producing a profile of HRQOL domain scores.

Table 6
Correlations between generic measures and the MLFQ for heart failure patients at baseline

| Index | MLHF | QWB-SA | HUI2 | HUI3 | EQ-5D | SF-6D |
|---|---|---|---|---|---|---|
| MLHF | 1.00 | −0.64 | −0.52 | −0.49 | −0.60 | −0.65 |
| QWB-SA | | 1.00 | 0.56 | 0.58 | 0.58 | 0.58 |
| HUI2 | | | 1.00 | 0.88 | 0.68 | 0.63 |
| HUI3 | | | | 1.00 | 0.67 | 0.63 |
| EQ-5D | | | | | 1.00 | 0.64 |
| SF-6D | | | | | | 1.00 |

All correlations were statistically significant: $P < 0.01$.

For both cataract and heart failure patients, the generic utility measures (EQ-5D, QWB-SA, HUI2, HUI3) tended to detect change in the same direction. The absolute differences captured by the measures varied. In the cataract study, the generic measures were able to capture change, but with a lower level of responsiveness than the disease-targeted measure. In the heart failure study, at least one generic measure was as responsive as the disease-targeted measure. Overall, there was probably a much weaker signal (i.e., less change to be detected) in the heart failure group.

Several other authors have reported differences in responsiveness between measures. Blanchard et al. [35] compared HUI2, HUI3, and SF-36 with a variety of disease-targeted measures for patients undergoing total-hip arthroplasty. They found the disease-targeted measures more responsive than the generic measures. However, similar to our results, the generic measures were also significantly responsive to change. In future analyses, we hope to report the associations between change captured by the self-report measures and clinical measures of change.

When disease-targeted measures are more responsive than generic measures, they provide important additional information. However, disease-targeted measures are not designed to be used for analyses that inform resource allocation decisions. Policy makers are faced with requests for resources from programs with very different specific objectives. The best way for them to choose between the competing alternatives is to apply measures that allow the comparison of outcomes in common units [36]. Although some investigators are now estimating "utilities" from disease-targeted measures [37,38], comparisons across studies can be difficult because of the potential for noncomparability of the measures.

The content of the different generic measures may help explain the differential responsiveness. For example, the QWB-SA and the HUI measures were more responsive to change after cataract surgery than were the EQ-5D and the SF-6D. One explanation for this greater responsiveness is the fact that the QWB-SA and HUI measures contain information about sensory functioning. The HUI measures include a component for sensory functioning, whereas the QWB-SA has a section on symptoms and problems. These symptoms include trouble seeing and other components of visual functioning. Other studies have confirmed the responsiveness of the HUI3 [39] and the QWB-SA [33] for patients with cataract disease. However, these measures have more items than some alternative tools. One of the major challenges in developing generic measures is to be both brief and comprehensive. When measures are too brief, they may sacrifice some comprehensiveness and responsiveness.

In summary, generic measures are capable of capturing changes between baseline and 1-month follow-up for patients undergoing cataract extraction with lens replacement. For heart failure patients, responsiveness was less well documented. Only the SF-6D showed significant change from baseline to 1 month, and differences between 1 and 6 months were only captured by the QWB-SA. On most of the measures, cataract patients gained most of their improvement by 1 month. At least on some measures (QWB-SA and SF-6D), heart failure patients continued to improve over the 6 months of study. However, for both clinical groups, the magnitude of change was not consistent across measures. Only the QWB-SA captured significant linear trends in both disease groups.

Preference-based measures are necessary to estimate QALYs for cost−utility analysis. Separate measures are available for this purpose, and there is no consensus on which measure is best. The competing measures capture similar information on change among patients undergoing cataract extraction or comprehensive care for heart failure. However, the measures are not equally sensitive to change, and the estimates of QALYs resulting from treatment may differ as a function of the choice of measurement instrument. More research is necessary to identify the sensitivity and specificity of leading preference-based generic measures of health outcome when applied in different clinical populations.

## References


[1] Hays RD, Kim S, Spritzer KL, Kaplan RM, Tally S, Feeny D, et al. Effects of mode and order of administration on generic health-related quality of life scores. Value Health 2009;12:1035−9.

[2] Ware JE Jr, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. J Clin Epidemiol 1998;51:903−12.

[3] Kosinski M, Keller SD, Hatoum HT, Kong SX, Ware JE Jr. The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: tests of data quality, scaling assumptions and score reliability. Med Care 1999;37:MS10−22.

[4] Stewart AL, Ware JE. Measuring functioning and well-being: the Medical Outcomes Study approach. Durham, NC: Duke University Press; 1992.

[5] Scott-Lennox JA, Wu AW, Boyer JG, Ware JE Jr. Reliability and validity of French, German, Italian, Dutch, and UK English translations of the Medical Outcomes Study HIV Health Survey. Med Care 1999;37:908−25.

[6] Keller SD, Ware JE Jr, Hatoum HT, Kong SX. The SF-36 Arthritis-Specific Health Index (ASHI): II. Tests of validity in four clinical trials. Med Care 1999;37:MS51−60.

[7] 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.

[8] Kind P. The performance characteristics of EQ-5D, a measure of health related quality of life for use in technology assessment. Annu Meet Int Soc Technol Assess Health Care 1997;13:81. [abstract].

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

[10] 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;43:203−20.

[11] Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol 1997;36:551−9.

[12] Gudex C, Dolan P, Kind P, Williams A. Health state valuations from the general public using the visual analogue scale. Qual Life Res 1996;5:521−31.

[13] 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;12:783−91.

[14] Kaplan R, Anderson J. The general health policy model: an integrated approach. In: Spilker B, editor. Quality of life and pharmacoeconomics in clinical trials. New York: Raven; 1996. p. 309−22.

[15] Kaplan RM. Decision making in medicine and health care. Annu Rev Clin Psychol 2005;1:525−56.

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

[17] Kaplan RM, Anderson JP, Patterson TL, McCutchan JA, Weinrich JD, Heaton RK, et al. Validity of the Quality of Well-Being Scale for persons with human immunodeficiency virus infection. HNRC Group. HIV Neurobehavioral Research Center. Psychosom Med 1995;57:138−47.

[18] Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. Health Serv Res 1976;11:478−507.

[19] Kaplan RM, Ganiats TG, Sieber WJ, Anderson JP. The Quality of Well-Being Scale: critical similarities and differences with SF-36. Int J Qual Health Care 1998;10:509−20. [see comments].

[20] Feeny D, Furlong W, Mulhern RK, Barr RD, Hudson M. A framework for assessing health-related quality of life among children with cancer. Int J Cancer Suppl 1999;12:2−9.

[21] Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. Health Utilities Index. Pharmacoeconomics 1995;7:490−502.

[22] Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. Med Care 1996;34:702−22.

[23] Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care 2002;40:113−28.

[24] Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. Ann Med 2001;33:375−84.

[25] Barr RD, Chalmers D, De Pauw S, Furlong W, Weitzman S, Feeny D. Health-related quality of life in survivors of Wilms' tumor and advanced neuroblastoma: a cross-sectional study. J Clin Oncol 2000;18:3280−7.

[26] Feeny D, Furlong W, Barr RD. Multiattribute approach to the assessment of health-related quality of life: Health Utilities Index. Med Pediatr Oncol 1998;(Suppl 1):54−9.

[27] Barr RD, Furlong W, Feeny D. Comments on health-related quality of life of adults surviving malignancies in childhood, Apajasalo et al. Eur J Cancer 1996;32A:1354−8. Eur J Cancer 1997;33:506−7.

[28] Mangione CM, Lee PP, Gutierrez PR, Spritzer K, Berry S, Hays RD. Development of the 25-item National Eye Institute Visual Function Questionnaire. Arch Ophthalmol 2001;119:1050−8.

[29] Mangione CM, Lee PP, Pitts J, Gutierrez P, Berry S, Hays RD. Psychometric properties of the National Eye Institute Visual Function Questionnaire (NEI-VFQ). NEI-VFQ Field Test Investigators. Arch Ophthalmol 1998;116:1496−504.

[30] Rector TS. A conceptual model of quality of life in relation to heart failure. J Card Fail 2005;11:173−6.

[31] Rector TS, Kubo SH, Cohn JN. Validity of the Minnesota Living with Heart Failure questionnaire as a measure of therapeutic response to enalapril or placebo. Am J Cardiol 1993;71:1106−7.

[32] van Tol BAF, Huijsmans RJ, Kroon DW, Schothorst M, Kwakkel G. Effects of exercise training on cardiac performance, exercise capacity and quality of life in patients with heart failure: a meta-analysis. Eur J Heart Fail 2006;8:841−50.

[33] Rosen PN, Kaplan RM, David K. Measuring outcomes of cataract surgery using the Quality of Well-Being Scale and VF-14 Visual Function Index. J Cataract Refract Surg 2005;31:369−78.

[34] Fryback DG, Dunham NC, Palta M, Hanmer J, Buechner J, Cherepanov D, et al. US norms for six generic health-related quality-of-life indexes from the National Health Measurement study. Med Care 2007;45:1162−70.

[35] Blanchard C, Feeny D, Mahon JL, Bourne R, Rorabeck C, Stitt L, et al. Is the Health Utilities Index valid in total hip arthroplasty patients? Qual Life Res 2004;13:339−48.

[36] Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996;276: 1172−7. [see comments].

[37] Misajon R, Hawthorne G, Richardson J, Barton J, Peacock S, Iezzi A, et al. Vision and quality of life: the development of a utility measure. Invest Ophthalmol Vis Sci 2005;46:4007−15.

[38] Peacock S, Misajon R, Iezzi A, Richardson J, Hawthorne G, Keeffe J. Vision and quality of life: development of methods for the VisQoL vision-related utility instrument. Ophthalmic Epidemiol 2008;15: 218−23.

[39] Asakawa K, Rolfson D, Senthilselvan A, Feeny D, Johnson JA. Health Utilities Index Mark 3 showed valid in Alzheimer disease, arthritis, and cataracts. J Clin Epidemiol 2008;61:733−9.