Comparison of EQ-5D and SF-6D utilities in mental health patients

Lamers, L.M.; Bouwmans, C.A.; van Straten, A.; Donker, M.C.; Hakkaart, L.

published in
Health Economics
2006

DOI (link to publisher)
10.1002/hec.1125

document version
Publisher's PDF, also known as Version of record

HEALTH ECONOMICS
Health Econ. **15**: 1229–1236 (2006)
Published online 20 April 2006 in Wiley InterScience (www.interscience.wiley.com). **DOI:**10.1002/hec.1125

Comparison of EQ-5D and SF-6D utilities in mental health patients

L. M. Lamers[a,b,*], C. A. M. Bouwmans[a], A. van Straten[b], M. C. H. Donker[b] and L. Hakkaart[a]
[a] *Institute for Medical Technology Assessment bv, Erasmus University Medical Center, Rotterdam, The Netherlands*
[b] *Department of Health Policy and Management, Erasmus University Medical Center, Rotterdam, The Netherlands*

Summary

Various preference-based measures of health are available for use as an outcome measure in cost–utility analysis. The aim of this study is to compare two such measures EQ-5D and SF-6D in mental health patients.

Baseline data from a Dutch multi-centre randomised trial of 616 patients with mood and/or anxiety disorders were used. Mean and median EQ-5D and SF-6D utilities were compared, both in the total sample and between severity subgroups based on quartiles of SCL-90 scores. Utilities were expected to decline with increased severity.

Both EQ-5D and SF-6D utilities differed significantly between patients of adjacent severity groups. Mean utilities increased from 0.51 at baseline to 0.68 at 1.5 years follow-up for EQ-5D and from 0.58 to 0.70 for SF-6D. For all severity subgroups, the mean change in EQ-5D utilities as well as in SF-6D utilities was statistically significant. Standardised response means were higher for SF-6D utilities.

We concluded that both EQ-5D and SF-6D discriminated between severity subgroups and captured improvements in health over time. However, the use of EQ-5D resulted in larger health gains and consequent lower cost–utility ratios, especially for the subgroup with the highest severity of mental health problems.

Received 4 February 2005; Accepted 20 February 2006

Keywords preference-based measures of health; EQ-5D; SF-6D; cost–utility analysis; mental health

Introduction

During the past decades, quality of life has become an important health outcome in economic evaluations. In cost–utility analysis, outcomes are typically measured in Quality-Adjusted Life Years (QALYs). This outcome measure captures both gains from reduced morbidity (quality) and mortality (quantity). The quality adjustment is based on a set of weights, called utilities, which reflect the desirability of the health states. For each possible health state, utilities should be measured on an interval scale, where 1 refers to full health and 0 refers to death [1]. Additionally, negative utilities may occur if very severe health states are evaluated as being worse than death [2]. Measuring utilities is complex and time-consuming. For this reason, generic preference-based instruments like EuroQol-5D (EQ-5D) [3], Short Form-6D (SF-6D) [4] or the Health Utility Index (HUI) [5,6] are attractive and widely used for cost–utility analysis.

The EQ-5D descriptive system consists of five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) with three levels each (no problems, some problems and extreme problems), thus defining 243 (3^5)

*Correspondence to: Institute for Medical Technology Assessment, Erasmus MC, PO Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: l.lamers@erasmusmc.nl

distinct health states. In 1993, the Measurement and Valuation of Health (MVH) Group at the Centre for Health Economics in York conducted a large-scale national (UK) study to elicit direct valuations of 3395 persons from the general public for 42 EQ-5D health states using the time trade-off method. Regression techniques were applied to these direct valuations to interpolate values for the 200 states for which no direct valuations had been elicited [7,8]. The resulting set of regression coefficients or tariff (the MVH A1 tariff) is widely used to calculate utilities for EQ-5D health states for cost–utility analyses of health care programmes and treatments. The utilities ranged from −0.59 to 1.00 [7].

Recently, the Sheffield Health Economic Group derived a preference-based measure of health from the SF-36 [4]. The SF-36 generates scores on eight dimensions of health: general health, physical functioning, role limitations (role-physical and role-emotional), social functioning, bodily pain, vitality and mental health. Originally the SF-36 was not developed for use in economic evaluations since it did not incorporate preference weights in its scoring system. The new preference-based measure, known as the SF-6D, is derived from 11 items of the SF-36 and is composed of six dimensions of health with four to six levels each. The SF-6D thus describes a total of 18 000 possible health states. Of these, 249 states were directly valued by a representative sample of 611 members of the general UK population, using the standard gamble method. Like the EQ-5D, regression models were estimated to predict single utility scores for all possible SF-6D health states, ranging from 0.30 to 1.00 [4].

Since the EQ-5D and the SF-6D differ in their descriptive systems and used different valuation methods to estimate their sets of utility scores, completion of both instruments by the same patient might be expected to result in different utilities. In other words, the different outcomes of cost–utility analyses may depend on the preference-based measure of health used. Brazier *et al.* [9] compared utility scores on the EQ-5D and the SF-6D in seven different patient groups. They found a substantial agreement between the mean utilities with on average some higher values generated by SF-6D. However, differences between the EQ-5D and the SF-6D preferences varied by patient group. Longworth and Bryan [10] compared utilities from both instruments in liver transplant patients. While the EQ-5D scores showed significant changes before and 12 months after transplantation, the SD-6D was not able to capture changes in utility following transplantation. The authors concluded that the SF-6D appeared not to describe health states at the lower end of the utility scale (close to or below zero), thus underestimating the magnitude of improvement in quality of life in this particular patient group. However, Longworth and Bryan [10] expected that the SF-6D to be more sensitive to smaller changes towards the top of the utility scale (close to 1) compared to EQ-5D.

Until now, no information about possible differences between the EQ-5D and the SF-6D utilities in patients with mental health problems was available, as previous studies did not include this population. The aim of this paper was to compare the EQ-5D and SF-6D utilities across patients suffering from mood and anxiety disorders, the most common mental health problems. To our knowledge, no study has yet compared differences in utilities on the EQ-5D and the SF-6D between subgroups of patients with a different degree of severity of illness. A second objective of our study was to assess the ability of the EQ-5D and SF-6D to discriminate between subgroups of patients by severity.

Methods

Data

Data were extracted from a large multicentre randomised trial in Mental Health Care Centres (MHC) in the Netherlands. This naturalistic trial assessed the cost-effectiveness of Brief Therapy (BT) compared to Cognitive Behavioural Therapy (CBT) and Care As Usual (CAU) in patients with mood and/or anxiety disorders (Hakkaart-van Roijen L, van Straten A, Al MJ, Rutten FFH, Donker MCH. The cost utility of brief psychological treatment for depression and anxiety. *Br J Psychiatry*, accepted for publication). Patients aged 18–65 with a diagnosis of a major depressive disorder, dysthymic disorder, panic disorder, social phobia, or generalised anxiety were included. Disorders were classified according the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) [11]. Comorbidity in these patients associated with other

DOI: 10.1002/hec

psychiatric diagnoses was allowed, except for psychotic or bipolar disorder (Hakkaart-van Roijen *et al.*, accepted for publication). The patient population was representative for the patients treated at MHCs in the Netherlands. The primary outcome was the prevalence of mood and anxiety disorders, as measured by the Composite Diagnostic Interview (CIDI). Secondary outcomes that were measured consisted of the Symptom Checklist (SCL-90) [12] and health-related quality of life measured by EQ-5D and SF-36.

The SCL-90 is a valid instrument to measure severity of neurotic illness [13,14]. It consists of 90 psychological symptoms, which are rated on a five-point scale, ranging from 1 (no distressed by the symptom) to 5 (extremely distressed by the symptom). By summing the item scores, a total score can be obtained that ranges from 90 (no distress) to 450 (extremely distressed), with 118 representing the mean score for the general Dutch population. The SCL-90 scores were used to a measure of severity of mental health problems. Four severity subgroups were formed using the quartiles of the SCL-90 scores.

Baseline data were available for 643 patients. Six-hundred and sixteen patients completed all the questions on the EQ-5D, SF-6D and SCL-90. The baseline data from these patients were used in the comparison of the EQ-5D and the SF-6D. The mean age was 36.8 years (median 35) and 61.2% of these patients were women. Complete EQ-5D and SF-6D were available at 1 and 1.5 year follow-up for 355 and 326 patients, respectively (Hakkaart-van Roijen *et al.*, accepted for publication).

Analyses

First, an overall comparison of the patient's self-rated health on the dimensions of the EQ-5D and the SF-6D was made. The degree of agreement between the dimensions of the EQ-5D and the SF-6D was examined using Spearman's rank correlation. Next, EQ-5D and SF-6D utilities were calculated and compared. The EQ-5D utilities were based on the MVH A1 tariff [7]. The SF-36 scores were translated into SF-6D utilities using the algorithm developed by Brazier *et al.* [4].

To assess the discriminative ability for severity of illness, the mean EQ-5D and SF-6D utilities were compared in severity subgroups based on the SCL-90 scores subgroups. The utilities of EQ-5D and SF-6D were also analysed separately for adjacent severity subgroups. We hypothesised that preferences would decrease with increasing severity. Differences in preferences between the subgroups on both the EQ-5D and the SF-6D were analysed using a one-sided student's *t*-test. Differences between EQ-5D and SF-6D within subgroups were analysed using a paired *t*-test.

For the severity subgroups, the mean improvements in EQ-5D and SF-6D utilities at 1 and 1.5 years follow-up i.e. follow-up utility minus baseline utility, were assessed. As a statistical measure of responsiveness the standardised response mean (SRM) was calculated using the following formula: SRM = mean difference/standard deviation of difference.

Results

Overall comparison

An inspection of the distribution of patients' scores across the dimensions of both instruments revealed that, in general, more respondents reported having no limitations on the EQ-5D compared to the SF-6D. Furthermore, fewer respondents were placed in the worst category (extreme problems) of the EQ-5D dimensions. Tables 1 and 2 show the self-reported health scores on the dimensions of EQ-5D and SF-6D at baseline. Patients responded differently on related dimensions of the instruments. The greatest differences occurred between the EQ-5D dimensions mobility and self-care and the related SF-6D dimension physical functioning. Seventy-eight percent of the respondents appeared to have no mobility problems and 93.5% had no problems with self-care. By contrast, only 18% of the respondents reported having no limitations in physical functioning on the SF-6D. Furthermore, 23% of the patients had no problems in performing usual activities (EQ-5D), while, respectively, 10.9 and 6.2% reported no limitations in the related SF-6D dimensions role limitation and social functioning. Surprisingly, given the population studied here, a rather high degree of divergence was also seen between the dimensions of both instruments covering mental functioning: almost 65% reported having problems most of the time or all of the time due to mental health (level 4

DOI: 10.1002/hec

Table 1. Distribution of EQ-5D scores per dimension ($n = 616$)

Level	Mobility (%)	Self-care (%)	Usual activities (%)	Pain/discomfort (%)	Anxiety/depression (%)
1. No problems	78.6	93.5	23.1	24.0	7.0
2. Some problems	20.3	6.2	69.2	66.2	59.6
3. Extreme problems	1.1	0.3	7.8	9.7	33.4

Table 2. Distribution of SF-6D scores per dimension ($n = 616$)

Level	Physical functioning (%)	Role limitation (%)	Social functioning (%)	Pain (%)	Mental health (%)	Vitality (%)
1 (no limitations)	18.0	10.9	6.2	16.4	0.3	1.0
2	21.6	5.0	10.7	10.4	1.8	6.2
3	19.5	33.3	34.4	29.5	33.1	31.3
4	8.4	50.8	32.8	23.2	45.3	44.5
5	30.0	NA	15.9	13.8	19.5	17.0
6	2.4	NA	NA	6.7	NA	NA

NA, not applicable.

Table 3. Correlations[a] between EQ-5D and SF-6D dimensions

	EQ-5D dimension				
SF-6D dimension	Mobility	Self-care	Usual activities	Pain/discomfort	Anxiety/depression
Physical functioning	**0.305**	**0.161**	0.194	0.309	0.058[b]
Role limitations	0.240	0.182	**0.329**	0.245	0.157
Social functioning	0.149	0.186	**0.342**	0.188	0.296
Pain	0.353	0.234	0.323	**0.570**	0.210
Mental health	0.105	0.081	0.202	0.112	**0.415**
Vitality	0.128	0.157	0.272	0.182	0.282

[a] Correlations between related dimensions are indicated in bold.
[b] Not statistically different from zero ($p > 0.05$).

and 5) on the SF-6D, compared to 33% who reported having extreme problems on the anxiety/depression dimension of the EQ-5D.

Rank correlations between the dimensions of the EQ-5D and the SF-6D, based on the patients' self-reported health, are presented in Table 3. Related dimensions based on the descriptive systems of both instruments are indicated in bold. One would expect to find the strongest correlations between the related dimensions on EQ-5D and SF-6D. Indeed, the highest correlation of 0.57 was found between the dimensions pain/discomfort (EQ-5D) and pain (SF-6D), followed by the correlation between the EQ-5D dimension anxiety/depression and the SF-6D dimension mental health. A low correlation was observed between the dimensions self-care and physical functioning.

In the total population, the mean EQ-5D and SF-6D utilities were 0.518 (SD 0.29) and 0.575 (SD 0.10), respectively. The median scores were 0.689 and 0.567. Both mean and median EQ-5D and SF-6D utilities were significantly different.

Comparison by severity

The study population was divided into four severity subgroups based on the quartiles of SCL-90 scores. The mean SCL-90 scores for these subgroups were 151, 198, 241 and 302. The mean

DOI: 10.1002/hec

Table 4. Mean, median, standard deviation of EQ-5D and SF-6D utility scores per severity subgroup based on SCL-90 scores

Severity subgroups based on SCL-90 scores	*N*	Mean	SD	Median
1st quartile (scores 95–178)				
EQ-5D index		0.701*	0.179	0.725*
SF-6D index	155	0.645*	0.092	0.626*
2nd quartile (scores 179–218)				
EQ-5D index	154	0.597*	0.237	0.689*
SF-6D index		0.595*	0.075	0.591*
3rd quartile (scores 219–263)				
EQ-5D index	153	0.480*	0.267	0.620*
SF-6D index		0.561*	0.074	0.547*
4th quartile (scores 264–404)				
EQ-5D index	154	0.294	0.308	0.255
SF-6D index		0.499	0.081	0.494

SD, standard deviation.
*Statistically significant difference between the group and the adjacent group of the next quartile ($p < 0.050$, one-sided).

number of DSM-IV diagnoses increased with severity from 1.35 in the first quartile to 2.35 in the fourth quartile. Of all patients, almost half had been diagnosed with a single disorder, 34% had two and over 19% of the patients had three or four diagnoses.

The mean and median utilities for severity subgroups are presented in Table 4. Overall, the SF-6D utilities were 0.057 higher than the mean EQ-5D utilities for the total population. The mean difference per severity subgroup ranged from 0.06 higher EQ-5D utilities in the low severity subgroups formed by the first quartile of SCL-90 scores to 0.21 lower EQ-5D utilities in the high severity subgroup formed by the fourth quartile of SCL-90 scores. The differences in mean EQ-5D and SF-6D utilities were statistically significant in all subgroups except for the subgroup formed by the second quartile of SCL-90 scores. Larger variances were found in the utilities on the EQ-5D compared to the SF-6D utilities.

The EQ-5D and SF-6D utilities were compared separately in adjacent severity subgroups. Both mean and median EQ-5D utilities decreased as severity increased. A relatively high difference was seen in the EQ-5D utilities between the subgroups formed by the third and fourth quartile of SCL-90 scores. Mean EQ-5D utilities differed significantly between all subgroups. The median EQ-5D utilities were systematically higher than the mean utilities with the exception of the median score in the highest severity subgroup. Similarly, the mean utilities of the SF-6D decreased with severity. Both mean and median utilities showed statistically significant differences between adjacent severity subgroups. The SF-6D median utilities were about the same or lower than the mean utilities. The patterns of the EQ-5D and the SF-6D utilities across the severity subgroups are shown in Figure 1.

The changes in utilities from baseline until 1 year and 1.5 years follow-up and the SRMs are presented in Table 5. For this analysis the data of 355 and 326 patients, respectively, with complete EQ-5D and SF-6D follow-up data were used. The mean EQ-5D and SF-6D utilities at baseline in the severity subgroups with complete follow-up were comparable with those in Table 4 for the total study population at baseline. In line with the results of the outcomes used in the clinical study, significant differences were found between the baseline and follow-up utilities on both instruments (Hakkaart-van Roijen *et al.*, accepted for publication). Overall, mean EQ-5D utilities increased from 0.513 at baseline to 0.680 after a 1.5-year follow-up and the SF-6D utilities increased from 0.577 to 0.701. The mean improvement in EQ-5D utilities was lower than for the SF-6D utilities in the low severity subgroup and higher for the two subgroups with the highest

DOI: 10.1002/hec

severity. The mean improvement in EQ-5D utilities increased with increasing severity. The standard deviations of the differences in utilities were lower for SF-6D than for the EQ-5D utilities. This resulted in SRMs for SF-6D utilities that were at least twice as high as the SRMs for EQ-5D utilities in all severity subgroups.

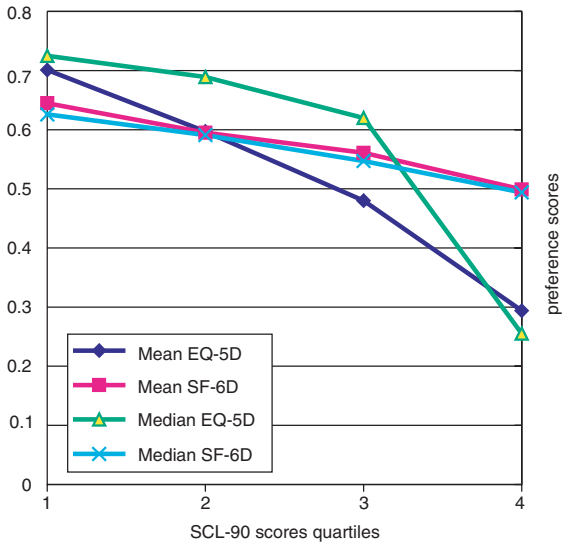

Figure 1. Patterns of EQ-5D and SF-6D utilities across severity subgroups

Discussion

In this paper, the ability of two different preference-based measures to assess the health-related quality of life in a group of patients suffering from the most common mental health problems was compared. The ability of the EQ-5D and the SF-6D to discriminate between severity subgroups of mental health patients was also studied. The results of the overall comparison showed that compared to SF-6D relatively more patients were found at the upper end of the scale (no problems) and fewer patients were found at the worst level on the EQ-5D dimensions, except for the dimensions of both instruments covering mental functioning. Almost 65% reported having problems most of the time or all of the time (the two highest levels) due to mental health on the SF-6D, compared to 33% who reported having extreme problems on the anxiety/depression dimension of the EQ-5D. Overall, these findings correspond to the results found by Brazier *et al.* [9] in a sample of patient with seven different medical conditions.

The different responses of patients on related dimensions of the instruments may (partly) be explained by differences in the classification systems of the dimensions of the instruments. With EQ-5D 243 health states can be distinguished, whereas SF-6D describes 18 000 different health states. In the algorithm developed by

Table 5. Mean improvement and SRM from baseline to 1 and 1.5 years follow-up for EQ-5D and SF-6D

Group based on SCL-90 scores quartiles	*n*	Mean improvement	SD difference	SRM	Mean improvement	SD difference	SRM
		EQ-5D: $T_{1\ year\ follow\text{-}up} - T_{baseline}$			*SF-6D*: $T_{1\ year\ follow\text{-}up} - T_{baseline}$		
1st quartile	91	0.073	0.273	0.267	0.093	0.149	0.621
2nd quartile	90	0.132	0.271	0.485	0.097	0.136	0.713
3rd quartile	92	0.173	0.349	0.496	0.096	0.138	0.699
4th quartile	82	0.240	0.377	0.636	0.100	0.122	0.823
Total	355	0.152	0.324	0.470	0.097	0.136	0.710
		EQ-5D: $T_{1.5\ year\ follow\text{-}up} - T_{baseline}$			*SF-6D*: $T_{1.5\ years\ follow\text{-}up} - T_{baseline}$		
1st quartile	76	0.094	0.223	0.423	0.127	0.146	0.872
2nd quartile	86	0.105	0.308	0.339	0.094	0.127	0.739
3rd quartile	85	0.163	0.423	0.386	0.138	0.162	0.852
4th quartile	79	0.311	0.403	0.771	0.139	0.158	0.878
Total	326	0.167	0.359	0.466	0.124	0.149	0.833

SD, standard deviation. SRM, standardised response mean.

DOI: 10.1002/hec

Brazier and colleagues to calculate SF-6D preferences, not all differences in SF-6D health states are translated into utility differences. In terms of utilities, 3000 different health states exist [4], which is still much more than the 243 states of EQ-5D. For some of the SF-6D dimensions, it is less apparent that the succeeding levels produce one continuum, because the various levels of one dimension are composed using more than one SF-36 item. For example, the physical functioning dimension ranks the health of a patient with some limitations in bathing and dressing as worse than that of a patient experiencing a lot of limitations in moderate activities.

To our knowledge, no previous studies have been carried out examining the possible differences in utilities on the EQ-5D and the SF-6D between subgroups of patients with a different degree of severity of illness. Both the EQ-5D and the SF-6D showed a decrease in utilities with increasing severity. For both instruments mean utilities were significantly different between all adjacent subgroups, thus demonstrating the ability of the EQ-5D and SF-6D to discriminate between severity subgroups of mental health patients. However, the difference in mean utility between the low and high severity subgroup is larger for EQ-5D than for SF-6D, which is mainly caused by the difference in mean utilities for the high severity group. This result seems to support the findings of Longworth and Bryan, who found that the SF-6D was less suitable to describe the health states at the lower end of the utility scale [10].

In order to explore the impact of using either the EQ-5D or SF-6D in cost–utility analyses, we examined changes in utilities for the patients on both instruments from baseline until end of treatment. Significant differences were found between the baseline and follow-up utilities on both instruments for the entire study population as well as the severity subgroups. Due to the smaller standard deviations of the differences in SF-6D utilities, the SRMs for the SF-6D were at least twice as high for SF-6D compared tot EQ-5D. However, in cost–utility analysis the mean improvements are used and not the SRMs. This means that for the total study population and three of the four severity subgroups the use of EQ-5D results in a more favourable cost–utility ratio. The greatest effect of the choice between EQ-5D and SF-6D on the cost–utility ratio was observed for the high severity subgroup, which had the largest difference in health gain. These results were comparable with the findings of Pickard *et al*. [15] in stroke.

Besides differences in their descriptive systems, the EQ-5D and SF-6D valuation studies showed differences that might (partly) explain these results. EQ-5D health states were valued using the time trade-off method (TTO), whereas the SF-6D valuation study made use of the standard gamble (SG). Various studies have shown that the SG generally produces higher values than the TTO [16,17]. The specific 'chained' variant of the SG method used in the SF-6D valuation study was expected to result in large overestimations of utilities [18–21]. The resulting preferences could then be corrected for this overestimation [22]. However, in the SF-6D valuation study no such correction was reported [4]. This might explain the relatively high preference score of 0.30 for the worst possible SF-6D health state compared to the score of −0.59 for the worst EQ-5D health state. The larger range of the EQ-5D utility scale provides a potential for the assessment of larger health gains and also caused the higher standard deviations of EQ-5D utilities compared to SF-6D.

In conclusion, both EQ-5D and SF-6D could discriminate between severity subgroups and captured improvements in health over time. However, the use of EQ-5D resulted in larger health gains, which will be translated to lower cost–utility ratios, especially for the subgroup with the highest severity of mental health problems. We concluded that the EQ-5D and the SF-6D are to some extent interchangeable in cost–utility analysis. This study illustrated that in populations or subgroups with severe mental health problems the results of cost–utility analyses can be affected by the choice of the instrument to measure health-related quality of life. For this population, the use of EQ-5D might result in more favourable cost–utility ratios.

Acknowledgements

The primary data were extracted from a study funded by the Dutch Organisation for Health Research and Development (ZonMW). The authors thank Han Bleichrodt for his comments on an earlier version of this paper. The opinions expressed in this paper are those of the authors and do not necessary reflect those of the aforementioned.

References

1. Drummond M, O'Brien B, Stoddard G, Torrance G. *Methods for the Economic Evaluation of Health Care Programmes* (2nd edn). Oxford University Press: Oxford, 1998.
2. Patrick D, Starks H, Cain K, Uhlmann R, Pearlman R. Measuring preferences for health states worse than death. *Med Decis Making* 1994; **14**: 9–18.
3. Brooks R. EuroQol: the current state of play. *Health Policy* 1996; **37**: 53–72.
4. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002; **21**: 271–292.
5. Feeny D, Furlong W, Torrance G *et al*. Multi-attribute and single-attribute utility functions for the health utility index mark 3 system. *Med Care* 2002; **40**: 113–128.
6. Feeny D, Furlong W, Boyle M, Torrance G. Multi-attribute health status classification systems: health utility index. *Pharmacoeconomics* 1995; **7**: 490–502.
7. Dolan P. Modelling valuations for EuroQol health states. *Med Care* 1997; **35**: 1095–1108.
8. Dolan P, Gudex C, Kind P, Williams A. A time trade-off method: results from a general population study. *Health Econ* 1996; **5**: 141–154.
9. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and the SF-6D across seven patient groups. *Health Econ* 2004; **13**(9): 873–884.
10. Longworth L, Bryan S. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Econ* 2003; **12**: 1061–1067.
11. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (4th edn). American Psychiatric Association: Washington, DC, 2000.
12. Arrindell W, Ettema J. *SCL-90. Manual for a Multidimensional Psychopathology Indicator*. Swets & Zeitlinge: Lisse, 2003 (in Dutch).
13. Kool S, Dekker J, Duijsens I, de Jonghe S. Major depression, double depression and personality disorders. *J Pers Disord* 2000; **14**: 274–281.
14. Olsen L, Mortensen E, Bech P. The SCL-90 and SCL-90R versions validated by item response in Danish community sample. *Acta Psychiatr Scand* 2004; **110**: 225–229.
15. Pickard AS, Johnson JA, Feeny DH. Responsiveness of generic health-related quality of life measures in stroke. *Qual Life Res* 2005; **14**: 207–219.
16. Dolan P. The measurement of health-related quality of life for use in resource allocation decisions in health care. In *Handbook of Health Economics*, Cuyler A, Newhouse J (eds). Elsevier: Amsterdam, 2000; 1723–1760.
17. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ* 2002; **11**: 447–465.
18. Llewellyn-Thomas H, Sutherland H, Tibshirani R. The measurement of patients' values in medicine. *Med Decis Making* 1982; **2**(4): 449–462.
19. Rutten-van Molken MP, Bakker C, van Doorslaer EK, van der Linden S. Methodological issues of patient utility measurement. Experience from two clinical trials. *Med Care* 1995; **33**(9): 922–937.
20. Bleichrodt H. Probability weighting in choice under risk: an empirical test. *J Risk Uncertain* 2001; **23**: 185–198.
21. Oliver A. The internal consistency of the standard gamble test: tests after adjusting for prospect theory. *J Health Econ* 2003; **22**(4): 659–674.
22. Bleichrodt H, Pinto J, Wakker P. Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Manage Sci* 2001; **47**: 1498–1514.

DOI: 10.1002/hec