



Published in final edited form as:

J Clin Epidemiol. 2011 May ; 64(5): 497–506. doi:10.1016/j.jclinepi.2010.04.010.

Five Preference-Based Indexes In Cataract And Heart Failure Patients Were Not Equally Responsive To Change

Robert M. Kaplan,

Department of Health Services, University of California, Los Angeles

Steven Tally,

Department of Family and Preventive Medicine, and Health Services Research Center, University of California, San Diego

Ron D. Hays,

Department of Medicine, University of California, Los Angeles

David Feeny,

Center for Health Research, Kaiser Permanente Northwest

Theodore G. Ganiats,

Family and Preventive Medicine, and Health Services Research Center, University of California, San Diego

Mari Palta, and

Department of Population Health Sciences, University of Wisconsin-Madison

Dennis G. Fryback

Department of Population Health Sciences, University of Wisconsin-Madison

Abstract

OBJECTIVE—To compare the responsiveness to clinical change of 5 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 SF-6D (based on responses scored from SF-36v2TM), Self-Administered Quality of Well-being scale (QWB-SA), the EQ-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 (VFQ-25) and heart failure patients completed the Minnesota Living with Heart Failure Questionnaire (MLHFQ). Responsiveness was estimated by the Standardized Response Mean (SRM).

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, while only the QWB-SA change was significant between 1 and 6 months.

© 2010 Elsevier Inc. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

CONCLUSIONS—Preference-based methods for measuring health outcomes are not equally responsive to change.

Keywords

Quality of Life; Measurement; Responsiveness; Cost-Utility Analysis

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 healthcare. Yet, comparisons of alternative healthcare investments are limited because the measures of health outcome used for these analyses are not standardized. In this paper we compare alternative methods for estimating health outcome. The measures we compare are known as preference-based measures of health-related quality of life (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 levels 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 paper, 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 to 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 following treatment. Vision impairment affects people of all ages with the primary concentration in the elderly. For cataract surgery, significant sudden and noteworthy change following intervention is expected. Heart failure is a significant health problem that affects the cardiovascular system and is particularly common in older adults. Improvements following treatment are often small and may be transitory.

Methods

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 vitreoretinal procedures. Patients with

traumatic cataract and with visual impairment so severe they are 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 three 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), unstable angina, or a coronary artery bypass graft (CABG) 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-month and 6-month follow-ups along with a postage paid self-addressed return envelope to each study participant who returned the baseline questionnaires¹²

Measures

The measures administered at baseline, one month, and six-months are described briefly below.

SF-6D—Perhaps the most commonly used outcome measure in the world today is the Medical Outcome Study Short Form-36 Health Survey (SF-36). The SF-36 grew out of work by the RAND Corporation in the Medical Outcomes Study (MOS).³ The SF-36v2™ includes eight health concepts: physical functioning, role limitations due to physical health problems, bodily pain, general health perceptions, vitality, social functioning, role limitations due to emotional problems, and mental health⁴. 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™^{5–7}.

Our study focuses on preference-based outcome measures. Although the SF-36v2™ is not a preference-based measure, Brazier and associates 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™ items⁸. The method is known as the SF-6D.

EQ-5D: The EQ-5D was developed by a collaborative group from Western Europe known as the EuroQol group.⁹ 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). For each domain, the respondent is asked to describe his or her health on that day using 3 response options (no problems, moderate problems, severe problems). The 5 domains combined with the 3 response options yields 5³ or, 243 unique health states. Adding perfect health and death gives 245 possible states¹⁰. Although the EQ-5D was originally validated in Europe, a scoring algorithm derived for the US general population is available and it was applied in this study. This scoring algorithm was derived from time tradeoff assessments of EQ-5D health states made by a population sample of some 4000 US adults in face-to-face household interviews¹¹

The EQ-5D is now used in a substantial number of clinical and population studies.^{12, 13} 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 have serious medical conditions. Information on the EQ-5D is available at: <http://www.euroqol.org>

Self-Administered Quality of Well-being Scale (QWB-SA): 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 4 domain scores are combined into a total score that provides a numerical point-in-time expression of well-being that ranges from zero (0) for dead to one (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 study¹⁴.

The self-administered QWB-SA has been shown to be highly correlated with the interviewer administered QWB and to retain the psychometric properties¹⁴. Extensive evaluations of reliability and validity have been published.^{15–20} Access to the measure and details about its development are available at <http://qwbsa.ucsd.edu>.

Health Utilities Index (HUI): The Health Utilities Index (HUI) is a family of health status and preference-based HRQOL measures.^{21, 22} Each member of the family includes a health status classification system, a preference-based multi-attribute 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 Health Utilities Index Mark 2 (HUI2) and Mark 3 (HUI3). HUI2 consists of 6 dimensions of health status: sensation (vision, hearing, speech), mobility, emotion, cognition, self care, and pain²². HUI3 includes 8 dimensions of health status: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain and discomfort, with five or six levels per attribute. Multiplicative multi-attribute utility functions based on community preferences have been estimated for HUI2²³ and HUI3²⁴. 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^{25–28}. Reference information on the HUI is available at: <http://www.fhs.mcmaster.ca/hug/> and <http://www.healthutilities.com/>

Disease Targeted Measures

NEI VFQ-25—Participants in the Cataract study were also evaluated using the National Eye Institute 25-Item Visual Functioning Questionnaire (VFQ-25). The VFQ-25 was developed by RAND and UCLA with support from the National Eye Institute. 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^{29, 30}.

MLHFQ—The MLHFQ assesses the extent to which heart failure affects daily life. The 21 MLHFQ items can be completed in 5–10 minutes. The content covers the most frequent and important ways heart failure affects daily functioning. THE MLHFQ yields an overall score and two subscores: physical and emotional. Support for reliability and validity of the MLHFQ is provided at www.mlhfq.org.^{31, 32}

Statistical Analysis

Number and percentage of cataract and 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 month 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 5 generic indexes and the respective disease-targeted measures are presented for cataract and heart failure patients separately.

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. The majority of the patients were white (87% for cataract and 79% for heart failure). The cataract patients tended to be female (59%), with most were 65 years or older. The heart failure patients tended to be male (67%) and somewhat younger; 78% were under age 65.

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 top portion of the table. At one 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, $SRM = 0.25$). Hence, the absolute differences that would be used to calculate quality-adjusted life years (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 6 patients, while the HUI3 would produce 0.52 QALYs, or one QALY for less than every 2 patients.

The largest SRMs were observed for change between baseline and the 1 month follow-up for the VFQ-25 ($SRM = 0.77$) and the HUI3 ($SRM = 0.25$). The SRMs for the other measures were smaller, and as noted above, the differences on the SF-6D were not statistically significant.

The lower portion of Table 3 shows changes in the cataract patients between one and six months. The analysis suggests that after one 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 <0.10 . Considering the three time points (baseline, one-month, six months), there was a significant linear trend for improved quality of life for the QWB-SA ($t=3.85$, $p<.0001$) HUI2 ($t=3.31$, $p<.001$) and HUI3 ($t=4.58$, $p<.0001$). There was also a strong linear trend for the VFQ-25 ($t=18.31$, $p<.0001$). Trends for the SF-6D ($t=0.78$, $p=.43$) and EQ-5D ($t=-1.15$, $p=.25$) were non-significant.

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 one month (top portion of table). After one month, the heart failure patients remained stable on all measures, except the

QWB-SA, which suggested some continued improvement (bottom portion of table). 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 while 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 Figure 2. Treatment for heart failure is expected to produce slow gradual gains³³ while cataract surgery is expected to produce change shortly after treatment³⁴. The figure summarizes mean scores on each index at baseline, one month, and six 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.³⁵ The figures show that cataract patients gain most of their improvement by one month (top two graphs). In contrast, heart failure patients continue to improve up to the six months evaluation on the QWB-SA and MLHFQ although the magnitude of improvement tends to be very modest (bottom section). Similar patterns were seen for the other measures (data not shown). Considering the three time points (baseline, one-month, six 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 ($t= 4.39$, $P <.0001$). The linear trend for the MLHFQ was comparable in strength to most of the generic measures ($t=-5.10$, $p<.0001$). The trend was non significant for the EQ-5D ($t =0.67$, $p =.50$), HUI2 ($t=0.70$, $p= 0.48$, and the HUI3 ($t=1.40$, $p=.16$).

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 one month and six months. The 1 month – 6 month SRM observed with the disease targeted MLHFQ was no larger (SRM = -0.26) than for the generic QWB-SA (SRM = 0.25). SRMs for the other indexes were not noteworthy.

Results reported above 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 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.

Discussion

At least five preference-based measures of health-related quality of life can be used for cost-utility analysis. Our analysis suggests that the five measures are not equally responsive to change following 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-36v2TM 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 8 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 multi-attribute 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.

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 and colleagues³⁶ compared HUI2 and 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.³⁷ Although some investigators are now estimating “utilities” from disease-targeted measures^{38, 39}, comparisons across studies can be difficult because of the potential for non-comparability 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 following 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 while 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 HUI3⁴⁰ and the QWB-SA⁴¹ 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 month were only captured by the QWB-SA. On the majority of 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-month 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 in order to identify the sensitivity and specificity of leading preference based generic measures of health outcome when applied in different clinical populations.

Acknowledgments

Supported by Grant P01AG020679 from the National Institute on Aging. Drs. Kaplan and Hays were also provided support by NIH grants 1 P01 AG020679-01A2, UCLA Claude D. Pepper Older Americans Independence Center, NIH/NIA 5P30AG028748, and 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 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.

References

1. Hays RD, Kim S, Spritzer KL, et al. Effects of mode and order of administration on generic health-related quality of life scores. *Value Health*. 2009; 12:1035–9. [PubMed: 19473334]
2. Hays RD, Kim S, Spritzer KL, et al. Effects of Mode and Order of Administration on Generic Health-Related Quality of Life Scores. *Value Health*. 2009
3. Ware JE Jr, Gandek B. Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *Journal of Clinical Epidemiology*. 1998; 51:903–12. [PubMed: 9817107]
4. 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. *Medical Care*. 1999; 37:MS10–22. [PubMed: 10335740]
5. Stewart, AL.; Ware, JE. *Measuring functioning and well-being: the medical outcomes study approach*. Durham: Duke University Press; 1992.
6. 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. *Medical Care*. 1999; 37:908–25. [PubMed: 10493469]
7. 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. *Medical Care*. 1999; 37:MS51–60. [PubMed: 10335743]
8. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*. 2002; 21:271–92. [PubMed: 11939242]
9. Kind P. The performance characteristics of EQ-5D, a measure of health related quality of life for use in technology assessment [abstract]. *Annual Meeting of International Society of Technology Assessment in Health Care*. 1997; 13:81.

10. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 2001; 33:337–43. [PubMed: 11491192]
11. 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. [PubMed: 15725977]
12. 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). *British Journal of Rheumatology*. 1997; 36:551–9. [PubMed: 9189057]
13. Gudex C, Dolan P, Kind P, Williams A. Health state valuations from the general public using the visual analogue scale. *Quality of Life Research*. 1996; 5:521–31. [PubMed: 8993098]
14. Kaplan RM, Sieber WJ, Ganiats TG. The Quality of Well-being Scale: Comparison of the interviewer-administered version with a self-administered questionnaire. *Psychology & Health*. 1997; 12:783–91.
15. 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.
16. Kaplan RM. Decision making in medicine and health care. *Annual Review of Clinical Psychology*. 2005; 1:525–56.
17. Kaplan RM, Anderson JP. A general health policy model: update and applications. *Health Serv Res*. 1988; 23:203–35. [PubMed: 3384669]
18. Kaplan RM, Anderson JP, Patterson TL, et al. Validity of the Quality of Well-Being Scale for persons with human immunodeficiency virus infection. HNRC Group. HIV Neurobehavioral Research Center Psychosomatic Medicine. 1995; 57:138–47.
19. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. *Health Serv Res*. 1976; 11:478–507. [PubMed: 1030700]
20. Kaplan RM, Ganiats TG, Sieber WJ, Anderson JP. The Quality of Well-Being Scale: critical similarities and differences with SF-36 [see comments]. *International Journal for Quality in Health Care*. 1998; 10:509–20. [PubMed: 9928590]
21. Feeny D, Furlong W, Mulhern RK, Barr RD, Hudson M. A framework for assessing health-related quality of life among children with cancer. *International Journal of Cancer Supplement*. 1999; 12:2–9.
22. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. *Health Utilities Index Pharmacoeconomics*. 1995; 7:490–502.
23. 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 Medical Care*. 1996; 34:702–22.
24. 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]
25. 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. [PubMed: 11491197]
26. 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. [PubMed: 10986061]
27. 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):54–9. [PubMed: 9659947]
28. 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*, 32A, No. 8, ppabcxyzpp1354–1358, 1996. *Eur J Cancer*. 1997; 33:506–7. [PubMed: 9155541]
29. 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. [PubMed: 11448327]
30. 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. *Archives of Ophthalmology*. 1998; 116:1496–504. [PubMed: 9823352]

31. Rector TS. A conceptual model of quality of life in relation to heart failure. *J Card Fail.* 2005; 11:173–6. [PubMed: 15812743]
32. 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. [PubMed: 8475878]
33. 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. *European Journal of Heart Failure.* 2006; 8:841–50. [PubMed: 16713337]
34. 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. [PubMed: 15767161]
35. 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; 45:1162–70. [PubMed: 18007166]
36. Blanchard C, Feeny D, Mahon JL, et al. Is the Health Utilities Index valid in total hip arthroplasty patients? *Qual Life Res.* 2004; 13:339–48. [PubMed: 15085906]
37. 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 [see comments]. *Jama.* 1996; 276:1172–7. [PubMed: 8827972]
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. [PubMed: 18780254]
39. Peacock R, Iezzi A, Richardson J, Wawthorne G, Keeffe J. Vision and Quality of Life: Development of Methods for the VisQoL Vision-Related Utility Instrument. *Ophthalmic Epidemiology.* 2008:15.
40. 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. [PubMed: 18359192]
41. Rosen, PN.; Galant, DM.; David, K.; Kaplan, RM. Measuring outcomes of cataract surgery using the quality of well-being scale and the VF-14. 2002. submitted for publication

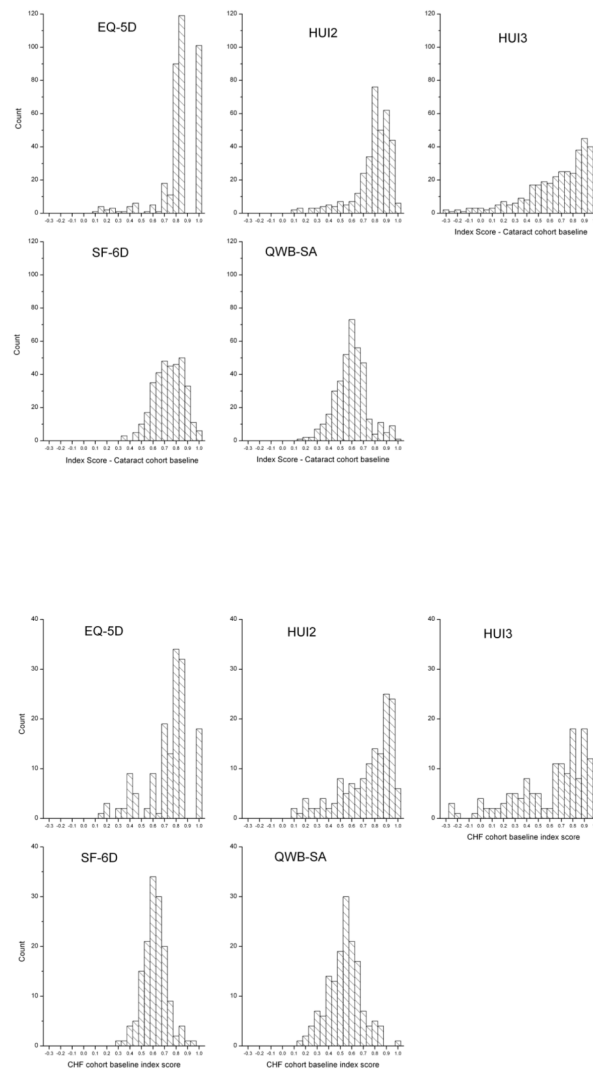


Figure 1. Distributions of baseline scores on 5 indexes at baseline for cataract (top) and heart failure patients (bottom).

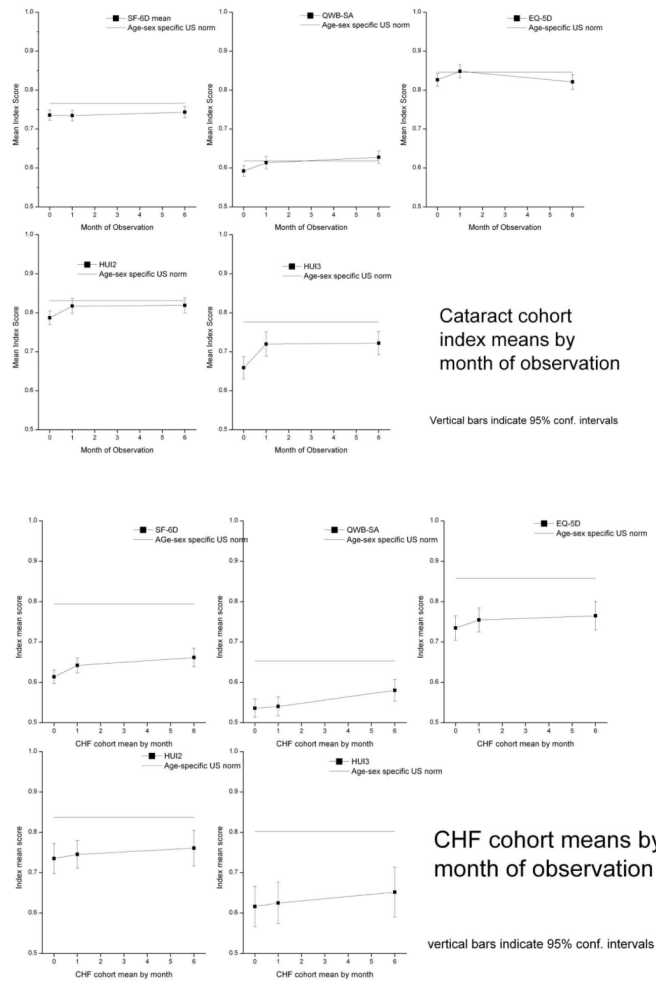


Figure 2. QWB-SA and HUI-3 Over Time in NHMS for Cataract Patients (Upper Panels) and Heart Failure (Lower Panels)

Table 1

Demographic characteristics of the samples

	Cataract patients, n (%)	Heart Failure patients, n (%)
Age 35–44	5 (1)	24 (15)
Age 45–64	115 (31)	101 (63)
Age 65–91	256 (68)	35 (22)
Race white	328 (87)	126 (79)
Race black	12 (3)	19 (12)
Race Asian	19 (5)	5 (3)
Race other	(1)	2 (1)
Race missing	13 (3)	8 (5)
Education <HS	21 (6)	20 (13)
Education HS graduate	60 (16)	45 (28)
Education some college	78 (21)	47 (29)
Education 2 year assoc. degree	27 (7)	12 (8)
Education 4 year college grad.	90 (24)	16 (10)
Edu. Master's degree	57 (15)	9 (6)
Edu. Doctorate or professional	34 (9)	6 (4)
Education missing	9 (2)	5 (3)
Female	222 (59)	52 (33)

Table 2

Number of patients in each group at each time point

	Baseline	1 month	6 months
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

Table 3

Changes in Cataract Patients at Between Baseline and One Month (top) and Between One Month and Six Months (Bottom) By Index

Difference (SD): 1 month - Baseline						
Index	N	Mean Difference	SD	t-value	SRM	
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): 6 months - 1 month						
Index	N	Mean Difference	SD	t-value	SRM	
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	

Table 4

Changes in Heart Failure Patients at Between Baseline and One Month (top) and Between One Month and Six Months (Bottom) by Index

Difference (SD): 1 month - Baseline						
Index	N	Mean Difference	SD	t-value	SRM	
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): 6 month - 1 month						
Index	N	Mean Difference	SD	t-value	SRM	
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	

Note: MLHFQ is scored so that a higher score is worse HRQOL.

Table 5

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

	VFQ-25	QWB-SA	HUI2	HUI3	EQ-5D	SF-6D	HAlex
VFQ-25	1.00	0.49	0.52	0.58	0.50	0.54	0.55
QWB-SA		1.00	0.53	0.60	0.55	0.54	0.58
HUI2			1.00	0.88	0.77	0.67	0.62
HUI3				1.00	0.7*	0.67	0.67
EQ-5D					1.00	0.68	0.65
SF-6D						1.00	0.69

Note—all correlations statistically significant: $p < .01$

Table 6

Correlations between generic measures and the MLFQ for heart failure patients at Baseline

	MLHF	QWB-SA	HUI2	HUI3	EQ-5D	SF-6D	HALex
MLHF	1.00	-0.64	-0.52	-0.49	-0.60	-0.65	-0.59
QWB-SA		1.00	0.56	0.58	0.58	0.58	0.55
HUI2			1.00	0.88	0.68	0.63	0.49
HUI3				1.00	0.67	0.63	0.47
EQ-5D					1.00	0.64	0.54
SF-6D						1.00	0.53

Note: All correlations statistically significant: $p < 0.01$