The University Of Sheffield.

School of
Health
And
Related
Research

Health Economics & Decision Science (HEDS)

Discussion Paper Series

Experience-based utility and own health state valuation: why do it and how to do it

Authors: John Brazier, Donna Rowen, Milad Karimi, Tessa Peasgood, Aki Tsuchiya, Julie Ratcliffe

Corresponding author: Donna Rowen

ScHARR, University of Sheffield,
Regent Court, 30 Regent Street, Sheffield, S1 4DA
Tel: (+44) (0)114 222 0728
Fax: (+44) (0)114 222 0710
Email: d.rowen@sheffield.ac.uk

Experience-based utility and own health state valuation: why do it and how to do it

John Brazier[1], Donna Rowen[1], Milad Karimi[1,2], Tessa Peasgood[1], Aki Tsuchiya[1,3], Julie Ratcliffe[4]

[1] School of Health and Related Research (ScHARR), University of Sheffield, UK, Regent Court, 30 Regent Street, S1 4DA

[2] Department of Population Health, Luxemburg Institute of Health, 1A-B, rue Thomas Edison, L-1445 Luxembourg

[3] Department of Economics, University of Sheffield, UK, Mappin Street, S1 4DT

[4] Health Economics Unit, Flinders University, Australia, Level 1 A Block, Repatriation General Hospital, Daws Road, Daw Park, South Australia 5041

Correspondence to: Donna Rowen, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, S1 4DA, UK. d.rowen@sheffield.ac.uk

ABSTRACT: In the estimation of population value sets for measures such as the EQ-5D, there is increasing interest in asking respondents to value their own health state, sometimes referred to as “experience-based utility values”, rather than hypothetical health states. Evidence shows that these experience-based utility values differ to hypothetical health state values. This may be attributed to many reasons. This paper first critically examines: why this difference is important, whether own visual analogue scale (VAS) or time trade-off (TTO) really measure experience-based utility values, the biases from current methods of collecting experience-based utility data, and the modelling of the data. Second, the paper reviews some of the normative arguments for and against using own health state valuation. Finally, the paper also examines other ways own health state values can be taken into account, such as including the use of *informed* general population preferences that may better take into account experience-based values.

Keywords: Experience-based utility; own health state valuation; hypothetical health state values; informed preferences

Acknowledgements: Previous versions of the paper were presented at the EuroQol Plenary Meeting in Berlin, Germany in June 2016 and at the Health Economists Study Group Meeting in Birmingham, UK in January 2017. We would like to thank all attendees for their comments and Chris Sampson in particular for his discussion of the paper.

Experience-based utility and own health state valuation: why do it and how to do it

1. Introduction

Historically, it was recognised that health state utility values could be obtained from a number of different sources including patients, their carers, health professionals and the community (see (1) for a detailed overview). Currently, health state value sets are usually obtained from members of the general public trying to imagine what the state would be like, mainly argued for on the basis that it is the general population who are the payers of healthcare. However, it has also been argued that values should be obtained from patients as they better understand what it is like to live in poorer health (2-4). Of course, utility values can be obtained directly from people whether or not they are patients. Own health state values can be obtained to value a person's state without a descriptive system like EQ-5D, or they can be used to estimate value sets for (generic) measures of health such as the EQ-5D. The latter has started to be described in the literature as "experience-based utility values", where general public respondents value their own current health state (4). These experience-based utility values are then modelled against their self-reported EQ-5D health state, for example, to estimate a value set for all states described by EQ-5D (e.g. (5-7)).

We critically examine a number of important issues: why the potential use of experience-based utility is important, the theory underpinning the extent to which own TTO measures experience-based utility values, the challenges of collecting own health state values utility values and the modelling of the data, and the normative arguments for and against using own health state utility values. Finally, the paper considers a number of alternative uses of own valuation focusing on the idea *informed* general population preferences that take into account experience-based utility values.

A clarification

There are two main considerations in this literature (see Table 1): *whose* values and *what* to value. The debate has tended to focus on the first of these: whether to use patient or general public values. The second consideration is whether they should value their own state or some hypothetical state. A (conventional) preference-based EQ-5D population value set asks a general public sample to value hypothetical states (cell 1), while another approach would be to ask patients to value their own current state (cell 4). The growing interest in experience-based utility values concerns cell (3). Of course, patients are also members of the general public and the strict public vs patient dichotomy is a false one (8) – but informative. A typical study in cell (4) will have a relatively small, clinically homogeneous sample, and will therefore not aim to elicit values for enough states to model the entire descriptive system, while a study in cell (3) will aim for a large, heterogeneous sample, with the aim to value all possible states in a descriptive system. Since there is little that is currently known that contrasts (1) against (3), we will draw on the established literature that compares between cells (1) and (4). A third consideration is whether to link values from any of these cells to a descriptive system e.g. EQ-5D, which is a comparatively recent development in the literature. Note that the term patient is used throughout the paper for simplicity, and it is that the values are based on experience, "experience-based" values, that is important not that the person is a patient per se.

Table 1: Categorising values by population and what is valued

	General public	Patients
Hypothetical states	(1)	(2)
Own current state	(3)	(4)

2. Does own VAS or own TTO measure experience-based utility?

Daniel Kahneman (9) distinguished between (stated) preference-based methods of valuation, such as willing-to-pay, standard gamble (SG) and time trade-off (TTO), which elicit what he calls 'decision-utility', and the hedonic and affective experiential methods associated with an outcome which he calls 'experienced-utility'. He and others have questioned the validity of decision-utility based methods due to systematic errors in our forecasts of how we will feel in future states (10). In its place, they have argued for a return to an interpretation of utility used by Bentham, where utility is the pleasure and pain experienced in each moment of time (11). This would imply a return to measuring utility directly and in cardinal form from the person experiencing it, and hence 'experienced-utility' (12).

Kahneman's use of the term 'experienced-utility' is very different to the way in which it is being used to describe a valuation based on own health. Where the valuation is using Visual Analogue Scales (VAS) we may consider this a proxy for current momentary experience, but where it is based on TTO, SG or another choice exercise, it is a preference expressed over hypothetical future experience, and a measure of 'decision' rather than 'experienced' utility in the way Kahneman has coined the terms.

Kahneman recommends measuring utility using moment-based happiness (9). The most direct method of doing so is the 'experienced sampling methodology' (13) in which participants are contacted at multiple random points during the day and asked about their how they are feeling and how happy they are at each time point. A more pragmatic and less intrusive solution is the 'day reconstruction method' (14) which asks people to recount different episodes of the previous day and how they felt during each episode. Data on affective experience is then combined: one possibility is to take the difference between the average positive feelings (or the most intense positive) and the average negative (or the most intense negative) (14) and this is used to estimate an area under the plotted curve to define experienced utility over time. Another possibility is to the use the proportion of time in which the most intense negative affect outweighs the most intense positive, referred to as a 'U-index' (15).

However, the methods for assessing experienced-utility in the sense used by Kahneman are criticised by economists as they do not require the respondent to make a sacrifice. In other words, there is no opportunity cost, so there are questions about the meaning behind the expression of feelings, whether they can be compared between individuals, and whether they provide ordinal or cardinal values. Similarly, there is the question of what respondents are valuing in an own TTO task to collect what is called in the literature '*experience-based* utility values'. The use of choice-based methods such as TTO to elicit values for experience-based utility require respondents to value their current state by imagining what it would be like to be in full health, which they may not have experienced for many years. For people who have lived in a chronic health state, for example chronic obstructive pulmonary disease or osteoarthritis, the task of imagining full health is potentially as difficult as a healthy member of the general population trying to imagine a poor health state. In addition they are being asked to imagine that they will remain in their current state for some specified period, such as 10 years in the case of Burstrom et al (7); and to imagine dying. This is a hypothetical task and raises the question about how realistic this would be for most respondents, indeed as noted above Kahneman has questioned people's ability to predict their experienced utility into the future. During the TTO task do respondents really imagine their current health state will not change or do they imagine some change over time? Do they consider the fact they may adapt, that their life circumstances may change and so forth? Do they include more than the EQ-5D state they have just completed (for example, their wider quality of life), and if so what else do they take into account? How do they deal with the idea of dying? Currently we do not know and there is a role for qualitative research to examine what is actually being valued with these tasks. However, what we can conclude is that own health state valuation using TTO is not the same as experienced based utility in the way Kahneman and colleagues describe it.

3. Does the choice of whose values matter?

The choice of whom to elicit values from is important because it may influence the resulting values. A number of empirical studies have been conducted which indicate that people with first-hand experience tend to (although not always) place higher values on dysfunctional health states than members of the general population who do not have similar experience, and the extent of this discrepancy tends to be much stronger when people value their own health state (16-18). However, there is some evidence suggesting that for mental health, values may be lower for people with experience of mental health problems (19), as contrary to the general population they tend to place a greater weight on mental health impairments relative to physical health impairments (20), though this research was not based on valuing own health.

Earlier empirical studies comparing patient and general population values tended to use relatively small sample sizes and focused on differences in patient and general population values for a single medical condition or type of health problem. More recently, studies have attempted to use larger sample sizes across general population or multiple patient samples, to compare across cells (1) and (4) of Table 1.

Mann et al*.* (5) estimated EQ-5D preference weights using regression analysis of own EQ-5D VAS data (cell 3) from patients diagnosed with eight different conditions (n=3,376), and compared the results to EQ-5D preference weights estimated using regression analysis of hypothetical VAS values elicited from the general population (n=2997; the MVH dataset, (21)). Compared with the general population model the decrements for anxiety/ depression were statistically significantly larger in the patient model but smaller for pain/discomfort and mobility. The magnitude of disagreement between the patient self-rated VAS model and the population VAS model was also found to vary depending upon the patient’s condition. Rand-Hendriksen et al. (22) examined differences in the relative importance attributed to EQ-5D dimensions between experience-based health valuations from patients (n=74,277) and hypothetical health valuations from the general population (n=3,773) for EQ-5D states using VAS in the United States. Self-care and pain/discomfort were found to be the most important dimensions for the hypothetical health valuations, whereas usual activities was the most important dimension for the experience-based health valuations. Little et al. (23) compared German experience-based (n=2,032, obtained from Leidl and Reitmeir (24)) and European hypothetical health valuations from the general population (n=6,870) for EQ-5D using VAS, finding that pain/discomfort was the most important dimension for the experience-based health valuations. Sun et al (6) generated EQ-5D preference weights by modeling own VAS using Chinese experience-based data (n=120,709) and showed anxiety/depression had the greatest impact on own VAS values.

Burstrom et al. (7) used TTO (alongside VAS) to estimate Swedish experience-based value sets for EQ-5D-3L health states using general population health survey data. A large sample of n=45,000 individuals was used in order to facilitate modeling of the experience-based TTO and VAS values in terms of the EQ-5D descriptive system. They found the anxiety/depression dimension to have the greatest impact on both TTO and VAS values.

In summary, although there are variations in the findings from these individual studies in terms of the relative impact of dimensions according to own health state valuation, the available evidence highlights the potential for systematic differences between hypothetical general population preferences and own health state values that could impact on the results of an economic evaluation. Therefore, the choice as to whose values are used may be critical when making resource allocation decisions.

4. Why do discrepancies in own health state values and hypothetical general population values exist?

There are a number of possible contributing factors for observed differences between patient and general population values. Earlier we argued that respondents in poor health states may find it hard to imagine full health. The literature has suggested additional possible explanations including: poor descriptions of health states (for the general population), use of different internal standards, or response shift and adaptation (1, 3). These are discussed in detail below.

Poor descriptions of health states

An important potential source of discrepancy is found when descriptions provided to the general population in cell (1) may not accurately describe the health state, even when these are produced using a health state descriptive system. Respondents can bring their own information to the valuation exercise by drawing upon their own personal experiences or limited knowledge. Given that the personal experiences of people with experience of a health condition and members of the general public are unlikely to be the same, it may mean that, in effect, they are evaluating different health states even when provided with identical descriptions of the health state to be valued. In addition, patient respondents in cell (4) will not necessarily be valuing the health state they are categorised into by a descriptive system (like EQ-5D). Respondents are asked to value their current health state as they see it and this may cover different dimensions to the measure being used. Furthermore, it has been suggested that general population respondents in (1) focus too much on ill health and ignore the remaining positive aspects of a person's life (3).

Changing standards

A well-known phenomenon in the psychometric literature is 'response shift', which refers to the possibility that individuals will change their internal standards for evaluating their own health in response to changes in their health (25). Response shift occurs due to changes in expectations. For example, an older person may rate his or her health according to their expectations of the best possible health for a person of their age rather than best possible health *per se*. Similarly, a patient may rate his or her health by comparing themselves with other patients rather than with healthy individuals. In either instance, response shift will contribute to discrepancies between patient and general population values in cells 1 and 4 for the same health states and, unlike the problem of incomplete or inaccurate health state descriptions, it is difficult to see how response shifts can be reduced or eliminated in practice. Indeed, it can be argued that response shift effects in health state valuation tasks conducted with patients should not be of concern, since these reflect adaptation and coping.

Adaptation to the state

Someone in a permanent and stable impaired health state is, depending on the health problem, likely to adapt over time, both physically and psychologically. Physical changes include the acquisition of new skills to help cope with a disability, such as learning to use a walking stick. Or, a person may change the things they do in order to limit the impact of their disability or illness. For example, someone who once played football may take up a sport that has a lower impact on their knees. There are also psychological adaptations that include a shift in the weight people place on different aspects of health and quality of life and, more fundamentally, a change in their view of what matters in life. In addition, people may lower their expectations of what they can achieve.

It is well established in the literature that people tend to under predict their ability to adapt to physical health conditions or impairment states (9, 26, 27). When general population respondents read the description of a state, their valuation may reflect a response to, say, going blind, rather than being blind for an extended period. In other words, they focus on the transition to the state rather than the longer term consequences. This focus results in the general population giving lower values compared with patient self-reported values for chronic states of health.

These explanations for the differences in values between the general public imagining the state (cell 1) and respondents valuing their own state (cell 4) may have normative implications that are examined in section 8.

5. What are the practical problems of collecting own health state values?

Asking respondents to value their own health state and to do so across a sufficiently wide range of health states of different type and severity raises two major practical problems. Firstly, respondents in poor health states may be unable or unwilling to undertake complex and quite intrusive valuation tasks. This may be due to their physical limitations. Furthermore, there are also mental health and cognitive problems that make completing a health state valuation task (e.g. TTO) more challenging than simply completing the EQ-5D classification system, particularly if it is to be self-completed (which tends to be the case with many of these surveys) only once, without any assistance from interviewers and without a practice question (which is difficult for own health valuation). Second, there will be ethical concerns with asking people in terminal or incurable conditions to imagine hypothetical scenarios involving return to full health, accompanied by either the risk of immediate death or shortening life. For these reasons it is not possible to ask some patients groups to complete health state valuation tasks for any state (hypothetical or own). These two factors will result in lower numbers in the poorer health states being sampled, as shown by the low numbers reporting more severe levels of EQ-5D in Table 2 (taken from (6, 7).

The data may also suffer from selection bias since people who are experiencing a health state that has impacted more on their utility may not wish to participate in elicitation surveys because of their health (even if they are able to). The magnitude of this type of selection bias is likely to vary not only by the severity of condition but also by dimension (e.g. mental health). This will introduce a complex pattern of bias with some (but not necessarily all) severe states having higher values than would be the case if it was a genuine random sample of those people in them. This may partly explain why own health state values are higher than the values of members of the general public imagining them. A major practical problem for the researcher is to obtain the right sample and one that is representative across the severity range across each health state. This is a survey design issue and a different approach is proposed in section 7.

Table 2: Observed distribution of EQ-5D own health in surveys used to provide experience-based utility data

EQ-5D dimension	**Level**	**Sweden (%) (n=46,169) (7)**	**China (%) (n=120,709) (6)**
Mobility	**1**	90.1	94.8
	2	9.8	4.8
	3	0.1	0.4
Self-care	**1**	98.4	96.8
	2	1.2	2.8
	3	0.4	0.4
Usual activities	**1**	91.2	95.2
	2	7.7	4.0
	3	1.1	0.8
Pain/discomfort	**1**	50.8	90.8
	2	45.1	8.8
	3	4.1	0.4
Anxiety/depression	**1**	66.5	93.6
	2	30.8	6.0
	3	2.7	0.4

6. Econometric issues arising in the elicitation of own health state values

The elicitation of experience-based utility values for EQ-5D in the literature has involved the use of regression analysis of large general population samples who value their own health using TTO (7) or VAS (6, 22, 24) to produce preference weights for every health state defined by the classification system. The sample sizes for these studies were large, for example 49,169 for Sweden (7), 74,277 for the US (22) and 120,709 for China (6). However, as these were general population samples the majority of respondents were in full health or mild health states. Table 2 reproduces the distribution of (three-level) EQ-5D responses in the Swedish and Chinese samples used to estimate experience-based utility preference weights for the EQ-5D. In both samples the proportion of responses in level 1 in each dimension is large, and 39.8% and 87% of respondents are in EQ-5D full health in the Swedish and Chinese samples respectively. However, the distribution for Sweden does have larger proportions of respondents at levels 2 and 3 for the pain/discomfort and anxiety/depression dimensions, although this is not observed in the Chinese sample.

The use of this methodology to produce a utility value for every state defined by the classification system requires that the sample size is extremely large, as each respondent can provide only one observation per time point, and the distribution of these states is highly skewed. In contrast, general population surveys eliciting preferences for hypothetical health states typically ask each respondent to value 8 or more health states, and these states are selected for design efficiency, thus saving on sample size.

The distribution of EQ-5D responses with the majority of responses at level 1 and a small proportion of responses at level 3 means that there are a large number of observations with TTO or VAS data for mild health states and only a small number of observations for severe health states. This creates problems for modelling the data to generate preference weights for every possible health state defined by the classification system. Even in the hypothetical health state valuation literature it is usually infeasible to value all health states defined by a health state classification. Health states for valuation are usually selected using by design as an orthogonal array (e.g. (28)), balance (e.g. (29)) or simulation of alternative selections (e.g. (30). All of these approaches select health states to ensure that the model estimated using the preference data is able to produce preference weights with acceptable margin of error for every possible health state defined by the classification system, and the selection of health states is an important component of any health state valuation study. However, the approach used by experience-based utility does not select health states in any way, nor does it necessarily include data on every health state as 148 and 167 unique health states were observed in the Swedish and Chinese samples respectively out of the 243 possible EQ-5D states. It is likely that this will detrimentally impact on the accuracy of any model used to estimate values for every health state defined by the classification system.

It may be argued that, while efficient designs aim to spread out the prediction error across dimension levels and thus across states, own valuation datasets contain more data on precisely those states for which information it is needed i.e. the most frequent states. At the extreme, there is no need to predict the value of health states that do not exist in the real world with the same (or even any) accuracy as those states that happen more often. Existing designs require some very rare states to estimate values for all states, but there may be ways to avoid these states in more sophisticated designs. Furthermore, the states that drive the results of cost-effectiveness models may not be the ones that arise with sufficient frequency in a general population sample.

7. How could own TTO valuation studies be designed?

An alternative approach to large scale general population experience-based utility surveys is to purposively sample people to provide own TTO values to ensure that values are obtained for health states that are informative for estimating a regression model estimating preference weights for every state described by the descriptive system. For example, a set of health states for say, the EQ-5D-5L could be selected using orthogonal array, balanced design or simulations of hypothetical states. Quotas could then be set for, say, 200 respondents in each health state and respondents from a sampling frame which ensures that respondents are representative of that state in terms of sociodemographic characteristics such as age and gender (which usually will not be the same as the general population).

To give an idea of the numbers involved, let us examine the numbers required in a general population survey using our experience from three online general population surveys. In a pool of 8,600 respondents across three online surveys of the general public (31, 32), we observed 586 unique EQ-5D-5L states broken down as follows: 11111 had a 35% share, the next two states covered another 20% of respondents and the next 20 states covered another 30%. Only the top six states had an n>200. The 200th ranked state had just n=3 observations. This would suggest that most EQ-5D-5L states (more than 90% of them) have a prevalence rate of 0.02% or lower in the general population. This suggests that in order to identify 200 individuals in a health state with a 0.02% prevalence, this would require screening 1 million individuals to find them (and not all of them will agree to be surveyed). There are a lot of assumptions in this rough calculation that we would not care to defend, but a crucial one is that an online survey using self-completed health state valuation task of own states is likely to suffer from the biases raised earlier. To achieve representative samples in the more severe states requires more directed sampling strategies and different modes of administration to ensure the more disabled or dissatisfied are not excluded.

The advantage of taking a designed approach is that the data would be informative for estimating the regression coefficients to produce weights for all health states defined by the descriptive system with acceptable levels of error, and the values for these would be experience-based not hypothetical. However, the challenges of finding people in more severe health states would be considerable, and the design would need to take into account the plausibility and prevalence of the more severe states selected in the study.

Also note that such a sample that extensively oversampled respondents in low-prevalent states will no longer be a representative sample, where all members of the general public had an equal chance of being recruited.

8. What are the normative arguments for using own health state values – and do we accept them?

The Washington Panel on the Cost Effectiveness in Health and Medicine (33) in advocating the use of general population values, argued that: '… the best articulation of society's preferences for a particular state would be gathered from a representative sample of fully informed members of the community' (this has been reaffirmed for the reference case in the update (34)). The Panel went on to use the notion of the 'veil of ignorance' to support the use of community values, where 'a rational public decides what is the best course of action when blind to its own self-interest, aggregating the utilities of persons who have no vested interest in particular health states seems most appropriate' (33). Patients tend to compare themselves to peers with similar health problems and this will influence their own health state valuations. They argue that the values of different patient groups are not comparable, whereas a general population sample provides a coherent set of values.

A key argument is that the general population pays for the service. However, whilst members of the general population may want to be involved in health care decision making, it is not clear that they want to be asked to value health states specifically (see for example (35)). At the very least, it does not necessarily imply that the current practice of using relatively uninformed general population preferences is optimal. A common argument for using patient values is the fact that patients understand the impact of their health on their well-being better than someone trying to imagine it (although they are having to imagine full health in most health state valuation tasks). Nevertheless, this does not imply that raw patient values should be used on their own to inform resource allocation decisions. This requires a value judgement that society wants to incorporate all the changes and adaptations that occur in patients who experience states of ill health over long periods of time. Some adaptation may be regarded as 'laudable', such as skill enhancement and activity adjustment, whereas cognitive denial of functional health, suppressed recognition of full health and lowered expectations may be seen as less desirable (2). Furthermore, there may be a concern that patient values are context based, reflecting comparisons with their recent experiences of ill health and the health of their immediate peers (3), which relates to response shift as discussed above.

9. Alternative approaches

One conclusion from the above discussion is that it may be difficult to justify the exclusive use of patient values, or the currently widely adopted practice of using values from relatively uninformed members of the general population. If it is accepted that ultimately it is the values of the general population that are required to inform resource allocation in a public system, it might be argued that respondents should be provided with more information on what the states are like for people experiencing them, in order that they can provide *informed* values. There are at least three different ways to achieve this:

1) Improve the descriptive systems e.g. include well-being dimensions that better reflect the impact on the lives of those experiencing the health states

2) Encourage more deliberation and reflection in the task

3) As 2) but provide more information on adaptation or experience-based values (e.g. through own TTO) to the general population sample undertaking the valuations

There is a fourth way proposed in a recent paper by Versteegh and Brouwer (4) of not trying to adjust general population values, but instead analysts provide the decision makers with results using both sets of values. This would mean providing two incremental cost-effectiveness ratios or net benefit values. However, this would add significantly to decision complexity and would imply the need for two threshold values. Unless the two analyses support the same decision, it further begs the question of how a decision maker decides which analysis to use or what weights to use if they were to be combined. It could result in inconsistencies across decisions that would be difficult to defend, particularly when decisions are appealed in a court of law. Here we argue that it is better to agree on a single value set for decision making. There is also a fifth way of using subjective well-being to re-weight classification systems such as EQ-5D using patient data (36). However the results are not anchored onto the 1-0 full health-dead scale required to generate QALYs, as currently there are no measures of subjective well-being that are valued using this scale.

Below we focus our discussion on options 2) and 3).

Using conventional valuation methods

Encouraging more deliberation and reflection has been proposed by a number of commentators on the subject (2, 3, 26, 27, 37) and there are examples of some studies which have attempted to operationalise this in practice.

A review of empirical studies attempting to inform general population respondents published in 2009 and earlier identified a total of 14 studies reporting upon methods used to elicit informed general population values for health states (38). Interventions used to inform the general population were categorised into the following: information to enrich the health state descriptions (n=7); simulation to reproduce the symptoms of the health state (n=2); opportunity to reflect and deliberate on the health state descriptions (n=2); and exercises to evoke adaptation to the health state (n=3).

The majority of the identified studies identified in the review attempted to generate informed general population values for health states by providing respondents with additional information through the use of audio recordings and videos. These studies appear to show general population values can be changed by the provision of additional data, though it is difficult to judge whether they were better informed other than the finding that values were closer to the patient values. For example, Clarke et al. (39) examined health state values for three Gaucher disease states by presenting information from patients currently living in the states through the use of multimedia equipment, finding no statistically significant differences in the utilities between the patients and the general population samples. Cunningham and Hunt (40) used descriptions of dentofacial deformities and photographs of dental patients corresponding to the health state descriptions, again finding no statistically significant differences in the utilities between the patients and the general population samples. Damschroder et al. (41) used an adaptation exercise when valuing states pertaining to pre-existing paraplegia and new onset of paraplegia, finding that completing an adaptation exercise statistically significantly increased utility values. McTaggart-Cowan et al. (26) explored the extent to which members of the general population changed their initial values for three rheumatoid arthritis states following an adaptation exercise, where they listened to recordings of patients discussing how they adapted. After undergoing the adaptation exercise, the respondents increased their values for the rheumatoid arthritis states, where younger and healthier individuals were more likely to increase their initial values after being informed.

However, one study found that discussion of preferences elicited in a group setting did not have a statistically significant impact on responses. Stein et al (42) used standard gamble with a panel of 15 members of the general population who valued 41 different health states five times over the course of six months. Following initial individual valuations, the group was given the opportunity to discuss the health states and following this discussion individuals were given the opportunity to change their health state values though with no reference to patient values or adaptation. Although no statistically significant differences were detected before and after the discussion, respondents indicated that the group discussion brought reassurance and cohesion to their responses.

Overall this evidence suggests that informed general population preferences can reduce the difference to own health state values though whether the values are therefore better informed is difficult to prove and further research is encouraged (26, 43, 44). The previous studies were concerned with small numbers of mainly condition specific states. It is not clear how practical these methods would be for valuing large numbers of generic states generated by measures such as EQ-5D-5L.

Using Citizen's juries and MCDA

Citizens' jury and multi-criteria decision analysis (MCDA) can also be used to elicit informed utility values from the general population, though, as far as we are aware, they have not been used to date to value health states and the methods may understandably need further consideration and adaptation to the context. A citizens' jury involves a small sample of participants deliberating on a topic and reaching a democratic recommendation (45). In the area of health policy a sample of general population participants i) are presented with a health policy dilemma, ii) review and examine evidence on the dilemma, with the presentation of the evidence undertaken by experts, iii) deliberate the dilemma to reach a consensus for a recommendation, which not all participants have to agree with (45). Participants in a citizens' jury are not typically expected to be informed prior to the process, rather they become informed during the process. The NICE Citizens' Council is an example of a citizens' jury that is routinely involved in health policy decision-making (though this is just one type of citizens' jury). The outcome of the citizens' jury is typically qualitative, meaning that strength of preference is not typically indicated. Therefore a citizens' jury would need to be combined with one of the other techniques, such as TTO. This could involve participants undertaking TTO tasks "cold" before becoming informed on the topic, undertaking the full deliberative process of the citizens' jury and then completing the same TTO survey again to provide informed responses (see for example (46)). Undertaking the same TTO survey before and after the deliberative process enables the assessment of impact of the deliberative process.

Multi-criteria decision analysis (MCDA) can be used to evaluate health states in a group setting by explicitly considering multiple criteria. This enables a structured approach as 1) criteria are determined, 2) scores reflecting the value of how health states perform according to each of multiple criteria is determined, and then 3) weights that reflect the relative importance of each of the different criteria are considered, before 4) a recommendation is reached (47). Different types of approaches can be used to score and weight the criteria: value measurement models; outranking models; and goal, aspiration or reference-level models (47). The scores and weights can be determined separately, for example using VAS, or the scores and weights can be determined simultaneously, for example using discrete choice experiments or TTO (see (48)). This could be similar to recent work on "personal utility functions" undertaken by Shah et al. (49).

10. Conclusions

This paper has critically examined the issue of experience-based utility, including consideration of the extent to which TTO and VAS tasks measure experience-based utility, the challenges of collecting and modeling experience-based utility values, and normative arguments for and against using experience-based utility values. The literature has involved the collection of large amounts of quite complex data from people who may not be actually valuing what we imagine they are valuing, and who are not providing values of health states across the severity range. The normative arguments for eliciting own TTO utility values require consideration, and the suitability of current experience-based data for informing policy should be questioned.

The paper proposed one approach for collecting own TTO utility values that would provide less biased data for modelling values to produce preference weights, by systematically selecting health states and purposively sampling respondents using a sampling frame to meet pre-defined quotas. Finally, the paper suggested alternative approaches of obtaining more *informed* general population preferences that take more account of patient views.

References

1. Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. Measuring and Valuing Health benefits for Economic Evaluation. 2nd Edition ed. Oxford: Oxford University Press: ; 2016.
2. Menzel P, Dolan P, Richardson J, Olsen A. The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. Social Science and Medicine. 2002;55:2149-1258.
3. Ubel P, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Quality of Life Research. 2003;12:599-607.
4. Versteegh M, Brouwer W. Patient and general public preferences for health states: a call to reconsider current guidelines. Social Science and Medicine. 2016;165:66-74.
5. Mann R, Brazier J, Tsuchiya A. A comparison of patient and general population weightings of EQ-5D dimensions. Health Economics. 2009;18:363-72.
6. Sun S, Chen J, Kind P, Xu L, Zhang Y, Burström K. Experience-based VAS values for EQ-5D-5L health states in a national general population health survey in China. Quality of Life Research. 2015;24:693-703.
7. Burstrom K, Sun S, Gerdtham U, Henriksson M, Johannesson M, Levin L, et al. Swedish experience-based value sets for EQ-5D health states. Quality of Life Research. 2014;23:431-42.
8. Dolan P. Whose preferences count? Medical Decision Making. 1999;19(4):482-6.
9. Kahneman D. Evaluation by moments: past and future. In: D K, AS T, editors. Choices, values and frames. 693–708. New York: Cambridge University Press and the Russell Sage Foundation; 2000.
10. Kahneman D, Sugden R. Experienced utility as a standard of policy evaluation. Environmental and resource economics,. 2005;32(1):161-81.
11. Kahneman D, Wakker P, Sarin R. Back to Bentham? Explorations of experienced utility. The quarterly journal of economics. 1997;112(2):375-405.
12. Dolan P, Kahneman D. Interpretations of utility and their implications for the valuation of health. Economic Journal. 2008;118:215-34.
13. Larson R, Csikszentmihalyi M. The experience sampling method. New Directions for Methodology of Social and Behavioral SciencE. 1983;15:41-56.
14. Kahneman D, Krueger A, Schkade D, Schwarz N, Stone A. A survey method for characterizing daily life experience: The day reconstruction method. . Science. 2004;306(5702):1776-80.
15. Kahneman D, Krueger A. Developments in the measurement of subjective well-being. The Journal of Economic Perspectives. 2006;20(1):3-24.
16. Tengs T, Wallace A. One thousand health related quality of life estimates. Medical Care. 2000;38:583-637.
17. Ratcliffe J, Brazier J, Palfreyman S, Michaels J. A comparison of patient and population values for health states in varicose veins patients. Health Economics. 2007;16:395-405.
18. Peeters Y, Stiggelbout A. Health state valuations of patients and the general public analytically compared: a meta-analytical comparison of patient and population health state utilities. Value in Health. 2010;13(306-309).
19. Pyne J, Fortney J, Tripathi S, Feeny D, Ubel P, Brazier J. How Bad Is Depression? Preference Score Estimates from Depressed Patients and the General Population. Health Services Research. 2009;44(4):1406-23.
20. Brazier J. Measuring and valuing mental health for use in economic evaluation. Journal of Health Services Research and Policy. 2008;13:70-5.
21. Williams A. The measurement and valuation of health: a chronicle. Discussion Paper 136, Centre for Health Economics: York Health Economics Consortium, University of York; 1995.
22. Rand-Hendriksen K, Augestad L, Kristiansen I, Stavem K. Comparison of hypothetical and experienced EQ-5D valuations: Relative weights of five dimensions. Quality of Life Research. 2012;21(6):1005-12.
23. Little M, Reitmeir P, Peters A, Leidl R. The impact of differences between patient and general population EQ-5D-3L values on the mean tariff scores of different patient groups. Value in Health. 2014;17:364-71.
24. Leidl R, Reitmeir P. A value set for the EQ-5D based on experienced health states: development and testing for the German population. Pharmacoeconomics. 2011;29:521–34.
25. Sprangers M, Schwartz C. Integrating response shift into health-related quality-of-life research: A theoretical model. Social Science and Medicine. 1999;48(1507-1515).

26. McTaggart-Cowan H, Tsuchiya A, O'Cathain A, Brazier J. Understanding the effect of disease adaptation information on general population values for hypothetical health states. Social Science and Medicine. 2011;72:1904-12.
27. Dolan P. Thinking about it: thoughts about health and valuing QALYs. Health Economics. 2011;20:1407-16.
28. Rowen D, Brazier J, Young T, Gaugris S, Craig B, King M, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. Value in Health. 2011;14(721-731).
29. Yang Y, Brazier J, Tsuchiya A, Young T. Estimating a Preference-Based Index for a 5-Dimensional Health State Classification for Asthma Derived from the Asthma Quality of Life Questionnaire. Medical Decision Making. 2011;31(281-291).
30. Rowen D, Mulhern B, Banerjee S, Van Hout B, Young T, Knapp M, et al. Estimating preference-based single index measures for dementia using DEMQOL and DEMQOL-Proxy. Value in Health. 2012;15(346-356).
31. Mulhern B, Bansback N, Hole A, Tsuchiya A. Using Discrete Choice Experiments with duration to model EQ-5D- 5L health state preferences: Testing experimental design strategies. Medical Decision Making. Forthcoming.
32. Mulhern B, Bansback N, Brazier J, Buckingham K, Cairns J, Devlin N, et al. Preparatory study for the revaluation of the EQ-5D tariff: methodology report. Health Technology Assessment. 2014;18(12).
33. Gold M, Siegel J, Russell L, Weinstein M. Cost-effectiveness in health and medicine. Oxford: Oxford University Press; 1996.
34. Sanders G, Neumann P, Basu A, Brock D, Feeny D, Krahn M, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. JAMA. 2016;316(10):1093-103.
35. Litva A, Coast J, Donovanc J, Eylesd J, Shepherde M, Tacchif J, et al. 'The public is too subjective': public involvement at different levels of health-care decision making. Social Science and Medicine. 2002;54:1825–37.
36. Mukuria C, Brazier J. Valuing the EQ-5D and the SF-6D health states using subjective well-being: A secondary analysis of patient data. Social Science and Medicine. 2013;77:97-105.
37. Fryback D. Whose quality of life? Or whose decision? Quality of Life Research. 2003;12:609–10.
38. McTaggart-Cowan H. Elicitation of informed general population health state utility values: a review of the literature. Value in Health. 2011;14(1153-1157).
39. Clarke A, Goldstein M, Michelson D. The effect of assessment method and respondent population on utilities elicited for Gaucher disease. Quality of Life Research. 1997;6:169–84.
40. Cunningham S, Hunt N. A comparison of health state utilities for dentofacial deformity as derived from patients and members of the general population. European Journal of Orthodontics. 2000;22:335-42.
41. Damschroder L, Zikmund-Fisher B, Ubel P. The impact of considering adaptation in health state valuation. Social Science and Medicine. 2005;61:267-77.
42. Stein K, Ratcliffe J, Round A, Milne R, Brazier J. Impact of discussion on preferences elicited in a group setting. Health and Quality of Life Outcomes. 2006;4:22.
43. Brazier J, Akehurst R, Brennan A, Dolan P, Claxton K, McCabe C, et al. Should patients have a greater role in valuing health states? . Applied Health Economics and Health Policy. 2005;4(4):210-08.
44. Robinson S, Bryan S. Does the process of deliberation change individuals' health state valuations? An exploratory study using the person trade-off technique. Value in Health. 2013;16:806-13.
45. Whitty J, Burton P, Kendall E, Ratcliffe J, Wilson A, Littlejohns P, et al. Harnessing the potential to quantify public preferences for healthcare priorities through citizens' juries. International Journal of Health Policy and Management. 2014;3:57-62.
46. Scuffham P, Ratcliffe J, Kendall E, Burton P, Wilson A, Chalkidou K, et al. Engaging the public in healthcare decision-making: quantifying preferences for healthcare through citizens' juries. BMJ Open. 2014;4.
47. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. Value in Health. 2012;15(1172-1181).
48. Karimi M, Brazier J, Paisley S. The effect of reflection and deliberation on health state values. Health Economics and Decision Science Discussion Paper 1610: University of Sheffield, UK; 2016.
49. Shah K, Devlin N, Mulhern B, van Hout B. Directly Eliciting Personal Utility Functions: A New Way To Value Health-Related Quality Of Life. Value in Health. 2015;18(3):A37.