University of Southern Denmark · Syddansk Universitet

Assessment of the Visual Analogue Scale (VAS) as a Valuation Method for Hypothetical Health States using the EuroQol (EQ-5D)

Kim U. Wittrup-Jensen(1,4), Jørgen Lauridsen(2), Kjeld M. Pedersen(3)

(1): Bayer HealthCare AG, kim.wittrup-jensen@bayerhealthcare.com

(2): Institute of Public Health – Health Economics, E-mail jtl@sam.sdu.dk

(3): Institute of Public Health – Health Economics, University of Southern Denmark, kmp@sam.sdu.dk

(4): The study was done when Kim U. Wittrup-Jensen was a PhD student at University of Southern Denmark

Health EconomicsPapers
2008:6

UNIVERSITY OF SOUTHERN DENMARK

A*bstract*

Background. In the clinical and economic evaluation of health care, the value of benefits gained should be determined from a public perspective. The Visual Analogue Scale (VAS), also called the Rating Scale, is very widely used for measuring health state preferences and it has been argued that the VAS elicits an individual's measurable value function. The link between an individual's measurable value function and his/her *von Neumann-Morgenstern utility function* has been explored, but the relationship between VAS and Time Trade-Off (TTO) scores has not been explored in the same detail.

Objectives. To assess whether it is possible from an empirical point of view to derive a set of valid EQ-5D tariffs using the VAS method. Second, to assess whether a possible threshold exists concerning how many inconsistencies to include/exclude in the modelling of VAS-based EQ-5D tariff. Third, to apply a power function transforming EQ-5D VAS valuations investigating the relation ship with EQ-5D TTO-based valuations. Fourth, to compare both transformed and non-transformed postal-based EQ-5D VAS tariffs with the corresponding set of Danish EQ-5D TTO tariffs. Finally, to test *Parducci's Range-Frequency (RF) model* to assess whether applying such a correction model transforms EQ-5D VAS values into EQ-5D TTO utilities.

Data and methods. In total 1,686 respondents aged 18 years or over completed the EQ-5D VAS valuation exercise. Based on consistent parameters, statistical significance, and level of R^2 values, a best-fitting model was chosen. Inconsistencies were assessed as strong and weak inconsistencies. A possible threshold for the number of inconsistencies to include was assessed by looking at how the significance of the mean values of the health states changed as more inconsistencies were included, and by looking at how inconsistencies included influenced the parameters in the regression analysis. A power function was applied to transform the predicted set of VAS tariffs.

Results. From an empirical point of view a model that fitted the data well, i.e. yielded consistent and significant parameters could be defined. There were no indications of the existence of a threshold for the number of inconsistencies to include in the modelling. Applying a power function to transform EQ-5D VAS tariffs did change the numeric values of the VAS tariffs, but did not result in a higher correlation with the EQ-5D TTO tariffs. Applying the RF model resulted in an almost perfect fit with the corresponding EQ-5D TTO tariffs.

Conclusions. The model presented in this study seems to predict the valuations of the health states for which there are no direct observations and it is possible to estimate a set of EQ-5D VAS-based tariffs. At present it is recommended that all inconsistencies are included in the modelling. Transformation of EQ-5D VAS tariffs does not necessarily make them EQ-5D TTO-based tariffs. The appliance of transformed EQ-5D VAS tariffs should be used as a last resort and even then they should be used with caution.

Introduction

The EuroQol (EQ-5D) questionnaire has been developed jointly by a group of European-based researchers with the intent of constructing a simple, self-administered instrument which provides a composite index score representing preferences for a given health state [Kind 1996]. Originally, it was the intention of the EuroQol Group to keep this new instrument very brief in order to ensure low respondent burden, i.e. when used alongside other measures of health-related quality of life instruments.

Since 1987 the EQ-5D questionnaire has undergone several changes, for example the number of dimensions was shortened from six to five. At present, a standardised questionnaire has evolved for the collection of health state values using the EQ-5D descriptive system. When filling out the EQ-5D classification system, the respondents are initially asked to describe their own health status using the EQ-5D profile and a Visual Analogue Scale (VAS). The EQ-5D consists of five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), with three ordinal levels in each dimension. The endpoints of the VAS are labelled 'best imaginable health state' and 'worst imaginable health state' anchored at 100 and 0, respectively. Respondents are asked to rate their own health state today using the VAS method. The method consists in drawing a line from an anchor box ('your own health today') to that point on the VAS which best represents respondents' own health on that given day [Johnson *et al.* 1998].

In addition, the standard EQ-5D questionnaire contains a section including descriptions of 16 additional health states, separated in two groups of eight, which are presented over two consecutive pages. Each of these two pages has a common format, with four health states printed in boxes on either side of the VAS. The respondents are, as in the VAS exercise where they rated their own health state, asked to draw a line from each health state to the VAS, indicating how good/bad they think the particular health state is. The respondents are asked to imagine *themselves* put in each health state and that the duration would be *one* year. In addition to recording values for the particular health states, respondents are also asked to indicate a value for the state 'death' by marking the VAS accordingly.

Thus far VAS valuations for EQ-5D health states have been carried out in several countries including the Netherlands [Essink-Bot *et al.* 1990; Essink-Bot *et al.* 1993], Norway [Nord 1991], Sweden [Björk & Norinder 1999], Japan [Hisashige *et al.* 1998; Ikeda *et al.* 2000], New Zealand [Devlin *et al.* 2000], Finland [Ohinmaa *et al.* 1995], the UK [Kind 1990; Dolan 1994, Abdalla & Russel 1994], the US [Johnson et al. 1998], Germany [Claes *et al.* 1998], Spain [Badia *et al.* 1999a] and Slovenia [Rupel & Rebolj 2000]. The majority of these surveys are randomised, yet not representative of the general populations in question.

In March 2001 the EQ-net project, funded by the European Union under the Biomed II scheme, ended. The aim of the project had been to develop the EQ-5D in key areas such as valuation, application, communication and translation. The primary effort of the EQ-net project had concentrated on harmonizing and integrating the results of the various European valuation projects carried out by EuroQol Group members over the last 10 years. Values for a subset of EQ-5D health states were found to be similar in a number of northern European countries [The EQ-net Biomed Group 2000]. The research, mostly funded nationally, was however fragmented and the Group recognized the desirability of integrating the results in a way that might produce a standardized set of European valuation data. Thus far the results suggest that values elicited using the VAS method appear to reflect common values throughout Europe.

Based on the direct valuations for EQ-5D health states collected in population surveys, scoring algorithms have been developed in order to weight individual health states for future respondents [Kind *et al.* 1994; Dolan *et al.* 1995]. Such weights can be estimated using a broad variety of econometric techniques. The advantage of such a weighting system is that it can be applied to future respondents without having them again complete the valuation exercise in the EQ-5D. Originally, the intent of the EuroQol Group was to derive weights for the EQ-5D applying VAS valuation techniques, because it is the simplest and the most appropriate for use in postal questionnaires [Brooks *et al.* 1996]. However, since there is theoretical evidence that the VAS technique does not provide cardinal utilities, the Group felt it was essential to explore other scaling techniques, e.g. the Time Trade-Off (TTO) and the Standard Gamble (SG) methods. Nonetheless, VAS values do provide an indication of the ordinal ranking of health states and can be used to compare the relative burdens of the health states as described in instruments like the EQ-5D [Johnson *et al.* 1998].

Objectives

As part of a large study concerning health status measurement in Denmark, a randomised sample of the adult Danish population completed the EuroQol (EQ-5D) questionnaire including the valuation exercise using the VAS method. This dataset offers a unique chance of investigating different features of the VAS as a valuation method, and subsequently, as part of the decision-making context of health care services. The objectives of this study are: 1) to compare EQ-5D VAS valuations elicited within a sample of the Danish population with similar results from other European countries described in the EQ-5D Biomed II project, 2) to study how many inconsistencies individuals display during VAS-based EQ-5D valuations and how these inconsistencies are handled in the process of predicting tariffs of non-directly estimated EQ-5D tariffs, 3) to estimate a national Danish set of EQ-5D tariffs by applying different model specifications and testing different econometric models, 4) a comparison of postal-based and interview-based EQ-5D VAS (direct) valuations, 5) to apply a power function in order to approximate EQ-5D VAS-based tariffs into TTO-based tariffs, between two national set of tariffs, and assess whether there is any justification for such an explicit transformation, and 6) to be the first to test the Parducci & Weddell Range Frequency (RF) correction method in order to see whether, and how, this improves the fit between EQ-5D VAS-based and EQ-5D TTO-based tariffs.

Data

The data

The survey company ACNielsen/AIM carried out a postal survey during the winter of 1999/-2000. The mailed questionnaire contained: 1) the EQ-5D self-classifier, 2) the 15-D questionnaire[1], 3) SF-36[2] (versions I & II), and d) background questions including the standard EQ-5D questions, with additional questions concerning utilisation of health services and health insurance. The procedure followed by ACNielsen/AIM was an initial telephone screening to check whether or not the respondent was willing to receive a mailed questionnaire. Out of 6,350 non-institutiona-lised persons aged 18 or more with, in principle, no upper age limit, who were contacted in this fashion, 1,356 replied that they did not wish to receive a questionnaire. Of the 4,996 people who then received the questionnaire, 1,663 did not return them, resulting in 3,331 completed questionnaires. It is a matter of debate how to calculate the response rate. Two possibilities exist: 3,331 vis-à-vis 4,996: 67 per cent, or 3,331 vis-à-vis 6,350: 53 per cent. In either case, the return rate is acceptable and fairly high compared to EQ-5D postal surveys carried out in the other countries.

Exclusion criteria in the survey

3,331 respondents responded to the VAS exercise in the postal-based survey. 1,686 respondents valued the standard EQ-5D health states in the valuation exercise.[3] All respondents were asked to value 8 health states over two rounds, i.e. 16 health states in total. The health states 11111 and 33333 and 'death' were part of both exercises.

The criteria for excluding respondents from the analysis could be separated into conditions which may affect rescaled data and conditions which may affect the analysis in general [Weijen *et al.* 1999].[4] In the latter category, the basis for excluding data was that data were interpreted as invalid. The starting point for excluding respondents was that the number should be as low as possible. The following criteria were applied:

i) all states given the same value
ii) less than three states valued

---

[1] See Sintonen (2001) for an in-depth presentation of the 15D questionnaire.

[2] See Ware & Sherbourne (1992), McHorney *et al.* (1993), Jenkinson *et al.* (1999) for an overview of the SF-36 versions I & II.

[3] Due to a split-sample design, the study contained seven different variations of the EuroQol (EQ-5D) questionnaire (results not reported here).

[4] For the former, there was no formal basis for excluding data. Excluding data may result in bias i.e. skewness of the rescaled data.

iii) 11111 or death not valued

iv) death valued higher than or equal to 11111.

Table 1 gives details on exclusions.

Table 1. Exclusion criteria and number of respondents excluded.

Criterion	**Number**	**Per cent**
All respondents in the data set	1686	100.0
All states given the same value	17	1.0
Less than three states valued	120	7.1
11111 or death not valued	354	21.0
Death valued higher than or equal to 11111	16	0.9
Number of respondents excluded from the data	507	30.1

With almost 30 per cent of the cases excluded, 1179 respondents remained for analysis in the survey, a response rate of 23.6 per cent (1,179/4,996 x 100) or 18.6 per cent (1,179/6,350 x 100), depending on how one interprets the appropriateness of the baseline.

Exclusion analysis

In order to investigate whether there were any differences in respondent characteristics with respect to those included and those excluded, we compared the two groups based on selected factors, as illustrated in Table 2. Before exclusion 58 per cent of the respondents were female, the same as after exclusion. The difference was not statistically significant. The average age was a little lower after exclusion (but statistically significant at the 5 per cent level). The elderly were the more likely to be excluded: before exclusion the over-65 age group comprised 20 per cent of the data, while after exclusion the proportion was 14 per cent. By looking both at age and how difficult the respondents found performing the exercise there was clearly a tendency that the older the respondent the more difficulties in understanding the exercise. This tendency was statistically significant. Finally, the respondents' average score on their self-reported health states was unchanged after exclusions, but this was not statistically significant.

Table 2. Characteristics of respondents before and after exclusion.

	Before exclusion (n = 1,686)	**After exclusion (n = 1,179)**	***p*-value**
Gender:			
Male	709 (42.1 %)	496 (42.1 %)	
Female	977 (57.9 %)	683 (57.9 %)	0.851[a]
Age:			
Mean (SD)	48.4 years (16.8 years)	45.6 years (15.6 years)	0.03[b]
Max. and Min.	(18 – 91 years)	(18 – 91 years)	
Age > 65 years	342 (20.3 %)	170 (14.4 %)	
Self-reported VAS score:			
Mean (SD)	85.0 (14.7)	86.0 (16.6)	0.253[b]
Median	90.0	90.0	
Max. and Min.	9 – 100	9 – 100	

[a] χ^2-test.
[b] Independent samples *t*-test.

Randomisation

As the survey aimed at estimating preferences for health reflecting those of the Danish population, it was important that the sample was randomised within the general population. To what degree this is the case cannot be measured directly. However, it can be measured indirectly by comparing socio-economic characteristics for the sample with the population as a whole to check for significant differences.

The survey was based on 1,179 respondents of whom around 57 per cent were women, which was over-representative of the percentage of women compared to the general population.[5] In addition, the survey was over-representative for the age group 30-59 years, and under-representative for the age group 60 years and above. See Table 3 for details.

Table 3. Randomisation judged by gender and age distributions. Per cent.

	General population (N = 4,127,847) (January 1^t 2000)	**VAS postal survey (n = 1,179, used for VAS estimation) (winter 1999/2000)**
Gender		
Male	48.9 %	42.1 %
Female	51.1 %	57.1 %
Age		
18 – 29 years	21.1 %	17.5 %
30 – 59 years	54.1 %	61.7 %

---

[5] In total 3,331 respondents returned the questionnaire and were subsequently embedded in the data set. However, due to a split-sample design, including different variations of the EuroQol questionnaire (results not reported here), only 1,179 respondents filled in the standard EuroQol (EQ-5D) questionnaire.

≥ 60 years	24.8 %	20.9 %

I was chose not to weight the sample, even though a difference between females and males was present, which one should be aware of when performing analyses where the gender factor plays a significant part. Performing a weighting of gender would not have had a significant impact on the results and moreover there could be adverse side effects such as an 'over-weighting' of cases, with error in measurement or other inaccuracies. Weighting was a possibility, but for operational reasons it was chosen not to undertake such an exercise.

Methods

Rescaling and transformations of directly valued health states

It was necessary to rescale the VAS values in such a way that the value for death was explicitly set at 0 and the health state 11111 to 1. In order to rescale the mean VAS values for each of the directly valued health states which the respondents valued, the following formula was applied:

$$X_{rescaled} = \frac{X_{VAS-value} - X_{death}}{X_{11111} - X_{death}} \tag{1}$$

This meant that the health state 11111 had a value of 1 and death a value of 0. The rescaled value for a health state was higher than zero if the health state was valued higher than death. For a health state valued as worse than death, the rescaled value would be negative. Rescaled VAS values larger than 1 would appear when the non-rescaled VAS values were valued at higher than 11111. In order for rescaled VAS values to be less than –1, it was necessary that the mean value of the non-rescaled VAS value plus the non-rescaled VAS value for the health state 11111 were lower than death. To continue with the rescaled VAS values lower/larger than –1/+1, these values have all been *truncated* to –1/+1.[6]

In order to take account of the presence of *floor* - and *ceiling- effects*, i.e. a tendency for the answers given by the respondents to cluster at the endpoints of the scale, a logarithmic transformation of the VAS values was performed, where upper and lower boundaries were fixed at 100.5 and -0.5, respectively. According to Weijnen *et al.* (1999) these limits would result in the lowest skewness of the VAS values. The transformations were based on a logarithmic transformation of the non-rescaled VAS values:

$$X_{transformation} = \ln\left(\frac{X_{VAS-value} + 0{,}5}{100{,}5 - X_{VAS-value}}\right) \tag{2}$$

---

[6] A truncation means that values below –1 and above +1 explicitly are given the value of –1 and +1, respective

where the interval was [+5.3; -5.3] and 11111 = +5.3, while the value for death was indeterminable and therefore varied across respondents.

A log-transformation of the rescaled VAS values, where the maximum and minimum limits were different from the transformed non-rescaled VAS values, was also performed. According to Weijnen *et al.* (1999) the upper and lower limits ought to be set at 1.01 and –1.01, respectively, as this would result in the lowest skewness. In this case the skewness would have a value of zero for a normal distribution. A distribution with a positive skewness would have a value larger than 1 and with a negative skewness a value lower than –1:

$$X_{trans-rescaled} = \ln\left(\frac{X_{rescaled} + 1{,}01}{1{,}01 - X_{rescaled}}\right) \tag{3}$$

with the interval [+5.3; -5.3], where 11111 = +5.3 and death = 0.

When performing a rescaling of the VAS values, the result was a distortion across the whole dataset. At the endpoints – for the values of 11111 and death – the skewness of the VAS values would be at its peak. Such results, however, were to be expected, as the purpose of the rescaling was to force the health state 11111 and death to take on the values 1 and 0, respectively. The results of the rescaling and transformation are illustrated in Table 4.

The rescaling was necessary in order to estimate VAS tariffs for the remaining non-directly valued EQ-5D health states. By truncating the health states 11111A and 11111B to the value of 1 and DeadA and DeadB to the value of 0, the estimated tariffs would automatically fall within the interval +1 to –1, where all negative health states were to be interpreted as being worse than death.[7] The last column in Table 4 shows rescaled values from a project estimating a common European EQ-5D value set. VAS values were estimated for a pooled data set from Britain, Germany, Finland, Holland, Sweden and Spain. By comparing the rescaled Danish VAS values from this study with the European set, it can be seen that there was a high degree of agreement. An immediate conclusion could be that Danish preferences for health do not differ significantly from preferences for health in other Western European countries.

---

[7] In the standard EuroQol (EQ-5D) questionnaire, the two health states 11111 & 33333 and the state 'dead' were all valued twice by the respondents.

Table 4. Mean, rescaled and transformed VAS values.

States	Rescaled VAS values	Transformed VAS values	Transformed rescaled VAS values	Rescaled EQ-net[1)]
11111A	1.00	3.97	5.30	1.00
11111B	1.00	3.97	5.30	1.00
11112	0.70	1.15	2.19	0.71
11121	0.76	1.43	2.48	0.76
11122	0.49	0.30	1.30	0.53
11211	0.79	1.67	2.74	0.76
12111	0.61	0.88	1.81	0.65
21111	0.76	1.45	2.48	0.75
21232	0.27	-0.61	0.72	0.28
22233	0.17	-1.21	0.53	0.12
22323	0.09	-1.44	0.24	0.05
32211	0.26	-0.68	0.74	0.21
33321	0.06	-1.73	0.21	0.04
33333A	-0.06	-3.23	-0.05	-0.14
33333B	-0.06	-3.15	-0.09	-0.21
DeadA	0.00	-2.46	0.00	0.00
DeadB	0.00	-2.44	0.00	0.00
Unconscious	-0.05	-3.11	-0.07	-0.09

Note: 1) The last column represents rescaled values obtained from the EQ net dataset . Source: The EQ-net Biomed Analysis Team (1999).

Handling inconsistency

All respondents valued the same EQ-5D health states. As the health states 11111 and 'death' were explicitly set at 1 and 0, respectively, and thus were not a part of the extrapolation of non-directly valued health states, they are not considered in the following analysis. Inconsistency was defined as *logical inconsistency*, i.e. inconsistency was present when a state which was logically worse was ranked higher than a logically better health state, for example if 11113 was valued higher than 11112, the respondent would be expressing a preference for 11113 over 11112. This particular form of inconsistency is normally referred to as *internal inconsistency*. However, it was important to be aware that a valuation could only be logically inconsistent within the same dimension. As respondents had different preferences across different dimensions, *ceteris paribus*, it cannot be said that 11113 was logically worse than e.g. 21111.

Table 5 illustrates how many inconsistencies the respondents showed within the survey – assessed as *weak* and *strong* inconsistencies. To obtain an expression of all possible inconsistencies that the respondents could display, see the illustration in appendix A. To be a weak inconsistency, two or more states were allowed to be given the same value, e.g. the health state 21232 should have a value ≥ state

11211. Strong inconsistencies were defined by the criterion that health states were not allowed to have the same value. In this case 21232 ought to be given a value > the health state 11211 in order to be an inconsistency. As shown in Table 5, more (35 per cent) respondents displayed no logical inconsistencies under the weak criterion than under the strong (18 per cent). This was to be expected from an intuitive point of view. Around 11 per cent and 21 per cent, under the weak and strong criteria respectively, displayed more than five logical inconsistencies.

Table 5. Distribution of weak and strong inconsistencies. (Per cent). n = 1,179.

Weak inconsistencies			Strong inconsistencies		
Pair-wise inconsi-Stencies	**Number of responses**	**Cumulative sum of responses**	**Pair-wise inconsi-stencies**	**Number of Responses**	**Cumulative sum of responses**
0	408 (34.6)	408 (34.6)	0	209 (17.7)	209 (17.7)
1	311 (26.4)	719 (61.0)	1	292 (24.8)	501 (42.5)
2	163 (13.4)	882 (74.4)	2	189 (16.0)	690 (58.5)
3	81 (6.9)	963 (81.3)	3	113 (9.6)	803 (68.1)
4	49 (4.2)	1,012 (85.5)	4	72 (6.1)	875 (74.2)
5	36 (3.1)	1,048 (88.6)	5	53 (4.5)	928 (78.7)
6	37 (3.1)	1,085 (91.7)	6	35 (3.0)	963 (81.7)
7	20 (1.7)	1,105 (93.4)	7	35 (3.0)	998 (84.7)
8	15 (1.4)	1,120 (94.8)	8	19 (1.6)	1,017 (86.3)
9	12 (1.1)	1,132 (95.9)	9	19 (1.6)	1,036 (87.9)
10	15 (1.4)	1,147 (97.3)	10	9 (0.8)	1,045 (88.7)
11-20	27 (2.3)	1,174 (99.6)	11-20	73 (6.2)	1,118 (94.9)
21-28	5 (0.4)	1,179 (100.0)	21-30	22 (1.8)	1,140 (96.7)
			31-39	39 (3.3)	1,179 (100.0)

An important issue concerning inconsistencies within the valuation of health was whether the inconsistencies had any effect on the respondent's valuation of health states. This was investigated by looking at the aggregated level, i.e. how the mean for every health state changed as more inconsistencies were taken into account or left out. In order to get an overview of how the inconsistencies influenced the directly-valued EQ-5D health states, the focus was on differences in mean values and rankings. The results are illustrated in Table 6.

Weak or strong inconsistencies?

In the following calculations, where the focus was on either the inclusion or exclusion of inconsistencies, the concept of *strong inconsistencies* is employed. Since no guidelines or even any pre-studies existed concerning whether one should make a distinction between weak and strong inconsistencies, we chose only to focus on strong inconsistencies. We felt that the valuations in the valuation exercise should be explicit, i.e. two health states could not be valued equally by being given the same score on a 0 to 100 scale. The distinction between weak and strong inconsistencies was included due to the importance of documentation in our method and also to document the substantial changes in the number of inconsistencies, whether or not one chooses to go along with either the weak or strong definition of inconsistencies.

Table 6. Mean health states values and rankings for the directly valued EQ-5D health states (rescaled) according to the number of (strong) inconsistencies in the individual respondent scores.

Mean values of $X_{rescaled}$ **Groups according to the number of inconsistencies in the individual responses**								
Health state	0	1	2	3	4	5	6	+7
11211	0.7931 (1)	0.7888 (1)	0.7865 (1)	0.7868 (1)	0.7806 (1)	0.7804 (1)	0.7809 (1)	0.7945 (1)
21232	0.2578 (7)	0.2339 (8)	0.2277 (8)	0.2238 (8)	0.2170 (8)	0.2126 (8)	0.2128 (8)	0.2651 (7)
11122	0.5183 (6)	0.5034 (6)	0.4955 (6)	0.4917 (6)	0.4872 (6)	0.4830 (6)	0.4821 (6)	0.4881 (6)
11121	0.7611 (3)	0.7524 (3)	0.7521 (3)	0.7511 (3)	0.7500 (3)	0.7475 (3)	0.7486 (3)	0.7582 (3)
22233	0.1780 (10)	0.1359 (9)	0.1250 (9)	0.1186 (9)	0.1115 (9)	0.1071 (9)	0.1051 (9)	0.1716 (9)
33333	-0.1058 (13)	-0.1223 (13)	-0.1253 (13)	-0.1242 (13)	-0.1269 (13)	-0.1320 (13)	-0.1314 (13)	-0.0566 (13)
33321	0.2001 (9)	0.0300 (11)	0.0348 (11)	0.0343 (11)	0.0290 (11)	0.0189 (11)	0.0150 (11)	0.0622 (11)
21111	0.7696 (2)	0.7665 (2)	0.7658 (2)	0.7578 (2)	0.7566 (2)	0.7544 (2)	0.7534 (2)	0.7619 (2)
UNC	-0.0530 (12)	-0.0602 (12)	-0.0685 (12)	-0.0752 (12)	-0.0838 (12)	-0.0935 (12)	-0.0970 (12)	-0.0521 (12)
12111	0.6733 (5)	0.6743 (5)	0.6613 (5)	0.6521 (5)	0.6454 (5)	0.6374 (5)	0.6318 (5)	0.6137 (5)
11112	0.7491 (4)	0.7172 (4)	0.7039 (4)	0.6925 (4)	0.6842 (4)	0.6798 (4)	0.6804 (4)	0.6958 (4)
32211	0.2566 (8)	0.2559 (7)	0.2503 (7)	0.2389 (7)	0.2304 (7)	0.2219 (7)	0.2200 (7)	0.2609 (8)
22323	0.1368 (11)	0.1147 (10)	0.1021 (10)	0.0915 (10)	0.0769 (10)	0.0648 (10)	0.0614 (10)	0.0915 (10)

Note: ‘UNC’ = unconscious.

Deciding on which of the above groups (including the group with zero inconsistencies) constitutes the appropriate sample from which to estimate the final EQ-5D tariff was essentially an arbitrary decision. However, some informal guidelines had been provided by Ohinmaa and Sintonen (1998). Using data from the Finnish population they tested the sensitivity of the mean values of EQ-5D health states from different samples, distinguished according to the number of inconsistencies in the individual responses. Compared to individuals with zero inconsistencies; the mean values of health states were increasingly different the greater the number of inconsistencies admitted. With more than three inconsistencies, almost 80 per cent of the average values were statistically different (at a five percentage level) from the group with no inconsistencies. Consequently Ohinmaa and Sintonen (1998) concluded that their estimates were seriously biased by responses with *more than three* inconsistencies, recommending that such responses be excluded. Furthermore, they argued that *“as the postal method [for eliciting health state values] will produce significantly more inconsistencies than the VAS studies using interviews”*, they recommended that *“there should be some guidelines [for other researchers] to exclude the most inconsistent respondents from the modelled data set.”*[8]

---

[8] That the postal-based VAS valuations of EQ-5D health states should be more biased than similar valuations from interview-based valuations contradicts findings by Dolan & Kind (1996), who compared inconsistencies between self-completed and interview administered VAS questionnaires and concluded that inconsistencies were more common in the latter type, which they suggested to be due to the former being completed by people who were more likely to answer them “logically” – that is, those who were less able to answer logically tended not to answer at all.

Looking at the results in Table 6, no such sensitivity in the mean values of EQ-5D health states was evident in our data. On the contrary, we found the opposite results, where respondents' mean valuations are very robust to the number of inconsistencies admitted. Using a *t*-test, almost all of our mean values of $X_{rescaled}$ were statistically the same across groups. Only one health state (33333) differed significantly, according to mean value, when we compared groups of 0 (zero) and +7 inconsistencies. Even though the numeric (mean) values changed according to how many inconsistencies were admitted, the rankings across groups were very stable. Health states 11211 and 33333 were ranked as the best and worst across all groups, respectively. The few health states where rank ordering changed between groups were the relatively severe health states.

The absence of an explicit *threshold* for how many inconsistencies, or none, to include in the estimation of EQ-5D tariffs, compared to the findings of Ohinmaa & Sintonen (1998), may be explained by our use of rescaled data instead of raw scores, as applied in the Finnish study. Devlin *et al.* (2000) first noted this and explained it as follows *"...given that rescaled data are the appropriate data for estimating a tariff, and given that t-values are not invariant to the data transformations introduced by rescaling (or not), then Ohinmaa and Sintonen's t-tests and the inferences they derive are invalid."*

As reported in this context, the difference between *t*-tests based on rescaled health state valuations and *t*-tests from raw data (as reported by Ohinmaa and Sintonen, 1998) could be appreciated by considering the usual *t*-statistic formula:

$$t = \frac{(\bar{X}_0 - \bar{X}_a)}{\sqrt{s_0^2 / n_0 + s_a^2 / n}}, \tag{4}$$

where $\mathbf{X}_0$ and $\mathbf{X}_a$ are the mean health state values for the group with no inconsistencies and its comparator group respectively, s_0 and s_a are their variances, and n_0 and n_a their sample sizes. As the two numeric values $\mathbf{X}_0$ and $\mathbf{X}_a$, as well as s^2_0 and s^2_a, depend on whether the EQ-5D health states have been rescaled or not, they each produce different *t*-statistics and hence different inferences.[9] In order to produce a valid tariff rescaling is necessary. We believe that our findings are correct and consequently no formal or explicit threshold exists concerning how many inconsistencies to include in the modelling.

---

[9] See next section for a detailed overview of the rescaling of the EQ-5D health states.

Estimating non-direct EQ-5D health states

Our approach was largely based on the same terminology used in Dolan (1997). However, rather than applying a generalised least-squares regression technique based on *ad hoc* assumptions, we explicitly tested both the random effects model (REM), the fixed effects model (FEM), and the Ordinary Least-Squares model (OLS). We also tested whether one model was superior to the others when applying explicit assumptions.

In our model each of the thirteen EQ-5D health states, including the state 'unconscious', for which we had data, could be represented by dummy variables designed to capture possible independent and interaction effects across the five dimensions of the EQ-5D health state classification system [Devlin *et al.* 2000]. Like Dolan (1997), we did not have any standard theoretical model that we could apply in order to direct how these dummies should be combined in an equation suitable for modelling directly valued EQ-5D health states. Therefore we decided to include all or none of the dummies from a particular set in the equation with which we experimented, rather than use different elements from different sets.

We applied a regression model where we explicitly assumed the functional form to be *additive*. The dependent variable was defined as $1 - X_{rescaled}$.[10] The reason for using this expression and not simply $X_{rescaled}$, was straightforward: Since the EQ-5D system values health states other than full health as negative deviations from a value of unity (where 11111 is defined as the value of 1), $X_{rescaled}$ can be represented as 1 minus the appropriate (linear) combination of dummy variables and their coefficients. In addition to applying an intercept, the independent variables were defined based on the EQ-5D ordinal structure.

By carefully checking the existing literature, we selected eleven different model specifications to be tested [Dolan 1997; Devlin *et al.* 2000; Rupel & Rebolj 2000]. An overview of the (eleven) different model specifications is illustrated in Figure 1, and a description of the independent variables is given in Figure 2. The basic form of the models can be expressed as:

$$1 - X_{rescaled} = \text{constant} + \mathbf{D}*\boldsymbol{\beta} + \text{error term} \qquad (5)$$

where **D** is the row vector of dummies and **β** represents the column vector of coefficients.

In the estimation of the EQ-5D tariffs, we decided to include the data containing *all* inconsistencies. However, whether this is the correct thing to do is not clear and will be discussed in the discussion section. By varying the number of inconsistencies included in the estimated model we looked at how the preferred model behaved.

---

[10] Dolan (1997) defined the dependent variable as $1 - S$, where S was the value given to a particular health state.

Figure 1. Overview of the eleven models used to elicit predicted estimates for non-directly valued EQ-5D states.

Name = f(x)
VAS1 = f(MO, SC, UA, PD, AD)
VAS2 = f(MO, SC, UA, PD, AD, N3)
VAS3 = f(MO, SC, UA, PD, AD, M2, S2, U2, P2, A2)
VAS4 = f(MO, SC, UA, PD, AD, M2, S2, U2, P2, A2, N3)
VAS5 = f(MO, SC, UA, PD, AD, M2, S2, U2, P2, A2, MOSC, MOUA, MOPD, MOAD, SCUA, SCPD, SCAD, UAPD, UAAD, PDAD)
VAS6 = f(MO, SC, UA, PD, AD, MOSC, MOUA, MOPD, MOAD, SCUA, SCPD, SCAD, UAPD, UAAD, PDAD)
VAS7 = f(F11, F21, F31, F41, F13, F23, F33, F43, F53)
VAS8 = f(MO, SC, UA, PD, AD, F11, F21, F31, F41, F13, F23, F33, F43, F53)
VAS9 = f(MO, SC, UA, PD, AD, F13, F23, F33, F43, F53)
VAS10 = f(MO, SC, UA, PD, AD, M2, S2, U2, P2, A2, N2, N3)
VAS11 = f(MO, SC, UA, PD, AD, N2, N3)

Figure 2. Definition of variables used in the regression analysis.

Variable	**Definition**
A	Intercept: indicator for any movement away from 11111
MO	1 if mobility is level 2; 2 if mobility is level 3; 0 otherwise
SC	1 if self-care is level 2; 2 if self-care is level 3; 0 otherwise
UA	1 if usual activities is level 2; 2 if usual activities is level 3; 0 otherwise
PD	1 if pain/discomfort is level 2; 2 if pain/discomfort is level 3; 0 otherwise
AD	1 if anxiety/depression is level 2; 2 if anxiety/depression is level 3; 0 otherwise
M2	1 if mobility is level 3; 0 otherwise
S2	1 if self-care is level 3; 0 otherwise
U2	1 if usual activities is level 3; 0 otherwise
P2	1 if pain/discomfort is level 3; 0 otherwise
A2	1 if anxiety/depression is level 3; 0 otherwise
MOSC	The product of MO and SC
MOSC	The product of MO and UA
MOPD	The product of MO and PD
MOAD	The product of MO and AD
SCUA	The product of SC and UA
SCPD	The product of SC and UA
SCAD	The product of SC and UA
UAPD	The product of UA and PD
UAAD	The product of UA and AD
PDAD	The product of PD and AD
F11	1 if the health state contains 1 dimension at level 1; otherwise 0
F21	1 if the health state contains 2 dimensions at level 1; otherwise 0
F31	1 if the health state contains 3 dimensions at level 1; otherwise 0

F41	1 if the health state contains 4 dimensions at level 1; otherwise 0
F13	1 if the health state contains 1 dimension at level 3; otherwise 0
F23	1 if the health state contains 2 dimensions at level 3; otherwise 0
F33	1 if the health state contains 3 dimensions at level 3; otherwise 0
F43	1 if the health state contains 4 dimensions at level 3; otherwise 0
F53	1 if the health state contains 5 dimensions at level 3; otherwise 0
N2	1 if any dimension is at level 2; otherwise 0
N3	1 if any dimension is at level 3; otherwise 0

Figure 2 gives an overview of the dummy variables and their definitions. Furthermore, two variables were specified: N2 and N3. These two variables should capture whether one of the levels within the five dimensions was at level 2 or 3, respectively.

The regression analysis

In total three sets of dummy variables were created:

1. Two dummy variables for every dimension, one that represented movement between the three levels, and one that represented movement from level 2 to level 3 (this allowed the effect of moving from level 1 to level 2 to be different from the effect of moving from level 2 to level 3).
2. Dummy variables allowing (first-order) interaction between the five dimensions.
3. Dummy variables that captured how many times a health state contains one or more dimensions which was on level 1 or level 3.

The analysis was conducted at the individual level, where each individual valued a discrete number of health states. In such a situation it was important to bear in mind that there was most likely to be a certain connection given relationship within (in-between) the valuations of the given health states. In other words, if a respondent valued a health state at below the mean compared to the valuations given by the rest of the sample, there was a tendency that the respondent would also value the remaining health states below the sample mean. This implied that the variance of the error term was partly determined by those respondents, who valued the health states and hence would remain constant. However, this violated one of the assumptions for using the *Ordinary Least-Squares* (OLS) *model*, so this method could not be applied here. Instead, we chose to apply a *Random Effects (RE) model*, which was specified as follows:

$$y_{it} = \alpha_1 + \beta_2 x_{2,it} + \beta_3 x_{3,it} + \ldots + \beta_K x_{K,it} + u_i + \varepsilon_{it} \quad (6)$$

where u_i was an individual specific random term representing the extent to which the intercept of the *i*'th respondent differed from the overall intercept, α_1. Rewriting (6) as

$$y_{it} = (\alpha_1 + u_{1,i}) + \beta_2 x_{2,it} + \beta_3 x_{3,it} + \ldots + \beta_K x_{K,it} + \varepsilon_{it} \quad (7)$$

or:

$$y_{it} = \beta_1 + \beta_2 x_{2,it} + \beta_3 x_{3,it} + \ldots + \beta_K x_{K,it} + \varepsilon_{it} \tag{8}$$

the RE model could equivalently be viewed as a *random intercept model*. The randomisation of the intercept β_1 may equally well be applied to any of the parameters by defining:

$$\beta_K = \alpha_K + u_{k,i} \tag{9}$$

where $u_{k,i}$ was the individual specific error term representing the extent to which the coefficient of x_k of the *i*'th respondent differed from the average fixed coefficient α_k. As shown, this *Random Coefficient* (RC) model was a general formulation covering the RE model as a special case in which only the intercept was randomised.

The RE model could be tested versus the RC model using a split-sample test. If the RE and RC coefficients were not significantly different, then the RE model was more efficient than the RC model, whereas significant differences indicated efficiency problems in the RE model. The split-sample test was performed by randomly splitting the sample of respondents into two equally sized sub-samples, estimating the RC model on sub-sample 1 and the RE model on sub-sample 2. Next, we calculated the test size (or parameter) $\mathbf{d}'\mathbf{V}^{-1}\mathbf{d}$ where **d** measured the distance between the RE and RC parameter vectors and V the sum of their covariance matrix. Finally, the test size was compared to a χ^2 distribution with *df* (degrees-of-freedom) equal to K.

Further, the efficiency of the RE model versus the OLS model was considered. If the RE model did not represent any improvement over the OLS model, then the RE model was less efficient than the OLS model. On the other hand, if the RE model represented an improvement, then the RE model was more efficient. Thus, the efficiency of the OLS model needed to be tested. Three independent tests were assessed. The first was a *Lagrange Multiplier* (LM) test, based on the OLS estimates only. Due to the conservative nature of the LM test (i.e. a relatively large tendency to reject the OLS model in finite sized samples), a *Likelihood Ratio* (LR) test was assessed, based on the difference between the OLS and the RE log likelihood values, and also a *Wald test,* based on the RE results only (calculated as the squared t value for significance of the estimated variance of u_i) was assessed.

An important consistency issue, which has so far frequently been ignored in empirical studies, is the condition for consistency of the RE model, i.e. that the individual effects are uncorrelated with the explanatory variables. For the present case this implied that the individual deviations from the average valuations of health states did not vary over the range of health states. Thus, if a respondent overestimated each of the health states presented by a given amount, then this amount should not change if the respondent was presented with any other health state. Clearly, this property was not automatically guaranteed. In the case of violation of the independence assumption, the *Fixed Effect (FE) model* provided consistent estimates. In the FE model specification, the random effects were replaced with a fixed effect for each respondent. Thus, the FE model was similar to an OLS model with a dummy variable for each respondent. On the other hand, if the RE model was consistent, one would expect the FE model to be strongly inefficient because many of the intercept parameters - one per respondent - may be without significant difference. In order to test the consistency of the RE model, the *Hausman-test* for equality of the RE and FE parameters was assessed.

The models were tested for misspecification using a *Ramsey RESET-test,* and a test for general heteroscedasticity was performed. Both tests are two-stage tests with a common first stage consisting of estimating the model in question. For the RESET-test, the second step consisted of re-estimating the model with the squared predicted values from the first stage as an additional explanatory variable, and using the significance of the *F*-value of this variable as indicative of functional misspecification. For the heteroscedasticity test, the second stage consisted of regressing the predicted values from the first stage on the squared residuals from the first stage, and using the significance of the *F*-value for this squared residual as indicative of heteroscedasticity.

In order to measure the extent of multicollinearity, the *condition number* (CN) for each model was estimated and we used the standard rule-of-thumb that a CN in excess of 20-30 indicates a problem.

Finally, to make a choice between different ways of representing a relationship between valuations and health states, it was required that the predicted valuations should be logically consistent. For a model with a specific set of explanatory variables, we further - if it was possible while maintaining the requested logical consistency - required efficiency and consistency of the model. This choice was based on the test statistics for the RE, FE, OLS, and RC models versus each other, as described above. Where there were conflicts between the logical consistency request and the test statistics, it was decided to choose the logically consistent model for further consideration, while reporting and discussing the statistically optimal model in order to shed light on possible problems in the chosen model.

Comparison of postal-based and interview-based VAS results

In order to assess the postal-based VAS results alongside the interview-based VAS results both directly-valued and predicted health state tariffs were applied.[11] The results found in the EQ-net Biomed II report, indicated that (pre-)ranked studies have higher VAS valuations in the health states that deviate by one level from full health. This ranking effect may be explained by the fact that, once the health states are visually ordered, the respondent will have a tendency to relate the scores for health states to those of ‘neighbouring’ health states [EQ-net Biomed Group 1999]. This may lead to a wider spread of (VAS) values over the whole range of possible scores. Since health states within the interview-based VAS study were rank ordered before the respondents were asked to value them, one would expect the mean of the VAS valuations in the postal-based study to be lower than the mean VAS valuations within the interview-based study. The correlation coefficients between VAS-based EQ-5D tariffs estimated within the two studies were expected to be high.

Applying a power function

In order to transform the VAS valuations into TTO valuations, a power function explaining the exponential relationship between VAS and TTO valuations at the individual level was applied. The power function has been described elsewhere and estimated as the relationship between interview-based VAS and TTO valuations of EQ-5D health states undertaken by the same group of individuals, but on an aggregated level, i.e. EQ-5D health state means.[12] The form of the power function was $f(u) = \text{VAS}_{\text{transformed}} = 1 - (1 - \text{VAS})^a$ where a was estimated to be 0.66 (adjusted R^2 was 0.83). Since the objective in this study was to estimate a set of VAS-based EQ-5D tariffs, the power function was applied directly to the estimated set of EQ-5D tariffs. However, whether this was the correct thing to do is unclear. This issue will be discussed in more detail in the discussion section.

Parducci & Weddel range-frequency (RF) method

As suggested by Robinson *et al.* (2001), the VAS values assigned to a particular health state are not independent of the remaining health states, a finding confirmed by Bleichrodt and Johannesson (1997). This implies that the VAS valuations had to be adjusted before entering a power function. The Parducci & Weddell range-frequency (R-F) model suggested that respondents would tend to spread out any given set of stimuli along the length of the scale [Parducci & Weddell 1986]. The essential idea of the R-F theory was that assessments of any particular stimuli represented a compromise between two principles: (1) the respective categories were assigned to successive equal sub-ranges of contextual stimuli (the range principle), and (2) the same number of contextual stimuli were assigned to the individual available categories (the frequency principle). The formula used to ‘correct’ the VAS valuations looked as follows:

$$Q_i^{VAS} = w \times \left(\frac{Q_i^C - Q_{\min}}{Q_{\max} - Q_{\min}} \right) + (1 - w) \times \left(1 - \frac{rankQ_i - 1}{N - 1} \right) \tag{10}$$

where Q_i^{VAS} denoted the correct VAS value, Q_i^{C} denoted the observed VAS value (not rescaled), $Q_{\max}$ is 1 (perfect health – top level), $Q_{\min}$ is 0 (worst level – bottom level), $\text{rank}Q_i$ denoted the ordinal ranking of health state Q_i within the set of stimuli, and N denoted the number of injury descriptions in the

---

[11] Estimated interview VAS-based EQ-5D tariffs were elicited *ad hoc* for this comparison and not reported elsewhere. Data were collected as part of the TTO study reported in Wittrup-Jensen *et al.* (2001).

[12] See Chapter One in this dissertation.

set (in this context 14). The parameter *w* represented the relative weighting between the two components, where the first component on the right hand side represented the effect of the value of the stimuli and the second component represented the effect of the rank of the stimulus. Schwartz (2001) suggested that *w* should be set at 0.5, implying equal weighting to the value and rank components.

Results

The directly valued VAS health states

As illustrated in Table 7, scores reported by respondents (in the postal-based survey) show signs of *convergent validity,* i.e. logically better health states are given higher values than logically worse health states. On average the health states 33333A and 33333B were valued lower than DeadA and DeadB. States 11111A and 11111B, had mean values of 96.6 and 94.4, respectively, close to 100. As seen from the minimum and maximum values, not all respondents valued the health states 11111A and 11111B as being the best health states. There were respondents who valued these health states as low as 10. The remaining states were all to be found within the interval 0–100, even DeadA and DeadB.

Table 7. Summary statistics for VAS data for the 18 health states with empirical scores. (n=1,179)

Health state	Mean	SD	Minimum	Maximum	Median	95 % CI	
11211	79.2	15.2	0	100	80	78.4	80.1
11111A	94.9	9.4	10	100	100	94.3	95.5
21232	37.0	21.4	0	100	32	35.8	38.3
11122	55.5	19.6	0	100	55	54.4	56.6
11121	76.4	15.9	0	100	80	75.5	77.3
22233	29.2	22.8	0	100	25	27.9	30.5
33333A	11.9	22.1	0	100	3	10.6	13.1
33321	21.7	21.0	0	100	17	20.5	22.9
21111	76.6	15.7	0	100	80	75.7	77.5
11111B	94.8	10.0	10	100	100	94.2	95.4
12111	66.1	20.4	0	100	70	65.0	67.3
11112	71.5	17.9	0	100	75	70.5	72.5
32211	37.1	22.3	0	100	35	35.8	38.4
33333B	12.0	21.6	0	100	5	10.7	13.2
22323	24.8	20.2	0	100	20	23.7	26.0
Unconscious	13.4	23.7	0	100	3	12.1	14.8
DeadA	14.7	15.1	0	75	10	13.9	15.6
DeadB	15.3	16.1	0	93	10	14.4	16.2

Predicting values for non-directly valued EQ-5D health states and the influence of inconsistencies

The RC model (R^2=0.03), which results in inconsistent parameters as some parameters have the wrong sign (positive), was estimated. The split-sample test showed a strong significant difference between the

RC and RE parameters, indicating efficiency problems in the RE model. However, since parameters in the RC model were logically inconsistent, we chose the RE model.

The LM test leads to a rejection of the OLS model, indicating that the OLS model is inefficient. Furthermore, the LR test indicated that the OLS model should be rejected compared to the RE model. Finally, the Wald test showed that the RE model was efficient and implicitly rejected the OLS model. The strength between the three tests is normally that Wald<LR<LM, implying that the LM test rejects the OLS model 'more easily' than both the LR and the Wald tests. The Hausman test showed that both Prob(H_0: RE)≈0 and Prob(H_0: OLS)≈0, which meant that both the RE model and the OLS model were rejected. The implications are that the FE model performed better, with regard to significance, than both the RE and OLS models. However, the FE model produced inconsistent parameters, as some parameters were negative. Therefore the RE model was chosen.

The RE model failed the Ramsey RESET test ($p<0.0001$), indicating that the model suffered from misspecification. In addition, the model also suffered from heteroscedasticity ($p<0.0001$). The RE model did not, however, suffer from multicollinearity, since the CN was estimated at 6.86.

From a statistical point of view, the FE model was superior to the RE model. However, some of the parameters in the FE model were negative. Hence in estimating Danish EQ-5D tariffs the RE model has been applied. Both misspecification, i.e. incorrect functional form and (omitted) variables, and heteroscedasticity in the RE model occurred, not surprisingly, as the power of the RESET tests increases, as the sample size increases. The presence of heteroscedasticity is likely to be one of the causes of inefficient parameter estimates.

First, the RE model was applied to all eleven models. As already noted in a previous section, choosing between the eleven estimated models hinges on the logical consistency of the parameters and on the significance of the parameters. When including all data, i.e. including all respondents no matter how many (strong) inconsistencies they displayed in the direct valuation of EQ-5D health states, it was found that VAS3, VAS4 and VAS5 had both inconsistent and insignificant parameters.[13] VAS6 and VAS10 had infinite log-likelihood estimates, for which the parameters could not be estimated. VAS8 and VAS9 had inconsistent parameters. VAS2 had consistent parameters, but the parameter UA ('usual activities') was insignificant at the 10 % level. VAS11 had consistent and significant parameters (p <

---

[13] For an explanation of the different models estimated (VAS1 – VAS11) please refer to Figure 2.

0.05). VAS1 showed consistent and significant parameters ($p < 0.01$). The R^2 value of VAS1 was 0.7210.

Nevertheless, estimating the models with all inconsistencies included could bias the parameters and the level of significance. Consequently all inconsistencies were excluded and all eleven models re-estimated. When all inconsistencies were excluded and only respondents with zero inconsistencies were accepted, both the VAS1 and VAS2 yielded consistent and significant parameters at the 1 per cent level. The R^2 values were 0.8468 and 0.8564, respectively. At the 5 per cent level, the VAS11 model also yielded consistent and significant parameters.

The results from VAS1 and VAS2 are shown in Table 8, and the results from the remaining nine models are shown in Appendix B. In order to locate a possible *threshold* within the VAS2 model, that is, at how many inconsistencies the parameters become insignificant, we tried to include inconsistencies in the model systematically, starting with zero inconsistencies. The threshold was located at *four* inconsistencies included, where all parameters were significant at a five-percentage level.

Table 8. Estimated parameters for VAS1 and VAS2 according to the number of inconsistencies in the individual respondent scores. (*p*-value).

	VAS1 (n = 1,170)		**VAS2 (n = 209)**	
Variable	All inconsistencies included	Zero inconsistencies included	All inconsistencies included	Zero inconsistencies included
a	0.2251 (<0.0001)	0.1910 (<0.0001)	0.2518 (<0.0001)	0.2112 (0.0001)
MO	-0.1255 (<0.0001)	-0.1458 (<0.0001)	-0.0239 (<0.0001)	-0.0682 (0.0001)
SC	-0.1117 (<0.0001)	-0.1054 (<0.0001)	-0.1324 (<0.0001)	-0.1212 (0.0001)
UA	-0.0639 (0.0001)	-0.0745 (<0.0001)	-0.0034 (<0.5149)*	-0.0285 (0.0002)
PD	-0.0777 (0.0001)	-0.1039 (<0.0001)	-0.0393 (<0.0001)	-0.0748 (0.0001)
AD	-0.0912 (0.0001)	-0.0676 (<0.0001)	-0.0689 (<0.0001)	-0.0505 (0.0001)
N3	-	-	-0.2865 (<0.0001)	-0.2185 (0.0001)
R^2	0.7210	0.8468	0.7325	0.8564

*($p > 0.05$).

Looking at the parameters illustrated in Table 8, it is difficult to reject either of the two models, as they are both acceptable for estimating EQ-5D. If all inconsistencies are included, VAS1 should be preferred and if zero inconsistencies are included, VAS2 should be preferred. Thus two sets of tariffs were calculated using the equations reproduced from Table 8. The dependent variable was $X_{rescaled}$. Table 9 illustrates the arithmetic by which these tariffs values were calculated.

Equation (I) from VAS1 (all inconsistencies included, n = 1,170):

$X_{rescaled} = 1 - 0.2251 - 0.1255*MO - 0.1117*SC - 0.0639*UA - 0.0777*PD - 0.0912*AD$

Equation (II) from VAS2 (zero inconsistencies included, n = 209):

$$X_{rescaled} = 1 - 0.2112 - 0.0682*MO - 0.1212*SC - 0.0285*UA - 0.0748*PD - 0.0505*AD - 0.2185*N3$$

The relative magnitude of the dummy coefficients expresses the relative importance to respondents (on average) of the five EQ-5D dimensions. In the first equation, when all respondents were included, mobility (*MO*) came ahead of self-care (*SC*), followed by anxiety/depression (*AD*), then pain/discomfort (PD), and finally usual activities (UA). This ordering was not the same as the ordering when zero in-consistencies were allowed as the ranking here was quite different, starting with self-care (*SC*), ahead of pain/discomfort (*PD*), followed by mobility (*MO*), then anxiety/depression (*AD*), and finally usual activities (*UA*). The differences were due to the number of respondents (inconsistencies) included. By comparison, Devlin *et al.* (2000) found the ordering: anxiety/depression (*AD*), pain/discomfort (*PD*), mobility (*MO*), self-care (*SC*), and usual activities (*UA*).

Table 9. Calculation of Danish VAS1 tariff for EQ-5D health state 11223.

	Full health (11111) =	**1.000**
minus	Constant term =	0.225
minus	Mobility (MO): level 1 =	0 * 0.126
minus	Self-care (SC): level 1 =	0 * 0.112
minus	Usual activities (UA): level 2 =	1 * 0.064
minus	Pain/discomfort (PD): level 2 =	1 * 0.078
minus	Anxiety/depression (AD): level 3 =	2 * 0.091
Equals	Tariff value for health state 11223 =	0.451

We used VAS1 and VAS2 to estimate two sets of tariffs, representing the best-fitting model when all inconsistencies were included and the best-fitting model when all inconsistencies were excluded, respectively. The two sets of tariffs are illustrated in appendix C and D. In order to make the comparison with some kind of *gold standard* the set of Danish TTO-based EQ-5D tariffs was used.[14] In the VAS1, around 5 percent of the tariffs were rated as being worse than dead. The corresponding number in the TTO was around 20 per cent. A comparison of the differences between the VAS1 and TTO tariffs shows that 55 per cent of the VAS1 tariff values were predicted at higher than or equal to the

---

[14] Please refer to Wittrup-Jensen *et al.* (2001).

corresponding TTO tariff values. Looking at the VAS2 EQ-5D tariffs, around 10 percent of the tariffs were rated as being worse than dead. 25 per cent of the VAS2 tariffs were predicted at higher than or equal to the corresponding TTO tariff. The correlations between the two sets of VAS tariffs and the TTO tariffs are illustrated in Table 10. As expected, there was a high and significant correlation between the tariffs estimated by the VAS method and the tariffs estimated by the TTO method, both with respect to the numeric value and the ranking.

Table 10. Correlation coefficients between Danish VAS1, VAS2 and TTO tariffs.

	Pearson correlation coefficient	**Spearman correlation coefficient**
VAS1 versus TTO	0.884*	0.866*
VAS2 versus TTO	0.879*	0.882*

*($p < 0.01$).

Predicted versus observed valuations

In total, direct valuations of thirteen EQ-5D tariffs were obtained. A comparison of the means of the observed (direct) valuations and the estimated (predicted) valuations provides a direct check of how well the tariffs approximated respondents' values. This is illustrated in Table 11. As summarised by the mean absolute differences, equation (I) differ from the mean values to a greater extent than equation (II). Although the majority of the differences are small, some are large e.g. –240 per cent for health state 33333. As indicated by the reported Pearson correlation coefficient, both equation I and II tariff values were closely (linearly) correlated with the corresponding mean values. Also, as indicated by the Spearman correlation coefficient, the rankings for both equation I and II tariff values were highly correlated with the corresponding mean values. For almost all health states the mean values were higher than the corresponding predicted values. The exceptions were, in equation I, health states 11122 and 12111, and in equation II, health states 11122, 32211, and 33321. Making the directly valued health states an explicit *gold standard*; this may indicate that the VAS1 and VAS2 models consecutively underestimated the (true) cardinal value of the EQ-5D health states.

Table 11. Comparison of directly valued EQ-5D scores and scores predicted by VAS1 and VAS2. Mean values.

VAS1 (all inconsistencies included)				VAS2 (zero inconsistencies included)		
Health state	Estimated value (eq. I)	Ob-served value	Difference (% of mean value)	Estimated value (eq. II)	Ob-served value	Difference (% of mean value)
21111	0.649	0.766	-0.117 (-15%)	0.599	0.770	-0.171 (-22%)
12111	0.663	0.661	0.002 (0%)	0.663	0.673	-0.010 (-2%)
11211	0.711	0.792	-0.081 (-10%)	0.758	0.793	-0.035 (-4%)
11121	0.697	0.764	-0.067 (-9%)	0.705	0.761	-0.056 (-7%)
11112	0.684	0.715	-0.031 (-4%)	0.716	0.750	-0.034 (-5%)
11122	0.606	0.555	0.051 (9%)	0.641	0.518	0.123 (24%)
21232	0.338	0.370	-0.032 (-9%)	0.129	0.258	-0.129 (-50%)
32211	0.347	0.371	-0.024 (-7%)	0.279	0.257	0.022 (9%)
22233	0.135	0.292	-0.157 (-54%)	0.065	0.178	-0.113 (-64%)
22323	0.149	0.248	-0.099 (-40%)	0.118	0.137	-0.019 (-14%)
33321	0.093	0.217	-0.124 (-57%)	0.065	0.028	0.037 (132%)
33333	-0.167	0.119	-0.286 (-240%)	-0.138	-0.106	-0.032 (30%)
Uncon.	-0.050	0.134	-0.084 (-137%)	-0.053	-0.053	0.000 (0%)
Mean absolute difference			-0.081			-0.032
Pearson's correlation coefficient			0.973*			0.975*
Spearman correlation coefficient			0.962*			0.924*

*(p < 0.01).

Postal-based versus interview-based VAS (direct) valuations

Table 12 illustrates the mean values of the directly assessed health states in the postal and interview-based VAS exercises. In addition, the rescaled mean values of the two studies are illustrated. There appeared to be quite a difference in the mean values of the directly valued health states between the postal and interview-based VAS exercises. Excluding the states 'death' and 'unconscious', nine out of thirteen health states were significantly different, which equals about 70 percent. As expected, it appears that the postal-based VAS valuations were lower than the valuations n the interview-based VAS study. However, this was only true for the relatively mild states, as respondents in the postal-based VAS study gave severe health states higher values than respondents in the interview-based VAS study.

Table 12. Comparison of VAS postal-based and interview-based mean value of directly valued health states.

	VAS_{postal}		$VAS_{interview}$	
Health state	**Mean**	**$Mean_{rescaled}$**	**Mean**	**$Mean_{rescaled}$**
11111	94.9 (n = 1170)	1.00	99.0 (n = 1332)	1.00
11112	71.5 (n = 1170)	0.70	81.8 (n = 659)*	0.75
11121	76.4 (n = 1170)	0.76	84.5 (n = 333)*	0.79
11122	55.5 (n = 1170)	0.49	71.1 (n = 333)*	0.64*
11211	79.2 (n = 1170)	0.79	86.2 (n = 673)*	0.83
12111	66.1 (n = 1170)	0.61	84.5 (n = 335)*	0.78*
21111	76.6 (n = 1170)	0.76	85.7 (n = 664)*	0.81
21232	37.0 (n = 1170)	0.27	37.7 (n = 340)	0.29
22233	29.2 (n = 1170)	0.17	20.8 (n = 340)*	0.10*
22323	24.8 (n = 1170)	0.09	21.9 (n = 335)	0.09
32211	37.1 (n = 1170)	0.26	30.4 (n = 335)*	0.22
33321	21.7 (n = 1170)	0.06	20.1 (n = 324)	0.06
33333	12.0 (n = 1170)	-0.06	4.1 (n = 1332)*	-0.10
Unconscious	13.4 (n = 1170)	-0.05	11.3 (n = 324)	-0.04
Dead	15.0 (n = 1170)	0.00	13.8 (n = 1332)	0.00

*($p < 0.01$)

The model VAS1 is the best-fitting model when predicting non-directly valued tariffs based on data collected both in the postal-based VAS study and in the interview-based VAS study. Hence, not surprisingly, both the coefficients representing the correlation of the values (Pearson correlation = 0.80) as well as the ranking (Spearman = 0.78) were significant ($p < 0.01$).

Transformation of VAS tariffs

Elsewhere a power function for the transformation of VAS valuations into TTO valuations at the aggregated (i.e. EQ-5D health state means) level has been estimated.[15] By using this power equation on the set of tariffs from the VAS1 model presented in appendix C, a 'transformed' set of tariffs was estimated. The transformed tariffs are illustrated in appendix E. Assessing the correlation coefficients between the set of transformed VAS-based EQ-5D tariffs and the Danish TTO-based EQ-5D tariffs, both the Pearson and Spearman coefficients remained significant ($p < 0.01$). Compared with the earlier coefficients presented in Table 7, the numeric changes were minor.

After the transformation, around 4 per cent of the VAS1-based EQ-5D tariffs were predicted at worse than death. Compared to the scenario including the untransformed VAS1-based EQ-5D tariffs illustrated in appendix C, this was only a minor change. Around 27 per cent of the transformed VAS1-based EQ-5D tariffs were predicted at higher than or equal to the TTO-based EQ-5D tariffs. Compared with the untransformed scenario, this was a decrease of approximately 10 per cent. Transforming the VAS1-based EQ-5D tariffs definitely does make a difference, *ceteris paribus*. However, how this difference would influence a cost-utility analysis is yet to be investigated.

---

[15] Refer to Chapter one in this dissertation.

The Parducci & Weddell R-F correction method

Adopting the R-F model leads to *corrected* VAS valuations, which all are lower than the observed VAS valuations. By applying the same regression analysis as shown above, it was found that $\forall = 0.99$ and R^2 (adjusted) is 0.96. Compared with the former power function, the correction of the observed VAS values resulted in higher values compared to the transformed VAS values. This brought them close to the (mean) TTO values and resulted in an almost perfect fit between the two valuation methods.

Discussion

Assessing data quality

The EQ-5D tariffs estimated in this study are based on direct VAS valuations of a sample of EQ-5D health states. The study was randomised within the Danish population. Problems and pitfalls that arise within such an exercise need to be addressed and assessed. In the standard EQ-5D questionnaire the time frame for respondents to imagine themselves in, within the VAS exercise, is one year. The respondents are also told that what happens after that is not known and should not be taken into consideration when valuating the health states. However, empirical studies show that respondents' VAS and/or TTO valuations are in fact affected by how the period of duration is specified [Gudex & Dolan 1995; Dolan 1996; Sackett & Torrance 1981; Sutherland *et al.* 1982; Ohinmaa & Sintonen 1994]. The findings are that as time spent in the health state increases, the mean health state valuations decrease.

Within the documented literature there appears to be indications of respondents being unwilling to use values at, or near, the endpoints of the VAS system, i.e. 0 and 100 for EQ-5D [Devlin *et al.* 2000]. The results of this behaviour are that the valuations are clustered somewhere between the endpoints. One implication noted by Badia *et al.* (1999b) is that when comparing valuation approaches for EQ-5D health states considerably more health states are valued worse than dead using the TTO method than using the VAS method. The reason, according to these authors, is that the VAS values are compressed into a much tighter valuation space than TTO values. The results reported here support these results, as the number of health states worse than dead, in the VAS1 tariffs, was 10 compared to 48 as reported in the Danish TTO-based set of EQ-5D tariffs. However, there is not enough evidence to conclude that respondents are unwilling to value health states at the extremes in the VAS exercise, as the differences between the VAS-based and TTO-based EQ-5D tariffs could be due to other factors. More research, preferably qualitative studies, ought to be conducted into assessing how respondents value health states when applying the VAS and TTO methods.

The response rate in this study was 53 per cent or 67 per cent, depending on how the response rate is calculated. This was fairly high by comparison with response rates found in similar studies across other countries [Essink-Bot *et al.* 1990; Brooks *et al.* 1991; Nord 1991; Nord *et al.* 1993; Ohinmaa *et al.* 1995; Johnson *et al.* 1998; Ikeda *et al.* 2000; Devlin *et al.* 2000; Rupel & Rebolj 2000]. The reason for nearly half (or one third of the respondents) not returning the postal questionnaire is likely to be related to responder burden imposed by both the length of the questionnaire and the burden of the valuation task. As the postal questionnaire, in addition to the EQ-5D questionnaire, also contained the 15D, the SF-36, and background questions and other tasks, this may have put an overload on some respondents. Although the self-completed VAS exercise is widely used in an international context, it is possible that some respondents had difficulties understanding the nature of the exercise was about, were thus unable to complete the exercise, and consequently did not return the questionnaire. As noted by Björk & Norinder (1999) this kind of health state valuation may be too complicated for a postal questionnaire. Such issues are difficult to examine with regard to a postal-based questionnaire, since one does not know why respondents fail to return the questionnaire. A possible way to find out why re-spondents fail to return the questionnaires could be by calling all respondents who fail to return the questionnaire and ask them why they did not return the questionnaire. However, this may give rise to other issues, for example some respondents may feel that their privacy is being invaded or that they are being monitored.

The quality of the data depends partly on the number of respondents/cases that have to be excluded. In order to use the direct valuations, they require to be rescaled. Around 30 per cent of the data had to be excluded for a variety of reasons, the most common being respondents who failed to score the health state 11111 or death in the VAS exercise. Similar findings are reported in Devlin *et al.* (2000). Although almost one third of the data had to be excluded, this did not jeopardise the randomisation of the sample in terms of either age or gender. However, the number of respondents 65 years and over was significantly lower after the exclusion, indicating that the majority of respondents being excluded were the elderly. It remains unknown whether there are any differences in terms of their health state preferences between the respondents excluded and the remaining respondents.

Danish VAS valuations compared to other European countries

We compared our mean values and the rescaled values of the directly valued health states with the valuations reported by the EQ-net Biomed Group and found a high degree of similarity between the two sets of valuations, both concerning the numeric valuations and the ranking. Hence preferences for health in Denmark may not differ significantly from those observed in other European countries. Johnson *et al.* (2000) compared VAS valuations for EQ-5D health states obtained in postal surveys in Finland and the United States. When applying a regression analysis, their estimates indicate that Finnish and US respondents did provide different preference valuations for different levels of health. However, the country-specific differences were not large and depended on the dimension and the level of problem on that dimension. These authors conclude that differences in health-state valuations are unlikely to have important implications when using the EQ-5D questionnaire in international studies.

Inconsistencies – an unavoidable fact?

A very important issue, and a central objective in the findings reported here, is how to handle inconsistencies in the valuation of direct VAS-based EQ-5D health states. In total, it was found that only around 17 per cent of the respondents displayed no inconsistencies at all and around 12 per cent had 10 inconsistencies or more. Our findings are supported by findings reported by Devlin *et al.* (2000). Within the documented literature, the number of inconsistencies reported in this study was quite large. Dolan & Kind (1996) reported an inconsistency rate of 9.3 per cent for a self-completed VAS postal survey in the UK. Badia *et al.* (1999b) reported a rate of 26 per cent for an interview-based VAS in a Catalan sample. However, other findings show a somewhat different result. Johnson *et al.* (1998) report a rate of 88 per cent within a self-completed postal VAS study in the US.

At present no *gold standard* exists with regard to how many inconsistencies to include in the modelling of non-directly valued EQ-5D health states or whether all inconsistencies should be excluded. Findings by Ohinmaa & Sintonen (1998) reported a threshold of *three* inconsistencies, indicating that respondents with more than three inconsistencies ought to be excluded from the data set for modelling purposes. These findings were not supported either by Devlin *et al.* (2000) or in our results. Badia *et al.* (1999b) concluded that the inconsistencies they analysed did not affect rankings in their final tariff of values. Our results support this conclusion.

We performed the regression analysis initially with all inconsistencies included. However, when only respondents with zero inconsistencies are included the choice of a 'best-fitting' model changed. The model we chose when all inconsistencies were included - VAS1 - still yielded consistent and significant parameters, but was out-performed by another model (VAS2) with a higher R^2 value. The VAS2 model, however, was 'only' based on 209 respondents and was not very robust or randomised for the general Danish population. When all inconsistencies were included, VAS2 resulted in consistent parameters, but with an insignificant *p*-value on the UA ('usual activities') parameter. By consecutively excluding inconsistencies the VAS2 model parameters all became significant ($p < 0.05$) with a maximum of four inconsistencies included. Al-though we have reported results from the VAS2 model, the VAS1 model is, from a purely statistical point of view, the best-fitting model for predicting a set of VAS based tariffs.

According to Ikeda *et al.* (2000) it seems difficult to argue for any particular point at which the parameters stabilise, so that additional exclusion of respondents based on inconsistency would seem unwarranted. Our findings show that the majority of the health state valuations remained fairly (and significantly) stable when including more inconsistencies. According to the modelling, the preferred model changed when excluding all inconsistencies versus including all inconsistencies, however, this left only 209 respondents in the dataset, which is very low and not randomised. More research in the field of whether to include inconsistencies is necessary to establish formal guidelines. At present, based on our findings, we recommend that all inconsistencies are included in the study, unless the study in question contains a dataset that after exclusion of all inconsistencies still remains randomised within the population of study.

Estimating Danish EQ-5D tariffs based on the VAS technique

Several studies have tried to elicit a national set of EQ-5D tariffs based on the VAS technique [Devlin *et al.* 2000, Ikeda et al. 2000; Rupel & Rebolj 2000]. In order to find an appropriate model for the valuation of non-directly EQ-5D health states, all three studies assessed several models, however, none were very explicit about the appropriate regression analysis. All three studies, for *ad hoc* reasons, adapted the best fitting regression model presented in Dolan (1997) without reporting explicitly on the tests. We believe that different kinds of regression models need to be more explicitly tested, and analysts should not merely choose the RE model, as reported in the documented literature.

Several different models were tested against each other by applying several kinds of econometric tests. We too had to reject the OLS model compared to the RE model. Nevertheless, the RE model showed some problems concerning efficiency and from a statistical point of view we found the FE model to be superior to the RE model. However, in the FE model some of the parameters were negative, which left us accepting the RE model. In conclusion, extensive testing resulted in a different preferred model than the one reported by previous studies. The FE model may report negative values due to the data set itself, and other datasets may not necessarily support our findings.

We therefore urge analysts wishing to estimate non-direct EQ-5D health states to be very explicit about their testing, as it is not initially given that the RE model will be the preferred econometric model to use.

Postal-based versus interview-based EQ-5D (direct) valuations

Applying the direct VAS valuations from the interview-based VAS exercise (not reported here), we compared them with the direct VAS valuations from the postal-based VAS valuations reported in this study. A direct comparison shows that, in general, the mean valuations based on the postal-based VAS exercise were lower than the mean valuations based on the interview-based VAS exercise. Our findings support those reported by the EQ-net Biomed Group (1999). The reason for the dissimilarities points in the direction that respondents value health states differently when they are given the chance to prerank the health states. If this is true, the differences are due to a *framing effect,* indicating that with a prior ranking exercise the respondents understand the exercise more clearly. What are the future directions then? Should we add a ranking exercise prior to the VAS exercise within the EQ-5D questionnaire? In order to say anything conclusive on this issue we need more empirical studies addressing whether our findings can be supported.

Transformation of VAS valuations

Torrance *et al.* (2001) recommend that VAS tariffs be transformed to either SG or TTO tariffs by applying a power function expressing the linear relationship between VAS and TTO valuations. Consequently, a power function estimated at the aggregated level is applied to transform the VAS tariffs.[16] This transformation does not change the correlation with the set of TTO-based EQ-5D tariffs significantly. In general, it is difficult to test empirically which set of tariffs should be applied in an economic evaluation. However, in economic theory there is nothing to indicate that VAS valuations become utilities when they are transformed by a power function. Such a transformation merely shows that a stable statistical fit between health states within the two valuation methods can be found. In this study, we applied the power function directly to the set of predicted VAS tariffs. However, if the objective is simply to investigate patient valuations of EQ-5D health states, the power function should be applied at the aggregated level, that is, mean valuations of the different health states. There are, to our knowledge, no documented studies that actually apply a power transformation (between VAS and TTO valuations) on VAS tariffs, which makes it hard for us to compare our results to external results.

In conclusion, we agree with Torrance *et al.* (2001) that VAS valuations ought to be transformed using a power function expressing the linear relationship between VAS valuations and either TTO, SG, or PTO valuations. However, we are not convinced that this transformation of VAS tariffs turns them into utilities, that is, an expression of individual preferences, and thus can be applied in the allocation of health care services. Although VAS tariffs are transformed by applying a power function, we urge that caution is shown in handling and interpreting the results stemming from such tariffs.

Applying the R-F correction method

As suggested by Parducci and Weddell (1984), VAS valuations are influenced by the remaining (VAS) valuations in the set. This has been confirmed by Bleichrodt and Johannesson (1997), who state that *"...the sensitivity of the rating-scale valuations to the number of preferred health states casts doubt on the common practice of valuing several health states simultaneously."* Their (Bleichrodt and Johannesson's) findings suggest that the rating scale may be useful only when health states are valued in isolation. Parducci and Weddell (1984) have developed a formula which should correct this. We applied this formula to our non-rescaled VAS valuations and used the corrected data to estimate a (new) power function. This resulted in $\forall = 0.99$ (adjusted $R^2 = 0.96$), which was almost a perfect fit. Cer-

---

[16] The power function is estimated in Chapter 1 in this dissertation.

tainly this correction of the VAS valuations made a difference as seen in the (new) estimated power function. This correction gives an almost perfect linear fit between the two valuation methods and may imply that a power transformation between the VAS-based EQ-5D tariffs and the TTO-based EQ-5D tariffs is no longer needed, since the difference between the two instruments is eliminated. If the theory behind the R-F model holds, this is certainly a beneficial correction of the VAS values, implying that the VAS-based EQ-5D tariffs may possibly be used as a substitute for TTO-based EQ-5D tariffs. However, it is still the case that the VAS valuations are based on a non-trade-off situation, whereas economists by training are most comfortable with trade-off/choice-based techniques for preference elicitation. We urge that focus on this correction method is increased.

References

1. Abdalla M & Russel I. *Tariffs for the EuroQol health states based on modelling the individual VAS and TTO data of the York study.* In O'Hanlon M & Buxton M. (eds) EuroQol Plenary Meeting, London, October 1994. Conference proceedings. HERG research report No 20. Health Economics Research Group, Brunel University. Uxbridge September 1995; 41-74.
2. Arrow KJ. *Social choice and individual values.* Wiley: New York. 1951.
3. Badia X, Monserrat S, Roset M, Herdman M. Feasibility, validity and test-retest reliability of scaling methods for health states: The visual analogue scale and the time trade-off. *Quality of Life Research* 1999a; 8(4): 303-10.
4. Badia X, Roset M, Herman M. Inconsistent responses in three preference-elicitation methods for health states. *Social Science and Medicine* 1999b; 49: 943-950.
5. Bass EB, Steinberg E, Pitt H, et al. Comparison of the rating scale and the standard gamble in measuring patient preferences for outcomes of gallstone disease. *Medical Decision Making* 1994; 14: 307-14.
6. Björk S & Norinder A. The weighting exercise for the Swedish version of the EuroQol. *Health Economics* 1999; 8(2): 117-26.
7. Bleichrodt H & Johannesson M. An experimental test of a theoretical foundation for rating scales valuations. *Medical Decision Making* 1997; 17: 208-16.
8. Boyd N, Sutherland H, Heasman K, et al. Whose utilities for decision analysis? *Medical Decision Making* 1990; 10: 58-67.
9. Brooks RG, Jendteg S, Lindgren B, Persson U, Bjørk S. EuroQol: Health-related quality of life measurement. Results of the Swedish questionnaire exercise. Health Policy 1991; 18: 25-36.
10. Brooks R, with the EuroQol Group. EuroQol: The current state of play. *Health Policy* 1996; 37: 53-72.
11. Claes C, Greiner W, Uber A, Graf von der Schulenburg JM. *The new German version of the EuroQol quality of life questionnaire.* In Rabin RE, Busschbach JJV, Charro Frank de, Essink-Bot ML, Bonsel GJ. (eds) EuroQol Plenary meeting Rotterdam 1997, 2 – 3 October. Discussion papers. Centre for Health Policy and Law, Erasmus University, Rotterdam. 1998.
12. Devlin NJ, Hansen P, Kind P, Williams A. *The health state preferences and logical inconsistencies of New Zealanders: A tale of two tariffs.* York Centre for Health Economics UK and University of Otago New Zealand. Discussion Paper No 180. 2000.
13. Dolan P, Gudex C, Kind P, Williams A. *A social tariff for EuroQol: Results form a UK general population survey.* York: Center for Health Economics, The University of York. Discussion Paper No 138. 1995.
14. Dolan P & Kind P. Inconsistency and health state valuations. *Social Science and Medicine* 1996; 42: 609-15.

15. Dolan P. Modelling valuations for health states: The effect of duration. *Health Policy* 1996; 38(3): 189-203.
16. Dolan P. Modeling valuations for EuroQol health states. *Medical Care* 1997; 35: 1095-1108.
17. Dolan P. The measurement of health-related quality of life for use in resource allocation decisions in health care. In Culyer AJ & Newhouse JP. (eds) *Handbook of Health Economics*. Elsevier North-Holland: Amsterdam. 2000.
18. Dyer JS & Sarin R. Measurable multiattribute value functions. *Operations Research* 1979; 27: 810-822.
19. Dyer J & Sarin R. Relative risk aversion. *Management Science* 1982; 28: 875-86.
20. Ellsberg D. Classic and current notions of measurable utility. *Economic Journal* 1954; 64: 528-56.
21. Elstein A, Holzmann G, Ravitch M, et al. Comparison of physicians' decisions regarding estrogen replacement therapy for menopausal women and decisions derived from a decision analytic model. *American Journal of Medicine* 1986; 80: 246-58.
22. Essink-Bot ML. Valuation of health states by the general public: Feasibility of standardised measurement procedure. *Social Science & Medicine 1990*; 31: 1201-6.
23. Essink-Bot ML, Stouthard M, Bonsel GJ. Generalizability of valuations on health states collected with the EuroQol questionnaire. *Health Economics* 1993; 2: 237-46.
24. Feeny D, Townsend M, Furlong W, et al. *Assessing health-related quality of life in prenatal diagnosis, comparing chorionic villi sampling amniocentesis: A technical report*. Hamilton, Ontario: Centre for Health Economics and Policy Analysis, McMaster University, 2000. Working Paper No 4. 2000.
25. Fishburn PC. Retrospective on the utility theory of von Neumann and Morgenstern. *Journal of Risk and Uncertainty* 1989; 2: 127-58.
26. Furlong W, Feeny D, Torrance GW, et al. *Multiplicative multi-attribute utility function for the Health Utilities Index Mark 3 (HUI3) system: A technical report*. Hamilton, Ontario: Centre for Health Economics and Policy Analysis, McMaster University. Woking Paper No 11. 1998.
27. Green WH. *Econometric analysis*. Prentice Hall: New York. 2000.
28. Gudex C & Dolan P. *Valuing health states: The effect of duration*. Discussion Paper 143, Centre for Health Economics, University of York. 1995.
29. Hisashige A, Mikasa H, Katayama T. Description and valuation of health-related quality of life among the general public in Japan by the EuroQol. *Journal of Medical Investigation* 1998; 45: 123-129.
30. Ikeda S, Sakai I, Tamura M, Tsuchiya A. *VAS valuations of hypothetical health states using EQ-5D in Japan*. 17th EuroQol Plenary Meeting, Pamplona, Spain, 28 – 29 September 2000.

31. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. *Journal of Epidemiology and Community Health* 1999; 53(1): 46-50.
32. Johnson JA, Coons SJ, Ergo A, Szava-Kovats G. Valuation of EuroQol (EQ-5D) health states in an adult US sample. *Pharmacoeconomics* 1998; 13(4): 421-33.
33. Johnson JA, Ohinmaa A, Murti B, Sintonen H, Coons SJ. Comparison of Finnish and U.S.-based visual analog scale valuations of the EQ-5D measure. *Medical Decision Making* 2000; 20: 281-289.
34. Keeney RL. A group preference axiomisation with cardinal utility. *Management Science* 1976; 23: 140-45.
35. Keeney RL & Raiffa H. *Decisions with multiple objectives. Preferences and value tradeoffs*. Cambridge University Press: Cambridge. 1976.
36. Kind P. *Measuring valuations for health states: A survey of patients in general practice*. York Centre for Health Economics Discussion Paper No 76. 1990.
37. Kind P, Gudex C, Dolan P, Williams A. Practical and methodological issues in the development of the EuroQol: The York experiment. *Adv Med Sociol* 1994; 5: 219-53.
38. Kind P. The EuroQol instrument: An index of health-related quality of life. In Spilker B. (eds) *Quality of life and pharmacoeconomics in clinical trials*. Lippincott-Raven Publishers: Phili-delphia. 1996.
39. Krabbe PFM, Essink-Bot ML, Bonsel GJ. The comparability and reliability of five health-state valuation methods. *Social Science and Medicine* 1997; 45: 1641-52.
40. Krantz DH, Luce RD, Suppes P, Tversky A. *Foundations of measurement*. Academic Press: New York. 1971.
41. Lipscomb J. Value preferences for health: Meaning measurement and use in program evaluation. In Kane RL & Kane RA. (eds) *Values and long term care*. Lexington Books: New York. 1982.
42. McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care* 1993; 31(3): 247-263.
43. Nord E. EuroQol: Health-related quality of life measurement. Valuations of health states by the general public in Norway. *Health Policy* 1991; 18(1): 25-36.
44. Nord E, Richardson J, Macarounas-Kirchman K. Social evaluation of health care versus personal evaluation of health states. Evidence of the validity of four health-state scaling instruments using Norwegian and Australian surveys. *International Journal of Technology Assessment in Health Care* 1993; 9: 463-478.
45. Ohinmaa A & Sintonen H. The effect of duration on the value given to the EuroQol states. In Busschbach J (eds.) *EuroQol conference proceedings*. Erasmus University: Rotterdam. 1994.
46. Ohinmaa A, Helala E, Sintonen H. *Modeling EuroQol values of Finnish adult population*. 12th EuroQol Plenary Meeting, Barcelona, Spain, 3 – 6 October 1995.

47. Ohinmaa A & Sintonen H. *Inconsistencies and modelling of the Finnish EuroQol (EQ-5D) preference values*. Paper presented at the 15th EuroQol Plenary Meeting. Hannover 1st and 2nd October. 1998.
48. Patrick D, Starks H, Cain K, Uhlmann R, Pearlman R. Measuring preferences for health states worse than death. *Medical Decision Making* 1994; 14: 9-18.
49. Poulton E. Models for biases in judging sensory magnitude. *Psychology Bulletin* 1979; 86: 777-803.
50. Read J, Quinn R, Berwick D, Fineberg H, Weinstein M. Preferences for health outcomes – comparisons of assessment methods. *Medical Decision Making* 1984; 4: 315-29.
51. Richardson J. Cost-utility analysis: What should be measured? *Social Science and Medicine* 1994; 39: 7-20.
52. Robinson A, Loomes G, Jones-Lee M. Visual analog scales, standard gambles and relative risk aversion. *Medical Decision Making 2001*; 21: 17-27.
53. Rupel V & Rebolj M. The Slovenian VAS tariff based on valuations of EQ-5D health states from the general population. 17th EuroQol Plenary Meeting, Pamplona, Spain, 28 – 29 September 2000.
54. Sackett DL & Torrance GW. The utility of different health states as perceived by the general public. *Journal of Chronic Diseases* 1981; 31: 697-704.
55. Schwartz A. Rating scales in context. *Medical Decision Making* 1998; 18: 236.
56. Sintonen H. The 15D instrument of health-related quality of life: Properties and applications. *Annals of Medicine* 2001; 33: 328-336.
57. Stevens SS. Issues in psychological measurement. *Psychological Review* 1971; 78: 426-50.
58. Sutherland HJ, Llewellyn-Thomas H, Boyd NF, Till JE. Attitudes towards quality of survival: The concept of maximal endurable time. *Medical Decision Making* 1982; 2: 299-309.
59. Torrance GW. Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Economic Planning Science* 1976; 10: 129-36.
60. Torrance GW, Feeny D, Furlong W. Visual Analog Scales: Do they have a role in the measurement of preferences for health states? *Medical Decision Making* 2001; 21: 329-34.
61. The EQ-net Biomed Group. Report on the EQ-net VAS database. Valuation task force. Centre for Health Policy and Law, Erasmus University, Rotterdam. 1999.
62. von Winterfeldt D & Edwards W. *Decision analysis and behavioural research*. Cambridge Universi-ty Press: Cambridge. 1986.
63. Ware JE & Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care* 1992; 30(6): 473-483.
64. Weijnen T, Nieuwenhuizen M, Oppe S, de Charro F. *A European EQ-5D value set. Myth or reality?* Report nr. 6 on the VAS values. The Rotterdam Analysis Team, 1999.

65. Wittrup-Jensen KU, Lauridsen JT, Gudex C, Brooks R, Pedersen KM. *Danish EuroQol tariffs estimated by the Visual Analogue Scale (VAS) and the Time Trade-Off (TTO).* Paper presented at the 6th EuroQol Group Meeting in Copenhagen, 6th/7th September 2001.
66. Wolfson AF, Sinclair AJ, Bombardier C, McGeer A. Preference measurements for functional status in stroke patients: Inter-rater and inter-technique comparisons. In Kane R & Kane R. (eds) *Values and long term care.* Lexington: Washington. 1982.

Appendix A

Tabel A.1. Logical consistencies in the VAS study.

j *i*	11211	11111	21232	11122	11121	22233	33333	33321	21111	BEV	12111	11112	32211	22323	Dead
11211	-	≤	≥	?	?	≥	≥	≥	?	?	?	?	≥	≥	?
11111		-	≥	≥	≥	≥	≥	≥	≥	≥	≥	≥	≥	≥	≥
21232			-	≤	≤	?	≥	?	≤	?	?	≤	?	?	?
11122				-	≤	≥	≥	?	?	?	?	≤	?	≥	?
11121					-	≥	≥	≥	?	?	?	?	?	≥	?
22233						-	≥	?	≤	?	≤	≤	?	?	?
33333							-	≤	≤	?	≤	≤	≤	≤	?
33321								-	≤	?	≤	?	≤	?	?
21111									-	?	?	?	≥	≥	?
UNC										-	?	?	?	?	?
12111											-	?	≥	≥	?
11112												-	?	≥	?
32211													-	?	?
22323														-	?
Dead															-

Note: although the focus is on logical *inconsistencies*, this appendix illustrates logical *consistencies*.

Appendix B

Table B. Estimates of models 3 to 11 including all inconsistencies.

Model Variable	3	4	5	6*	7	8	9	10*	11
A	0.8921 (<0.0001)	0.9670 (<0.0001)	0.9670 (<0.0001)	.	-0.0496 (<0.0001)	-0.0496 (<0.0001)	0.9915 (<0.0001)	.	0.8406 (<0.0001)
MO	-0.1211 (<0.0001)	-0.2053 (<0.0001)	-0.2053 (<0.0001)	.		0.0361 (<0.0001)	-0.2480 (<0.0001)	.	-0.0516 (<0.0001)
SC	-0.2679 (<0.0001)	-0.3522 (<0.0001)	-0.3522 (<0.0001)	.		-0.1108 (<0.0001)	-0.3950 (<0.0001)	.	-0.1405 (<0.0001)
UA	-0.0886 (<0.0001)	-0.1729 (0.5149)	-0.1729 (<0.0001)	.		0.0396 (<0.0001)	-0.2219 (<0.0001)	.	-0.0120 (0.0272)
PD	-0.1519 (<0.0001)	-0.2080 (<0.0001)	-0.2080 (<0.0001)	.		0.0624 (<0.0001)	-0.1900 (<0.0001)	.	-0.0515 (<0.0001)
AD	-0.2242 (<0.0001)	-0.2709 (<0.0001)	-0.2709 (<0.0001)	.		-0.0295 (<0.0001)	-0.2954 (<0.0001)	.	-0.0820 (<0.0001)
M2	-0.0320 (0.0548)	-0.0321 (0.0539)	-0.2943 (<0.0001)					.	
S2	0.3249 (<0.0001)	0.5498 (<0.0001)	-0.2368 (0.0030)					.	
U2	-0.0242 (0.3028)	-0.0151 (0.5197)	0.2471 (<0.0001)					.	
P2	0.1108 (<0.0001)	0.1012 (<0.0001)	0.3634 (<0.0001)					.	
A2	0.3912 (<0.0001)	0.5316 (<0.0001)	0.2694 (<0.0001)					.	
MOSC			0 .	.					
MOUA			0 .	.					
MOPD			0 .	.					
MOAD			0 .	.					
SCUA			0 .	.					
SCPD			0 .	.					
SCAD			0 .	.					
UAPD			0 .	.					
UAAD			0 .	.					
PDAD			0 .						
F11					0.1125 (<0.0001)	0.1204 (<0.0001)			
F21					0.1087 (<0.0001)	0.2865 (<0.0001)			
F31					0.5376 (<0.0001)	0.5048 (<0.0001)			
F41					0.7749 (<0.0001)	0.7752 (<0.0001)			
F13					0.2023 (<0.0001)	0.0235 (0.2302)	0.4009 (<0.0001)		
F23					0.1829 (<0.0001)	0.1636 (<0.0001)	0.9933 (<0.0001)		
F33					0 .	0 .	0.9910 (<0.0001)		
F43					0 .	0 .	0 .		
F53					-0.0045 (0.6475)	0 .	1.6548 (<0.0001)		
N2								.	-0.2196 (0.0001)
N3		0.2622 (<0.0001)						.	-0.2196 (0.0001)
R^2	0.7477	0.7486	0.7486	.	0.7394	0.7473	0.7466	.	0.7331

*Model could not be computed due to infinite log-likelihood estimates.

Appendix C

Table C. VAS1 (n=1170) and TTO tariffs for the EQ-5D health care classification system – absolute values and differences.

State	VAS1	TTO*	difference	state	VAS1	TTO*	difference	state	VAS1	TTO	difference
11111	1.000	1.000	0.000	13231	0.331	0.251	0.080	23121	0.347	0.579	-0.232
11112	0.684	0.818	-0.134	13232	0.240	0.183	0.057	23122	0.256	0.511	-0.255
11113	0.593	0.519	0.074	13233	0.149	-0.117	0.266	23123	0.165	0.211	-0.046
11121	0.697	0.824	-0.127	13311	0.423	0.551	-0.128	23131	0.269	0.245	0.024
11122	0.606	0.756	-0.150	13312	0.332	0.483	-0.151	23132	0.178	0.177	0.001
11123	0.515	0.456	0.059	13313	0.241	0.183	0.058	23133	0.087	-0.122	0.209
11131	0.619	0.490	0.129	13321	0.345	0.489	-0.144	23211	0.361	0.594	-0.233
11132	0.528	0.422	0.106	13322	0.254	0.421	-0.167	23212	0.270	0.525	-0.255
11133	0.437	0.123	0.315	13323	0.163	0.121	0.042	23213	0.179	0.226	-0.047
11211	0.711	0.838	-0.127	13331	0.267	0.155	0.112	23221	0.283	0.531	-0.248
11212	0.620	0.770	-0.150	13332	0.176	0.087	0.089	23222	0.192	0.463	-0.271
11213	0.529	0.471	0.058	13333	0.085	-0.213	0.298	23223	0.101	0.164	-0.063
11221	0.633	0.776	-0.143	21111	0.649	0.833	-0.184	23231	0.205	0.198	0.007
11222	0.542	0.708	-0.166	21112	0.558	0.765	-0.207	23232	0.114	0.129	-0.015
11223	0.451	0.409	0.042	21113	0.467	0.465	0.002	23233	0.023	-0.170	0.193
11231	0.555	0.442	0.113	21121	0.571	0.771	-0.200	23311	0.297	0.498	-0.201
11232	0.464	0.374	0.090	21122	0.480	0.703	-0.223	23312	0.206	0.430	-0.224
11233	0.373	0.075	0.298	21123	0.389	0.403	-0.014	23313	0.115	0.130	-0.015
11311	0.647	0.743	-0.096	21131	0.493	0.437	0.056	23321	0.219	0.436	-0.217
11312	0.556	0.674	-0.118	21132	0.402	0.369	0.033	23322	0.128	0.367	-0.239
11313	0.465	0.375	0.090	21133	0.311	0.069	0.242	23323	0.037	0.068	-0.031
11321	0.569	0.680	-0.111	21211	0.585	0.785	-0.200	23331	0.141	0.102	0.039
11322	0.478	0.612	-0.134	21212	0.494	0.717	-0.223	23332	0.050	0.034	0.016
11323	0.387	0.313	0.074	21213	0.403	0.418	-0.015	23333	-0.041	-0.266	0.225
11331	0.491	0.347	0.144	21221	0.507	0.723	-0.216	31111	0.523	0.475	0.048
11332	0.400	0.278	0.122	21222	0.416	0.655	-0.239	31112	0.432	0.407	0.025
11333	0.309	-0.021	0.330	21223	0.325	0.355	-0.030	31113	0.341	0.107	0.234
12111	0.663	0.823	-0.160	21231	0.429	0.389	0.040	31121	0.445	0.413	0.032
12112	0.572	0.755	-0.183	21232	0.338	0.321	0.017	31122	0.354	0.345	0.009
12113	0.481	0.456	0.025	21233	0.247	0.021	0.226	31123	0.263	0.045	0.218
12121	0.585	0.761	-0.176	21311	0.521	0.689	-0.168	31131	0.367	0.079	0.288
12122	0.494	0.693	-0.199	21312	0.430	0.621	-0.191	31132	0.276	0.011	0.265
12123	0.403	0.393	0.010	21313	0.339	0.322	0.017	31133	0.185	-0.289	0.474
12131	0.507	0.427	0.080	21321	0.443	0.627	-0.184	31211	0.459	0.427	0.032
12132	0.416	0.359	0.057	21322	0.352	0.559	-0.207	31212	0.368	0.359	0.009
12133	0.325	0.060	0.265	21323	0.261	0.260	0.001	31213	0.277	0.060	0.217
12211	0.599	0.776	-0.177	21331	0.365	0.293	0.072	31221	0.381	0.365	0.016
12212	0.508	0.707	-0.199	21332	0.274	0.225	0.049	31222	0.290	0.297	-0.007
12213	0.417	0.408	0.009	21333	0.183	-0.074	0.257	31223	0.199	-0.003	0.202
12221	0.521	0.713	-0.192	22111	0.537	0.770	-0.233	31231	0.303	0.031	0.272
12222	0.430	0.645	-0.215	22112	0.446	0.702	-0.256	31232	0.212	-0.037	0.249
12223	0.339	0.346	-0.007	22113	0.355	0.402	-0.047	31233	0.121	-0.336	0.457
12231	0.443	0.380	0.063	22121	0.459	0.708	-0.249	31311	0.395	0.331	0.064
12232	0.352	0.311	0.041	22122	0.368	0.640	-0.272	31312	0.304	0.263	0.041
12233	0.261	0.012	0.249	22123	0.277	0.340	-0.063	31313	0.213	-0.036	0.249
12311	0.535	0.680	-0.145	22131	0.381	0.374	0.007	31321	0.317	0.269	0.048
12312	0.444	0.612	-0.168	22132	0.290	0.306	-0.016	31322	0.226	0.201	0.025
12313	0.353	0.312	0.041	22133	0.199	0.006	0.193	31323	0.135	-0.098	0.233
12321	0.457	0.618	-0.161	22211	0.473	0.722	-0.249	31331	0.239	-0.065	0.304
12322	0.366	0.549	-0.183	22212	0.382	0.654	-0.272	31332	0.148	-0.133	0.281
12323	0.275	0.250	0.025	22213	0.291	0.355	-0.064	31333	0.057	-0.432	0.489
12331	0.379	0.284	0.095	22221	0.395	0.660	-0.265	32111	0.411	0.412	-0.001
12332	0.288	0.216	0.072	22222	0.304	0.592	-0.288	32112	0.320	0.344	-0.024
12333	0.197	-0.084	0.281	22223	0.213	0.292	-0.079	32113	0.229	0.044	0.185
13111	0.551	0.695	-0.144	22231	0.317	0.326	-0.009	32121	0.333	0.350	-0.017
13112	0.460	0.626	-0.166	22232	0.226	0.258	-0.032	32122	0.242	0.282	-0.040
13113	0.369	0.327	0.042	22233	0.135	-0.041	0.176	32123	0.151	-0.018	0.169
13121	0.473	0.632	-0.159	22311	0.409	0.627	-0.218	32131	0.255	0.016	0.239
13122	0.382	0.564	-0.182	22312	0.318	0.558	-0.240	32132	0.164	-0.052	0.216
13123	0.291	0.265	0.026	22313	0.227	0.259	-0.032	32133	0.073	-0.352	0.425
13131	0.395	0.299	0.096	22321	0.331	0.564	-0.233	32211	0.347	0.364	-0.017
13132	0.304	0.230	0.074	22322	0.240	0.496	-0.256	32212	0.256	0.296	-0.040
13133	0.213	-0.069	0.282	22323	0.149	0.197	-0.048	32213	0.165	-0.003	0.168
13211	0.487	0.647	-0.160	22331	0.253	0.231	0.022	32221	0.269	0.302	-0.033
13212	0.396	0.579	-0.183	22332	0.162	0.162	0.000	32222	0.178	0.234	-0.056
13213	0.305	0.279	0.026	22333	0.071	-0.137	0.208	32223	0.087	-0.066	0.153
13221	0.409	0.585	-0.176	23111	0.425	0.641	-0.216	32231	0.191	-0.032	0.223
13222	0.318	0.516	-0.198	23112	0.334	0.573	-0.239	32232	0.100	-0.100	0.200
13223	0.227	0.217	0.010	23113	0.243	0.274	-0.031	32233	0.009	-0.399	0.408
32311	0.283	0.269	0.014								
32312	0.192	0.200	-0.008								
32313	0.101	-0.099	0.200								
32321	0.205	0.206	-0.001								

32322	0.114	0.138	-0.024		
32323	0.023	-0.161	0.184		
32331	0.127	-0.127	0.254		
32332	0.036	-0.196	0.232		
32333	-0.055	-0.495	0.440		
33111	0.299	0.283	0.016		
33112	0.208	0.215	-0.007		
33113	0.117	-0.084	0.201		
33121	0.221	0.221	0.000		
33122	0.130	0.153	-0.023		
33123	0.039	-0.146	0.185		
33131	0.143	-0.113	0.256		
33132	0.052	-0.181	0.233		
33133	-0.039	-0.480	0.441		
33211	0.235	0.236	-0.001		
33212	0.144	0.167	-0.023		
33213	0.053	-0.132	0.185		
33221	0.157	0.173	-0.016		
33222	0.066	0.105	-0.039		
33223	-0.025	-0.194	0.169		
33231	0.079	-0.160	0.239		
33232	-0.012	-0.229	0.217		
33233	-0.103	-0.528	0.425		
33311	0.171	0.140	0.031		
33312	0.080	0.072	0.008		
33313	-0.011	-0.228	0.217		
33321	0.093	0.078	0.015		
33322	0.002	0.009	-0.007		
33323	-0.089	-0.290	0.201		
33331	0.015	-0.256	0.271		
33332	-0.076	-0.324	0.248		
33333	-0.167	-0.624	0.457		
Dead	[0.000]	[0.000]	[0.000]		
Uncon.	[-0.050]	[-0.293]	[0.243]		

Note. *TTO tariffs obtained from Chapter 1 reported in this dissertation.

Appendix D

Table D. VAS2 (n = 209) and TTO tariffs for the EQ-5D health care classification system – absolute values and differences.

State	VAS2	TTO*	difference	state	VAS2	TTO*	difference	state	VAS2	TTO*	difference
11111	1.000	1.000	0.000	13222	0.151	0.516	-0.365	22333	0.043	-0.137	0.180
11112	0.716	0.818	-0.102	13223	0.087	0.217	-0.130	23111	0.248	0.641	-0.393
11113	0.418	0.519	-0.101	13231	0.140	0.251	-0.111	23112	0.184	0.573	-0.389
11121	0.705	0.824	-0.119	13232	0.076	0.183	-0.107	23113	0.120	0.274	-0.154
11122	0.641	0.756	-0.115	13233	0.012	-0.117	0.129	23121	0.173	0.579	-0.406
11123	0.343	0.456	-0.113	13311	0.268	0.551	-0.283	23122	0.109	0.511	-0.402
11131	0.396	0.490	-0.094	13312	0.204	0.483	-0.279	23123	0.045	0.211	-0.166
11132	0.332	0.422	-0.090	13313	0.140	0.183	-0.043	23131	0.098	0.245	-0.147
11133	0.268	0.123	0.146	13321	0.193	0.489	-0.296	23132	0.034	0.177	-0.143
11211	0.758	0.838	-0.080	13322	0.129	0.421	-0.292	23133	-0.030	-0.122	0.092
11212	0.694	0.770	-0.076	13323	0.065	0.121	-0.056	23211	0.226	0.594	-0.368
11213	0.396	0.471	-0.075	13331	0.118	0.155	-0.037	23212	0.162	0.525	-0.363
11221	0.683	0.776	-0.093	13332	0.054	0.087	-0.033	23213	0.098	0.226	-0.128
11222	0.619	0.708	-0.089	13333	-0.010	-0.213	0.203	23221	0.151	0.531	-0.380
11223	0.321	0.409	-0.088	21111	0.599	0.833	-0.234	23222	0.087	0.463	-0.376
11231	0.374	0.442	-0.068	21112	0.535	0.765	-0.230	23223	0.023	0.164	-0.141
11232	0.310	0.374	-0.064	21113	0.237	0.465	-0.228	23231	0.076	0.198	-0.122
11233	0.246	0.075	0.171	21121	0.524	0.771	-0.247	23232	0.012	0.129	-0.117
11311	0.502	0.743	-0.241	21122	0.460	0.703	-0.243	23233	-0.052	-0.170	0.118
11312	0.438	0.674	-0.236	21123	0.162	0.403	-0.241	23311	0.204	0.498	-0.294
11313	0.374	0.375	-0.001	21131	0.215	0.437	-0.222	23312	0.140	0.430	-0.290
11321	0.427	0.680	-0.253	21132	0.151	0.369	-0.218	23313	0.076	0.130	-0.054
11322	0.363	0.612	-0.249	21133	0.087	0.069	0.018	23321	0.129	0.436	-0.307
11323	0.299	0.313	-0.014	21211	0.577	0.785	-0.208	23322	0.065	0.367	-0.302
11331	0.352	0.347	0.005	21212	0.513	0.717	-0.204	23323	0.001	0.068	-0.067
11332	0.288	0.278	0.010	21213	0.215	0.418	-0.203	23331	0.054	0.102	-0.048
11333	0.224	-0.021	0.245	21221	0.502	0.723	-0.221	23332	-0.010	0.034	-0.044
12111	0.663	0.823	-0.160	21222	0.438	0.655	-0.217	23333	-0.074	-0.266	0.192
12112	0.599	0.755	-0.156	21223	0.140	0.355	-0.215	31111	0.184	0.475	-0.291
12113	0.301	0.456	-0.155	21231	0.193	0.389	-0.196	31112	0.120	0.407	-0.287
12121	0.588	0.761	-0.173	21232	0.129	0.321	-0.192	31113	0.056	0.107	-0.051
12122	0.524	0.693	-0.169	21233	0.065	0.021	0.044	31121	0.109	0.413	-0.304
12123	0.226	0.393	-0.167	21311	0.321	0.689	-0.368	31122	0.045	0.345	-0.300
12131	0.279	0.427	-0.148	21312	0.257	0.621	-0.364	31123	-0.019	0.045	-0.064
12132	0.215	0.359	-0.144	21313	0.193	0.322	-0.129	31131	0.034	0.079	-0.045
12133	0.151	0.060	0.091	21321	0.246	0.627	-0.381	31132	-0.030	0.011	-0.041
12211	0.641	0.776	-0.135	21322	0.182	0.559	-0.377	31133	-0.094	-0.289	0.195
12212	0.577	0.707	-0.130	21323	0.118	0.260	-0.142	31211	0.162	0.427	-0.265
12213	0.279	0.408	-0.129	21331	0.171	0.293	-0.122	31212	0.098	0.359	-0.261
12221	0.566	0.713	-0.147	21332	0.107	0.225	-0.118	31213	0.034	0.060	-0.026
12222	0.502	0.645	-0.143	21333	0.043	-0.074	0.117	31221	0.087	0.365	-0.278
12223	0.204	0.346	-0.142	22111	0.599	0.770	-0.171	31222	0.023	0.297	-0.274
12231	0.257	0.380	-0.123	22112	0.535	0.702	-0.167	31223	-0.041	-0.003	-0.038
12232	0.193	0.311	-0.118	22113	0.237	0.402	-0.165	31231	0.012	0.031	-0.019
12233	0.129	0.012	0.117	22121	0.524	0.708	-0.184	31232	-0.052	-0.037	-0.015
12311	0.385	0.680	-0.295	22122	0.460	0.640	-0.180	31233	-0.116	-0.336	0.220
12312	0.321	0.612	-0.291	22123	0.162	0.340	-0.178	31311	0.140	0.331	-0.191
12313	0.257	0.312	-0.055	22131	0.215	0.374	-0.159	31312	0.076	0.263	-0.187
12321	0.310	0.618	-0.308	22132	0.151	0.306	-0.155	31313	0.012	-0.036	0.048
12322	0.246	0.549	-0.303	22133	0.087	0.006	0.081	31321	0.065	0.269	-0.204
12323	0.182	0.250	-0.068	22211	0.577	0.722	-0.145	31322	0.001	0.201	-0.200
12331	0.235	0.284	-0.049	22212	0.513	0.654	-0.141	31323	-0.063	-0.098	0.035
12332	0.171	0.216	-0.045	22213	0.215	0.355	-0.140	31331	-0.010	-0.065	0.055
12333	0.107	-0.084	0.191	22221	0.502	0.660	-0.158	31332	-0.074	-0.133	0.059
13111	0.312	0.695	-0.383	22222	0.438	0.592	-0.154	31333	-0.138	-0.432	0.294
13112	0.248	0.626	-0.378	22223	0.140	0.292	-0.152	32111	0.301	0.412	-0.111
13113	0.184	0.327	-0.143	22231	0.193	0.326	-0.133	32112	0.237	0.344	-0.107
13121	0.237	0.632	-0.395	22232	0.129	0.258	-0.129	32113	0.173	0.044	0.129
13122	0.173	0.564	-0.391	22233	0.065	-0.041	0.106	32121	0.226	0.350	-0.124
13123	0.109	0.265	-0.156	22311	0.321	0.627	-0.306	32122	0.162	0.282	-0.120
13131	0.162	0.299	-0.137	22312	0.257	0.558	-0.301	32123	0.098	-0.018	0.116
13132	0.098	0.230	-0.132	22313	0.193	0.259	-0.066	32131	0.151	0.016	0.135
13133	0.034	-0.069	0.103	22321	0.246	0.564	-0.318	32132	0.087	-0.052	0.139
13211	0.290	0.647	-0.357	22322	0.182	0.496	-0.314	32133	0.023	-0.352	0.375
13212	0.226	0.579	-0.353	22323	0.118	0.197	-0.079	32211	0.279	0.364	-0.085
13213	0.162	0.279	-0.117	22331	0.171	0.231	-0.060	32212	0.215	0.296	-0.081
13221	0.215	0.585	-0.370	22332	0.107	0.162	-0.055	32213	0.151	-0.003	0.154
32221	0.204	0.302	-0.098								
32222	0.140	0.234	-0.094								
32223	0.076	-0.066	0.142								
32231	0.129	-0.032	0.161								
32232	0.065	-0.100	0.165								

32233	0.001	-0.399	0.400								
32311	0.257	0.269	-0.012								
32312	0.193	0.200	-0.007								
32313	0.129	-0.099	0.228								
32321	0.182	0.206	-0.024								
32322	0.118	0.138	-0.020								
32323	0.054	-0.161	0.215								
32331	0.107	-0.127	0.234								
32332	0.043	-0.196	0.239								
32333	-0.021	-0.495	0.474								
33111	0.184	0.283	-0.099								
33112	0.120	0.215	-0.095								
33113	0.056	-0.084	0.140								
33121	0.109	0.221	-0.112								
33122	0.045	0.153	-0.108								
33123	-0.019	-0.146	0.127								
33131	0.034	-0.113	0.147								
33132	-0.030	-0.181	0.151								
33133	-0.094	-0.480	0.386								
33211	0.162	0.236	-0.074								
33212	0.098	0.167	-0.069								
33213	0.034	-0.132	0.166								
33221	0.087	0.173	-0.086								
33222	0.023	0.105	-0.082								
33223	-0.041	-0.194	0.153								
33231	0.012	-0.160	0.172								
33232	-0.052	-0.229	0.177								
33233	-0.116	-0.528	0.412								
33311	0.140	0.140	0.000								
33312	0.076	0.072	0.004								
33313	0.012	-0.228	0.240								
33321	0.065	0.078	-0.013								
33322	0.001	0.009	-0.008								
33323	-0.063	-0.290	0.227								
33331	-0.010	-0.256	0.246								
33332	-0.074	-0.324	0.250								
33333	-0.138	-0.624	0.486								
Death	[0.000]	[0.000]	[0.000]								
Uncon.	[-0.061]	[-0.293]	[0.232]								

Note. *TTO tariffs obtained from Chapter 1 reported in this dissertation.

Appendix E

Table E. Transformed VAS1 (n = 1170) and TTO tariffs for the EQ-5D health care classification system – absolute values and differences.

state	VAS1	TTO*	difference	state	VAS1	TTO*	difference	state	VAS1	TTO*	difference
11111	1.000	1.000	0.000	13231	0.233	0.251	-0.018	23121	0.245	0.579	-0.334
11112	0.532	0.818	-0.286	13232	0.166	0.183	-0.017	23122	0.177	0.511	-0.334
11113	0.447	0.519	-0.071	13233	0.101	-0.117	0.218	23123	0.112	0.211	-0.099
11121	0.545	0.824	-0.279	13311	0.304	0.551	-0.247	23131	0.187	0.245	-0.059
11122	0.459	0.756	-0.297	13312	0.234	0.483	-0.249	23132	0.121	0.177	-0.056
11123	0.380	0.456	-0.077	13313	0.166	0.183	-0.017	23133	0.058	-0.122	0.181
11131	0.471	0.490	-0.019	13321	0.244	0.489	-0.245	23211	0.256	0.594	-0.338
11132	0.391	0.422	-0.031	13322	0.176	0.421	-0.245	23212	0.188	0.525	-0.338
11133	0.316	0.123	0.193	13323	0.111	0.121	-0.010	23213	0.122	0.226	-0.104
11211	0.559	0.838	-0.279	13331	0.185	0.155	0.030	23221	0.197	0.531	-0.334
11212	0.472	0.770	-0.298	13332	0.120	0.087	0.033	23222	0.131	0.463	-0.332
11213	0.392	0.471	-0.079	13333	0.057	-0.213	0.270	23223	0.068	0.164	-0.096
11221	0.484	0.776	-0.292	21111	0.499	0.833	-0.334	23231	0.141	0.198	-0.057
11222	0.403	0.708	-0.305	21112	0.417	0.765	-0.348	23232	0.077	0.129	-0.053
11223	0.327	0.409	-0.082	21113	0.340	0.465	-0.126	23233	0.015	-0.170	0.185
11231	0.414	0.442	-0.028	21121	0.428	0.771	-0.343	23311	0.208	0.498	-0.290
11232	0.337	0.374	-0.037	21122	0.351	0.703	-0.352	23312	0.141	0.430	-0.288
11233	0.265	0.075	0.190	21123	0.278	0.403	-0.126	23313	0.077	0.130	-0.053
11311	0.497	0.743	-0.246	21131	0.361	0.437	-0.076	23321	0.151	0.436	-0.285
11312	0.415	0.674	-0.260	21132	0.288	0.369	-0.081	23322	0.086	0.367	-0.281
11313	0.338	0.375	-0.037	21133	0.218	0.069	0.149	23323	0.025	0.068	-0.043
11321	0.426	0.680	-0.254	21211	0.440	0.785	-0.345	23331	0.095	0.102	-0.006
11322	0.349	0.612	-0.263	21212	0.362	0.717	-0.355	23332	0.033	0.034	0.000
11323	0.276	0.313	-0.037	21213	0.289	0.418	-0.129	23333	-0.027	-0.266	0.239
11331	0.360	0.347	0.013	21221	0.373	0.723	-0.350	31111	0.386	0.475	-0.089
11332	0.286	0.278	0.008	21222	0.299	0.655	-0.356	31112	0.312	0.407	-0.095
11333	0.216	-0.021	0.238	21223	0.228	0.355	-0.127	31113	0.241	0.107	0.133
12111	0.512	0.823	-0.311	21231	0.309	0.389	-0.080	31121	0.322	0.413	-0.091
12112	0.429	0.755	-0.326	21232	0.238	0.321	-0.083	31122	0.251	0.345	-0.094
12113	0.351	0.456	-0.104	21233	0.171	0.021	0.149	31123	0.182	0.045	0.137
12121	0.440	0.761	-0.321	21311	0.385	0.689	-0.305	31131	0.261	0.079	0.181
12122	0.362	0.693	-0.331	21312	0.310	0.621	-0.311	31132	0.192	0.011	0.181
12123	0.289	0.393	-0.105	21313	0.239	0.322	-0.083	31133	0.126	-0.289	0.415
12131	0.373	0.427	-0.054	21321	0.320	0.627	-0.307	31211	0.333	0.427	-0.094
12132	0.299	0.359	-0.060	21322	0.249	0.559	-0.310	31212	0.261	0.359	-0.098
12133	0.228	0.060	0.169	21323	0.181	0.260	-0.079	31213	0.193	0.060	0.133
12211	0.453	0.776	-0.323	21331	0.259	0.293	-0.034	31221	0.271	0.365	-0.094
12212	0.374	0.707	-0.334	21332	0.190	0.225	-0.035	31222	0.202	0.297	-0.095
12213	0.300	0.408	-0.108	21333	0.125	-0.074	0.199	31223	0.136	-0.003	0.139
12221	0.385	0.713	-0.329	22111	0.398	0.770	-0.372	31231	0.212	0.031	0.181
12222	0.310	0.645	-0.335	22112	0.323	0.702	-0.379	31232	0.146	-0.037	0.182
12223	0.239	0.346	-0.107	22113	0.251	0.402	-0.151	31233	0.082	-0.336	0.418
12231	0.320	0.380	-0.059	22121	0.333	0.708	-0.375	31311	0.282	0.331	-0.049
12232	0.249	0.311	-0.062	22122	0.261	0.640	-0.379	31312	0.213	0.263	-0.051
12233	0.181	0.012	0.169	22123	0.193	0.340	-0.148	31313	0.146	-0.036	0.183
12311	0.397	0.680	-0.283	22131	0.271	0.374	-0.103	31321	0.222	0.269	-0.047
12312	0.321	0.612	-0.290	22132	0.202	0.306	-0.104	31322	0.156	0.201	-0.046
12313	0.250	0.312	-0.062	22133	0.136	0.006	0.130	31323	0.091	-0.098	0.190
12321	0.332	0.618	-0.286	22211	0.345	0.722	-0.378	31331	0.165	-0.065	0.230
12322	0.260	0.549	-0.290	22212	0.272	0.654	-0.382	31332	0.100	-0.133	0.233
12323	0.191	0.250	-0.059	22213	0.203	0.355	-0.152	31333	0.038	-0.432	0.470
12331	0.270	0.284	-0.014	22221	0.282	0.660	-0.378	32111	0.295	0.412	-0.117
12332	0.201	0.216	-0.015	22222	0.213	0.592	-0.379	32112	0.225	0.344	-0.119
12333	0.135	-0.084	0.219	22223	0.146	0.292	-0.146	32113	0.158	0.044	0.113
13111	0.411	0.695	-0.284	22231	0.222	0.326	-0.104	32121	0.235	0.350	-0.116
13112	0.334	0.626	-0.292	22232	0.156	0.258	-0.103	32122	0.167	0.282	-0.115
13113	0.262	0.327	-0.065	22233	0.091	-0.041	0.133	32123	0.102	-0.018	0.120
13121	0.345	0.632	-0.288	22311	0.293	0.627	-0.333	32131	0.177	0.016	0.160
13122	0.272	0.564	-0.292	22312	0.223	0.558	-0.335	32132	0.112	-0.052	0.163
13123	0.203	0.265	-0.062	22313	0.156	0.259	-0.103	32133	0.049	-0.352	0.400
13131	0.282	0.299	-0.016	22321	0.233	0.564	-0.331	32211	0.245	0.364	-0.119
13132	0.213	0.230	-0.018	22322	0.166	0.496	-0.331	32212	0.177	0.296	-0.119
13133	0.146	-0.069	0.215	22323	0.101	0.197	-0.096	32213	0.112	-0.003	0.116
13211	0.356	0.647	-0.290	22331	0.175	0.231	-0.055	32221	0.187	0.302	-0.115
13212	0.283	0.579	-0.296	22332	0.110	0.162	-0.052	32222	0.121	0.234	-0.113
13213	0.213	0.279	-0.066	22333	0.047	-0.137	0.185	32223	0.058	-0.066	0.124
13221	0.293	0.585	-0.291	23111	0.306	0.641	-0.335	32231	0.131	-0.032	0.162
13222	0.223	0.516	-0.293	23112	0.235	0.573	-0.338	32232	0.067	-0.100	0.167
13223	0.156	0.217	-0.061	23113	0.168	0.274	-0.106	32233	0.006	-0.399	0.405
32311	0.197	0.269	-0.071								
32312	0.131	0.200	-0.069								
32313	0.068	-0.099	0.167								
32321	0.141	0.206	-0.066								
32322	0.077	0.138	-0.061								

32323	0.015	-0.161	0.177		
32331	0.086	-0.127	0.213		
32332	0.024	-0.196	0.220		
32333	-0.036	-0.495	0.459		
33111	0.209	0.283	-0.074		
33112	0.143	0.215	-0.073		
33113	0.079	-0.084	0.163		
33121	0.152	0.221	-0.069		
33122	0.088	0.153	-0.065		
33123	0.026	-0.146	0.172		
33131	0.097	-0.113	0.209		
33132	0.035	-0.181	0.215		
33133	-0.026	-0.480	0.455		
33211	0.162	0.236	-0.074		
33212	0.098	0.167	-0.070		
33213	0.035	-0.132	0.167		
33221	0.107	0.173	-0.067		
33222	0.044	0.105	-0.061		
33223	-0.016	-0.194	0.178		
33231	0.053	-0.160	0.213		
33232	-0.008	-0.229	0.221		
33233	-0.067	-0.528	0.461		
33311	0.116	0.140	-0.023		
33312	0.054	0.072	-0.018		
33313	-0.007	-0.228	0.221		
33321	0.062	0.078	-0.015		
33322	0.001	0.009	-0.008		
33323	-0.058	-0.290	0.232		
33331	0.010	-0.256	0.266		
33332	-0.050	-0.324	0.275		
33333	-0.107	-0.624	0.517		
Death	[0.000]	[0.000]	[0.000]		
Uncon.	[-0.033]	[-0.293]	[0.260]		

Note. *TTO tariffs obtained from chapter one reported in this dissertation.