Fair interpersonal utility comparison in the context of health valuation studies: early results of a new multi-step preference aggregation procedure

Paul Schneider*[,1], Ben van Hout[1], John Brazier[1]

[1] *ScHARR, University of Sheffield, UK*

Abstract

Background

The time trade-off (TTO) method is widely used to elicit individual preferences over (EQ-5D 3L, 5L, and Y) health states. Resulting utility values, measured on a scale that is anchored at full health and dead, are commonly aggregated across individuals, in order to derive a social value set. Here, we argue that this aggregation procedure is problematic, because it fails to take into account that TTO utility values are not only determined by an individual's preference for the quality, but also for the quantity of life. As a consequence, when health state utilities are aggregated across different individuals, individuals with a wider range of utility values have – simply for arithmetic reasons – more weight in the estimation of the social value set. Inspired by relative utilitarianism, we propose an alternative aggregation procedure, to equalise individuals' weights in the estimation of social health state values. We apply the alternative procedure to an EQ-5D 3L data set and compare the results against the conventional social tariff.

Methods

We used UK EQ-5D 3L data from the 1993 measurement and valuation of health study. As far as feasible, we replicated the main effects regression model and estimated social tariffs using the conventional and the proposed alternative utility aggregation procedure. Differences between both approaches were assessed by comparing model coefficients, as well as absolute and relative social health state values. In addition, we used linear and polynomial regression models to investigate to what extent individuals' utility ranges are associated with influence on the (relative) social tariff.

Results

2,985 participants were included in the analysis. The alternative aggregation procedure yielded different results than the conventional method: The main effects model coefficients differed between 0% and 8%, and social values were generally higher (mean error = 0.022) in the alternative tariff, and 156 (64.2%) health states had a different (relative) rank. For the conventional method, we found a strong, non-linear relationship between an individuals' utility ranges and their influence on the (relative) social value set (R2=0.18), while for the alternative method, the relationship was much weaker (R2=0.04).

Conclusion

The conventional TTO preference aggregation procedure gives individuals with a wider range of utility values systematically more weight in the estimation of the EQ-5D 3L social tariff. The proposed alternative method can resolve this issue and provides significantly different social health state values.

---

*School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, 30 Regent St, Sheffield S1 4DA, UK. Email: p.schneider@sheffield.ac.uk

1 Introduction

The measurement and valuation of health-related quality of life (HRQOL) is an integral component of health economic evaluation. The methods that are being used to value the changes in the quality and the quantity of life due to health interventions have evolved significantly over the last decades and have now reached a high level of sophistication. However, while much attention has been paid to psychometric and statistical aspects, little consideration was given to their normative underpinnings and implications. In fact, a coherent theoretical basis for the measurement and valuation of health is missing.

In previous work we have argued that the aggregation of health state utilities derived from time trade-off method (TTO) - which is the most widely used method to elicit health state preferences - might yield *unfair* social outcomes. This study aims to extend our theoretical work by testing our hypotheses and the proposed solution empirically.

The remainder of this paper is divided into three main sections. In the next section (2), we outline our theoretical framework, summarise our criticism of the conventional TTO method, and describe an alternative utility aggregation mechanism. In section 3, we report on an empirical application of the proposed method on EQ-5D 3L valuation data from the UK, and assess whether the results confirmed our hypothesis. Finally, empirical and theoretical findings are discussed together in section 4.

Glossary of key concepts

Health state utilities: a cardinal measure of preference, measured on a scale anchored at full health (=1) and dead (=0). May include negative values for states worse than dead.

Social tariff: a function which maps all health states to a social value, which represents the collective preference of a group of individuals.

Relative social tariff: maps health states to relative social values - i.e. the value of one state compared to all other (excluding dead), normalised between the best (=1) and the worst health state (=0).

Utility range: An individual's range of utility values, i.e. the distance between their highest (1, assigned to full health) and the lowest utility (a value between 1 and -1).

Influence: The effect an individual's preference function has on the (relative) social tariff, given the preferences of all other individuals.

2 Theoretical framework

2.1 The conceptual basis of social value sets

HRQOL is a social construct. It cannot be measured (purely) objectively, but requires making subjective value judgements. The EQ-5D and other instruments to measure and value health are therefore based on individual preferences, elicited through TTO exercises.

The TTO method measures health state utilities by identifying points of indifference between shorter periods in full health and usually a fixed period in an ill-health state. The resulting utility values lie on a scale with full health and being dead as anchor points at one and zero respectively. States considered worse than being dead have negative values, which, purely for practical reasons, have a lower bound of minus one (that's achieved by a simple transformation).

When different individuals are asked to state their preferences over a number of health states, there can be considerable disagreement between individuals (is mild pain better or worse than severe mobility impairment?). To determine the social value of a health state, which is supposed to represent the preference of the group of individuals as a whole, some form of compromise must be reached. To this end, utilities are aggregated across individuals, usually by taking the arithmetic average. When cardinal aggregation procedures, such as the average, are applied, utility values of different individuals are implicitly assumed to be directly comparable with each other - averaging over values that are measured on different scales, such as degrees Celsius and degrees Fahrenheit, would simply not be meaningful.

Here, we argue that TTO health state utilities might indeed not be interpersonally comparable. Even though they appear to be measured with a common yardstick, the units of measurement might differ between individuals.

2.2 Interpersonally (in)comparable health state utilities

In a previous working paper, we have discussed the conceptual problems with the TTO utility scale and their implications in more detail (**schneidervanHout**). Here, we confine ourselves to present the main argument. It starts from the premise that there is no scientific way to quantify and compare the intensity of preferences between different individuals. The measurement of health states preferences on a 1-0 scale between full health and dead is thus assumed not to capture utilities (or HRQOL) in some natural units, but rather that it is imposed on normative grounds, as a matter of fairness. The value of one year in full health is set to be 1 for all individuals because everyone's preferences should, in principle, carry equal weight in the determination of the social tariff. Whether some individuals are able to enjoy being in full health more than others are not obviously considered relevant.

The TTO utility aggregation procedure resembles the concept of relative utilitarianism, which normalises every individual's utility scale between their best (=1) and their worst alternative (=0) (Isbell 1959). This method seems intuitively attractive, as it expresses the democratic ideal that each individual is "to count for one and no more than one" (Rawls 2009, p. 283). The approach has also been widely discussed in the fields of social choice theory and welfare economics - see for example Hausman (1995, p. 489), who argues (for different reasons) that relative utilitarianism is the only "correct way of making interpersonal comparisons".

In contrast to relative utilitarianism, however, the TTO method does not assign a zero value to the worst alternative, but to 'being dead' - which we consider to be qualitatively different from a health state. This has considerable implications for the comparability of health state utilities between individuals.

Empirical health valuation studies show that, while some individuals only trade small proportions of their remaining life time for gains in HRQOL, others are willing to give up a large share for relatively minor improvements in their health state (see figure 1 for an illustration of the differences between in-

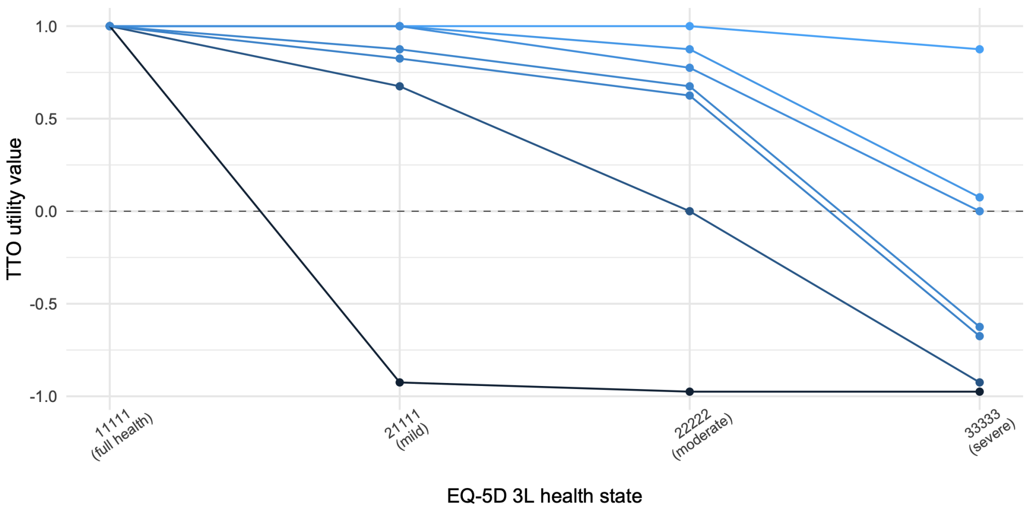

Figure 1: Figure 1: Differences between individuals' utility ranges. Shown are the health state utility values of a selected sample of eighth participants from the MVH study. Each line represents the preferences of one individual over four EQ-5D 3L health states.

dividuals' utility ranges). Moreover, some individuals consider certain health states worse than being dead, while others do not.

As a consequence, the effective range of utility values, that is the distance between the best state and the worst health state, differs considerably between individuals. In fact, utility ranges can vary between zero (i.e. 1 to 1), in those individuals who refuse to trade any survival time, and two (i.e. 1 to -1), in those who consider certain health states much worse than being dead. In general, the more willing an individual is to trade survival time (in full health) for gains in HRQOL, and the more dis-utility they derive from poor health, the wider their utility range. This raises the question whether TTO utility values have the same meaning for different individuals.

2.3 Utility range and influence

Differences in the utility range are problematic, because individuals with a wider utility range have – simply for arithmetic reasons – more leverage in the estimation of the (average) social tariff. More specifically, if social values are being estimated for more than one health state, individuals with a wider utility range have more influence on the relative cardinal ordering of health states. This means, they have disproportionate say over the question how much better or worse being one year in state i is compared to one year in state j.

These differences in the influence between individuals should be considered illegitimate, because they are not caused by genuine differences in the evaluation of the HRQOL, but are due to differences in the preferences for survival time, relative to HRQOL - one may call this the exchange rate between units of quality and units of quantity of life. There does not seem to be any compelling reason, why individuals who are generally more willing to trade survival time (in full health) for gains in HRQOL should have more say in the relative ordering of health states in the social tariff.

To illustrate the point, consider a utility comparison between two individuals: Alice and Bob. Suppose that for Alice, all health states have TTO values between 1 and 0.9, while for Bob, they range between 1 and -1. Because of the wider utility range, Bob has much more say in the relative value set, when utilities are aggregated across individuals to derive a social tariff. We argue that this should be considered unfair, because it cannot be assumed that Alice does not notice much difference between any of the health states, or that her HRQOL in all those states is high. Instead, it seems more plausible that she has a higher rate of substitution between quality and quantity of life: her opportunity costs for giving up lifetime might be higher, for example because she has dependents to care for. It would then be illegitimate to give her less weight in the estimation of a social tariff than Bob.

To avoid this type of bias, we propose an alternative aggregation procedure, which aims to equalise individuals' influences in the estimation of the relative social tariff.

2.4 Model

Inspired by relative utilitarianism, we propose a *Multi-step Utility Aggregation procedure* (MUAP). The procedure consists of four steps, which are outlined below. Subsequently, we provide a simple example to illustrate how the method works. For convenience, we first introduce some basic notations.

Let $H \in \{h_1, h_2, \ldots, h_k\}$ denote a set of k health states, for which social values are to be determined for a group of n individuals. Individual j's health states preferences, denoted $u_j(H) \in \{u_j(h_1), u_j(h_2), \ldots, u_j(h_k)\}$, are measured on the TTO scale, anchored at full health (=1) and dead (=0). By definition, full health has a value of 1, which is everyone's highest utility value ($u(h_{full}) = 1 = \max u(H)$), while the lowest value can take different values with $1 > \min u_j(H) \geq -1$ (we have to define $\min u_j(H) < 1$ to avoid division by zero). Finally, j's utility range is given by $r_j = \max u(H) - \min u_j(H)$.

The conventional aggregation method (CAM)

The conventional TTO health state utility aggregation method (CAM) defines the social tariff $F(.)$ simply as the average utility, as shown below.

$$F(H) = \frac{\sum_{j=1}^{n} u_j(H)}{n} \implies F(h_i) = \frac{\sum_{j=1}^{n} u_j(h_i)}{n}$$

Multi-step utility aggregation procedure (MUAP)

The MUAP is based on the idea that a social tariff consists of two main components, for which a social welfare function must be defined: a relative tariff, which determines the (relative) value of one health states compared to another, and a scaling factor, which anchors the relative tariff on the full-health-dead scale and thereby determines the social rate of substitution between quality and quantity of life.

To derive a relative social tariff, everyone's utilities are normalised between their best (full health) and their worst health state. Normalised preferences and then aggregated by taking the average. The scaling factor is determined by the average utility range. The relative social tariff and the scaling factor are then combined to derive the alternative social tariff $S(.)$. A formal description of the four steps are given below.

The MUAP - Step-by-step instructions

1. **Min-Max Normalisation**
   For every individual $j \in n$, health state utilities are normalised between the highest value ($= 1 = u_j(h_{full})$, by definition) and the lowest value, so that everyone's utilities range from 1 (best) to 0 (worst).

$$u'_j(H) = \frac{u_j(H) - \min u_j(H)}{1 - \min u_j(H)} \quad (1)$$

2. **Relative social tariff**
   Normalised utility values are aggregated across individuals to derive the *relative* social tariff.

$$S'(H) = \frac{\sum_{j=1}^{n} u'_j(H)}{n} \quad (2)$$

3. **Scaling factor**
   Individual utility ranges ($r_j = 1 - \min u_j(H)$) are aggregated across individuals to derive the scaling factor.

$$R = \frac{\sum_{j=1}^{n} r_j}{n} \quad (3)$$

4. **Rescaling**
   Finally, the relative social tariff are re-scaled to the original full-health-dead scale, using the average scaling factor R, and re-anchored at full health with $1 - R$.

$$S(H) = R * S'(H) + 1 - R \quad (4)$$

Notations:
H: set of all health states $h_1, h_2, \ldots, h_k$
$u'_j(H)$: individual j's normalised preferences
$S'(.)$ relative social tariff
$u_j(H)$: individual j's health state preferences
$\min_j(H)$ / $\max_j(H)$: j's lowest/highest utility
$S(.)$ final alternative social tariff.

2.5 A simple example

To demonstrate the conceptual difference between the CAM and the MUAP, consider the example from above (see section 2.3). Suppose *Alice* and *Bob* state their preferences over four health states: h_{full} (full health); h_a (immobile); h_b (pain); h_0 (immobile and pain). Table 1 shows the utility values of both individuals and the corresponding CAM and MUAP social value sets. The results are visually illustrated in figure 2.

It can be seen that Alice and Bob both consider h_{full} the best and h_0 the worst state, yet they have divergent preferences for the other two states: Alice prefers h_a over h_b (0.967 vs 0.933), while Bob prefers h_b over h_a (-

0.333 vs. 0.333). When utilities are aggregated using the CAM, the social tariff indicates that, as a group, the Alice and Bob prefer h_b (0.633) over h_a (0.317). However, Alice's relative preference for h_a (i.e. compared to other health states) is exactly as strong as Bob's relative preference for h_b (0.667 vs 0.667). It is only because Bob has a wider utility range than Alice (2.00 vs 0.1) that he has more influence on the social value set. When the MUAP is applied, influences are re-weighted and, as a group, Alice and Bob become indifferent between h_a and h_b (0.475 and 0.475). The social values for h_{full} and h_0 are the same in both tariffs.

Table 1. Individual and social values

		h^*	h_a	h_b	h_0	r
CAM	u_{Alice}	1	0.967	0.933	0.900	
	u_{Bob}	1	-0.333	0.333	-1.000	
	F	**1**	**0.317**	**0.633**	**-0.050**	
MUAP	u'_{Alice}	1	0.667	0.333	0.00	0.100
	u'_{Bob}	1	0.333	0.667	0.00	2.00
	S' / $[R]$	1	0.500	0.500	0.000	[1.050]
	S	**1**	**0.475**	**0.475**	**-0.050**	

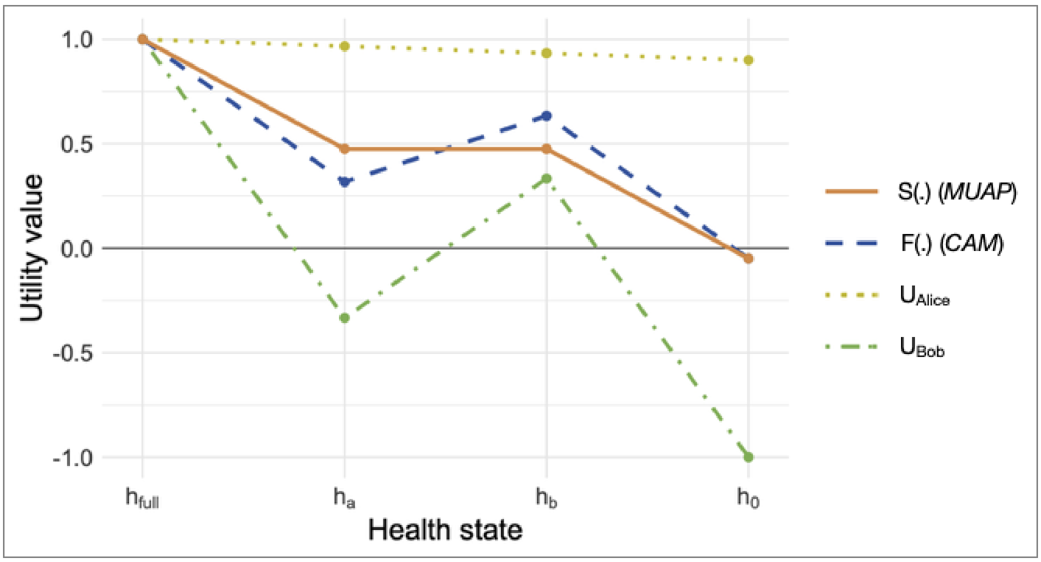

Figure 2: Shown are the two individual preference profiles (Alice = yellow; Bob = green) and the correspodning CAM (blue) and MUAP social value sets (orange).

3 Empirical application

3.1 Objective

In order to test the proposed MUAP empirically, we applied it to an EQ-5D 3L health state valuation data set from the UK [4]. We investigated the following two research question, which immediately follow from the preceding theoretical discussion:

1. What is the impact of using the proposed MUAP, instead of the CAM, to derive a social tariff on the UK social EQ-5D 3L value set?
2. To what extent are individuals' utility ranges associated with individuals' influences on the CAM and MUAP relative UK EQ-5D 3L social value set?

3.2 Methods

3.2.1. The EQ-5D-3L Instrument

The EQ-5D 3L instrument is a generic preference-based instrument for the measurement and valuation of health. It consists of 243 health state, which are defined along five dimensions: mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD). Each dimension has three levels: no problems (1), some problems (2), and severe problems (3). Health states can accordingly be referred to by a 5-digit code. '12321', for example, denotes a state with no mobility problems, some problems with self-care, unable to perform usual activities, some pain/discomfort, and no anxiety/depression; '11111' denotes full health; and '33333' denotes the (objectively) worst health state.

3.2.2. Setting and data

For this study, we used the data set that was used the UK EQ-5D 3L value set from the 1993 Measurement and Valuation of Health (MVH) Study (Dolan 1997; "MVH original data" n.d.; Group et al. 1995). We aimed to reproduce the methods of the original study, but had to make two adjustments:

Firstly, the study sample originally consisted of 2,997 individuals, who were representative of the UK adult population. To avoid zero-division, we had to exclude 12 individuals who assigned a utility of 1 to all states, i.e. who had a utility range of 0.

Secondly, in the MVH, 42 of the 243 EQ-5D-3L health states were valued. Each respondent were asked to complete 12 TTO exercises, which included the worst health state '33333' and 11 other states. To be able to apply the MUAP, and to simplify the comparison of the statistical models for the MUAP and the CAM, we added 'full health' ('1111') with a utility of 1 as a 13^{th} health state for each individual.

3.2.3. Statistical modelling of the CAM and MUAP social tariff

Since not all EQ-5D 3L health states were valued by every individual, a statistical model needs to be fitted to the data to estimate the average utility value for all 243 EQ-5D 3L health states. The CAM tariff is represented by the main effects model proposed by Dolan (Dolan 1997). This well-known OLS regression model contains 12 coefficients, which are (relatively) easy to interpret: There are two dummy variables for each dimension (e.g. MO2 and MO3 for mobility), to represent the move between levels one (no problems) and two (some problems), and between two and three (severe problems). A constant (α) is subtracted for any move away from full health, and N3 represents an additional utility decrement for having severe problems on at least one dimension. The dependent variable was (one minus) the utility given to the health states. By convention, the social value for full health ('11111') is set to 1, even though the statistical models would predict a lower value.

$$1-\hat{y} = \alpha + b_1MO2 + b_2MO3 + b_3SC2 + b_4SC3 + b_5UA2 + b_6UA3 + b_7PD2 + b_8PD3 + b_9AD2 + b_{10}AD3 + b_{11}N3 + \in$$

To represent the alternative MUAP social tariff, we adjusted the main effects model corresponding to the steps described above (section 2.4): First, for each individual, utility values were Min-Max normalised between full health (=1) and the lowest utility value (=0), which could be assigned to '33333' or any other health state. Secondly, to estimate the relative social tariff, the main effects regression model was fitted with (one minus) the normalised health state utility values as the dependent variable (i.e. replacing $1-\hat{y}$ with $1-\hat{y}'$). Thirdly, the scaling factor R was computed as the average individual utility range. Finally, to derive the alternative social tariff, estimated relative social values for all health states were re-scaled, using the scaling factor.

The scaling factor can also be applied to the MUAP model itself, in order to rescale the beta coefficient values. This allows estimating final social health state values directly, and it also enables a comparison of the relative importance of the explanatory variables between the CAM and the MUAP model.

3.2.4. Analysis

Differences between the CAM and the MUAP social value set were investigated in two ways. Firstly, we compared the coefficients of the CAM model with the rescaled coefficients of the MUAP model. Secondly, we compared the absolute social values for all 243 EQ-5D 3L health states, as well as their relative rankings.

Moreover, we assessed differences in individuals' influences on the relative social tariffs in both approaches. Our theoretical framework predicts that, in the CAM, individuals with a wider utility range have more influence on the relative social value set, while in the MUAP, this association is not expected. In the absence of an established measure of 'influence', we defined it as the sum of changes in the social value set that occur due to the inclusion/exclusion of a particular individual. This means, we re-fitted the (CAM and MUAP) main effects models without the individual in question, and assessed to what extent the estimates for the 243 relative social health state values change as a result. The sum of the absolute differences between the estimates with and without the individual in the model were then used as the measure of influence. The relationship between utility ranges and influence was assessed by visual inspection and and tested statistically using linear and polynomial regression models.

3.2.5. Data and code availability

The 1993 MVH dataset is available on the UK Data Service ("MVH original data" n.d.). The principal investigators kindly allowed us to publish the data that were used to generate the results of this study alongside our R source code on an data repository, where it is openly available for reuse and modification [**???**].

3.3 Results

3.3.1. Comparison between CAM and MUAP social tariffs

A total of 2,985 participants were included in the analysis. Table 2 shows the estimates for main effects models using the CAM and the MUAP approach. For most parameters, the differences were small. However, in the rescaled MUAP model, the beta for AD3 was 6% higher compared to the CAM model, and the beta for N3 was 8% lower, respectively.

Table 2. Main effects model - CAM and MUAP parameter estimates

	CAM	MUAP		MUAP/CAM
Variable	b (95%CI)	Normalised b (95%CI)	Rescaled b (95%CI)	Ratio
Intercept	0.05 (0.04; 0.06)	0.03 (0.03; 0.04)	0.05 (0.04-0.06)	1.01
MO2	0.07 (0.06; 0.08)	0.04 (0.04; 0.05)	0.07 (0.06-0.08)	0.99
MO3	0.24 (0.23; 0.25)	0.15 (0.14; 0.16)	0.24 (0.23-0.25)	1.01
SC2	0.12 (0.11; 0.13)	0.07 (0.06; 0.08)	0.11 (0.10-0.12)	0.96
SC3	0.11 (0.09; 0.12)	0.07 (0.06; 0.08)	0.11 (0.10-0.12)	1.05
UA2	0.04 (0.02; 0.05)	0.02 (0.02; 0.03)	0.04 (0.03-0.05)	1.04
UA3	0.06 (0.04; 0.07)	0.03 (0.03; 0.04)	0.06 (0.04-0.07)	1.01
PD2	0.13 (0.12; 0.14)	0.08 (0.07; 0.08)	0.12 (0.11-0.13)	0.96
PD3	0.26 (0.24; 0.27)	0.16 (0.15; 0.17)	0.26 (0.25-0.27)	1.00
AD2	0.08 (0.07; 0.10)	0.05 (0.04; 0.06)	0.08 (0.07-0.09)	0.97
AD3	0.16 (0.15; 0.18)	0.11 (0.10; 0.11)	0.17 (0.16-0.18)	1.06
N3	0.28 (0.27; 0.30)	0.16 (0.15; 0.17)	0.26 (0.25-0.28)	0.92
R^2	0.51	0.59	0.51	
Observations	38,805	38,805	38,805	

We estimated the UK social EQ-5D 3L value set using the CAM and the MUAP approach. The absolute social values and ranks for all 243 EQ-5D 3L health states are provided in table S1 in the appendix. A visual overview and comparison of the two social tariffs as a wholes are provided in Figure 3.

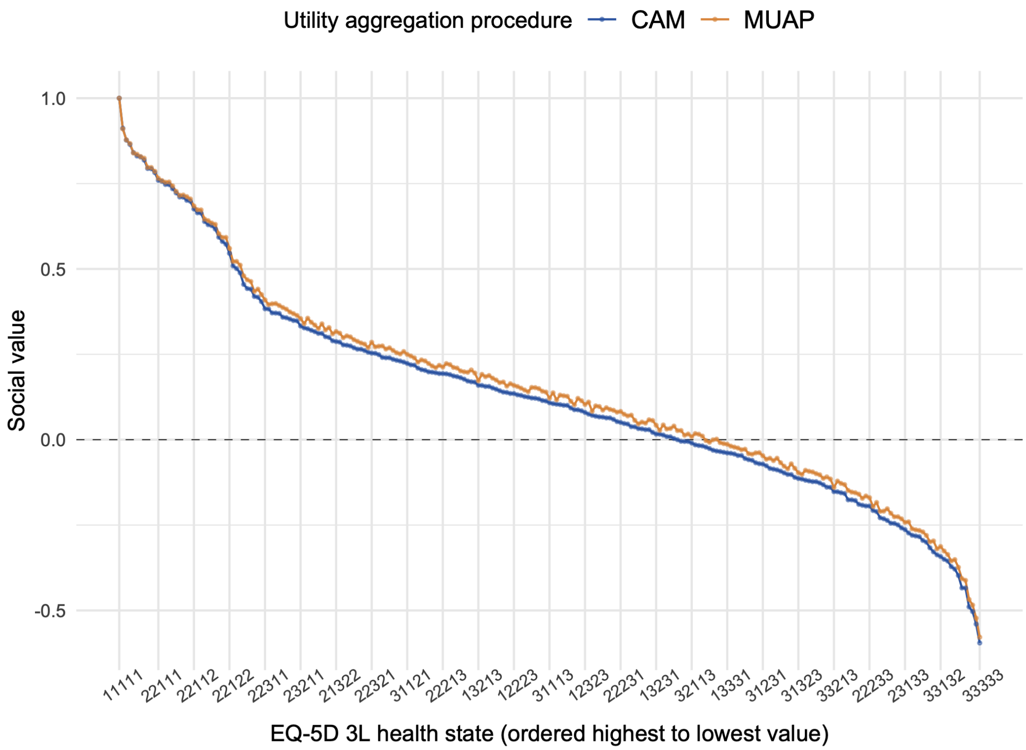

Figure 3: Comparison between the conventional and the alternative UK social EQ-5D 3L tariff. The blue line indicates the CAM social values for all 243 health states (labels are shown for 25), ordered from the highest to the lowest value. The orange line depicts the corresponding MUAP social values.

The MUAP yielded different, and generally higher social values for EQ-5D 3L health states than the CAM: The mean (and mean absolute) difference between social values was 0.022 (SD = 0.008). Differences tended to be smaller for high and low ranked states, and larger for intermediate states. The maximum absolute difference of 0.036 was observed for state '22132' (CAM=0.003; MUAP=0.040). To provide a comparison, in the CAM, the average distance from one EQ-5D 3L health state to the next nearest state (rank i to $i+1$) was 0.007 (SD = 0.010).

Differences in absolute social values often resulted in changes of the ordering of health states. A total of 156 (64.2%) health states had a different (relative) rank in the MUAP tariff. The highest rank change of +6 occurred for health state '31213'.

3.3.2. Utility ranges and influence

Individuals' utility ranges varied between 0.01 and 1.98, with an average (SD) of 1.62 (0.36) and a median (IQR) of 1.73 (0.55). Only 134 (4%) of the participants had a utility range of less than 1, i.e. did not consider any health state to be worse than dead. Interestingly, only 2,205 (74%) participants assigned the lowest utility value to the objectively worst health state ('33333'), while for 780 (26%) participants, at least one health state had a lower value.

Table 3 shows descriptive statistics for the measure of individuals' influences on the relative social tariff using the CAM and the MUAP approach. On average, a single individual had less influence in the MUAP model (0.024 versus 0.021), and measures of dispersion indicated that influences were more evenly distributed across individuals. However, influence in the CAM was highly correlated with influence in the MUAP: The Spearman rank and Pearson correlation coefficients were 0.85 and 0.84, respectively.

Table 3. Dispersion of influence

	CAM	MUAP	$\frac{MUAP}{CAM}$
Mean	0.024	0.021	
Standard dev.	0.011	0.009	0.835
Minimum	0.004	0.004	
Maximum	0.092	0.064	
Range (max-min)	0.088	0.060	0.684
Max/min ratio	22.873	17.892	0.782
Q_{25}	0.016	0.015	
Median Q_{50}	0.022	0.019	
Q_{75}	0.029	0.026	
IQR (Q_{75}-Q_{25})	0.013	0.011	0.835

For both, the CAM and the MUAP, the association between between individuals' utility ranges and their influences on the relative social value set was modeled using linear regression models with linear, quadratic, and cubic terms. Model estimates are provided in table 4 and a visual illustration of the relationships is given in figure 3.

For the CAM, we found a clear, non-linear relationship between utility range an influence on the relative social values of health states. The three-degree polynomial model explained 18% of the variance of the influence between individuals. For the MUAP, beta coefficients were also highly statistically significant (p¡.01) and had a similar pattern as for the CAM model, but the effect was considerably smaller and, overall, individuals' utility ranges explained only a minor proportion of the variance (4%). For both approaches, model coefficients indicate that individuals with a narrow utility range had more influence. However, these estimates are based on a small group of individuals.

Table 4. Association between utility ranges and influence on relative social tariffs - beta (95% CI)

	CAM linear	CAM square	CAM cubic	MUAP linear	MUAP sqaure	MUAP cubic
Intercept	0.02 (0.02; 0.02)	0.02 (0.02; 0.02)	0.02 (0.02; 0.02)	0.02 (0.02; 0.02)	0.02 (0.02; 0.02)	0.02 (0.02; 0.02)
Range	0.15 (0.13; 0.17)	0.15 (0.13; 0.17)	0.15 (0.13; 0.17)	-0.02 (-0.04; -0.01)	-0.02 (-0.04; -0.01)	-0.02 (-0.04; -0.01)
Range^2		0.18 (0.16; 0.20)	0.18 (0.16; 0.20)		0.09 (0.07; 0.11)	0.09 (0.07; 0.11)
Range^3			0.09 (0.07; 0.11)			0.04 (0.02; 0.06)
R^2	0.06	0.16	0.18	0.00	0.04	0.04
N	2,985	2,985	2,985	2,985	2,985	2,985

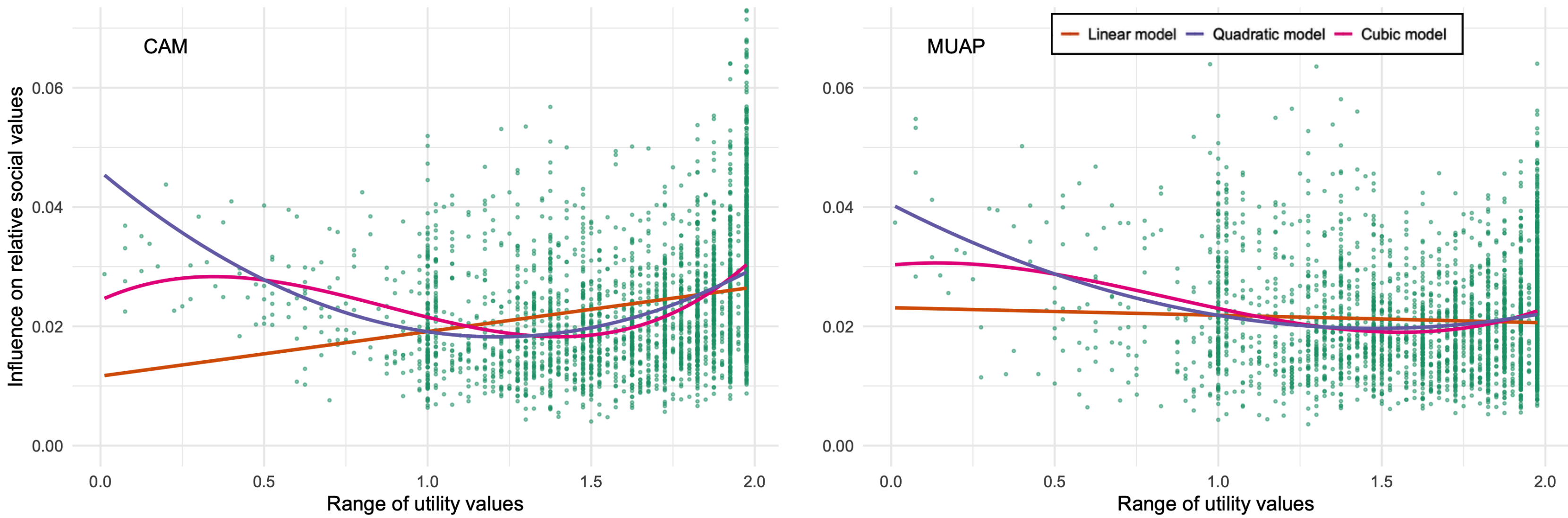

Figure 4: Relationship between individuals' utility ranges and their influences on the relative social tariff for the CAM (left) and the MUAP (right). Each point represents one participant (n=2,985). Lines show the linear (orange), quadratic (purple), and cubic model (magenta) estimates.

4 Discussion

4.1 Main findings

In this paper, we have outlined potential flaws in the procedure that is commonly used to aggregate individual TTO health state utilities into social value sets for use in economic evaluations. Our main theoretical argument is that the conventional method could potentially be unfair, because it does not ensure that every individual has an equal say in the outcome. Instead, social value sets might disproportionally represent the preferences of those individuals who place a low value on survival time, compared to HRQOL.

Using the well-known data set of the MVH study, we found evidence that confirms our hypothesis and showed that individuals' individuals with a wider range of utility values have indeed systematically more weight in the derivation of relative social health states values. The relationship was more complex than expected: not only did individuals with a high utility range had more influence, but also individuals with very low utility ranges. However, only few individuals had such low utility ranges (n = 134; 4%), and the effect is likely due to the unusual preference profiles of these particular individuals, and not a result of the low utility range itself.

Inspired by the concept of relative utilitarianism, we proposed an alternative utility aggregation method, the MUAP, to resolve the problem of inequtiable distribution of influence. The method equalises individuals' weights in the estimation of the relative social tariff, by normalising everyone preferences between their best and worst options. The feasibility of the MUAP was demonstrated on EQ-5D 3L data. We were able to show that the alternative method resolves to a large extent the excess influence of individuals with wide utility ranges. Individuals with very low utility ranges still seem to have disproportionate influence, but, again, this is likely caused by their unusual relative preference profiles.

The UK social EQ-5D 3L tariff derived through the MUAP differed from the tariff derived through the CAM. Health states generally had higher social values in the MUAP and many had a different relative rank (156; 64.2%).

Even though differences between MUAP and CAM social values for any given health state might seem small - even the maximum absolute difference was only 0.036 - in relation to the 1-0 TTO scale, these differences should be considered relevant. If a social EQ-5D 3L tariff derived through the MUAP, instead of the CAM, were used to value the health outcomes in some economic evaluations, aggregate differences between QALY estimates could be significant.

It should be noted that difference between the CAM and the MUAP social tariff were only expected if individuals' utility ranges are associated with certain preferences for dimension weights. For example, individuals with a wider range might put more weight on mobility and less on pain/discomfort. The changes we observed in the beta coefficients between the CAM and the (rescaled) MUAP model provide evidence for the existence of such associations: when the influence of individuals with wide utility ranges was equalised, the weight assigned to severe AD (anxiety/depression) increased by 6%, while the weight for N3 decreased by 8%. Unfortunately, we were unable to further investigate the underlying mechanism, since individual-level preference functions cannot be estimated reliably from the MVH data.

4.2 Further considerations

The MUAP approach links directly to two other recently proposed alternative utility rescaling/re-weighting methods: Jakubczyk and Golicki (2019) suggest to equalise utilities based on 'worst fears'. Their approach seems similar to ours, but they consider 'being dead' a health state (while we do not), so that in their method, utilities are normalised between full health and either 'being dead' or the worst health state, depending on which has a lower utility.

A whole series of alternative measurement scales for HRQOL were suggested by Sampson, Devlin, and Parkin (2018). They argue that dead should not be considered relevant in the derivation of the social value of health states, and, therefore, they reject the use of 'being dead' in the TTO altogether. Instead, they recommend a range of other outcomes, such as being unconscious, the worst health state, or minimum endurable quality of life, which could be used in TTO exercises instead. Unfortunately, the works of Jakubczyk et al. and Sampson et al. have not yet been fully published, so that their approaches cannot be examined in more detail at this point.

Our approach might also be helpful for addressing some conceptual problems in the development of the EQ-5D-Y and other instruments for the valuation of health in children. Empirical studies found that TTO values are higher for health states experienced by a 10-year-old child compared to health states experienced by an adult. At the same time, values derived through the VAS valuation techniques tend to be lower for states experienced by children. These results are difficult to interpret and might even seem paradoxical. Within our theoretical framework, however, they can be easily integrated. The TTO value can be higher in children, even though their HRQOL is judged to be lower, if the rate of substitution between quality and quantity of life is also higher: this means, in children, people might be willing to trade fewer units of survival time (in full health), to gain a unit of HRQOL. Differences in the valuation of health states in children and adults could then be explained by differences in the respective scaling factors. To harmonise the valuation systems in children and adults, one scaling factor would need to be applied to both, or the value of survival time needs to be higher in children, so that of one year in full health is worth 1.3 QALY, for example.

4.3 Limitations

This study has several limitations which should be taken into account when interpreting our findings.

In order to test the MUAP to the EQ-5D 3L data, we had to make some minor adjustments to the methodology that was used in the original MVH study. For the min-max-normalisation to be applicable, it is essential that the TTO value for the best and the worst health state are known for each individual. We assumed that in the context of the EQ-5D 3L descriptive system, it would not be necessary to elicit the preferences for all health states: full health has, by definition, the highest value (=1). The inclusion of full health as an additional health state for each individual might nevertheless be problematic, as it complicates the interpretation of the constant/alpha term in the main effects models, which is usually taken to be any divergence from full health.

For the worst state, we were confident that it was captured in the data set, since all participants valued '33333', which is objectively dominated by all other states. However, 780 (26%) participants assigned the lowest TTO value to one of the other 11 health states they evaluated - which we then used as the worst health state for the normalisation procedure. If we take this result at face value, and accept that individuals may consider another state to be worse than '33333', we would, in principle, have to elicit an individuals' preferences for all states to ensure that the lowest TTO value is observed for everyone.

Finally, it is important to emphasise that our theoretical framework is only concerned with equalising the influences on the *relative* social tariff. To transform relative into absolute values, that is to rescale normalised utilities back to the full health-dead scale, we used a scaling factor. Even though this factor is also undoubtedly a social value – it was defined as the average utility range - it is not entirely clear what properties it has or whether comparisons of utility ranges between individuals are at all permissible.

4.4 Conclusion and outlook

This is the first application of our alternative utility aggregation method. We have outlined the theoretical advantages that the method has over the conventional approach and demonstrated its feasibility using an EQ-5D 3L dataset from the UK. Even though the differences in the resulting social value set appear modest, in practice, the impact on aggregate QALY and subsequent cost-effectiveness estimates could be significant.

Even though several open questions remain, we think our model provides a promising approach to provide more equitable social value estimates. While this paper focuses exclusively on TTO values, our approach should be equally applicable to utilities elicited through the standard gamble method. More research is needed to better understand the practical and normative implications of this and other utility aggregation methods. Those methods should be tested empirically in other data sets and descriptive systems. Ideally, studies should seek to investigate the relationship between individual-level preference functions and the social value set.

In addition, it seems crucial to establish a more sound theoretical foundation for the valuation of health in general. In this context, it is worth mentioning that in Welfare Economics, Social Choice Theory and certain parts of philosophy, the problem of interpersonal utility comparisons appears to have been addressed with much more sophistication than in health economics. In these fields, considerable effort has been devoted to rigorously investigate which types of utility comparisons are permissible, within and between individuals, for different sets of assumptions. For an introduction and overview, readers may refer to Fleurbaey and Hammond (2004) and d'Aspremont and Gevers (2002). We are convinced that research on the valuation of health would benefit from a closer consideration of this extensive body of work.

Acknowledgement

We are sincerely grateful to Koonal Shah, Robert Sugden, Simon McNamara, Nicolas Silva-Illanes and the participants of the 2020 HESG meeting in Newcastle for helpful discussions of the ideas in this paper. This work was supported by the Wellcome Trust Doctoral Training Centre in Public Health Economics and Decision Science (108903/Z/19/Z) and the University of Sheffield through a PhD scholarship. The usual disclaimer applies.

References

d'Aspremont, Claude and Louis Gevers (2002). "Social welfare functionals and interpersonal comparability". In: *Handbook of social choice and welfare* 1, pp. 459–541.

Dolan, Paul (1997). "Modeling valuations for EuroQol health states". In: *Medical care*, pp. 1095–1108.

Fleurbaey, Marc and Peter J Hammond (2004). "Interpersonally comparable utility". In: *Handbook of utility theory*. Springer, pp. 1179–1285.

Group, MVH et al. (1995). "The measurement and valuation of health: Final report on the modelling of valuation tariffs". In: *Centre for Health Economics, University of York*. URL: `https://www.york.ac.uk/media/che/documents/reports/MVH%5C%20Final%5C%20Report.pdf`.

Hausman, Daniel M (1995). "The impossibility of interpersonal utility comparisons". In: *Mind* 104.415, pp. 473–490.

Isbell, JR (1959). *Absolute games, in" Contributions to the Theory of Games, IV"(AW Tucker and RD Luce, Eds.)*

Jakubczyk, M K and D Golicki (2019). "What if 0 is Not Equal to 0? Interpersonal Utility Anchoring Using the Worst Fears". In: *Patient 12: 429. https://doi.org/10.1007/s40271-019-00368-9*.

"MVH original data" (n.d.). In: (). URL: `https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=3444`.

Rawls, John (2009). *A theory of justice*. Harvard university press.

Sampson, Christopher J, Nancy Devlin, and David W Parkin (2018). "Drop dead: is anchoring at 'dead' a theoretical requirement in health state valuation?" In: *Conference: 35th EuroQol Group Scientific Plenary*.

Appendix

Table S1: Full UK social EQ-5D 3L value sets for the CAM and the MUAP.

Rank	Health state	CAM Value	MUAP Value	Difference Abs. (rank)	Rank	Health state	CAM Value	MUAP Value	Difference Abs. (rank)
1	11111	1.000	1.000	0.00 (-)	62	21322	0.287	0.317	0.03 (-1)
2	11211	0.912	0.910	-0.00 (-)	63	23112	0.286	0.312	0.03 (-1)
3	21111	0.877	0.878	0.00 (-)	64	23311	0.278	0.299	0.02 (+2)
4	11112	0.865	0.867	0.00 (-)	65	11131	0.277	0.304	0.03 (-1)
5	21211	0.841	0.840	-0.00 (-)	66	13221	0.274	0.301	0.03 (-1)
6	12111	0.831	0.836	0.00 (-)	67	31112	0.269	0.293	0.02 (-)
7	11212	0.828	0.829	0.00 (-)	68	13312	0.265	0.288	0.02 (-)
8	11121	0.819	0.824	0.01 (-)	69	12213	0.265	0.283	0.02 (+1)
9	12211	0.794	0.797	0.00 (-)	70	31311	0.261	0.280	0.02 (+1)
10	21112	0.793	0.797	0.00 (-)	71	21313	0.256	0.270	0.01 (+4)
11	11221	0.782	0.786	0.00 (-)	72	22321	0.253	0.285	0.03 (-3)
12	22111	0.759	0.765	0.01 (-)	73	11223	0.253	0.272	0.02 (+1)
13	21212	0.757	0.759	0.00 (-)	74	23212	0.249	0.274	0.02 (-1)
14	21121	0.747	0.754	0.01 (+1)	75	12322	0.241	0.274	0.03 (-3)
15	12112	0.747	0.755	0.01 (-1)	76	11231	0.240	0.266	0.03 (+1)
16	11122	0.735	0.743	0.01 (-)	77	23121	0.240	0.269	0.03 (-1)
17	22211	0.723	0.727	0.00 (-)	78	32111	0.235	0.261	0.03 (-)
18	21221	0.711	0.716	0.00 (+1)	79	31212	0.233	0.255	0.02 (+1)
19	12212	0.710	0.716	0.01 (-1)	80	22113	0.230	0.251	0.02 (+1)
20	12121	0.701	0.712	0.01 (-)	81	13122	0.227	0.258	0.03 (-2)
21	11222	0.698	0.705	0.01 (-)	82	31121	0.223	0.250	0.03 (-)
22	22112	0.676	0.684	0.01 (-)	83	13321	0.219	0.245	0.03 (-)
23	12221	0.664	0.673	0.01 (-)	84	21123	0.218	0.240	0.02 (-)
24	21122	0.664	0.673	0.01 (-)	85	12313	0.210	0.227	0.02 (+2)
25	22212	0.639	0.646	0.01 (-)	86	21131	0.205	0.234	0.03 (-1)
26	22121	0.630	0.641	0.01 (-)	87	23221	0.203	0.231	0.03 (-1)
27	21222	0.627	0.635	0.01 (-)	88	32211	0.199	0.223	0.02 (-)
28	12122	0.617	0.631	0.01 (-)	89	11323	0.198	0.216	0.02 (+3)
29	22221	0.593	0.603	0.01 (-)	90	13113	0.196	0.211	0.01 (+5)
30	12222	0.580	0.592	0.01 (+1)	91	23312	0.194	0.218	0.02 (-)
31	11311	0.572	0.592	0.02 (-1)	92	22213	0.194	0.213	0.02 (+1)
32	22122	0.546	0.560	0.01 (-)	93	11132	0.193	0.223	0.03 (-4)
33	22222	0.509	0.522	0.01 (+1)	94	13222	0.191	0.220	0.03 (-4)
34	21311	0.501	0.522	0.02 (-1)	95	31221	0.187	0.212	0.03 (-1)
35	11312	0.489	0.511	0.02 (-)	96	11331	0.185	0.210	0.03 (-)
36	12311	0.455	0.480	0.02 (-)	97	21223	0.182	0.201	0.02 (+1)
37	11321	0.443	0.468	0.03 (-)	98	31312	0.177	0.199	0.02 (+1)
38	13111	0.441	0.463	0.02 (-)	99	12123	0.172	0.197	0.03 (+1)
39	11113	0.419	0.434	0.01 (+1)	100	22322	0.170	0.204	0.03 (-3)
40	21312	0.417	0.441	0.02 (-1)	101	21231	0.168	0.195	0.03 (-)
41	13211	0.404	0.425	0.02 (-)	102	13213	0.159	0.173	0.01 (+5)
42	22311	0.383	0.409	0.03 (-)	103	12131	0.159	0.191	0.03 (-1)
43	11213	0.383	0.396	0.01 (+2)	104	11232	0.156	0.184	0.03 (-)
44	21321	0.371	0.398	0.03 (-)	105	23122	0.156	0.188	0.03 (-2)
45	12312	0.371	0.399	0.03 (-2)	106	32112	0.151	0.180	0.03 (-1)
46	23111	0.370	0.393	0.02 (-)	107	23321	0.148	0.175	0.03 (-1)
47	11322	0.359	0.387	0.03 (-)	108	32311	0.143	0.167	0.02 (+1)
48	13112	0.357	0.382	0.03 (-)	109	31122	0.139	0.169	0.03 (-1)
49	31111	0.353	0.374	0.02 (-)	110	22313	0.138	0.157	0.02 (+2)
50	13311	0.349	0.369	0.02 (-)	111	13322	0.135	0.164	0.03 (-1)
51	21113	0.348	0.364	0.02 (-)	112	12223	0.135	0.159	0.02 (-1)
52	23211	0.333	0.355	0.02 (+1)	113	31321	0.131	0.156	0.02 (-)
53	11313	0.327	0.340	0.01 (+2)	114	33111	0.130	0.151	0.02 (+2)
54	12321	0.325	0.356	0.03 (-2)	115	21323	0.126	0.145	0.02 (+3)
55	13212	0.320	0.344	0.02 (-1)	116	23113	0.125	0.141	0.02 (+4)
56	31211	0.316	0.336	0.02 (+1)	117	12231	0.122	0.153	0.03 (-3)
57	21213	0.311	0.325	0.01 (+2)	118	21132	0.121	0.152	0.03 (-3)
58	13121	0.311	0.339	0.03 (-2)	119	23222	0.119	0.149	0.03 (-2)
59	12113	0.302	0.321	0.02 (+1)	120	32212	0.115	0.142	0.03 (-1)
60	22312	0.299	0.328	0.03 (-2)	121	21331	0.113	0.139	0.03 (-)
61	11123	0.290	0.310	0.02 (+2)	122	31113	0.108	0.122	0.01 (+4)

Table S1: Full UK social EQ-5D 3L value sets for the CAM and the MUAP (continued).

Rank	Health state	CAM Value	MUAP Value	Difference Abs. (rank)	Rank	Health state	CAM Value	MUAP Value	Difference Abs. (rank)
123	32121	0.106	0.137	0.03 (-1)	184	33122	-0.084	-0.054	0.03 (-)
124	13313	0.104	0.117	0.01 (+4)	185	12133	-0.086	-0.061	0.03 (+1)
125	31222	0.103	0.130	0.03 (-2)	186	22332	-0.088	-0.054	0.03 (-3)
126	11332	0.101	0.129	0.03 (-2)	187	33321	-0.092	-0.067	0.02 (-)
127	22123	0.100	0.127	0.03 (-2)	188	23323	-0.097	-0.078	0.02 (+1)
128	33211	0.093	0.113	0.02 (+2)	189	32313	-0.102	-0.085	0.02 (+2)
129	23213	0.088	0.102	0.01 (+4)	190	23132	-0.102	-0.071	0.03 (-2)
130	22131	0.087	0.121	0.03 (-3)	191	23331	-0.110	-0.084	0.03 (-1)
131	21232	0.085	0.114	0.03 (-2)	192	31323	-0.114	-0.097	0.02 (+3)
132	12323	0.080	0.103	0.02 (-)	193	33113	-0.115	-0.102	0.01 (+4)
133	12132	0.075	0.110	0.04 (-2)	194	31132	-0.119	-0.090	0.03 (-2)
134	31213	0.071	0.083	0.01 (+6)	195	33222	-0.121	-0.093	0.03 (-2)
135	32221	0.069	0.099	0.03 (-1)	196	13332	-0.123	-0.095	0.03 (-2)
136	12331	0.067	0.097	0.03 (-1)	197	12233	-0.123	-0.099	0.02 (-1)
137	13123	0.066	0.087	0.02 (+1)	198	31331	-0.127	-0.103	0.02 (-)
138	23322	0.064	0.094	0.03 (-2)	199	21333	-0.132	-0.113	0.02 (+1)
139	22223	0.064	0.089	0.03 (-2)	200	23232	-0.139	-0.109	0.03 (-1)
140	32312	0.059	0.086	0.03 (-1)	201	32123	-0.140	-0.115	0.02 (-)
141	13131	0.053	0.081	0.03 (+1)	202	33213	-0.152	-0.140	0.01 (+3)
142	22231	0.051	0.083	0.03 (-1)	203	32131	-0.153	-0.121	0.03 (-1)
143	31322	0.047	0.075	0.03 (-)	204	31232	-0.155	-0.128	0.03 (-1)
144	33112	0.046	0.070	0.02 (+1)	205	22133	-0.158	-0.131	0.03 (-1)
145	12232	0.038	0.072	0.03 (-1)	206	33322	-0.176	-0.149	0.03 (-)
146	33311	0.038	0.057	0.02 (+1)	207	32223	-0.176	-0.153	0.02 (-)
147	23313	0.033	0.046	0.01 (+4)	208	12333	-0.178	-0.155	0.02 (-)
148	11133	0.032	0.051	0.02 (+1)	209	32231	-0.189	-0.160	0.03 (-)
149	13223	0.029	0.049	0.02 (+1)	210	13133	-0.192	-0.172	0.02 (+2)
150	21332	0.029	0.058	0.03 (-4)	211	23332	-0.194	-0.165	0.03 (-1)
151	32122	0.022	0.056	0.03 (-3)	212	22233	-0.194	-0.170	0.02 (-1)
152	13231	0.016	0.042	0.03 (+1)	213	33313	-0.207	-0.196	0.01 (+1)
153	31313	0.016	0.027	0.01 (+4)	214	31332	-0.211	-0.184	0.03 (-1)
154	32321	0.014	0.043	0.03 (-2)	215	13233	-0.229	-0.210	0.02 (+2)
155	33212	0.009	0.031	0.02 (+1)	216	32323	-0.232	-0.209	0.02 (-)
156	22323	0.008	0.033	0.02 (-1)	217	32132	-0.236	-0.202	0.03 (-2)
157	22132	0.003	0.040	0.04 (-3)	218	32331	-0.245	-0.215	0.03 (-)
158	33121	-0.000	0.027	0.03 (+1)	219	33123	-0.245	-0.226	0.02 (+1)
159	22331	-0.005	0.027	0.03 (-1)	220	22333	-0.250	-0.226	0.02 (-1)
160	11233	-0.005	0.013	0.02 (+3)	221	33131	-0.258	-0.232	0.03 (-)
161	23123	-0.005	0.017	0.02 (-)	222	23133	-0.263	-0.242	0.02 (+1)
162	32113	-0.010	0.009	0.02 (+3)	223	32232	-0.273	-0.241	0.03 (-1)
163	32222	-0.015	0.018	0.03 (-3)	224	31133	-0.280	-0.261	0.02 (-)
164	12332	-0.017	0.016	0.03 (-2)	225	33223	-0.282	-0.264	0.02 (-)
165	23131	-0.018	0.010	0.03 (-1)	226	13333	-0.284	-0.266	0.02 (-)
166	31123	-0.022	-0.003	0.02 (+2)	227	33231	-0.295	-0.270	0.02 (-)
167	13323	-0.026	-0.007	0.02 (+2)	228	23233	-0.300	-0.280	0.02 (-)
168	13132	-0.031	-0.000	0.03 (-1)	229	31233	-0.316	-0.299	0.02 (+1)
169	22232	-0.033	0.002	0.03 (-3)	230	32332	-0.328	-0.297	0.03 (-1)
170	31131	-0.035	-0.009	0.03 (-)	231	33323	-0.337	-0.320	0.02 (+1)
171	33221	-0.037	-0.012	0.03 (-)	232	33132	-0.342	-0.313	0.03 (-1)
172	13331	-0.039	-0.013	0.03 (-)	233	33331	-0.350	-0.326	0.02 (-)
173	21133	-0.040	-0.019	0.02 (-)	234	23333	-0.355	-0.336	0.02 (-)
174	23223	-0.042	-0.022	0.02 (-)	235	31333	-0.372	-0.355	0.02 (+1)
175	33312	-0.046	-0.024	0.02 (-)	236	33232	-0.379	-0.351	0.03 (-1)
176	32213	-0.046	-0.029	0.02 (+1)	237	32133	-0.398	-0.374	0.02 (-)
177	23231	-0.055	-0.028	0.03 (-1)	238	33332	-0.434	-0.407	0.03 (-)
178	31223	-0.058	-0.041	0.02 (+2)	239	32233	-0.434	-0.412	0.02 (-)
179	11333	-0.060	-0.043	0.02 (+2)	240	32333	-0.490	-0.468	0.02 (-)
180	13232	-0.068	-0.039	0.03 (-1)	241	33133	-0.503	-0.484	0.02 (-)
181	32322	-0.070	-0.038	0.03 (-3)	242	33233	-0.540	-0.522	0.02 (-)
182	31231	-0.071	-0.047	0.02 (-)	243	33333	-0.595	-0.578	0.02 (-)
183	21233	-0.077	-0.057	0.02 (+2)					