CURRENT OPINION

Pharmacoeconomics 2007; 25 (2): 93-106
1170-7690/07/0002-0093/$44.95/0

Measuring Preferences for Cost-Utility Analysis

How Choice of Method May Influence Decision-Making

Christine M. McDonough and *Anna N.A. Tosteson*

Dartmouth Medical School, Multidisciplinary Clinical Research Center in Musculoskeletal Diseases and the Center for the Evaluative Clinical Sciences, Lebanon, New Hampshire, USA

Abstract

Preferences for health are required when the economic value of healthcare interventions are assessed within the framework of cost-utility analysis. The objective of this paper was to review alternative methods for preference measurement and to evaluate the extent to which the method may affect healthcare decision-making. Two broad approaches to preference measurement that provide societal health state values were considered: (i) direct measurement; and (ii) preference-based health state classification systems.

Among studies that compared alternative preference-based systems, the EQ-5D tended to provide larger change scores and more favourable cost-effectiveness ratios than the Health Utilities Index (HUI)-2 and -3, while the SF-6D provided smaller change scores and less favourable ratios than the other systems. However, these patterns may not hold for all applications. Empirical evidence comparing systems and decision-making impact suggests that preferences will have the greatest impact on economic analyses when chronic conditions or long-term sequelae are involved. At present, there is no clearly superior method, and further study of cost-effectiveness ratios from alternative systems is needed to evaluate system performance.

Although there is some evidence that incremental cost-effectiveness ratio (ICER) thresholds (e.g. $US50 000 per QALY gained) are used in decision-making, they are not strictly applied. Nonetheless, as ICERs rise, the probability of acceptance of a new therapy is likely to decrease, making the differences in QALYs obtained using alternative methods potentially meaningful.

It is imperative that those conducting cost-utility analyses characterise the impact that uncertainty in health state values has on the economic value of the interventions studied. Consistent reporting of such analyses would provide further insight into the policy implications of preference measurement.

As pressures to contain costs of medical care have escalated, cost-utility analysis (CUA) has received both critical acclaim and scrutiny as a methodology to inform decision-makers regarding the economic value of healthcare interventions.[1-6] The number of published CUAs has grown steadily over the past 20 years,[3] and use by policy-makers appears to be increasing.[7,8] In some jurisdictions, including the UK, Australia and the Canadian provinces of Ontario and British Columbia, a formal role for CUA in pharmaceutical coverage decisions is mandated.[7,9] In other jurisdictions, such as the US, explicit use of CUA is more limited.[1,10-13]

A key source of initial resistance to CUA was concern about the validity and comparability of results between studies.[1,7,14,15] Issues of comparability have the potential to undermine a fundamental strength of CUA, which is meant to facilitate valid economic comparisons across a wide spectrum of interventions. Studies highlighting discrepancies between methods[15-22] underscore the importance of understanding the potential impact that methodological differences may have on decision making.

Development of a reference case by the *US Panel on Cost-Effectiveness in Health and Medicine*[15] provided one methodological standard for CUA. The perspective recommended for the reference case was societal, and the methods for valuing health outcomes included use of a generic health state classification system and community preferences, with sensitivity analysis to include patient preferences for studies of specific conditions.[15] Since publication of the reference case, there is evidence of improving quality in CUA methodology and reporting.[3] Support for use of reference case criteria to improve comparability is indicated by recent recommendations within the US regulatory environment.[23] Formal guidelines in other jurisdictions demonstrate that, although differences exist, there may be emerging consensus on key points noted in the reference case.[24,25] Despite these positive developments, the extent to which methodological differences in CUA may affect policy decisions remains uncertain. In this paper, we explore how alternative methods that fulfill reference case criteria for preference measurement, thereby yielding 'societal health state values', may affect decision making. We describe potential sources of variation in societal health state values, highlight relevant studies and discuss research assessing the decision-making impact.

1. Methods for Estimating Societal Health State Values

The QALY is the most commonly used measure for health in CUA.[26] QALYs combine the attributes of length and quality of life (QOL) into a single measure. The length of time in each health state is weighted according to an associated 'health state value', on a scale with 1 representing best imaginable health and 0 representing death. In our discussion, we use 'preference' as a general term reflecting the desirability of a health state and 'health state value' (HSV) to connote the numerical strength of preference for a health state.

We consider two general approaches to HSV measurement: (i) direct preference elicitation for relevant health states; and (ii) preference-based health state classification systems.[27-29]

1.1 Direct Preference Measurement

The standard gamble, time trade-off (TTO) and rating scale are commonly used preference elicitation methods with unique characteristics.[15,30-32] Differences between preferences obtained by these methods are well documented,[21,33-35] with HSVs typically highest for the standard gamble and lowest for the rating scale.[36-38] One study of dialysis patients provided evidence that patient preferences from the standard gamble resulted in cost-effectiveness ratios that were higher by $US5916 per QALY than when TTO values were used. Although this is one example of how measurement approach may affect cost-effectiveness results, individual patient rather than societal HSVs were used.[21]

Direct measurement of societal HSVs is a resource-intensive endeavour, requiring development of relevant health state descriptions and access to a representative population sample. Since this is often not feasible, researchers may use condition, age- and gender-specific values from a published source.[39]

1.2 Preference-Based Health State Classification Systems

Preference-based health state classification systems define each respondent's health state based on a questionnaire, and assign a societal HSV with a scoring algorithm that incorporates preferences from a general population sample. This approach allows researchers to use societal HSVs with minimal resources compared with direct preference measurement.

The most widely used systems[19] include the EQ-5D,[40-42] the Health Utilities Index (HUI),[43,44] the Quality of Well-Being Scale (QWB)[45] and the SF-36-derived SF-6D.[46,47] The three basic steps in developing a system are (i) classifying health; (ii) eliciting population preferences for a subset of health states; and (iii) developing a scoring algorithm to assign values for the full range of health states. We address potential sources of variation and related research for each step.

1.2.1 Descriptive System

A generic self-report health status questionnaire is the basis for most systems. Most systems include attributes for pain, physical function, social or role function and anxiety/depression, but differ in the number of response levels and how these are described and weighted. Some systems include other attributes, such as hearing and vision. Other differences include the perspective used in assessing health status. For example, the HUI questions ask about respondents' functional capacity,[43,48,49] while the EQ-5D and others ask about actual performance. The reference period also varies among systems, from 'today' to 'over the past 4 weeks'.

Each instrument characterises a unique number of health states based on the numbers and levels of attributes included in the questionnaire. It is unknown how many health states are needed to describe health adequately; however, the EQ-5D has 243 health states, HUI-2 has 24 000 and HUI-3 has 972 000.

Taken together, these aspects of the descriptive systems would have an impact on the psychometric properties of the system, including the ability to measure health status numerically, and to detect meaningful changes in health. Inadequacy in the descriptive systems may result in ceiling and floor effects, inability to measure key attributes of health (validity) and inability to measure important change (responsiveness).

1.2.2 Preference Measurement

In the development of preference-based systems, a subset of the unique system-defined health states are valued by a sample of the population as the basis for estimating values for the full range of health states. Choice of direct preference elicitation method implicitly incorporates differences noted in section 1.1 into the estimation of each system's societal HSVs. The EQ-5D offers value sets based on both TTO and rating scale measurements.[50] The SF-6D employs the standard gamble[46,47] and the QWB uses category scaling[45,51] methods. The developers of HUI and others, for example, argued that the rating scale is not appropriate for use in CUA. They used the standard gamble and a transformation of rating scale preference measurements for HUI-2 and -3.[36] Debate continues about the merits of various elicitation methods[52] and transformation[53,54] of rating scale values, v, into utilities, u, using a power curve such as $u = 1 - (1 - v)^{2.3}$.

Comparing CUA results using alternative preference measurement methods for the same system can demonstrate the isolated impact of preference elicitation methods. A recent study comparing elicitation methods for the EQ-5D and SF-6D concluded that this component alone can contribute to differences in HSV of up to 0.31, and may impact on cost-effectiveness ratios.[55] Conner-Spady et al.[56] compared HSVs from the EQ-5D using the TTO and visual analogue scale (VAS) in 436 patients with joint replacements, and found a lower baseline mean, wider range, more negative HSVs and larger QALY gains for TTO than VAS-based value sets, indicating a more favourable cost-effectiveness ratio for TTO relative to VAS when using the EQ-5D. The QALY gain reported for TTO-based preference weights was 5.14 versus 3.64 for the VAS using a 10-year time horizon. If the cost of joint replacement were $US7000,[57] the incremental cost-effectiveness ratio (ICER) would vary only slightly across methods (from $US1362 to $US1923 per QALY gained) and would be unlikely to influence decision-making.

1.2.3 Source of Community Preferences

The representativeness of population samples varies among systems.[58] Furthermore, eliciting and scoring population preferences for health state classification systems are resource-intensive, and HSVs are not available for every population. Therefore, preferences from a different population from that of interest are sometimes used. The extent to which differences in population preferences would contribute to variation in CUA results has been explored. In a study comparing rating scale valuations of EQ-5D health states by Finnish and US general population samples, small differences were noted that the authors concluded would not impact EQ-5D HSVs in international studies.[59] A study comparing TTO HSVs between general US and UK population samples[60] reported higher HSVs for the US sample. Differences in EQ-5D TTO HSVs between UK and Spanish populations[61] suggest that cultural differences may influence health state valuation.

1.2.4 Scoring Methods

Using directly measured preferences from a sample of the general population, a statistical model is fitted to estimate HSVs for the remaining health states. The HUI scoring system is based on multi-attribute utility theory[62] and a multiplicative function that captures interactions among attributes and allows characterisation of single attribute utility functions for levels within each attribute.[36,63] EQ-5D with York preference weights uses an additive model that includes level of severity, movement away from perfect health, and a term (the N3 term) to account for interaction between attributes when any attribute is at the worst level.[41,42] Scoring algorithms have been developed for the EQ-5D for various populations. US TTO-based preference weights are now available for the EQ-5D (EQ-5D-US) using a random effects model.[64] SF-6D scoring is very similar to that of the EQ-5D N3 model, with a smaller decrement when any attribute is at its worst level.[65] The QWB utilises an additive function that does not allow for interaction among the attributes.[45] The range of each scale and whether they include states worse than death (e.g. negative scores) varies.

Few studies compared CUA results using different scoring algorithms for the same system. Conner-Spady et al.[56] compared HSVs and QALYs gains estimated by EQ-5D with and without the N3 term. They reported that the N3 terms resulted in lower baseline mean and effect size for values, and more QALYs gained in a sample of patients with joint-replacements.

2. Empirical Evidence

We searched the international published literature (See the Appendix for search terms), and found no studies that directly investigated the impact of preference measurement method on policy decisions. Though a systematic review of policy decisions was beyond the scope of this paper, we included all identified studies that provided head-to-head comparisons of the most commonly used preference-based systems or addressed impact of system choice.

2.1 Comparisons of Preference-Based Health State Classification Systems

A review[19] of 23 published cost-utility analyses conducted alongside clinical trials found that 20 studies utilised a preference-based health state classification system to estimate QALYs. The HUI and EQ-5D were the most commonly used systems, with 16 using the EQ-5D. The authors suggested that different systems could qualitatively impact CUA results, and called for greater reporting transparency.

Table I summarises studies that have reported cross-sectional comparisons between systems. These studies found varying differences in mean HSVs across systems and offer insight into system characteristics that may contribute to variation in HSVs.[39,66-84] Generally, correlations between values from alternative systems were moderate to strong, indicating that they measure the same construct. The SF-6D demonstrated floor effects and a limited range of available scores. The EQ-5D did not provide HSVs between 0.88 and 1, and provided lower HSVs for similar health states than other systems. The HUI-3 was limited in characterising diminished mobility other than ambulation. These characteristics may be important for instrument choice relative to the condition of interest.

Evidence of differences in HSVs from cross-sectional studies is reason for concern, but longitudinal studies are necessary to understand system performance when measuring change in health. Fewer studies[56,76,87-101] report longitudinal head-to-head comparisons of preference-based systems (table II). EQ-5D estimates were generally largest, followed by HUI-3, -2 and finally, SF-6D. Interactions have been noted between HSVs and level of improvement.[91] Pickard et al.[87] reported that change in SF-6D HSVs correlated with mental status, while change in HUI and EQ-5D related more to daily function and disability. No consistent pattern of correlation was evident in the studies we reviewed.[87,95]

Comparisons of ICERs obtained using alternative systems were less common. Thomas et al.[102] reported CUA results for chronic low back pain acupuncture treatment using the SF-6D and EQ-5D. The ICER for the SF-6D was £4241 (95% CI 191, 28 026) per QALY compared with £3598 (95% CI 189, 22 035) for the EQ-5D. Neumann et al.[97] compared the results of a cost-effectiveness model for Alzheimer's drug treatment using the HUI-2 and -3, and found lower mean utility scores using the HUI-3 versus HUI-2, resulting in ICERs of $US9000/QALY for the HUI-3 and $US11 000/QALY for the HUI-2 with a duration of drug effect of 18 months. The authors noted that, while the difference in ICERs was slight for this analysis of drug treatment, it could be substantial for a disease prevention drug.

Overall, the EQ-5D tended to have larger changes in HSVs, which would generally translate to more favourable cost-effectiveness ratios when using the EQ-5D compared with the HUI-2 or -3 (see table II). Similarly, in studies that included the EQ-5D and SF-6D, the EQ-5D had larger changes. Comparing the HUI-2 or -3 and SF-6D, the SF-6D tended to have smaller changes. These patterns are generally consistent with the results of cross-sectional comparisons, yet it is difficult to identify a superior system.

2.2 Potential Impact of System Choice on Decision-Making

Recent commentaries illustrate the concern that both lack of data on preferences and variation in methods will impact healthcare decisions.[103,104] We found no empirical studies that address this question

Table I. Comparison of mean health state values in cross-sectional comparisons of systems

Study (population)	Number of subjects	Systems studied	Baseline mean (SD)	Correlation coefficient				
				EQ-5D	HUI-2	HUI-3	SF-6D	15D[a]
Luo et al.[85]	4 048	EQ-5D	0.87 (0.13)		0.67[b]	0.66[b]		
(general adult US population)		HUI-2	0.86 (0.32)			0.87[b]		
		HUI-3	0.81 (0.38)					
Petrou and Hockley[70]	14 736	EQ-5D	0.84 (0.23)				0.70[b]	
(general adult UK population)		SF-6D	0.80 (0.15)					
Barton et al.[72]	915	EQ-5D	0.79 (0.23)			0.44[c]	0.50[c]	
(hearing impairment)		HUI-3	0.56 (0.15)				0.41[c]	
		SF-6D	0.77 (0.08)					
McDonough et al.[74]	2 097	EQ-5D	0.39 (0.33)		0.68[d]	0.67[d]	0.69[d]	
(spine disorders)		HUI-2	0.59 (0.22)			0.78[d]	0.67[d]	
		HUI-3	0.45 (0.27)				0.72[d]	
		SF-6D	0.57 (0.12)					
Espallargues et al.[73]	209	EQ-5D	0.72 (0.22)	NR	NR	NR	NR	NR
(macular degeneration)		HUI-3	0.34 (0.28)					
		SF-6D	0.66 (0.14)					
Marra et al.[79]	313	EQ-5D	0.66 (0.13)	NR	NR	NR	NR	NR
(rheumatoid arthritis)		HUI-2	0.71 (0.19)					
		HUI-3	0.53 (0.29)					
		SF-6D	0.63 (0.24)					
Feeny et al.[75]	264	HUI-2	0.92 (0.13)	NR	NR	NR	NR	NR
(teen survivors low birthweight)		HUI-3	0.84 (0.16)					
Maddigan et al.[83]	372	HUI-2	0.78 (0.17)	NR	NR	NR	NR	NR
(type II diabetes)		HUI-3	0.64 (0.29)					
Luo et al.[82]	114	EQ-5D	0.75 (0.21)			0.45[d]		
(rheumatic disease)		HUI-3	0.76 (0.17)					
O'Brien et al.[84]	246	HUI-3	0.61 (0.12)				0.58[b]	
(cardiac disease)		SF-6D	0.58 (0.16)					
Schulz et al.[80]	29	EQ-5D	0.81 (0.12)			0.32[d]		
(benign prostatic hyperplasia)		HUI-3	0.79 (0.12)					
Brazier et al.[65]	2 605	EQ-5D	0.59 (NR)				0.66[b]	
(combined patient groups)[e]		SF-6D	0.63 (NR)					
Stavem et al.[81]	397	EQ-5D	0.81 (0.23)					0.78[d]
(epilepsy)		15D[e]	0.88 (0.12)					
Belanger et al.[86]	1 477	EQ-5D	0.83 (NR)			0.69[b]		
(general Canadian population)		HUI-3	0.85 (NR)					

a 15D is a preference-based health state classification system for which data were reported compared with more commonly used systems.[81]

b Pearson's correlation coefficient.

c Kendall's tau for n = 863.

d Spearman's rank correlation coefficient.

e This study reported values for patient groups: chronic obstructive pulmonary disease, osteoarthritis, irritable bowel syndrome, low back pain, leg ulcers, menopausal symptoms and aged >65 years.

HUI = Health Utilities Index; **NR** = not reported.

by directly investigating policy decisions relative to different measurement methods. Studies addressing the potential influence of economic evaluation on policy decisions have considered thresholds for favourable cost effectiveness to characterise qualitative changes in study results. Chapman et al.[17] assessed the affect of quality adjustment on analyses reporting both life years and QALYs published before 1998 and found qualitatively similar results for both QALYs and life years for the majority of analyses. However, quality adjustment had an important impact, causing estimates to cross thresholds for dominance of $US50 000 or $US100 000 in 18% of cases. This could be interpreted as evidence that, in the majority of cases, method choice would not be likely to have a qualitative impact on the results of the studies, and therefore on downstream policy decisions. Use of QALYs had the greatest impact for analyses assessing palliative treatments, chronic diseases and situations where there may be long-term adverse effects. This may indicate that for this subset of diseases, differences in estimated HSVs carried over a lifetime may have a qualitative impact on economic evaluation results and related policy decisions.

Using the same database, Schackman et al.[18] examined the results of sensitivity analyses on health-related QOL for pharmaceutical CUAs to determine the proportion of times that specific cost-effectiveness thresholds (i.e. $US20 000, $US50 000 and $US100 000 per QALY gained) were crossed. In 31% of sensitivity analyses, the ICER exceeded a threshold and in 13% of analyses, the ratio fell below a threshold. This indicates that if the magnitude of preference method-related variation reached the levels modeled in sensitivity analyses, there would be qualitative differences in results, as defined by crossing thresholds. The use of thresholds for policy decisions among a broad range of organisations is not well documented. We found some indication that thresholds of £20 000–30 000 are loosely used, and that overall, as ICERs increase the likelihood of acceptance probably decreases.[105-108]

3. Discussion

The paucity of published evidence limits our ability to draw conclusions concerning how preference measurement methods affect policy decisions. The ICERs available for comparison do not represent large enough changes to impact decision-making. However, evidence from comparisons of preference-based systems support a wide range of variation in estimates. Based on the evidence reviewed here, it appears that choice of preference measure may contribute to qualitatively different ICERs under some circumstances. As ICERs rise, the probability of acceptance appears to decrease, making the differences in QALYs obtained using alternative methods potentially meaningful. This is especially important for treatments with long-term consequences and ICERs around common thresholds. In our review, the EQ-5D tended to provide more favourable cost-effectiveness ratios than the HUI, while the SF-6D provided less favourable ratios than the other systems. Whether these patterns will hold for all applications depends on each system's ability to measure change across the full range of health.

To assess the impact of HSV on CUA, our review focused on societal HSVs, estimated with the most commonly used preference-based health state classification systems. However, evidence shows that in practice, the majority of CUAs do not yet meet reference case criteria.[20,109,110] Evidence from reviews of published CUAs indicate increasing use of preference-based systems and community preferences.[20,110] Among CUAs published between 1998 and 2001, 23% used preference-based systems and 36% used direct preference measurement such as standard gamble, TTO and rating scale. Utilities were community-based in 27% of estimates.[110]

Arguments for using community preferences hold that the population potentially affected by the decision should be polled. From a position of uncertainty about their own future health (i.e. 'the veil of

Table II. Studies addressing longitudinal comparisons in mean health state values provided by the most commonly used preference-based health state classification systems

Study (population)	Subjects (n)	Systems studied	Baseline mean (SD)	Change in health state value mean (SD)
Langfitt et al.[101] (chronic epilepsy)	64	EQ-5D	0.76 (0.26)	0.11 (0.28)
		EQ-5D-US	0.82 (0.18)	0.07 (0.21)
		HUI-2	0.78 (0.18)	0.03 (0.15)
		HUI-3	0.61 (0.30)	0.08 (0.27)
		SF-6D	0.70 (0.14)	0.08 (0.16)
Kaplan et al.[95] (rheumatoid arthritis)	628	EQ-5D	0.56 (0.25)	0.10 (0.27)
		HUI-2	0.64 (0.20)	0.06 (0.18)
		HUI-3	0.43 (0.27)	0.09 (0.24)
		SF-6D (VAS)	0.43 (0.16)	0.06 (0.16)
		SF-6D (SG)	0.81 (0.10)	0.03 (0.09)
Pickard et al.[87] (stroke)	98	EQ-5D	0.31 (0.38)	0.32 (0.38)
		HUI-2	0.51 (0.20)	0.12 (0.23)
		HUI-3	0.19 (0.30)	0.25 (0.32)
		SF-6D	0.55 (0.09)	0.13 (0.15)
Stavem et al.[99] (HIV/AIDS)	60	EQ-5D	0.77 (0.26)	0.02 (NR)
		SF-6D	0.73 (0.17)	0.03 (NR)
		15D[a]	0.86 (0.14)	0.02 (NR)
Thoma et al.[100] (breast reduction surgery)	41	HUI-2	(NR)	0.06 (0.14)
		HUI-3	(NR)	0.12 (0.19)
Hatoum et al.[88] (coronary artery disease)	184	HUI-3	0.63 (0.29)	0.15 (0.28)
		SF-6D	0.67 (0.12)	0.08 (0.13)
Feeny et al.[89] (total hip arthroplasty)	63	HUI-2	0.62 (0.19)	0.22 (0.07)
		HUI-3	0.62 (0.32)	0.23 (0.04)
		SF-6D	0.61 (0.10)	0.10 (0.02)
Holland et al.[98] (elderly patients)	123	EQ-5D	0.61 (0.29)	–0.16 (0.34)
		AQoL[b]	0.45 (0.27)	–0.12 (0.24)
Longworth and Bryan[96] (liver transplant)	183	EQ-5D	0.52 (0.33)	0.09 (0.37)
		SF-6D	0.61 (0.12)	0.01 (0.28)
Bosch et al.[94] (intermittent claudication)	87	HUI-2	0.70 (0.20)	0.10 (NR)
		HUI-3	0.66 (0.20)	0.11 (NR)
		SF-6D	0.66 (0.09)	0.08 (NR)
Bosch and Hunink[92] (intermittent claudication)	88	EQ-5D	0.57 (0.25)	0.22 (NR)
		HUI-3	0.66 (0.20)	0.11 (NR)
Suarez-Almazor et al.[93] (low back pain)	37	EQ-5D	0.38 (0.33)	–4.8 (17.4)[c]
		HUI-2	0.49 (0.19)	–1.8 (16.1)[c]

a 15D is a preference-based health state classification system for which data were reported compared with most commonly used systems.[81]

b AQoL is a preference-based health state classification system for which data were compared with most commonly used systems.[76,98]

c 3-month results, rescaled to 1–100, with 100 corresponding to best health.

AQoL = Assessment of Quality of Life; **HUI** = Health Utilities Index; **NR** = not reported; **SG** = standard gamble; **VAS** = visual analogue scale.

ignorance') they would value policy decisions with the most benefit for society as a whole.[15,111] However, patients or population subgroups may be in a better position to assess the impact of condition-related health changes than a general population sample.[15,112] This is controversial in light of evidence that patients provide higher values than other groups for their health states,[15,33,67,68,75,111,113-115] and the possibility that this could cause undervaluation of preventive intervention[67] and treatment. In addition, there is evidence that differences in valuations between patients and other groups varies with other factors,[116,117] such as severity and chronicity of illness.[118,119]

As a practical matter, the choice of methods may be dictated by resource availability, making off-the-shelf tools attractive. To facilitate CUAs when primary data are not available, the *Panel on Cost-Effectiveness in Health and Medicine*[15] recommended construction of a national catalogue of 'off-the-shelf' community preferences for health states. Work is ongoing in this area and includes population-based HSVs using multiple instruments,[39,85,86,120,121] mapping of health status data to HSVs and catalogues or registries from the published literature.[20,122] Several methods have been explored to estimate HSVs from the SF-36,[66,123-126] the SF-12,[123,127-136] condition-specific tools[137-139] and national survey instruments.[140-142] Some studies indicate that estimates from mapping may have important limitations[95,123,143] while others emphasise the fairly strong correlations between approaches.[23] A study investigating potential decision-making impact of algorithm choice compared ICERs from various SF-36 and -12 algorithms and the SF-6D.[144] They reported ICERs for an asthma cohort ranging from $US30 769 to $US63 492 per QALY, and for a stroke cohort of $US27 972 to $US50 000 per QALY. Many of the reported ICERs did not have overlapping confidence intervals. The growth of methods for estimating preferences using existing data is expanding the range of possible CUA applications, and warrants continued comparative study. In addition, efforts have been made to make published CUA more accessible through registries,[145-148] some of which include information on methodology. These efforts are important steps in improving availability and transparency of CUAs.

Finally, the impact of preference measurement method should be considered in the context of the role of CUA in policy decisions. Empirical evidence is limited; however, one study[149] indicated that economic evaluation in the US has a larger impact on decisions at the organisational level than population and patient/provider levels. Moreover, it appears that economic evaluation is used to varying degrees in different organisations in conjunction with legal, political and regulatory considerations.[149,150]

4. Conclusion

It is unknown to what extent the choice of preference measurement method impacts health policy decision-making. The existing evidence points toward potential impact for a subset of situations, most likely for ICERs near established thresholds when chronic diseases and/or long-term health effects are involved. At present, there is no clearly superior method for estimating societal preferences; however, alternative systems produce a wide range of HSVs. The ability to convert HSVs from one system to another would greatly improve comparability of CUAs. A recent publication of ongoing research using MEPS (Medical Expenditure Panel Survey) and National Health Interview Survey data uses linear regression to standardise estimates from various systems, and to convert estimates from one system to those of another.[151] Related research provided population-based norms.[152] Ongoing work will further our understanding of system responsiveness.[23]

Psychometric comparisons between systems were more common than comparisons of impact on ICERs and/or policy decisions. Explicit study of cost-effectiveness ratios obtained from alternative preference-based systems is needed to improve our understanding of potential policy implications. We encourage researchers who have access to comparative data to publish these estimates. Sensitivity analyses are suggested to address uncertainty associated with estimates of health effect, including variation due to measurement method. An in-depth systematic review of decisions made by policy makers relative to alternative measurement method would be a timely and important contribution to the literature. Research in these areas combined with ongoing dialogue about standardisation and conversion of estimates hold promise for attaining consistency of estimates for CUA.

Acknowledgements

This paper was funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (P60-AR048094) and the National Institute on Aging (R01-AG12262). The authors have no conflicts of interest that are directly relevant to the content of this article.

Appendix

1. Literature Search Strategy

The following key words were used in Medline to identify papers addressing the topic of preference measurement method and policy decision-making. In addition, selected authors, references from papers known from previous work and known websites were hand-searched. Searches were limited to English language and carried out between December 2005 and June 2006.

- Cost-utility analysis
- Cost-effectiveness analysis
- Economic evaluation
- Methods
- Attitude to health
- Health status indicators
- Health status
- Quality of life
- Quality-adjusted life years
- QALYs
- Time trade off
- Standard gamble
- Rating scale
- Visual analogue scale
- Utilities
- Health state values
- Values
- Preferences
- Preferences for health states
- Preference classification systems
- Preference-based
- EQ-5D
- HUI
- SF-6D
- SF
- Quality of well-being.

References

1. Luce BR. What will it take to make cost-effectiveness analysis acceptable in the United States? Med Care 2005; 43 (7 Suppl.): II44-8
2. Neumann PJ. Why don't Americans use cost-effectiveness analysis? Am J Manag Care 2004; 10: 308-12
3. Neumann PJ, Greenberg D, Olchanski NV, et al. Growth and quality of the cost-utility literature, 1976-2001. Value Health 2005; 8 (1): 3-9
4. Neumann PJ, Divi N, Beinfeld MT, et al. Medicare's national coverage decisions, 1999-2003: quality of evidence and review times. Health Aff 2005; 24 (1): 243-54
5. Thorpe KE. The rise in health care spending and what to do about it. Health Aff 2005; 24 (6): 1436-45
6. Foote SB, Neumann PJ. The impact of Medicare modernization on coverage policy: recommendations for reform. Am J Manag Care 2005; 11 (3): 140-2
7. Bloom BS. Use of formal benefit/cost evaluations in health system decision making [see comment]. Am J Manag Care 2004; 10 (5): 329-35
8. Dickson M, Hurst J, Jacobzone S. Survey of pharmacoeconomic assessment activity in eleven countries. Health working papers. Paris: OECD, 2003
9. Siegel J. Cost-effectiveness analysis in US healthcare decision-making: where is it going? Med Care 2005; 43 (7): II-1-14
10. Tunis SR. Economic analysis in healthcare decisions. Am J Manag Care 2004; 10 (5): 301-4
11. Aspinall SL, Good C, Glassman PA, et al. The evolving use of cost-effectiveness analysis in formulary management within the Department of Veteran Affairs. Med Care 2005; 43 (7 Suppl.): II20-6
12. Neumann PJ. Evidence-based and value-based formulary guidelines. Health Aff 2004; 23 (1): 124-34

13. Guyatt GH, Baumann M, Pauker S, et al. Addressing resource allocation issues in recommendations from clinical practice guideline panels. Suggestions from an American College of Chest Physicians Task Force. Chest 2006; 129: 182-7
14. Drummond M, Sculpher M. Common methodological flaws in economic evaluations. Med Care 2005; 43 (7 Suppl.): II5-14
15. Gold MR, Siegel JE, Russell LB, et al. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996
16. Chapman RH, Stone PW, Sandberg EA, et al. A comprehensive league table of cost-utility ratios and a sub-table of "panel-worthy" studies. Med Decis Making 2000; 20: 451-67
17. Chapman RC, Berger MLMD, Weinstein MC, et al. When does quality-adjusting life-years matter in cost-effectiveness analysis? Health Econ 2004; 13: 429-36
18. Schackman BR, Gold HT, Stone PW, et al. How often do sensitivity analyses for economic parameters change cost-utility analysis conclusions? Pharmacoeconomics 2004; 22 (5): 293-300
19. Richardson G, Manca A. Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. Health Econ 2004; 13: 1203-10
20. Bell CM, Chapman RH, Stone DA, et al. An off-the-shelf help list: a comprehensive catalog of preference scores from published cost-utility analyses. Med Decis Making 2001; 21: 288-94
21. Hornberger JC, Redelmeier DA, Petersen J. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. J Clin Epidemiol 1992; 45 (5): 505-12
22. Nord E. Toward quality assurance in QALY calculations. Int J Qual Health Care 1993; 9 (1): 37-45
23. Committee to Evaluate Measures of Health Benefits for Environmental Health and Safety Regulation. Valuing health for regulatory cost-effectiveness analysis. Washington, DC: National Academy of Sciences, 2006
24. Hjelmgren J, Berggren F, Andersson F. Health econonomic guidelines: similarities, differences, and some implications. Value Health 2001; 4 (3): 225-50
25. National Institute of Clinical Excellence. Technology appraisal methods N0515 [online]. Available from URL: http://www.nice.org.uk/page.aspx?.o=201973 [Accessed 2006 Jan 25]
26. Greenberg D, Pliskin JS. Preference-based outcome measures in cost-utility analyses: a 20-year overview. Int J Technol Assess Health Care 2002; 18 (3): 461-6
27. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. Ann Rev Public Health 2000; 21: 587-611
28. Brazier J, Deverill M, Green C, et al. A review of the use of health status measures in economic evaluation. Health Technol Assess 1999; 3 (9): 57-81
29. Kopec JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. J Clin Epidemiol 2003; 56 (4): 317-25
30. Drummond MF, O'Brien B, Stoddart GL, et al. Methods for the economic evaluation of health care programmes. 2nd ed. New York: Oxford University Press, 1997
31. Green C, Brazier J, Deverill M. Valuing health-related quality of life: a review of health state valuation techniques. Pharmacoeconomics 2000; 17 (2): 151-65
32. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ 1986; 5 (1): 1-30
33. Dolan P. Valuing health-related quality of life: issues and controversies. Pharmacoeconomics 1998; 15 (2): 119-27
34. Nord E. Methods for quality adjustment of life years. Soc Sci Med 1992; 34 (5): 559-69
35. Froberg DG, Kane RL. Methodology for measuring health-state preferences II: scaling methods. J Clin Epidemiol 1989; 42 (5): 459-71
36. Torrance GW, Furlong W, Feeny D, et al. Multi-attribute preference functions: health utilities index. Pharmacoeconomics 1995; 7 (6): 503-20
37. van Osch SMC, Wakker PP, van den Hout WB, et al. Correcting biases in standard gamble and time tradeoff utilities. Med Decis Making 2004; 24: 511-7
38. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. Health Econ 2002; 11: 447-56
39. Fryback DG, Dasbach EJ, Klein R, et al. The Beaver Dam Health Outcomes Study: initial catalogue of health-state quality factors. Med Decis Making 1993; 13 (2): 89-102
40. Brooks R. EuroQol: the current state of play. Health Policy 1996; 37 (1): 53-72
41. Dolan P. Modelling valuations for EuroQol health states. Med Care 1997; 35 (11): 1095-108
42. Dolan P, Roberts J. Modelling valuations for EQ-5D health states: an alternative model using differences in valuations. Med Care 2002; 40 (5): 442-6
43. Feeny D, Furlong W, Boyle M, et al. Multi-attribute health status classification systems: health utilities index. Pharmacoeconomics 1995; 7 (6): 490-502
44. Torrance GW, Furlong W, Feeny D, et al. Multi-attribute preference functions: health utilities index. Pharmacoeconomics 1995; 7 (6): 503-20
45. Kaplan RM, Anderson JP. A general health policy model: update and applications. Health Serv Res 1988; 23 (2): 203-35
46. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002; 21 (2): 271-92
47. Brazier J, Usherwood T, Harper R, et al. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol 1998; 51 (11): 1115-28
48. Asada Y. Medical technologies, nonhuman aids, human assistance, and environmental factors in the assessment of health states. Qual Life Res 2005; 14: 867-74
49. Ware JE, Brook RH, Davies AR, et al. Choosing measures of health status for individuals in general populations. Am J Public Health 1981; 71 (6): 620-5
50. EuroQol Group. EQ-5D [online]. Available from URL: http://www.euroqol.org/web/users/valuation.php [Accessed 2006 Jan 31]
51. Kaplan RM, Ganiats TG, Sieber WJ, et al. The quality of well-being scale: critical similarities and differences with SF-36 [see comment]. Int J Qual Health Care 1998; 10 (6): 509-20

52. Parkin D, Devlin N. Is there a case for using visual analogue scale valuations in cost-utility analysis? Health Econ 2006; 15: 653-64
53. McCabe CJ, Stevens KJ, Brazier J. Utility scores for the health utilities index mark 2: an empirical assessment of alternative mapping functions. Med Care 2005; 43 (6): 627-35
54. Stevens KJ, McCabe CJ, Brazier J. Mapping between visual analogue scale and standard gamble data: results from the UK Health Utilities Index 2 valuation survey. Health Econ 2006 May; 15 (5): 527-33
55. Tsuchiya A, Brazier J, Roberts J. Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. J Health Econ 2006; 25: 334-46
56. Conner-Spady B, Voaklander DC, Suarez-Almazor ME. The effect of different EuroQol weights on potential QALYs gained in patients with hip and knee replacement. 17th Plenary Meeting of the EuroQol Group; 2000 Sep 28-29; Pamplona
57. Burns AW, Bourne RB, Chesworth BM, et al. Cost effectiveness of revision total knee arthroplasty. Clin Orthop Relat Res 2006; 446: 29-33
58. Fryback DG. A US valuation of the EQ-5D. Med Care 2005; 43 (3): 199-200
59. Johnson JA, Ohinmaa A, Murti B, et al. Comparison of Finnish and US-based visual analog scale valuations of the EQ-5D measure. Med Decis Making 2000; 20 (3): 281-9
60. Johnson JA, Luo N, Shaw JW, et al. Valuations of EQ-5D health states: are the United States and United Kingdom different? Med Care 2005; 43 (3): 221-8
61. Badia X, Montserrat R, Herdman M, et al. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. Med Decis Making 2001; 21: 7-16
62. Keeney R. A group preference axiomatization with cardinal utility. Manag Sci 1976; 23: 140-5
63. Torrance GW. Preferences for health outcomes and cost-utility analysis. Am J Manag Care 1997; 3 Suppl.: S8-20
64. 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 (3): 203-20
65. Brazier J, Roberts J, Tsuchiya A. A comparison of the EQ-5D and SF-6D across seven patient groups. Proceedings of the 18th Plenary Meeting of the Euroqol Group; 2001 Sep 6-7; Odense
66. Hollingworth W, Deyo RA, Sullivan SD, et al. The practicality and validity of directly elicited and SF-36 derived health state preferences in patients with low back pain. Health Econ 2002; 11 (1): 71-85
67. Gabriel SE, Kneeland TS, Melton LJ, et al. Health-related quality of life in economic evaluations for osteoporosis: whose values should we use? Med Decis Making 1999; 19 (2): 141-8
68. Krahn M, Ritvo P, Irvine J, et al. Patient and community preferences for outcomes in prostate cancer: implications for clinical policy. Med Care 2003; 41 (1): 153-64
69. Macran S, Weatherly H, Kind P. Measuring population health: a comparison of three generic health status measures. Med Care 2003; 41 (2): 218-31
70. Petrou S, Hockley C. An investigation into the empirical validity of the EQ-5D and SF-6D on hypothetical preferences in a general population. Health Econ 2005; 14: 1169-89
71. Fisk JD, Brown MG, Sketris IS, et al. A comparison of health utility measures for the evaluation of multiple sclerosis treatments. J Neurol Neurosurg Psychiatry 2005; 76: 58-63
72. Barton GR, Bankart J, Davis AC. A comparison of the quality of life of hearing-impaired people as estimated by three different utility measures. Int J Audiol 2005; 44: 157-63
73. Espallargues M, Czoski-Murray CJ, Bansback NJ, et al. The impact of age-related macular degeneration on health status utility values. Invest Ophthalmol Vis Sci 2005; 46 (11): 4016-23
74. McDonough CM, Grove MR, Tosteson TD, et al. Comparison of EQ-5D, HUI, and SF-36-derived societal health state values among spine patient outcomes research trial (SPORT) participants. Qual Life Res 2005; 14 (5): 1321-32
75. Feeny D, Furlong W, Saigal S, et al. Comparing directly measured standard gamble scores to HUI2 and HUI3 utility scores: group- and individual-level comparisons. Soc Sci Med 2004; 58: 799-809
76. Hawthorne G, Richardson J, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. Ann Med 2001; 33 (5): 358-70
77. Elvik R. The validity of using health state indexes in measuring the consequences of traffic injury for public health. Soc Sci Med 1995; 40 (10): 1385-98
78. de Vries SO, Kuipers WD, Hunink MG. Intermittent claudication: symptom severity versus health values. J Vasc Surg 1998; 27 (3): 422-30
79. Marra CA, Esdaile JM, Guh D, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. Med Care 2004; 42 (11): 1125-31
80. Schulz MW, Chen J, Woo HH, et al. A comparison of techniques for eliciting patient preferences in patients with benign prostatic hyperplasia. J Urol 2002; 168 (1): 155-9
81. Stavem K, Bjornaes H, Lossius MI. Properties of the 15D and EQ-5D utility measures in a community sample of people with epilepsy. Epilepsy Res 2001; 44 (2-3): 179-89
82. Luo N, Chew L, Fong K, et al. A comparison of the EuroQol-5D and the Health Utilities Index Mark 3 in patients with rheumatic disease. J Rheumatol 2003; 30 (10): 2268-74
83. Maddigan SL, Feeny D, Johnson JA. A comparison of the Health Utilities Indices Mark 2 and Mark 3 in type 2 diabetes. Med Decis Making 2003; 23: 489-501
84. O'Brien BJ, Spathe M, Blackhouse G, et al. A view from the bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. Health Econ 2003; 12: 975-81
85. Luo N, Johnson JA, Shaw JW, et al. Self-reported health status of the general adult US population as assessed by the EQ-5D and Health Utilites Index. Med Care 2005; 43 (11): 1078-86
86. Belanger A, Berthelot J-M, Guimond E, et al. A head-to-head comparison of two generic health status measures in the household population: McMaster Health Utilities Index (Mark 3) and the EQ-5D. Ottawa: Statistics Canada, Health Analysis and Modelling Group, 2000. Final Revision April 2000

87. Pickard AS, Johnson JA, Feeny DH. Responsiveness of generic health-related quality of life measures in stroke. Qual Life Res 2005; 14 (1): 207-19
88. Hatoum HT, Brazier JE, Akhras KS. Comparison of the HUI3 with the SF-36 preference based SF-6D in a clinical trial setting. Value Health 2004; 7 (5): 602-9
89. Feeny D, Wu L, Eng K. Comparing Short Form 6D, Standard Gamble, and Health Utilities Index Mark 2 and Mark 3 utility scores: results from total hip arthroplasty patients. Qual Life Res 2004; 13: 1659-70
90. Conner-Spady B, Suarez-Almazor ME. A comparison of preference-based health status tools in patients with musculoskeletal disease. 18th Plenary Meeting of the EuroQol Group; 2001 Sep 6-7; Odense
91. Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. Med Care 2003; 41 (7): 791-801
92. Bosch JL, Hunink M. Comparison of the Health Utilities Index Mark 3 (HUI3) and the EuroQol EQ-5D in patients treated for intermittent claudication. Qual Life Res 2000; 9: 591-601
93. Suarez-Almazor M, Kendall C, Johnson J, et al. Use of health status measures in patients with low back pain in clinical settings. Comparison of specific, generic and preference-based instruments. Rheumatology 2000; 39: 783-90
94. Bosch JL, Halpern EF, Gazelle GS. Comparison of preference-based utilities of the Short-Form 36 Health Survey and Health Utilities Index before and after treatment of patients with intermittent claudication. Med Decis Making 2002; 22 (5): 403-9
95. Kaplan R, Groessl EJ, Sengupta N, et al. Comparison of measured utility scores and imputed scores from the SF-36 in patients with rheumatoid arthritis. Med Care 2005; 43 (1): 79-87
96. Longworth L, Bryan S. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. Health Econ 2003; 12 (12): 1061-7
97. Neumann PJ, Sandberg EA, Araki SS, et al. A comparison of HUI2 and HUI3 utility scores in Alzheimer's disease. Med Decis Making 2000; 20 (4): 413-22
98. Holland R, Smith RD, Harvey I, et al. Assessing quality of life in the elderly: a direct comparison of the EQ-5D and AQoL. Health Econ 2004; 13 (8): 793-805
99. Stavem K, Froland SS, Hellum KB. Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. Qual Life Res 2005; 14 (4): 971-80
100. Thoma A, Sprague S, Veltri K, et al. Methodology and measurement properties of health-related quality of life instruments: a prospective study of patients undergoing breast reduction surgery. Health Qual Life Outcomes 2005 Jul 22; 3: 44
101. Langfitt J, Vickrey B, McDermott M, et al. Validity and responsiveness of generic preference-based HRQOL instruments in chronic epilepsy. Qual Life Res 2006; 15: 899-914
102. Thomas KJ, MacPherson H, Ratcliffe J, et al. Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain. Health Technol Assess 2005; 9 (32): 1-109
103. Stiggelbout AM. Health state classification systems: how comparable are our ratios? Med Decis Making 2006; 25: 223-5
104. Neumann PJ. Health utilities in Alzheimer's disease and implications for cost-effectiveness analysis. Pharmacoeconomics 2005; 23 (6): 537-41
105. National Institute of Clinical Excellence. Technology appraisal methods N0515 [online]. Available from URL: http://www.nice.org.uk/page.aspx?.o=201973 [Accessed 2006 Jan 25]
106. Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Econ 2004; 13: 437-52
107. Pearson SD, Rawlins MD. Quality, innovation, and value for money: NICE and the British National Health Service. JAMA 2005; 294 (20): 2618-22
108. Henry DA, Hill SR, Harris A. Drug prices and value for money: the Australian pharmaceutical benefits scheme. JAMA 2005; 294 (20): 2630-2
109. Rasanen P, Roine E, Sintonen H, et al. Use of quality-adjusted life years for the estimation of effectiveness of health care: a systematic literature review. Int J Technol Assess Health Care 2006; 22 (2): 235-41
110. Brauer C, Rosen AB, Greenberg D, et al. Trends in the measurement of health utilities in published cost-utility analyses. Value Health 2006; 9 (4): 213-8
111. De Wit GA, Busschbach JJV, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? Health Econ 2000; 9: 109-26
112. Greenberg D, Pliskin JS. Using health state classification systems for utility elicitation in the elderly. Med Decis Making 2006; 25: 220-2
113. Nord E, Pinto JL, Richardson J, et al. Incorporating societal concerns for fairness in numerical valuations of health programmes [see comment] (published erratum appears in Health Econ 1999 Sep; 8 (6): 559). Health Econ 1999; 8 (1): 25-39
114. Polsky D, Willke RJ, Scott K, et al. A comparison of scoring weights for the Euroqol derived from patients and the general public. Health Econ 2001; 10: 27-37
115. Feeny D, Blanchard C, Mahon JL, et al. Comparing community-preference-based and direct standard gamble utility scores: evidence from elective total hip arthroplasty. Int J Technol Assess Health Care 2003; 19 (2): 362-72
116. Sullivan PW, Gushchyan V. Systematic differences in subjective vs societal preferences. Qual Life Res 2005; 14 (9): 2011
117. Souchek J, Byrne MM, Kelly PA, et al. Valuation of arthritis health states across ethnic groups and between patients and community members. Med Care 2005; 43 (9): 921-8
118. Insinga RP, Fryback DG. Understanding differences between self-ratings and population ratings for health in the EuroQOL. Qual Life Res 2003; 12 (6): 611-9
119. McPherson K, Myers J, Taylor WJ, et al. Self-valuation and societal valuations of health state differ with disease severity in chronic and disabling conditions. Med Care 2004; 42 (11): 1143-51
120. Sullivan PW, Lawrence WF, Gushchyan V. A national catalogue of preference-based scores for chronic conditions in the United States. Med Care 2005; 43 (7): 736-49
121. Sullivan PW, Ghushchyan V. Preference-based EQ-5D index scores for chronic conditions in the United States. Med Decis Making 2006 Jul/Aug; 26 (4): 410-20

122. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. Med Care 2000; 38 (6): 583-637
123. Sherbourne C, Unutzer J, Schoenbaum M, et al. Can utility-weighted health-related quality-of-life estimates capture health effects of quality improvement for depression? Med Care 2001; 39 (11): 1246-59
124. Lee TA, Hollingworth W, Sullivan SD. Comparison of directly elicited preferences to preferences derived from the SF-36 in adults with asthma. Med Decis Making 2003; 23: 323-34
125. Fryback DG, Lawrence WF, Martin PA, et al. Predicting quality of well-being scores from the SF-36: results from the Beaver Dam Health Outcomes Study. Med Decis Making 1997; 17 (1): 1-9
126. Shmueli A. Subjective health status and health values in the general population. Med Decis Making 1999; 19 (2): 122-7
127. Lundberg L, Johannesson M, Isacson DG, et al. The relationship between health-state utilities and the SF-12 in a general population. Med Decis Making 1999; 19 (2): 128-40
128. Franks P, Lubetkin EI, Gold MR, et al. Mapping the SF-12 to preference-based instruments. Med Care 2003; 41 (11): 1277-83
129. Franks P, Lubetkin EI, Gold MR, et al. Mapping the SF-12 to the Euroqol EQ-5D Index in a national US sample. Med Decis Making 2004; 24: 247-54
130. Sengupta N, Nichol MB, Wu J, et al. Mapping the SF-12 to the HUI3 and VAS in a managed care population. Med Care 2004; 42 (9): 927-37
131. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004; 42 (9): 851-9
132. Sanderson K, Andrews G, Corry J, et al. Using the effect size to model change in preference values from descriptive health status. Qual Life Res 2004; 13 (7): 1255-64
133. Lenert LA, Sherbourne CD, Sugar C, et al. Estimation of utilities for the effects of depression from the SF-12. Med Care 2000; 38 (7): 763-70
134. Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D preference scores from the SF-12 Health Survey in a nationally representative sample. Med Decis Making 2004; 24 (2): 160-9
135. Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. Med Decis Making 2006; 26 (1): 18-29
136. Sullivan PW, Ghushchyan V. Mapping the EQ-5D Index from the SF-12: US general population preferences in a nationally representative sample. Med Decis Making 2006; 26: 401-9
137. Lenert LA, Sturley AP, Rapaport MH, et al. Public preferences for health states with schizophrenia and a mapping function to estimate utilities from positive and negative symptom scale scores. Schizophr Res 2004; 71 (1): 155-65
138. Kind P, Macran S. Eliciting social preference weights for functional assessment of cancer therapy-lung health states. Pharmacoeconomics 2005; 23 (11): 1143-53
139. Brazier J, Kolotkin RL, Crosby RD, et al. Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. Value Health 2004; 7 (4): 490-8
140. Gold MR, Franks P, McCoy KI, et al. Toward consistency in cost-utility analyses: using national measures to create condition-specific values [see comment]. Med Care 1998; 36 (6): 778-92
141. Gold M, Franks P, Erickson P. Assessing the health of the nation: the predictive validity of a preference-based measure and self-rated health. Med Care 1996; 34 (2): 163-77
142. 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
143. Feeny D. As good as it gets but good enough for which applications? Med Decis Making 2006; 26: 307-9
144. Pickard AS, Wang Z, Walton SM, et al. Are decisions using cost-utility analyses robust to choice of SF-36/SF-12 preference-based algorithm? Health Qual Life Outcomes 2005; 3 (1): 11
145. Harvard Center for Risk Analysis, Harvard School of Public Health. Cost-effectiveness analysis registry [online]. Available from URL: http://www.hsph.harvard.edu/cearegistry/ [Accessed 2006 Feb 2]
146. The Office of Health Economics, The International Federation of Pharmaceutical Manufacturers' Associations. Health Economic Evaluation Database (HEED) [online]. Available from URL: http://www.ohe-heed.com/ [Accessed 2006 Feb 2]
147. The United Kingdom National Health Service. Economic evaluation database [online]. Available from URL: http://www.york.ac.uk/inst/crd/nhsdhp.htm [Accessed 2006 Feb 2]
148. The United Kingdom National Health Service. European Network on Health Economics Evaluation Databases [online]. Available from URL: http://www.cordis.lu/data/PROJ_FP5/ACTIONeqDndSESSIONeq112482005919ndDOCeq622ndTBLeqEN_PROJ.htm [Accessed 2006 Feb 2]
149. van Elden ME, Severens JL, Novak A. Economic evaluations of healthcare programmes and decision making: the influence of economic evaluations on different healthcare decision-making levels. Pharmacoeconomics 2005; 23 (11): 1075-82
150. Taylor RS, Drummond MF, Salkeld G, et al. Inclusion of cost effectiveness in licensing requirements of new drugs: the fourth hurdle. BMJ 2004; 329 (7472): 972-5
151. Franks P, Hanmer J, Fryback DG. Relative disutilities of 47 risk factors and conditions assessed with seven preference-based health status measures in a national U.S. sample: toward consistency in cost-effectiveness analysis. Med Care 2006; 44 (5): 478-85
152. 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

Correspondence and offprints: Dr *Anna N.A. Tosteson*, HB7505 Clinical Research, Dartmouth-Hitchcock Medical Center, One Medical Center Dr, Lebanon, NH 03756, USA.
E-mail: anna.n.a.tosteson@dartmouth.edu

