Interpersonal comparability of health state utilities: why it is unfair to measure preferences in units of full-health-time, and what we can do about it

Paul Schneider[1*]

[1] *School of Health and Related Research (ScHARR),*
University of Sheffield, UK

Abstract

Individual health state preferences are commonly measured using the time trade-off (TTO) method. Elicited health state utilities are then aggregated across individuals to derive a social value set, which is used in economic evaluations to value health outcomes in terms of QALYs. Underlying this procedure is the notion of fairness: the value of one year in full health and the value of being dead is the same for everyone, which presumably aims to ensure equality between individuals. On closer inspection, however, the current methodology fails to make utilities interpersonally comparable.

This work is motivated by the observation that in empirical health valuation studies utility differences between individuals are mainly driven by the their willingness to trade survival time for quality of life: while some refuse to give up any lifetime (non-traders), others consider time with slightly impaired health not worth living (high-traders). When utilities are aggregated across individuals, high-traders have significantly more influence on the resulting cardinal ordering of health states; or put more generally, when utilities are measured in units of *full-health-time*, health state preferences are contaminated by preferences over survival time, and vice versa.

Here, we argue that this property of health state utility comparisons is neither necessary nor desirable. We demonstrate that, it is not sufficient to assert that the utility difference between one year in full health and being dead is the same for all individuals; to be able to make utilities in all respects interpersonally comparable, it is also required that the utility difference between full health and the worst possible health state is the same for all. We propose a simple, multi-step procedure for enabling a fair comparison of health state utilities across individuals, while retaining the properties of the QALY scale, where 1 is full health and 0 is the equivalent to being dead.

*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 concept of health-related quality of life (hrqol) is central to health economic evaluations [1, 2]. It is taken to be a measure of the social value of being in a given health state. Combined with (survival) time, it allows the computation of QALYs, which are widely used to inform health policy decision-making [3, 4].

Hrqol is commonly assessed using the time trade-off method (TTO) [5]. The widely used EQ-5D 3L and 5L instrument are partly or entirely based on it [6–8]. The general procedure is the following: a group of individuals (e.g. patients, general public) is surveyed and their preferences over health states are elicited using the TTO method. Subsequently, health state utilities are aggregated across individuals, in order to derive a social value set, which maps health states to (social) hrqol values. The method of choice to aggregate utilities is the arithmetic average [9, 10]. For this, and other cardinal aggregation procedures, to yield fair social value sets, utilities need to be measured on the same scale; they need to be fully interpersonally comparable [11, 12]. To avoid confusion, we should state that this paper is not concerned with the comparability of QALYs, but of the utility values that determine how many QALYs any given health states is worth [13, 14].

In this paper, we challenge the often uncritically accepted assumption that TTO health states utilities are comparable across different individuals [15]. We argue that when 'being dead' is conceptualised as a state with a utility of zero, the elicited utility values do not only reflect an individual's assessment of the hrqol of health states, but also of the individual's preference for survival time [16–18]. This causes the measurement scale for hrqol to differ between individuals, which, in turn, makes TTO health state utilities interpersonally incomparable. Aggregating them may yield inequitable social value sets. We thus propose a simple multi-step aggregation procedure, in which every individual's preference has equal weight in the outcome.

The remainder of this paper is structured as follows: First, we describe the general concept of TTO and our perspective on the utility aggregation procedure (section 2). Subsequently, we outline conceptual flaws in the current TTO methodology and demonstrate how these create interpersonally incomparable utility values (section 3). In section 4, we discuss the role of being dead for the measurement of hrqol, before we finally propose an alternative multi-step aggregation procedure (section 5) and discuss some further considerations and the next steps (section 6).

Glossar

Health state utilities: Utility of living in some state, measured on the TTO scale, in units of full-health-time. May include evaluations of aspects other than health.

Health-related quality of life (hrqol): The part of the health state utility that is attributable to the health component

Relative value of health states: The value of a health state, relative to another, expressed as a multiple or a fraction, i.e. measured on a hrqol ratio scale

2 The time trade-off method

2.1 Elicitation of individual health state preferences

TTO is a choice-based method to elicit individual health state preferences under certainty. While the specific methodology varies between different protocols and applications [5, 6, 8, 19], the following description may outline the general concept (also see figure 1).

In a TTO exercise, individuals are asked to choose between living t years in some health state h_i (and then be dead), and $t - k$ (with $k \geq 0$) years in a state of full health, denoted $h*$ (and then be dead). By definition, a value of 1 is assigned to full health. The value of any other state is deter-mined by guiding the individual through a series of choices, in which the value of k is adaptively increased or decreased, until a point is identified, at which the individual is indifferent between the two alternatives. The value of $\frac{(t-k)}{t}$ at the point of indifference reflects the proportion of the utilities that would be derived from being in full-health for the same amount of time – which then interpreted as the hrqol of state h_i. For example, being indifferent between 10 years in state h_i (moderate pain) and 6 years in $h*$ (full health), would

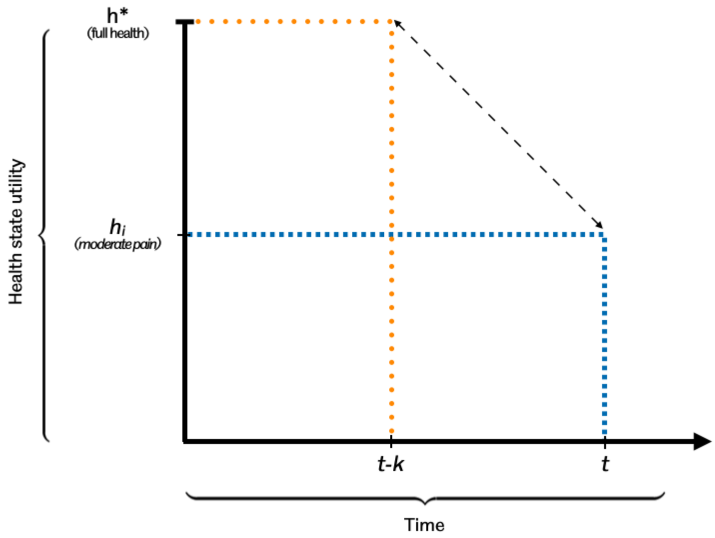

Figure 1: Time trade-off method for states better than dead (adapted from Torrance[21]).

Equality at the point of indifference between t years in state h_i and $t-k$ years in full health ($h*$): $u(h^*) * t - k = p(h_i) * t$ With constant proportional trade-off and $u(h^*) = 1$, the utility of state h_i is: $u(h_i) = \frac{t-k}{t}$

yield a utility of $\frac{10-4}{10} = 0.6.$, i.e. for a given amount of time, being in state h_i is assumed to provide 60% of the utility of being in full health.

When the point of indifference is at $k = t$, the state is taken to be equal to 'being dead'. If the individual is willing to give up additional full health time to avoid being in h_i, the state is assumed to be worse than dead and gets assigned a negative value – however, the way in which negative values are elicited differs between protocols [19]. Furthermore, consistent valuation rests on the constant proportional trade-off assumption, which means that individuals trade the same pro-portion of survival time to gain hrqol, irrespective of the absolute amount of time [20].

2.2 Interpersonal health state utility comparisons

When health state preferences are elicited from a group of individuals, there can be reasonable disagreement about whether some health state h_i is better or worse than another health state h_j, not least about the exact utility values. To determine the health state preferences of the group as a whole, TTO utilities are aggregated, usually by taking the average, into a set of social hrqol values – one single value for each health state.

Cardinal aggregation procedures, such as the arithmetic average, can only be applied when utilities are fully interpersonally comparable [11, 12]. Analogous to the in-comparability of degrees Celsius and degrees Fahrenheit, the aggregation of utility values that are measured on different scales would not be meaningful [22]. The TTO method aims to achieve interpersonal comparability by anchoring the utility scale at full health (=1) and 'being dead' (=0). For these two states it is defined that preference levels are the same for every single individual. The values for all other health states are determined by making relative comparisons with respect to these two, so that by implication, a one-unit change means the same for all individuals.

Supposedly, the rationale for the normalisation of utilities between 1 and 0 is not that different individuals are actually assumed to experience full health in exactly the same way. Rather, it is posited on normative grounds, as a matter of fairness. Whether some individuals are able to enjoy being in full health more or less should play no role in the social evaluation of health states. Instead, all individuals should carry equal weight [12, 23]. The TTO method differs in this regard from other preference elicitation methods, such as Willingness-to-Pay, in which preference intensities can systematically vary be-tween individuals, based on their ability to pay, for example, so that the preferences of some individuals have more weight than the preferences of others.

3 (Un)fair utility comparisons

3.1 Distortions in the hrqol measurement scale

Having outlined that we believe the underlying principle of interpersonal health state utility is, or at least should be, equality, we now proceed to demonstrate that the current methodology fails to achieve it.

We should begin to lay out our argument by noting that health outcomes can be evaluated along two dimensions: survival time and hrqol. Even though the QALY combines both into a uni-dimensional measure, conceptually, they refer to different aspects. The question how valuable it is to go from state h_i to state h_j only requires an evaluation of the hrqol dimension – an evaluation of the survival time is not necessary. The question how valuable is it to gain one additional year of life in in state h_i includes an evaluation of the state's hrqol, but it may also involve the evaluation of many other aspects that have nothing or little to do with health (meaning, family, contentment, etc). As a result, individuals can derive positive utilities from living in state h_i, even if they derive no utilities from the state of their health and even if it caused severe suffering. Making choices in TTO exercises, between living t years in state h_i and $t - k$ years in h_j, therefore, require an additional, third type of evaluation, namely about the **R**ate of **S**ubstitution between health-related **Q**uality and **Q**uantity of life (RSQQ): 'how valuable is a unit change in hrqol compared to a unit change in survival time?'.

The relevance of these conceptual distinctions become apparent in light of the results from empirical research [16–18]. An analysis of preference data from the UK EQ-5D 3L health valuation study [14, 24] suggests that the RSQQ greatly differs across individuals: while some individuals refuse to give up any time for gains in hrqol, others seem to be willing to sacrifice a large proportion of their remaining life years for relatively minor improvements; and, overall, only very few individuals consider the worst health state to have a value of zero. The differences in the RSQQ actually explain a considerable proportion of the variability in utility values between individuals, whereas the relative order of health states is much more consistent. Figure 3.1 illustrates the respective phenomenon in a sample of nine participants.

Since the TTO method measures utilities in units of full-health-time equivalents, which include both, hrqol and quantity of life, it is impossible to infer the RSQQ from a single value. A value of 0.9 for state h_i, for example, could mean that the individual thinks the hrqol of that state is high; but it could also mean that the hrqol is low, and the individual is just unwilling to give up much lifetime to gain hrqol; or it might be something in-between. However, if we consider the full health states preferences of two individuals and find that for one individual, all utilities range between 1 and 0.9, and for the other, they range between 1 and 0, it appears unlikely that individual one is unable to notice much differences in the health component of these states. It seems rather more plausible that even though the hrqol of some states is

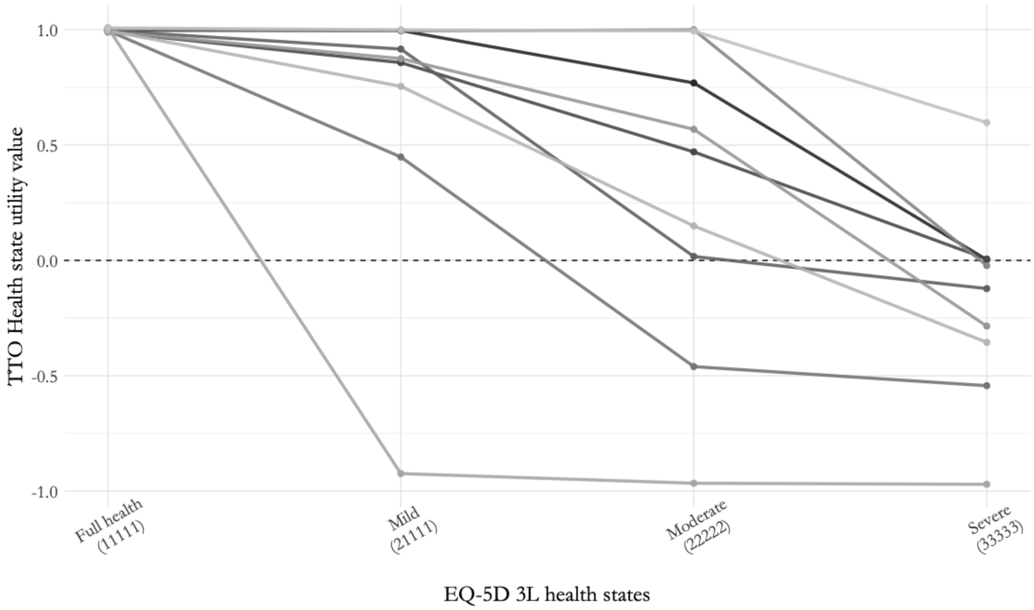

Figure 2: EQ-5D 3L preferences profiles of four individuals with different quality-quantity substitution rates. The figure shows the TTO preferences of nine individuals over four EQ-5D 3L health states. Each line represents the preference profile of one individual. The health states have a hierarchical order and are correspondingly labeled as full heath (EQ-5D 3L code '11111'); minor health problems ('21111'); medium health problems ('22222'); and worst health state ('33333').

very low for individual one, it is not worth sacrificing many years of life for them.

3.2 Implications for interpersonal utility comparisons

The problem with the TTO method is that differences in the RSQQ are not captured as such. Since utilities are measured in units of full-health-time, with 'being dead' as an anchor point at zero – not only for the time, but also for the hrqol dimension – the RSQQ is fixed at 1:1. Consequently, differences in the RSQQ can only be (mis)interpreted as differences in hrqol values. This distorts the measurement scale and makes utility values interpersonally incomparable. When utility values are nevertheless aggregated across different individuals, the weight of an individual's hrqol 'measurements' in the estimation of the social hrqol value set depends on their RSQQ. This is the case, because the RSQQ determines the effective range of utility values: Individuals who place a (relatively) higher value on survival time (e.g. because they value other aspects than health in life), have, ceteris paribus, higher values with a narrower range than individuals who place a (relatively) higher value on hrqol. Yet, individuals with a wider range of utility values have – simply for arithmetic reasons – more leverage, i.e. they have more influence on the (relative) social hrqol values of health states.

When negative health state utilities are used for states considered worse than dead, the problem is further aggravated, since the effective range of values is no longer bound between 1 and 0. Moreover, the nature of negative utilities is different from their positive counterparts, in that they are not measured as a proportion of the utilities derived from full health. Rather, they reflect actual preference intensities and thereby defy the rules of utility normalisation. Although the range of negative values is often restricted, either by the experimental setup or by (arbitrarily) rescaling values to a minimum of -1, conceptually, negative values have no lower limit [25]. Depending on the individual's ability to derive disutility from poor health, negative utilities may take values up to minus infinity. Besides the conceptual inconsistency that negative values introduce into the valuation of health, we would argue that the wide, potentially infinite, range of utility values of individuals with negative utilities pose a serious problem for the interpersonal comparability of health state utilities [25, 26].

We argue that there is no theoretical justification for why an individual's RSQQ, or their ability to derive negative utilities from poor health, should be taken into account in the derivation of (relative) social hrqol values of health states. Doing so violates the notion that everyone's preferences should be equally important. To make interpersonal utility comparisons fairer, we propose an alternative aggregation procedure, which overcomes this problem, by (temporarily) normalising individuals' utilities between full health and the worst health state. However, before we go on to outline our method in more detail, it will be necessary to demonstrate that such alternative normalisation procedure is legitimate, which means to demonstrate that 'being dead' is – contrary to the conventional understanding – not a natural zero point on the hrqol measurement scale.

4 The role of being dead in the valuation of health

If ‘being dead’ were a health state with an absolute hrqol value of zero, our proposal had little substance. The normalisation between full health and dead would then provide a valid yardstick to measure hrqol, and the concept of a RSQQ would be futile. However, there are several arguments for rejecting this conception.

First of all we want to make an ontological argument by noting that ‘being dead’ is an oxymoron in it-self. To ‘be’ dead actually means nothing less than to be not. When an individual dies, they cease to exist, so that it might just not bear any meaning to say ‘someone is dead for one year’. The question how many utilities are derived from this state of ’not being’ during this period seem equally pointless.

Suppose dead were a state of being, it still does not follow that dead should also be considered a health state. A move from one health state to another usually involves changes in the health status, which may or may not have impact on other domains. Moving from some health state to dead, however, involves changes in literally all aspects of life., including relationships, income, and, of course, being alive itself [16–18]. These additional aspects, and the valuation thereof, do not seem relevant for the measurement of hrqol [27]. Comparing health states with ‘being dead’ does, therefore, provide no useful information about hrqol, but only about the relative value of changes in health compared to changes in survival time.

Even if ‘being dead’ were a health state, there is no conclusive argument for why every individual should assign a hrqol value of zero to it. Although it is common practice to assume this were the case, a theoretical foundation for this is missing [27–29]. Notwithstanding, only recently it was proposed that being dead were a ‘natural’ zero point on the ratio scale on which hrqol is measured [28]. It is not within the scope of this paper to provide a full technical response, so we will confine ourselves to challenge this proposition on a conceptual level.

The ‘natural’ zero point of a ratio scale, e.g. for measuring physical quantities, such as mass or time, has an unambiguous meaning. It defines the point at which the quantity is completely absent, where there is no mass or no time. Yet, when ’being dead’ is set to zero on the hrqol measurement scale, it seems have different properties. For some individuals, it divides utilities into positive and negative values, while for others it does not [8]. In fact, individuals who value being alive irrespective of their health, cannot even move to any health state with a zero value (while they are alive). Such a scale does not seem to measure hrqol in some natural units, but in relation to some external reference. In this regard, it appears rather similar to the zero points in degrees Celsius or degrees Fahrenheit, which divide the (interval) scale for temperature into negative and positive values, at different, arbitrarily chosen points [22].

To conclude, conceptualising dead as a health state and setting it to zero is by no means ‘natural’. It thus does not seem a priori illegitimate to use a different scale to measure hrqol. Of course, to be useful for the valuation of health outcomes, hrqol values have to be scaled to the full health-dead scale of the QALY at some point. However, for the comparison and aggregation of hrqol values across individuals, normalising health state utilities between the best and the worst health state appears to be a much better because fairer method.

5 Alternative multi-step health state utility aggregation procedure

5.1 Overview and notations

The alternative procedure that we propose to aggregate individual health state preferences into a social hrqol value set is based on the principle of relative utilitarianism. The aim is to give every-one’s preferences equal weight in the outcome. The procedure consists of four simple steps that we will outline below. and then apply on a simple example to demonstrate its advantage over the ’traditional’ method. For convenience, we shall introduce some notations.

Suppose there are n individuals $\{1, 2, \ldots, n\}$ with preferences over health states, measured on the TTO utility scale $u_j\left(h_i\right) : \mu\{1 \leq \mu \geq -1\}$. Further, let $H\{\in h_1, h_2, \ldots, h_k\}$ be a set of mutually exclusive health states.

$u_j\left(H\right) \ \in \{u_j\left(h_1\right), u_j\left(h_2\right), \ldots, u_j\left(h_k\right)\}$ then denotes the set of individual j's preferences over all health states in H, and $\min u_j(H)$ denotes the lowest, and $\max u_j(H)$ denotes the highest value in $u_j(H)$. To avoid division by zero during the normalisation procedure, we define that $\min u_j(H) < \max u_j(H) = 1$.

5.2 Model formulation

The 'traditional' social welfare function $U\left(h_i\right) : \mu 1\{\leq \mu \geq -1\}$ maps health states to a social hrqol value. The function is given by:

$$U(h_i) = \frac{\sum u_1\left(h_i\right), u_2\left(h_i\right), \ldots, u_n\left(h_i\right)}{n} \tag{1}$$

Our alternative social social welfare function $S\left(h_i\right) : \mu 1\{\leq \mu \geq -1\}$ can be derived through the following four simple steps:

1. **Normalisation**: the health state utility values of each individual are normalised between their best and their worst health state, so that everyone's values have the same range (1-0).

$$u_j'(h_i) = \frac{\sum_{j=1}^{n} u_j h_i}{n} \tag{2}$$

2. **Aggregation**: A provisional social preference function $S'(.)$ is derived by aggregating the normalised utilities across individuals.

$$S'(h_i) = \frac{\sum_{j=1}^{n} u_j' h_i}{n} \tag{3}$$

3. **Scaling factor**: An individual's RSQQ, denoted r can be interpreted as a scaling factor, which expands or shrinks the range of utilities. These RSQQ are also aggregated across individuals to determine the social RSQQ, denoted R. In addition, an anchor point A is required, to be able to relate the normalised hrqol scale to the full-health-dead-scale.

$$r_j = \min u_j(H) - \max u_j(H) \qquad \text{(Individual RSQQ)} \tag{4}$$

$$R = \frac{\sum_{j=1}^{n} r_j}{n} \qquad \text{(Social RSQQ)} \tag{5}$$

$$A = 1 - R \qquad \text{(Anchor point)} \tag{6}$$

4. **Rescaling**: The social RSQQ R is used to rescale the provisional social preference function $S'(.)$ back to the full health-dead scale between 1 and 0.

$$S(h_i) = R * S'(h_i) + A \tag{7}$$

5.3 A simple example

Suppose a TTO is conducted among two individuals (1 and 2), to derive a social value set for a simple descriptive system, with two attributes (pain and mobility) and two levels (no problems, problems): h^* (full health); h_i (immobile); h_j (pain); h_0 (immobile and pain). Further suppose that individual 1's utility for h_i is higher than for h_j (0.9 vs 0.8), while individual 2's preferences are the other way around (0.66 vs. 0.33). By definition, h^* has a value of 1, and both individuals also agree that h_0 is dominated by the other states, yet individual 2 assigns it a much lower value (0.00 vs 0.7). When their utilities are aggregated using the traditional method, the social values for h_i and h_j are 0.62 and 0.73, indicating that,as a group, the two individuals prefer h_j over h_i. The social values for h^* and h_0 are 1 and 0.35, respectively.

In our interpretation of health state preferences, the aggregate results are unfair towards individual 1. When their preferences are normalised between the best and the worst state, it can be seen that – relative to the other health states – individual 1's preference intensity for h_i is exactly as strong as individual 2's preference for h_j (0.66 vs 0.66). It is only because individual 2 has a higher RSQQ (1.00 vs 0.30) that makes their preference for h_j seem stronger. When our alternative aggregation procedure is applied, the individuals' relative health states preferences receive equal weight in the derivation of the social preference function. The respective alternative social hrqol values for h_i and h_j are 0.68 and 0.68. The values for h^* and h_0 are unchanged.

Table 1 shows the health state utility values of both individuals and the corresponding traditional and alternative social value sets. In addition, the results are visually illustrated depicted in Figure 3.

Table 1. Individual and social health state preferences

		h^*	h_i	h_j	h_0	r
'Traditional' utility aggregation procedure	u_1	1	0.90	0.80	0.70	
	u_2	1	0.33	0.66	0.00	
	U	1	0.62	0.73	0.35	
Alternative utility aggregation procedure	u'_1	1	0.66	0.33	0.00	0.30
	u'_2	1	0.33	0.66	0.00	1.00
	S' / $[R]$	1	0.50	0.50	0.00	[0.65]
	S	**1**	**0.68**	**0.68**	**0.35**	

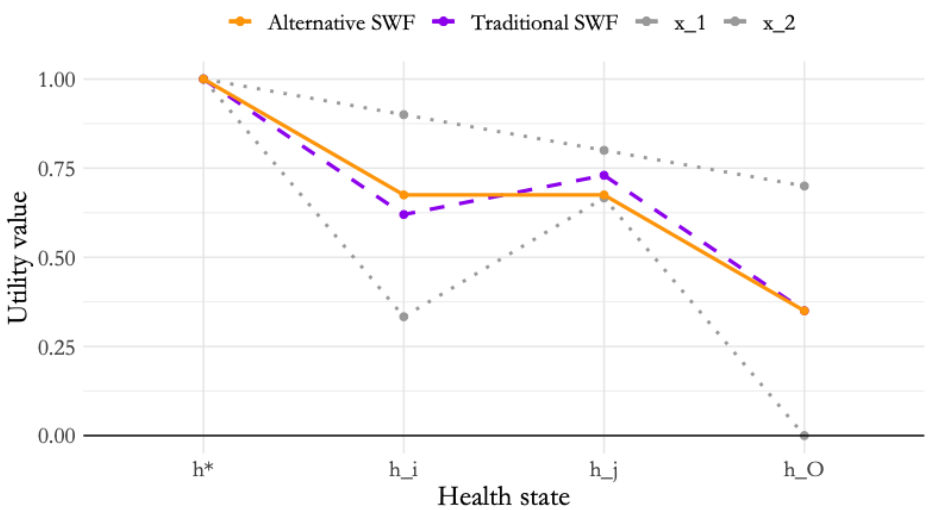

Figure 3: Overview of preference profiles. Shown are the two individual preference profiles (gray), the traditional social welfare function (purple), and our alternative social welfare function (orange).

6 Discussion

We have outlined potential flaws in the methodology that is commonly used to compare and aggregate individual utilities in the context of TTO health valuation studies. The main argument of our paper is that the current method in unfair, because it fails to give every individual equal say in the outcome. Social value sets then mainly represent the preferences of those individuals who place a lower value on survival time, compared to hrqol. The alternative multi-step aggregation procedure that we propose aims to overcome this problem, by separating the aggregation of hrqol from the aggregation of the RSQQ. Even though this paper focuses exclusively on utilities from TTO, we think our approach equally applies to utilities elicited through the standard gamble method.

Currently, our critique of the traditional method is only conceptual. We have not yet empirically tested whether our alternative method yields any different results. It should be noted that this can only be expected, if the health state preferences of individuals with a high RSQQ systematically differ from those with a low RSQQ (e.g. individuals with a high RSQQ care about mobility, while individuals with lower RSQQ being free from pain). The next step will thus be to apply our alternative method on actual preference data and to compare the results with the traditional method.

Our conceptual approach links directly to other alternative utility rescaling/reweighting methods. Jakubczyk et al.[30], for example, propose to equalise utilities based on 'worst fears'. For each individual, utilities are normalised between full health and either 'being dead' or the worst health state, which ever has a lower utility. Their method and rationale have not yet been published, so that their approach can currently not be examined in more detail. A series of alternative measurement scales for hrqol were suggested by Sampson et al.[27]. They argue that dead should not be considered relevant at all in the derivation of the social value of health states, and, therefore, reject the use of 'being dead' 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, their work has also not been published yet.

For the purpose of this paper, we have accepted the arithmetic mean as a valid aggregation method. From a welfare economic perspective the average might indeed be considered the method of choice. However, it must be recognized that health state utilities are actually not aggregated to decide which outcome is socially preferred, but only to derive a social value set, which is then used to inform this decision. It seems unclear whether the utilitarian justification for maximising the sum of utilities is applicable in this case, or whether other aggregation methods would be more appropriate [9, 10, 31–33].

With regard to the broader theoretical context of this paper, it is also worth mentioning that, over the last decades, the general problem of interpersonal utility comparisons has been addressed by numerous scholars in the fields of Welfare Economics, Social Choice Theory and certain parts of philosophy. Much effort has been devoted to studying if, and if so how, utilities can be compared within and between individuals. A vast range of approaches has been proposed, ranging from no comparability, which leaves almost no room for evaluations beyond the Pareto criterion, over ordinal comparability, for which Arrow famously proved that aggregation is inherently problematic, to cardinal comparability, which can be further divided into different types (e.g. level, unit). We are convinced that research on the valuation of health could benefit from a closer consideration of this extensive body of work. Interested readers may refer to [12] and [34] for an introduction and overview of the field and the historic context.

Acknowledgement

This work was funded by the Wellcome Trust through a PhD scholarship.

References

1. Karimi, M. & Brazier, J. Health, health-related quality of life, and quality of life: what is the difference? *Pharmacoeconomics* **34,** 645–649 (2016).
2. Brazier, J. Valuing health states for use in cost-effectiveness analysis. *Pharmacoeconomics* **26,** 769–779 (2008).
3. Loomes, G. & McKenzie, L. The use of QALYs in health care decision making. *Social science & medicine* **28,** 299–308 (1989).
4. Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L. & Torrance, G. W. *Methods for the economic evaluation of health care programmes* (Oxford university press, 2015).
5. Torrance, G. W. Social preferences for health states: an empirical evaluation of three measurement techniques. *Socio-economic planning sciences* **10,** 129–136 (1976).
6. Dolan, P., Gudex, C., Kind, P. & Williams, A. The time trade-off method: results from a general population study. *Health economics* **5,** 141–154 (1996).
7. Devlin, N., Shah, K., Feng, Y., Mulhern, B. & Van Hout, B. Valuing health-related quality of life: an EQ-5D-5L value set for England. 2016. *URL: www. ohe. org/publications/valuing-healthrelated-quality-life-eq-5d-5l-value-set-england (accessed 10 June 2016)* (2017).
8. Attema, A. E., Edelaar-Peeters, Y., Versteegh, M. M. & Stolk, E. A. Time trade-off: one methodology, different methods. *The European Journal of Health Economics* **14,** 53–64 (2013).
9. Devlin, N., Shah, K., Buckingham, K., *et al.* What is the normative basis for selecting the measure of 'average'preferences for use in social choices. *OHE research paper. Office of Health Economics, London* (2017).
10. Dewitt, B., Davis, A., Fischhoff, B. & Hanmer, J. An approach to reconciling competing ethical principles in aggregating heterogeneous health preferences. *Medical Decision Making* **37,** 647–656 (2017).
11. Grabisch, M., Marichal, J.-L., Mesiar, R. & Pap, E. Aggregation functions: means. *Information Sciences* **181,** 1–22 (2011).
12. Fleurbaey, M. & Hammond, P. J. in *Handbook of utility theory* 1179–1285 (Springer, 2004).
13. Carr-Hill, R. A. Assumptions of the QALY procedure. *Social science & medicine* **29,** 469–477 (1989).
14. Dolan, P. The measurement of individual utility and social welfare. *Journal of health economics* **17,** 39–52 (1998).

15. Hausman, D. M. The impossibility of interpersonal utility comparisons. *Mind* **104,** 473–490 (1995).
16. Jakubczyk, M., Golicki, D. & Niewada, M. The impact of a belief in life after death on health-state preferences: True difference or artifact? *Quality of Life Research* **25,** 2997–3008 (2016).
17. Barry, L. *et al.* Euthanasia, religiosity and the valuation of health states: results from an Irish EQ5D5L valuation study and their implications for anchor values. *Health and quality of life outcomes* **16,** 152 (2018).
18. Augestad, L. A., Rand-Hendriksen, K., Stavem, K. & Kristiansen, I. S. Time trade-off and attitudes toward euthanasia: implications of using 'death'as an anchor in health state valuation. *Quality of Life Research* **22,** 705–714 (2013).
19. Tilling, C., Devlin, N., Tsuchiya, A. & Buckingham, K. Protocols for time tradeoff valuations of health states worse than dead: a literature review. *Medical Decision Making* **30,** 610–619 (2010).
20. Pliskin, J. S., Shepard, D. S. & Weinstein, M. C. Utility functions for life years and health status. *Operations research* **28,** 206–224 (1980).
21. Torrance, G. W. Measurement of health state utilities for economic appraisal: a review. *Journal of health economics* **5,** 1–30 (1986).
22. Rossi, M. *Interpersonal comparisons of utility: The epistemological problem.* PhD thesis (London School of Economics and Political Science (United Kingdom), 2009).
23. Isbell, J. *Absolute games, in" Contributions to the Theory of Games, IV"(AW Tucker and RD Luce, Eds.)* 1959.
24. Kind, P., Hardman, G. & Macran, S. UK populations norms for EQ-5D: University of York. *Centre for Health Economics* (1999).
25. Lamers, L. M. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. *Medical care* **45,** 238–244 (2007).
26. De Charro, F., Busschbach, J., Essink-Bot, M.-L., van Hout, B. & Krabbe, P. in *EQ-5D concepts and methods: A developmental history* 171–179 (Springer, 2005).
27. Sampson, C. J., Devlin, N. & Parkin, D. W. Drop dead: is anchoring at 'dead' a theoretical requirement in health state valuation? *Conference: 35th EuroQol Group Scientific Plenary* (2018).
28. Roudijk, B., Donders, A. R. T. & Stalmeier, P. F. Setting Dead at Zero: Applying scale properties to the QALY model. *Medical Decision Making* **38,** 627–634 (2018).

29. Miyamoto, J. M., Wakker, P. P., Bleichrodt, H. & Peters, H. J. The zero-condition: a simplifying assumption in QALY measurement and multiattribute utility. *Management Science* **44,** 839–849 (1998).
30. Jakubczyk, M. K. & Golicki, D. What if 0 is Not Equal to 0? Interpersonal Utility Anchoring Using the Worst Fears. *Patient 12: 429. https://doi.org/10.1007/s40271-019-00368-9* (2019).
31. Schneider, P. Social tariffs and democratic choice–do population-based health state values reflect the will of the people? (2019).
32. Prieto, L. & Sacristán, J. A. Problems and solutions in calculating quality-adjusted life years (QALYs). *Health and quality of life outcomes* **1,** 80 (2003).
33. Forman, E. & Peniwati, K. Aggregating individual judgments and priorities with the analytic hierarchy process. *European journal of operational research* **108,** 165–169 (1998).
34. d'Aspremont, C. & Gevers, L. Social welfare functionals and interpersonal comparability. *Handbook of social choice and welfare* **1,** 459–541 (2002).