WORKING PAPERS IN ECONOMICS

No.12/11

LENE LUNDE

CAN EQ-5D AND 15D BE USED INTERCHANGEABLY IN ECONOMIC EVALUTIONS? ASSESSING QUALITY OF LIFE IN POST-STROKE PATIENTS

Department of Economics
UNIVERSITY OF BERGEN

Can EQ-5D and 15D be Used Interchangeably in Economic Evaluations?

Assessing Quality of Life in Post-Stroke Patients

By

Lene Lunde

University of Bergen
Department of Economics
Hermann Fossgt. 6
5007 Bergen, Norway
Telephone: +4755589238
Fax: +4755589210
Email: lene.lunde@econ.uib.no

Abstract

This paper provides an analysis on whether 15D and EQ-5D can be used interchangeably as health outcome measures in cost-effectiveness studies. The empirical performance of the two multi attribute utility instruments is tested in terms of feasibility, utility score, linear relationship and agreement by using a novel Norwegian data set. There is also provided an analysis of how the instruments rank individuals in terms of health status, and how they distinguish between an individual with a certain trait and not. The results show that EQ-5D and 15D cannot be used interchangeably in economic evaluations, and that EQ-5D is likely to give a more favourable cost utility ratio than 15D. The utility scores generated from the two instruments are significantly different from each other, even though they correlate well. The instruments also rank individuals in terms of health status differently.

Key words: Multi-attribute utility instruments, empirical performance, EQ-5D, 15D, post-stroke patients

JEL codes: I10 I19

1. Introduction

Using QALYs as an outcome measure in economic evaluations permits comparison of the cost-effectiveness of diverse health care interventions. This comparability is considered one of the main advantages of cost utility analysis. However, the comparability of the analysis may be jeopardised because of diverse approaches, or different instruments, used to elicit health utility. Health utility can be elicited directly from patients by using standard techniques, such as time trade-off or standard gamble or, indirectly, by using questionnaires that incorporate preference weights from community members. Ranges of different questionnaires to measure health preferences indirectly are developed and frequently used. We often call them multi-attribute utility instruments (MAU), or preference-based measures. These instruments consist of two interrelated components that vary from instrument to instrument. The first component is a classification system of health, and the second component is a system of utility weights. The utility weights are used to convert the descriptive system into a summary index score. There is so far no gold standard and few guidelines on which instrument are the most appropriate to use. The instrument used varies, and it is therefore important to be aware that they could have significant implications for the estimated QALY gains in economic evaluations and, thus, also for the comparability of different analyses. Lack of empirical evidence may be a reason that there is no gold standard.

In this paper, we focus on EQ-5D and 15D, two among many MAUs. We analyse and compare the empirical performance of the two instruments, and investigate whether they can be used interchangeably in cost utility analysis. Can the effectiveness of an intervention where EQ-5D utilities are used be compared with the effectiveness of an intervention measured with 15D utilities? In the previous literature, there has been a lot of focus on the comparative performance of EQ-5D and SF-6D (Barton et al., 2008; Brazier et al., 2004; Lamers et al., 2006; Longworth Seymour et al., 2010) while there has not been much emphasis on empirical evidence of EQ-5D and 15D's comparative performance, even though they differ in their measurement characteristics. The first impression is that 15D seems to have a very rich and sensitive classification system, while it looks like EQ-5D has a poor classification system that does not cover all aspects of health. Nevertheless, the EQ-5D is much more frequently used, while 15D nearly seems to be forgotten. EQ-5D is, for instance, recommended for use in cost-effectiveness analyses by the National Institute of Health and Clinical Excellence (NICE) in the UK, and by the health care insurance board in the Netherlands. In an earlier review of Brazier et al., (1999) concluded that we need head-to-head comparisons of different MAUs across a range of conditions and severity levels in order to assess the implications for the interpretation and comparability of economic analyses.

Stroke is the most common reason for serious disability and need for long-term public health care services in Western societies (Fjærtoft et al., 2007; Hannerz et al.,2001). Annually, around 15 million people suffer from stroke worldwide. Five million of them die, and another five million are left permanently disabled (Mackay et al., 2004). Previous studies have compared the EQ-5D and 15D for several diseases (Saarni et al., 2006; Sintonen, 2001a; Stavem et al., 1999, 2001, 2005), but this is the first study investigating the comparative performance of EQ-5D and 15D when assessing HRQoL among post-stroke patients. The importance of economic evaluations of stroke-related care and rehabilitation is likely to increase in the future due to an aging population and, thereby, increased incidence rates (Hannerz et al., 2001). This will strain the health care system in acute services, nursing and care, rehabilitation and, thus the resources spent within this field.

From the earlier studies comparing EQ-5D and 15D, we know that 15D, in general, tends to give a slightly higher mean and median utility score than EQ-5D. However, in these studies, only one set of utility weights for EQ-5D has been explored. A range of utility weights have been developed for EQ-5D; therefore, we will analyse the comparative performance of EQ-5D and 15D by using two sets of utility weights for EQ-5D. We are interested in finding out whether value sets make any difference for the comparative performance. We use a novel Norwegian data set and analyse the differences between 15D and EQ-5D in terms of feasibility and utility score. It is time-consuming and costly to gather health-related quality-of-life data. The instruments' feasibility is therefore important. A highly functional preference-based instrument should be easy to complete for all kinds of patients. Stroke victims may be paralysed or have cognitive impairments, and may have difficulties with completing such a questionnaire. We analyse the instruments' feasibility by looking at different response and completion rates. If EQ-5D and 15D are to be used interchangeably, they have to produce same utility scores, or the scales at least must be a linear transformation of each other. The distribution of utility scores and summary statistics are fully explored and compared. Further, we assess the association between the two instruments and their sensitivity, and compare the ranking of patients' health status level by each instrument.

The structure of the paper is as follows. The EQ-5D and the 15D are briefly described in section 2. Patient recruitment, procedures, and description of the data and variables are presented in section 3. We provide the results of the analysis in section 4; first, we look at the instruments' feasibility; secondly, the total utility scores and their properties. The discussion of findings in section 5 is followed by conclusions.

2. 15D and EQ-5D

EQ-5D consists of five attributes that cover mobility, self-care, usual activity, pain/discomfort, and anxiety/depression. Each attribute has three levels, going from 1, indicating no problems, to 3, indicating extreme problems. Two hundred and forty-five possible health states (dead and unconscious included) are defined through this instrument. We make use of two different value sets[1], henceforth referred to as EQ-TTO (the British value set based on time trade-off) and EQ-VAS (the Finnish value set, based on VAS). When EQ-TTO was developed, the time trade-off was applied to 2997 adults in the UK to measure their health preferences for in total 43 different health states. However, each respondent valued no more than 13 hypothetical health states. They were given the opportunity to value them as worse as or better than being dead and valued each health state one time. A generalised least-squares regression technique was used to derive the scoring function (Dolan, 1997). The model is additive and the minimum possible score is negative (-0.59). When EQ-VAS was developed, 928 randomly selected healthy Finns valued hypothetical health states using VAS. A weighting for age and sex was done in the sample due to an overrepresentation of the elderly. Each respondent valued a subset of 46 states with duration of one year on the standard VAS. Dead and unconscious were included. They valued each health state once, except for three: the best possible (11111), the worst possible (33333), and dead. The final model used to derive an additive utility function is based on ordinary least square and logit transformation of the individual data (Szende et al., 2007). The lowest possible score in this system is also negative (-0.011).

The 15D measures HRQoL using 15 different attributes: mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort/symptoms,

[1] Several systems of country-specific utility weights (value sets) have been developed to assign scores to the defined health states. Direct techniques such as the time trade-off or the VAS have been used to develop the different scoring algorithms.

depression, distress, vitality, and sexual function. Each attribute has five levels, where level 1 describes a situation with no problems, and level 5, a situation with severe problems. Consequently, 15D defines an enormous number of health states (5^{15}). Only one set of utility weights has been developed for the 15D. These weights were elicited from several representative samples of the Finnish adult population through a three-stage valuation procedure, using a VAS (Sintonen, 2001b). Respondents first indicated the relative importance of each dimension on a VAS scale, and then valued each level on each dimension (including dead and unconscious) using a VAS. The scoring function is also additive, but in contrast to the two selected value sets in EQ-5D, the lowest possible score is 0.11.

3. Methods

3.1. Recruitment

At least six months after stroke onset, all surviving stroke[2] patients admitted to a stroke unit at a large (Norwegian) hospital[3], between January, 2006 and February, 2009 were sent a questionnaire. The reason for this time restriction was that the validity of the patients' own judgements of HRQoL can be doubtful, especially during the acute phase after stroke and significant changes in health conditions (Van Exel et al., 2004). After the most acute phase, the health condition may be assumed to have stabilised. All patients surviving at least six months after the stroke occurred were included in order to reduce selection problems. Non-responders were sent a postal reminder at least six weeks after the first invitation. We allowed the patients' carers to complete the questionnaire (with the agreement of the patient) due to high rates of paralysis amongst stroke victims. Patients were categorised as disabled if they

[2] Patients suffering from ischemic stroke, haemorrhagic stroke, and transient ischemic attack (TIA).

[3] Haukeland University Hospital, one of the largest hospitals in Norway.

were not capable of completing the questionnaire themselves. Ethical approval was obtained from the regional committee for medical and health professional research ethics in western Norway and from Norwegian Social Science Data Services. A pilot study was carried out at the beginning of the survey[4].

3.2. Data and variables

In addition to the HRQoL data from EQ-5D and 15D, information was collected concerning the patients' gender, age, marital status, education level, income level, occupational status, and body mass index (BMI). The sample was categorised into three different age groups: (i) ≤ 66 years, (ii) 67 to 79 years, and (iii) ≥ 80. Marital status is categorised into three categories: (i) married, (ii) widowed, and (iii) single. We split education and income level into two groups. Those with a college or university degree are said to have higher education, while those who earn more than NOK[5] 500 000 annually are classified as high-income. Occupational status is categorised into three different groups: (i) employed, (ii) retired, and (iii) receivers of disability pension[6]. Finally, the BMI value is categorised into four groups in line with the World Health Organization's classification: (i) underweight (BMI<18.5), (ii) normal range (BMI 18.5-24.99), (iii) overweight (25≤ BMI <30), and (iv) obese (BMI ≥ 30)

A range of health status indicators like self-assessed health, assessment of overall health using the visual analogue scale (VAS), and self-reported information about chronic diseases, functional problems, and pain were included in the questionnaire. The self-assessed health question has five standard categories, which we collapsed into three categories: (i) very good or good health, (ii) mediocre health, and (iii) poor or very poor health. The VAS assessment

[4] In the pilot, 50 patients were randomly selected to receive the invitation and the questionnaire. After the pilot, we made some minor changes to the questionnaire.
[5] NOK 500 000 is equivalent to USD 97087, 40. USD 1 = NOK 5,15
[6] Only three individuals were unemployed, and we did not, therefore, include a group for them.

of overall health is categorised into these groups: (i) good VAS (≥ 80), (ii) average VAS (≥40 and <80), and (iii) poor VAS (<40). We created binary variables for having any chronic diseases, functional problems or pain where 1 indicates severe or some problems related to these aspects, and 0 indicates no problems.

Further, the survey contained three different clinical measures that are widely used to identify common symptoms and health indicators of post-stroke patients. Performance of daily activities related to mobility and self-care is measured by the Barthel Index (BI). The index is categorised into three different groups after Sulter et al., (1999): (i) independent, with minimal assistance (BI ≥ 85), (ii) assisted independence (60≤ BI <85), and (iii) dependent (BI < 60) Fatigue is measured by the fatigue severity scale (FSS), and indicated by a binary variable equal one when fatigue is present. Anxiety and depression are measured by the Hospital Anxiety and Depression Scale (HADS). Depression and anxiety are categorised in a binary variable where one indicates a definite case of depression or anxiety. We constructed a binary variable for each of the three types of stroke considered (ischemic, intracerebral haemorrhage and subarachnoid haemorrhage). In addition, we include three variables indicating time elapsed in months since the stroke: i) <12 months, ii) 12-18 months, and iii) more than 18 months.

3.3. Procedure

We investigate the instruments' feasibility with the overall response rate of the survey, and the completion rates for EQ-5D and 15D. The three sets of utility scores computed with the instruments are compared visually, by box plot, and by summary statistics. Differences between distributions are tested with a Kolmogorov-Smirnov test. Summary statistics are also reported for quintiles. Differences between scores are tested by using paired t-tests and signed rank tests. The sensitivity of each instrument is investigated with a test of construct validity[7]; the validation is done by extreme groups and in line with Streiner et al., (1995). The association between the two instruments is measured by the intraclass correlation coefficient (ICC), by Spearman rank correlation, and by scatter plots. The ICC is derived from a two-way mixed model with an absolute agreement definition and average measure according to Shrout et al., (1979). The extent of the Spearman rank correlation between scores is described[8] in line with Pickard et al., (2005). To check whether the individuals are ranked equally by each instrument, we estimate an ordinary least-square regression, with rank as the dependent variable and the socioeconomic variables, health status, and clinical indicators as covariates. The effects of the covariates on the rank from each regression should not be significantly different if the individuals within each group have the same rank. All analyses used standard statistical software (Intercooled Stata 10.1 and SPSS 15). *P-values* less than 0.05 are statistically significant.

4. Results

4.1. Feasibility

Of the 698 patients invited, 408 (58.5%) returned the questionnaire (304 after first invitation, and 104 after the reminder). 333 individuals (81.6%) completed the 15D, whereas 345 of the respondents (84.6%) completed the EQ-5D. The difference in completion rates is not significant.

[*Table 1 about here*]

---

[7] Construct validity is measured by a scale's ability to distinguish between utility scores within one group where some respondents have a certain trait and others do not.

[8] The correlation is defined as absent (<0.20), poor (0.20-0.34), moderate (0.35-0.50), and strong (>0.50).

Table 1 presents differences in completion rates between 15D and EQ-5D for subgroups where the differences are significant at the 5% level. Completion rates are, in general, higher for EQ-5D than for 15D apart from those who are married and who do not have any problems with pain. When investigating which attributes the respondents failed to answer (table 2), we find the majority of missing values in the attribute of sexual activity and symptoms/pain for 15D. All of the respondents answered the questions concerning mobility, hearing, and vitality. In EQ-5D, we find the majority of the missing values in the attribute of pain and anxiety. There are, relative to the other sections, many missing values in the attribute of usual activity in EQ-5D, compared to 15D's corresponding attribute.

[*Table 2 about here*]

4.2. Utility Scores

Figures 1-3 show the distribution of the EQ-5D and 15D scores. The Kolmogorov-Smirnov test[9] confirms that all distributions differ from each other. However, they all have a negative skew in common[10]. We find the smallest skew in the distribution for the EQ-VAS scores. Figure 2 illustrates that there is a gap between the best ill-health state and perfectly healthy. In fact, when tabulating the score, we do not find any utility scores between 0.8 and 1.00 in EQ-VAS, while we do not find any between 0.88 and 1.00 in EQ-TTO.

[Figure 1-3 about here]

[9] The null hypothesis that the distributions of 15D and EQ-VAS are equal is rejected (p<0.0001). The hypotheses that the distribution of 15D and EQ-TTO are equal are also rejected (p < 0.0001), and we find the same when testing the equality for the distributions of EQ-TTO and EQ-VAS.

[10] The distribution of 15D has a skewness = -0.94 (p<0.0001), the skewness of EQ-TTO = -1.46 (p<0.0001), and the skewness of EQ-VAS = -0.09 (p<0.5).

Figure 4 represents a box plot of the total utility scores computed with each of the MAU instruments. The median scores clearly differ by instrument and value set. 15D has the highest median score and EQ-VAS, the lowest. The boxes of EQ-TTO and EQ-VAS are a lot wider than the box for 15D. This indicates that the within-instrument variability is larger in EQ-TTO than in 15D, where the box is relatively small and the variation correspondingly moderate. We find most outliers with EQ-TTO, while there are none with EQ-VAS and few with 15D.

[Figure 4 about here]

The summary statistics in table 3 confirm the differences between scores. The mean 15D score is significantly higher than the mean EQ-VAS and EQ-TTO scores ($p < 0.0001$). In addition, the mean EQ-TTO score is significantly higher than the mean EQ-VAS score ($p < 0.0001$). Since the distributions of the utility scores are non-normal, we included a non-parametric test of the differences in utility scores. The signed-rank test confirms that the differences are significant. With respect to minimum values and the range of utility scores, there are large differences. EQ-TTO is particularly different from the others because of the low negative minimum value and the wide range. We can clearly see that the score of 15D is higher than the scores of EQ-TTO and EQ-VAS across the 25th, 50th, and 75th quantiles of the distribution, and the EQ-TTO score is higher than the EQ-VAS score across these quantiles.

[Table 3 about here]

When investigating the mean scores across quintiles of the three different distributions, we find the same pattern. The mean score of 15D is higher than the mean score of EQ-TTO and EQ-VAS across all quintiles, except for the fifth (table 4). The difference between the mean scores is largest in the first quintile and decreasing over higher quintiles. However, the differences in median are not decreasing with the quintiles. Except for the first quintile, the EQ-TTO value is higher than the EQ-VAS value. In the fifth quintile, they are equal for both mean and median. Except for the fifth quintile, we find the smallest range of scores in 15D. The range of EQ-VAS scores is larger than the range of EQ-TTO scores between the second and fourth quintiles.

[Table 4 about here]

4.3. Sensitivity

Table 5 reports the mean and median utility scores for 15D, EQ-TTO, and EQ-VAS by subgroups identified by health status and clinical indicators. Both instruments detect significant differences between those with a certain trait and not, and can therefore be said to be sensitive. However, the magnitude of the differences in utility scores varies between the instruments and value sets. For all subgroups, the EQ-5D shows larger differences than the 15D does. In 10 out of 11 cases, EQ-TTO detects larger differences than EQ-VAS.

[*Table 5 about here*]

4.4. Agreement and association

According to the intraclass correlation coefficient (ICC), we find fair to good agreement between 15D and EQ-TTO (ICC= 0.75)[11]. The Spearman rank correlation indicates a strong association between 15D and EQ-5D (= 0.69 and 0.68). Figures 5 and 6 are scatter plots of individuals' 15D scores against their EQ-TTO and EQ-VAS scores, respectively. Clearly, the 15D scores are more similar to the EQ-TTO scores than the EQ-VAS scores. The figures indicate ceiling effects in EQ-5D.

[11] In valuing the ICC, we used these criteria to describe the size of the correlation: poor reproducibility (< 0.40), fair to good reproducibility (0.40-0.75), and excellent reproducibility (>0.75) [31].

[Figure 5 & 6 about here]

Eighty-one (23.48%) individuals have perfect health according to EQ-5D, while according to 15D only twenty-five (7.53%) individuals have perfect health. This difference is statistically significant (*p=0.000)*. There are relatively few values in the lower end of the scales, and we do not find any evidence of floor effects. None of the individuals has the lowest possible scores; however, there are some negative values. These occur only in EQ-TTO, while it is also possible in EQ-VAS. There is a significant variation in the number of low values. Of the 15D scores, 2.4% are below 0.5, whereas 15.7% and 20.3% of EQ-TTO and EQ-VAS scores, respectively, are below 0.5.

4.5. Rank

Table 6 reports the regression coefficients and standard errors of the rank regression of 15D, EQ-TTO, and EQ-VAS. The coefficients show the effect of the covariates on the rank of each instrument's score. There should not be any significant differences between the coefficients from each regression if the instruments rank individuals equally after utility score. We find significantly different effects between individuals with a poor VAS value, and individuals that are of normal weight when looking at the results of EQ-TTO and 15D. Employed individuals, those with good self-assessed health, and those who have a poor VAS value have significantly different effects on the rank of EQ-VAS and 15D. This suggests that individuals are not ranked similarly by 15D and EQ-5D scores.

[Table 6 around here]

5. Discussion

EQ-5D had a higher completion rate than 15D for the whole sample and for most of the subgroups, with some exceptions. However, this study indicates that the “total workload” of completing 15D might not be the reason for low completion rates, but rather one single attribute, in this case the attribute of sexual activity. The main reason for not fully completing the EQ-5D questionnaire seems to be the attribute of depression and anxiety. In the corresponding attributes of 15D, we find almost no missing values. The definitions of anxiety and depression are broader in 15D than in EQ-5D. This may lead to more answers. The five different levels in 15D are also more moderate and sophisticated than in EQ-5D. This may also be the reason that EQ-5D has so many more missing values in the attribute for usual activities compared to the corresponding attribute in 15D.

Earlier research has shown that when there is a floor or ceiling effect in one instrument and not in another, this leads to a weaker association between instruments at the extremes of the distribution, and to differences in precision relative to other parts of the distribution (Seymour et al., 2010). This is not exactly the case in our study. We identified a relatively large ceiling effect in EQ-5D compared to 15D. However, the distributions of scores are more similar in the upper part of the distribution, while there are many more inequalities at the lower end of the distributions. We did not identify any floor effects in any of the instruments. Nevertheless, there are large differences in how many patients scored less than 0.5. Furthermore, earlier studies have shown that the mean scores for different MAUs tend to be more similar in healthier populations than in those with higher morbidity (Conner-Spady et al., 2003) This seems to be the case here. The differences in mean and median scores are largest in the lowest quintile, and smallest in the highest quintile. This illustrates that the association between different instruments and value sets may differ at different points in the distribution. The large differences in the lower quintiles may be attributable to differences in the utility weights. When one instrument allows for negative values and another does not, there will be large differences because of different scoring ranges.

Different scoring rods and ranges might not be a problem as long as we are aware of them. However, according to our rank regression, we found evidence that patients with different characteristics are ranked differently by the instruments. We found more differences in ranking between 15D and EQ-VAS, than between 15D and EQ-TTO. This indicates that the scoring algorithms not only influence the scale of scores, but also how individuals are ranked after score. Furthermore, it suggests that there are probably more similarities between the scoring algorithms of EQ-TTO and 15D than between EQ-VAS and 15D since they rank the individuals more similarly. Since the utility weights of both EQ-VAS and 15D are based on VAS and preferences of the Finnish population, we expected their utility values to be closer to each other than the values of 15D and EQ-TTO. The scoring ranges between EQ-VAS and 15D are also more alike than EQ-TTO and 15D. Nevertheless, the mean, median, and the lower quartile of EQ-TTO are closer to 15D than EQ-VAS. Only the upper quartile of the EQ-VAS is closer to the 15D than the EQ-TTO. This confirms that the association between instruments and value sets differ at unique points in the distribution. But most importantly, it illustrates that the scoring algorithms are more decisive for differences in utility score than method of valuation and institutional settings where the preferences are measured.

Both instruments, irrespective of value set, distinguish between individuals with a certain trait and those without. However, EQ-TTO and EQ-VAS find much larger differences than 15D. The distribution of EQ-5D is heavily skewed towards the top, and the patients have, in general, higher values with 15D than with EQ-5D. 15D largely avoids ceiling effects due to the number of possible health states. Because the utility indices derived by 15D are initially high, and there are few scores below 0.5, there may be less room for great differences between those with a certain trait and those without. Or, framed a different way, there is less room for a large health improvement in 15D. This is an important finding because the effect of a health care intervention can then potentially be underestimated by 15D or overestimated by others.

7. Conclusion

In this paper, we have compared the empirical use of 15D and EQ-5D. Importantly, this has demonstrated that EQ-5D and 15D cannot be used interchangeably in economic evaluations. Both of them are feasible to use when measuring HRQoL in stroke patients. EQ-5D results in a slightly higher completion rate than 15D. In addition, it seems that 15D is more feasible for men than for women, for those who are married, and for those without any problems with pain. EQ-5D seems to be less feasible for individuals with depression and anxiety and problems with pain. As in earlier studies (Saarni et al., 2006; Sintonen, 2001a; Stavem et al., 1999, 2001, 2005), we have found the utility scores of the instruments to be highly correlated even though the instruments are scaled differently. However, we find that it is not only the scale that differs between the instruments; the ranking of individuals after utility scores also differs. This constitutes a problem for the comparability of QALY analysis. A health improvement will then probably be valued very differently. When the scales from each instrument are not proportional to each other, it is not enough to adjust and standardise the scores before the QALYs are calculated. We also have to keep in mind that the differences of the utility scales vary across the distributions. We found 15D to give consistently higher scores than EQ-5D, as in earlier findings. We found this to be irrespective of value set used in EQ-5D. Even though there are large ceiling effects in EQ-5D, this instrument will probably detect larger health gains related to an intervention than 15D. The reason is that all of the 15D values are concentrated at the upper end of the distribution, and then there is less room for large improvements. We therefore conclude that EQ-5D can potentially get more favourable cost utility ratios than 15D. A further examination must be done to investigate whether these differences in health gains are large enough to alter the result of a cost utility analysis. The use of two different value sets in EQ-5D has demonstrated that the valuation method is less decisive for the different scoring scales. The country in which the valuation is done is also of less importance. The construction of the scoring algorithms creates the largest differences between 15D, EQ-VAS, and EQ-TTO. Due to these differences, we must be very careful with the interpretation and comparisons of economic evaluations where these instruments are used. Furthermore, we cannot use them interchangeably. Comparisons of studies where EQ-5D is used with different value sets must also be done carefully due to the diverse construction of scoring algorithms.

Acknowledgements

The author would like to thank the division of neurology at Haukeland University Hospital, and especially chief physician Dr. Halvor Næss for facilitating the data collection. Thank you to Dr. Jan Brøgger also for helping out with data technicalities. I am also grateful for advice and comments by my supervisors, Professor Jan Erik Askildsen and Professor John Cairns. Finally, the author would like to thank the research network Health Economics Bergen and the Department of Economics, University of Bergen for financial support.

References

Barton G.R, Sach T.H, Avery A.J, Jenkinson C, Doherty M, Whynes D.K, Muir K.R. 2008. A comparisons of the performance of EQ-5D and SF-6D for individuals aged ≥ 45 years. *Health Economics* 17:815-32

Brazier J, Deverill M, Green C, Harper R, Booth A. 1999. A review of the use of health status measures in economic evaluation. *Health Technology Assessment* vol. 3: NO9

Brazier J, Roberts J, Tsuchiya A, Busschbach J. 2004. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics* 13:873-884

Conner-Spady B, Suarez-Almazor M. 2003. Variation in the estimation of Quality-adjusted Life-years by Different Preference-based Instruments. *Medical Care* 41(7): 791-801

Dolan P.1997. Modelling valuations for EuroQol health states. *Medical Care* 35(11): 1095-1108

Fjærtoft H, Indredavik B. Kostnadsvurderinger ved hjerneslag. 2007. *Tidsskr Nor Lægeforen* 6 (127):744-747

Hannerz H, Nielsen M.L. 2001. Life Expectancies Among Survivors of Acute Cerebrovascular Disease. *Stroke* 32:1739-1744

Lamers L.M, Bouwmans C.A.M, van Straten A, Donker M.C.H, Hakkaart L. 2006. Comparisons of EQ-5D and SF-6D utilities in mental health patients. *Health Economics* 15:1229-1236

Longworth L, Bryan S. 2003. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics* 12:1061-1067

Mackay J, Mensah G. Atlas of Heart Disease and Stroke. www.who.org [1 March 2008]

Pickard AS, Johnson JA, Feeny DH. 2005. Responsiveness of generic health-related quality of life measure in stroke. *Quality of Life Research* 14: 207-219

Saarni SI, Härkänen T, Sintonen H, Suvisaari J, Koskinen S, Aromaa A, Lönnquist J. 2006. The impact of 29 chronic conditions on health related quality of life: A general population survey in Finland using 15D and EQ-5D. *Quality of Life Research* 15:1403-1414

Seymour J, McNamee P, Scott A, Tinelli M. 2010. Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. *Health Economics* 19:683-696

Shrout P.E, Fleiss J.L. 1979. Intraclass correlation: Uses in Assessing Rater Reliability. *Psychological Bulletin* 86(2): 420-428

Sintonen H. 2001a. Comparing properties of the 15D and the EQ-5D in measuring health-related quality of life. *Archives of Hellenic Medicine* 18(2):156-160

Sintonen H. 2001b. The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine* 33(5): 328-336

Stavem K. 1999. Reliability, validity and responsiveness of two multi attribute utility measures in patients with chronic obstructive pulmonary disease. *Quality of Life Research* 8: 45-54

Stavem K, Bjørnæs H, Lossius MI. 2001. Properties of the 15D and the EQ-5D utility measures in a community sample of people with epilepsy. *Epilepsy Research* 44: 179-189

Stavem K, Frøland SS, Hellum KB. 2005. Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. *Quality of Life Research* 14:971-980

Streiner DL, Normann G.R. Health Measurements Scales.1995. A practical Guide to Their Development and Use. *Oxford University Press, Second Edition*

Sulter G, Steen C, De Keyser J. 1999. Use of the Barthel Index and Modified Rankin Scale in Acute Stroke Trials. *Stroke* 30: 1538-1541

Szende A, Oppe M, Devlin N, eds. 2007. EQ-5D value sets inventory, comparative review and user guide. *Euroqol Group Monographs Volume 2, Springer*

Van Exel N.J.A, Scholte op Reimer W.J.M, Koopmanschap M.A. 2004. Assessment of post-stroke quality of life in cost-effectiveness studies: The usefulness of the Barthel Index and the EuroQol-5D. *Quality of Life Research* 13: 427-433

Tables

Subgroups	*15D*	*EQ-5D*
Female	70 %	82 %
Married	**89 %**	**84 %**
Widow	63 %	81 %
Single	72 %	86 %
Normal weight	77 %	84 %
No handicap	80 %	93 %
No pain	**93 %**	**80 %**
Pain	80 %	85 %
Fatigue	77 %	83 %
Non-anxiety	83 %	88 %
Time 3	78 %	97 %

Table 1. Subgroup differences in completion rates
Higher completion rates for 15D than for EQ-5D indicated in bold

15D		**EQ-5D**	
Attribute	**Missing values**		**Missing values**
1. Mobility	-	***1. Mobility***	4
2. Vision	1	***2. Self Care***	6
3. Hearing	-	***3. Usual Activities***	17
4. Breathing	4	***4. Pain/Discomfort***	21
5. Sleeping	2	***5. Anxiety/Depression***	35
6. Eating	1		
7. Speech	3		
8. Elimination	3		
9. Usual act.	4		
10. Mental funct.	1		
11.Symptoms/pain	8		
12. Depression	3		
13. Distress	2		
14. Vitality	-		
15. Sexual f.	57		

Table 2. Missing values for 15D and EQ-5D by attribute

Stats	***15D***	***EQ-VAS***	***EQ-TTO***
N	333	345	345
Mean	.83	.67	.70
p 25	.74	.52	.62
median	.86	.68	.76
p75	.94	.88	.79
Sd	.14	.23	.30
Min	.29	.03	-.43
Max	1	1	1
Range	0.71	0.97	1.43

Table 3. Summary statistics, utility scores

1. quintile	*Mean*	*p50*	*sd*	*min*	*max*	*Range*
15D	.61	.64	.095	.2858	.7118	.426
EQ-TTO	.21	.22	.262	-.429	.587	1.016
EQ-VAS	.38	.37	.150	.025	.711	.461
2. quintile						
15D	.77	.77	.030	.7156	.813	.0974
EQ-TTO	.66	.69	.037	.587	.691	.104
EQ-VAS	.55	.52	.060	.466	.626	.160
3. quintile						
15D	.85	.86	.024	.8145	.8904	.0759
EQ-TTO	.76	.76	.030	.708	.796	.088
EQ-VAS	.65	.68	.068	.537	.731	.194
4. quintile						
15D	.92	.92	.0184	.8904	.9567	.0663
EQ-TTO	.88	.85	.0616	.812	1	.188
EQ-VAS	.77	.74	.1163	.635	1	.365
5. quintile						
15D	.98	.99	.0166	.9574	1	.0426
EQ-TTO	1	1	0	1	1	0
EQ-VAS	1	1	0	1	1	0

Table 4. Summary statistics, utility score for quintiles
All mean differences are significant at 1% level, except for the differences in mean score between 15D and EQ-TTO, which are significant at the 5% level

		15D		***EQ-TTO***		***EQ-VAS***	
		Mean	Median	Mean	Median	Mean	Median
Good SAH	Yes	0.92	0.93	0.80	0.78	0.86	0.85
	No	0.73	0.75	0.54	0.53	0.54	0.69
	Diff.	**0.19**	**0.18**	**0.26**	**0.25**	**0.32**	**0.16**
Good VAS	Yes	0.93	0.95	0.89	1	0.84	1
	No	0.78	0.79	0.61	0.70	0.59	0.58
	Diff.	**0.15**	**0.16**	**0.28**	**0.30**	**0.25**	**0.42**
Diseases	Yes	0.74	0.76	0.56	0.69	0.55	0.54
	No	0.88	0.90	0.81	0.85	0.75	0.71
	Diff.	**-0.14**	**-0.14**	**-0.25**	**-0.11**	**-0.20**	**-0.17**
Chronic diseases	Yes	0.74	0.76	0.55	0.69	0.54	0.53
	No	0.87	0.90	0.78	0.85	0.73	0.68
	Diff.	**-0.13**	**-0.14**	**-0.23**	**-0.16**	**-0.19**	**-0.15**
Handicap	Yes	0.74	0.75	0.53	0.66	0.53	0.54
	No	0.89	0.90	0.83	0.85	0.75	0.68
	Diff.	**-0.15**	**-0.15**	**-0.30**	**-0.19**	**-0.22**	**-0.14**
Pain	Yes	0.72	0.72	0.48	0.62	0.50	0.52
	No	0.83	0.85	0.78	0.76	0.70	0.68
	Diff.	**-0.11**	**-0.13**	**-0.30**	**-0.14**	**-0.20**	**-0.16**
Disable	Yes	0.68	0.71	0.44	0.59	0.48	0.53
	No	0.86	0.89	0.76	0.80	0.72	0.68
	Diff.	**-0.18**	**-0.18**	**-0.32**	**-0.21**	**-0.24**	**-0.15**
Independent	Yes	0.86	0.88	0.77	0.80	0.72	0.68
	No	0.64	0.65	0.25	0.26	0.38	0.37
	Diff.	**0.22**	**0.23**	**0.52**	**0.54**	**0.34**	**0.31**
Fatigue	Yes	0.76	0.78	0.59	0.69	0.58	0.57
	No	0.91	0.94	0.85	0.85	0.80	0.74
	Diff.	**-0.15**	**-0.16**	**-0.26**	**-0.16**	**-0.22**	**-0.17**
Depressed	Yes	0.58	0.57	0.26	0.25	0.38	0.36
	No	0.85	0.87	0.74	0.76	0.70	0.68
	Diff.	**-0.27**	**-0.30**	**-0.48**	**-0.51**	**-0.32**	**-0.32**
Anxiety	Yes	0.65	0.65	0.41	0.57	0.44	0.45
	No	0.84	0.87	0.72	0.76	0.69	0.68
	Diff.	**-0.19**	**-0.22**	**-0.31**	**-0.19**	**-0.25**	**-0.23**

Table 5. Mean and median utility scores across different groups divided after clinical and health status indicators. All differences are significant at 1% level

	Rank of 15D		*Rank of EQ-TTO*		*Rank of EQ-VAS*	
Male	3.385	(7.265)	10.428	(9.829)	13.758	(9.705)
Age3	-3.841	(12.602)	-0.234	(17.050)	2.982	(16.835)
Age2	7.747	(9.748)	4.504	(13.188)	-6.882	(13.021)
Married	6.942	(6.977)	9.927	(9.439)	18.896*	(9.320)
High education	-7.927	(7.699)	-2.441	(10.416)	-0.920	(10.285)
High income	20.833	(10.847)	13.589	(14.676)	13.388	(14.490)
Employed	12.785	(9.708)	-5.364	(13.135)	**-12.580**	**(12.969)**
Normal weight	-0.887	(6.690)	**-14.172**	**(9.051)**	-3.950	(8.936)
Good sah	49.799***	(8.977)	40.019**	(12.146)	**31.610****	**(11.992)**
Poor sah	-22.488	(13.061)	-6.434	(17.671)	-11.821	(17.447)
Disable	-35.492**	(11.032)	-30.861*	(14.926)	-28.551	(14.737)
Good vas	37.388***	(8.188)	41.118***	(11.078)	41.970***	(10.938)
Poor vas	-1.988	(15.563)	**-44.720***	**(21.056)**	**-24.211**	**(20.790)**
Dependent	-27.243	(23.638)	-24.397	(31.981)	-37.104	(31.577)
Fatigue	-28.170***	(7.969)	-18.144	(10.782)	-18.334	(10.646)
Depressed	-8.026	(17.065)	2.573	(23.088)	-20.121	(22.796)
Anxiety	-45.839***	(13.152)	-33.103	(17.794)	-48.549**	(17.569)
Ischemic	-15.313	(9.232)	-8.237	(12.490)	-4.616	(12.332)
Haemorrhagic	-31.269*	(15.113)	-22.342	(20.446)	-34.947	(20.188)
Time1	-14.905	(8.693)	-16.996	(11.762)	-13.251	(11.613)
Time2	-5.590	(10.337)	0.489	(13.985)	-0.016	(13.808)
_cons	150.649***	(15.827)	146.185***	(21.413)	142.631***	(21.142)

Table 6. Regression of the rank of 15D, EQ-TTO, and EQ-VAS

Standard errors in parentheses

Significant differences between the 15D coefficients and the EQ-5D coefficients in bold

Significant effects on the rank: * p<0.05, ** p<0.01, *** p<0.001

Figures

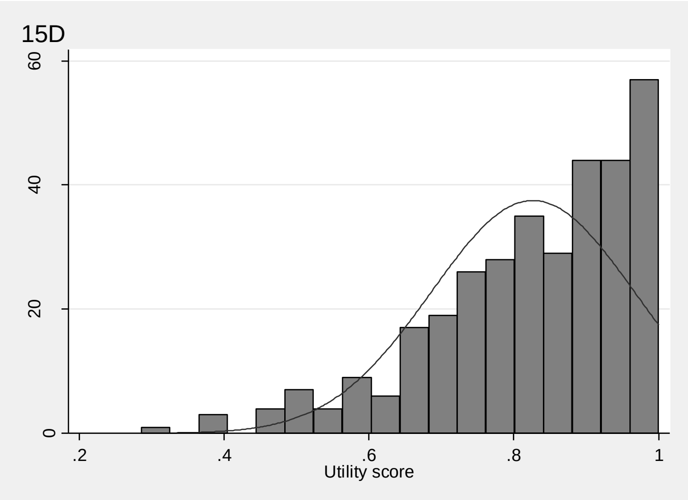

Figure 1. Distribution of 15D scores with the corresponding normal curve

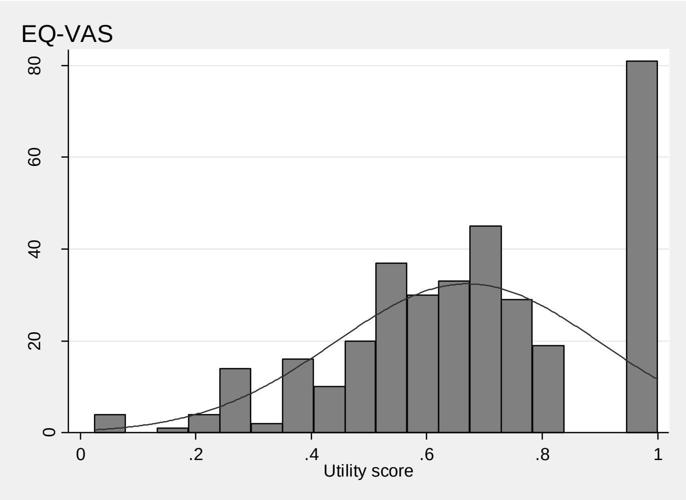

Figure 2. Distribution of EQ-VAS scores with the corresponding normal curve

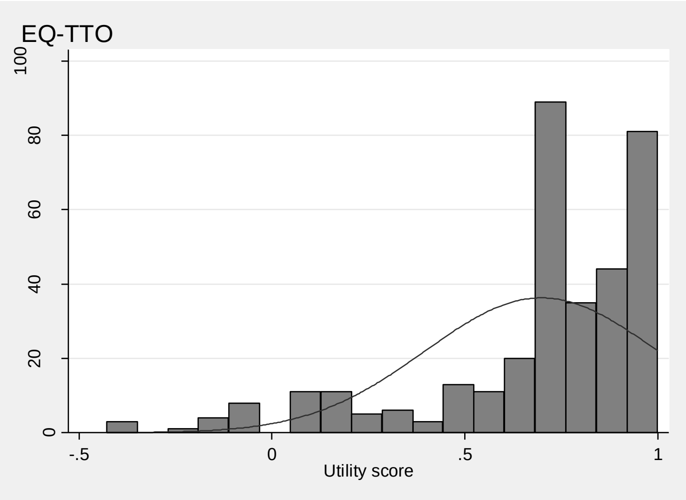

Figure 3. Distribution of EQ-TTO scores with the corresponding normal curve

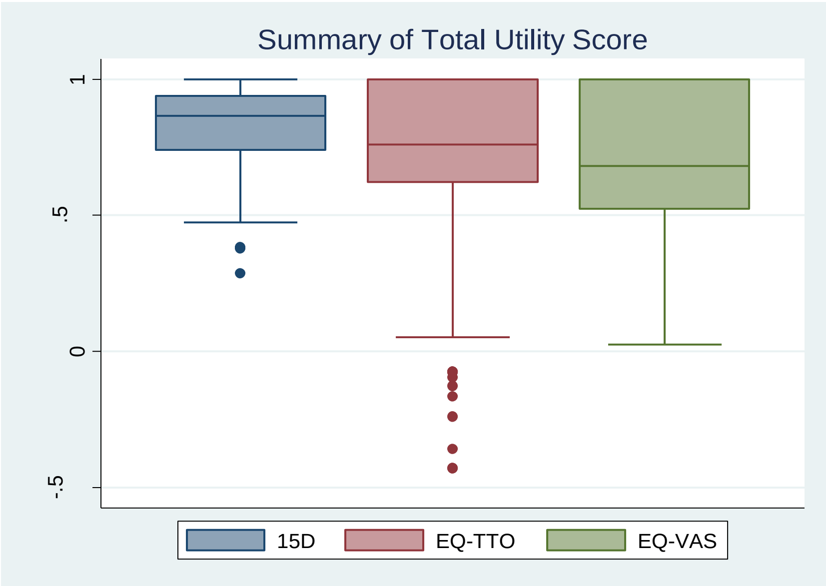

Figure 4. Summary of total utility scores

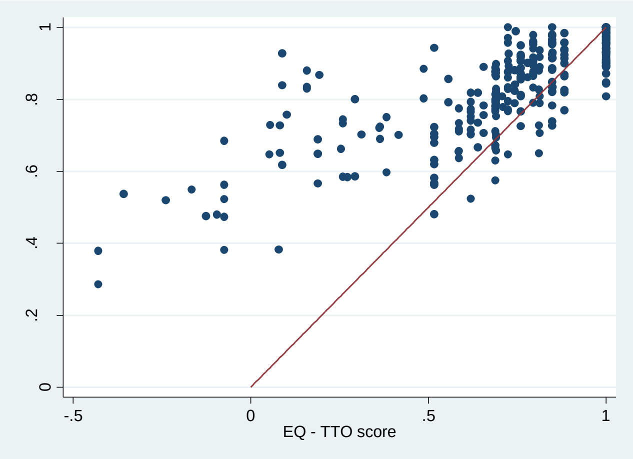

Figure 5. Scatter plot of 15D scores against EQ-TTO scores

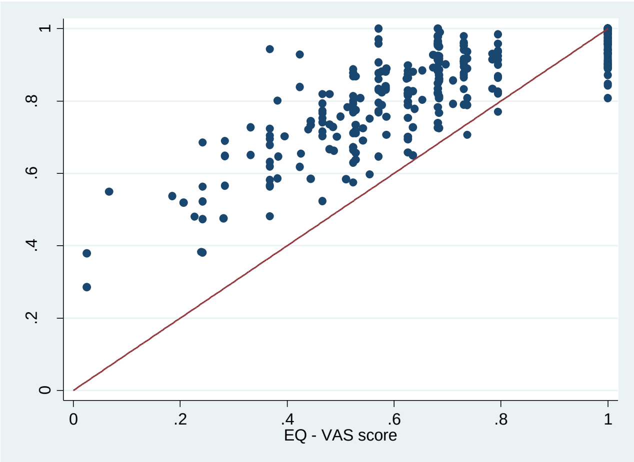

Figure 6. Scatter plot of 15D scores against EQ-VAS scores

Department of Economics
University of Bergen
Fosswinckels gate 14
N-5007 Bergen, Norway
Phone: +47 55 58 92 00
Telefax: +47 55 58 92 10
http://www.svf.uib.no/econ