**Comparing increments in utility of health. An individual-based approach**

**Abstract**

Many economic evaluations of health care changes rely on quality-adjusted life year (QALY) estimates. Notably, though, the QALY approach values health *states* rather than *changes* in health states. Hence, a gain in utility of health is only indirectly valued through an *ex ante* preference elicitation of health states and the subsequent subtraction of health state values from one another, rather than being valued directly. There is therefore an underlying assumption that individuals, from an ex ante perspective *ceteris paribus,* would be indifferent between equal utility increments from health states with different baseline utilities. We ran an experiment on a sample of UK individuals, using established preference elicitation methods to test an assumption underpinning the indirect valuation approach. The results indicate that the subtraction approach could lead to sub-optimal resource allocations and suggest that a new approach which values health changes directly would better reflect individual preferences. This paper provides the foundations for a method to achieve this.

**1. Introduction**

A cost-utility analysis involves a comparison of the quality-adjusted life years (QALYs) gained for a given intervention with the incremental costs where the QALY measure captures both quality and quantity gains (Drummond et al., 2005). Using the standard QALY procedure, the change in health utility resulting from a health care intervention is indirectly approximated through *ex ante* preference elicitation on health state and subsequent subtraction of health state values from one another (Nord et al., 2010). This subtraction method will in practice be carried out using QALYs based on average health state values for several individuals (often a sample of the general population). The method implicitly assumes that on average individuals, from an ex ante perspective *ceteris paribus,* would be indifferent between an incremental gain in their health utility from a health state associated with a higher utility and an equally sized incremental gain in their utility from a health state associated with a lower utility. If gains of equal utility increments are not empirically found to be equally preferable by the individual on average, an assumption underlying the subtraction method is called into question and might suggest that a new approach, one that values health changes directly, would better reflect individual preferences.

The aim of this paper is to develop a method that would allow us to measure individual-based preferences over utility increments from different baselines. Note that this is a different issue from the one of equity when the potential recipient are different individuals (see e.g. Dolan et al., (2005); Gyrd-Hansen (2004); Lancsar et al., (2011); Exel et al., (2015)). We elicit our data using existing utility scores and face-to-face interviews on a sample of UK individuals. By asking individuals to make direct comparisons between equal utility increments (from health states with different baseline utilities) we obtain rankings at the individual level, allowing for a direct test of an assumption which underlies the subtraction methodology. A test of the subtraction method is implicitly a test of the intra-person interval property as defined in Torrance (2006 p. 1074); “a gain of equal utility increments anywhere on the scale should be equally preferable *for the individual whose utilities are being represented.* For example, if an individual’s utilities are A = 0.2, B = 0.4, C = 0.6 and D = 0.8, the person should be indifferent to whether the change is from A to B or from C to D”. According to Torrance (2006) and to the best knowledge of the authors, the intra-person interval property has, so far, not been tested. Although not central to this paper, we show that our data has potential to measure the strength of such individual-based preferences and hence indicate the size of the potential bias when using the subtraction method. We leave to the end of the paper, a discussion as to how the findings in this paper may be translated into a health policy tool.

The paper proceeds as follows. In Section 2, we describe the empirical study and analytical approach. Section 3 presents the results whereas the final section reflects on the implications for future research and policy formation.

**2. The framework for comparing increments in utility**

In the experiment, we follow the suggestion by Torrance and test whether individuals looking through a “veil of ignorance”, not knowing which outcomes and choices might occur to them in the future, would agree in advance that a change for them from 0.2 to 0.4 is equally preferable to a change from 0.6 to 0.8 (Torrance 2006). Hence, we compare two *changes* at a time (Gain X and Gain Y) from two proposed levels of individual’s baseline utility of health (A and C). The changes are represented using either the visual analogue scale (VAS), the HUI Mark 3 (HUI:3) (Horsman et al., 2003) descriptive system or the EQ-5D-5L. Using the VAS, we ask respondents for an individual assessment of the changes directly. However, for the HUI:3 and EQ-5D-5L, the health state change is described using the descriptive system and associated individual scores are assumed to be the population scores generated by the HUI:3 algorithm (Horsman et al., 2003) and the official crosswalk EQ-5D-5L values for the UK ([http://www.euroqol.org](http://www.euroqol.org))). The ranking of the two changes constitute Stage 1 of our analytical approach. By keeping Gain X fixed and varying Gain Y until the individual identifies their indifference point between the two utility increments from different proposed baselines, we also illustrate (as a Stage 2) how strength of individual preferences can be elicited for different gains and baselines providing an indication of the size of the potential bias with using the subtraction method. EQ-5D is a widely used generic health status measurement method and in the UK, preference values are derived using the Time-Trade-Off (TTO) method. Throughout the paper, we consider the elicited preference values to be a proxy for individual utility even though only preferences measured using Standard Gamble(SG) are based directly on von Neumann Morgenstern utility theory. The TTO method is framed in a riskless world (Bleichrodt 2002) and as this paper takes the existing application of the QALY framework as a starting point, our method is designed in a world of certainty to correspond to this. However, it would be relatively straightforward to extend it to a world of uncertainty.

**2.1. Experimental design**

A “common” experiment was designed varying only in terms of the method used to represent health states; the VAS, HUI:3 or EQ-5D-5L. Due to the potential for fatigue, if subjects were asked to answer all variants, the experiment was applied on two separate samples; Sample I; students (from the Universities of York and Newcastle) and Sample II; members of the general public (Newcastle-upon-Tyne). Subjects in Sample I were randomised into either VASSor HUIS, whereas subjects in Sample II were randomised into either VASP or EQ5DP.

**TABLE 1 AROUND HERE**

**2.2. Experimental procedures**

A total of 170 subjects participated in the experiment. All experimental sessions were face-to-face interviews attended by only one participant and the interviewer. An example of the experimental instructions is included in the online Appendix. Prior to the experiment itself, a piloting phase was carried out in order to test for comprehension. Following the pilot, experimental instructions were further amended and more training exercises included. This process allowed us to verify that the experimental instructions were understood by as many participants as possible, a crucial feature of any new method and, as much as is possible, to establish that subjects understood the concept of a health *change*. Core procedures, questions etc. were identical across the samples. However, a more extensive introduction and explanation was required for Sample II to ensure they understood the task. Thus, while the experiment generates treatment-specific results, we refrain from any direct comparisons due to this difference in information communication. Nevertheless, we are still able to draw broad qualitative conclusions from the two sets of responses. After the warm up and training exercises, the actual preference elicitation exercise was carried out on both Sample I and II. This is described in the following section.

**2.3. Eliciting Preferences over Changes in Health states**

Table 2 contains the full set of incremental changes offered to respondents. In each question, the respondent was shown two health changes, Gain X and Gain Y, described in terms of an initial baseline health and a final health state. States corresponding to Gain X are labelled “A” and “B”, while those corresponding to Gain Y are labelled “C and “D”. The utility scores for these different states are contained in Table 2. It can be observed that the initial baseline state in Gain X is always better than the initial baseline state in the corresponding Gain Y. In all variants (e.g. EQ5D; HUI; VAS), six questions were asked. In the question identifiers, the subscript refers to the specific utility gain (offered in both X and Y) that a respondent was asked to compare i.e. Q10.25 means that, in Question 1, respondents had to choose between an increment of 0.25, in this case from baselines of 0.75 (A) and 0.5 (C). In Q20.25 and Q30.25, while the utility gains are the same it can been seen that the initial baseline states differ. The same principle applies in the Q40.5-Q60.5. In all questions, a respondent was asked to indicate which of X or Y offered the “better” change. These responses will be reported as Stage 1 results. Examples of the visual presentation of Gain X are shown in the online Appendix (for VASS+P, HUI and EQ5D)). It was explicitly emphasised that the respondent should imagine herself in the health states described and choose the better one for themselves personally, as opposed to choosing on behalf of others.

**TABLE 2 AROUND HERE**

Following this, different alternatives to health state D were presented to the respondent. If Gain X is preferred to Gain Y, D is improved (and hence Gain Y becomes better) while if Gain Y is preferred to Gain X, D is made worse (and hence Gain Y becomes worse.). We denote D\* as the health state that makes the respondent indifferent between X and (the “new”) Y, which we denote Y\*. D\* and Y\* will be reported as Stage 2 results and discussed later. To identify D\*, a total of 21 VAS cards were made available to the respondents (increments of 0.1). For HUIS ­and EQ5DP, 18 combinations were chosen to represent health related quality of life utilities at equal intervals of around 0.06, ranging from 0 to 1. These utility levels are reported in Table A1-2 in the Appendix and were generated by the HUI:3 algorithm and taken from the official UK values for the EQ-5D-5L. Care was taken to ensure that each of the 18 combinations dominated the adjacent lower card (i.e. either equal or better than for the corresponding health dimension). Due to this requirement, the utility of the initial baseline health states vary slight across the different variants. Baseline health states are marked in italics in Table A1-2. If the respondent did not explicitly indicate indifference between any two specific health gains i.e. did not identify a D\* health state, for the purposes of the subsequent analysis, we assume a point of indifference exactly halfway between the two health gains where the preference changed from X to Y or Y to X. The ordering of questions asked was randomised. Participants were encouraged to ask questions throughout the interview with respect to procedures and the interviewer was instructed to clarify any misunderstandings.

**2.4. Analytical approach**

We report the experimental data in two stages. Stage 1 is based on whether each subject was indifferent between Gain X and Gain Y or whether one was preferred. The main research question we set out to test is RQ1;

*RQ 1; Are individuals, on average, indifferent between “equal” increments in health from different baseline utility states?*

We test using unpaired two-sided t-tests, whether the proportion of subjects who stated that Gain X and Gain Y were equally preferred is significantly different from the proportion who indicated a preference for either X or Y. If subjects were not indifferent between the gains, we analyse individual preferences in more details in RQ1a +b.

*RQ1a; Do individuals on average prefer an “equal” increment in health from a poorer baseline utility state?*

We test using one-sided t-tests, whether the proportion of subjects who preferred Gain Y (i.e. an equivalent gain from a poorer starting point) was significantly different from the proportion of subjects who did not prefer Gain Y.

*RQ1b; Do individuals on average prefer an “equal” increment in health from a better baseline utility state?*

For completeness, we test, using one-sided t-tests whether the proportion of subjects who preferred Gain X (i.e. an equivalent gain from a better baseline) was significantly different from the proportion of subjects who did not prefer Gain X.

Taken as a whole, if we find that on average individuals rank “equal” increments differently, we can reject the intra-person interval property which raises some doubt regarding the subtraction method. In addition, we report and examine Stage 2 results i.e. individuals’ strength of preferences as elicited in the experiment. Hence, we report the average health state D\* identified in the two samples and the associated Gain Y\*.

**3. Results**

Tables 3 and 4 report the experimental Stage 1 data for the six questions. We first state the number of individuals who indicated strict preference for either Gain Y, Gain X or were indifferent. In both samples and across all health state measurement methods we can reject that Gain X and Gain Y were equally preferred with p-values <0.01.

**TABLE 3 AROUND HERE**

**TABLE 4 AROUND HERE**

Turning to RQ1a, we test whether subjects prefer a change from a poorer baseline utility of health to an equally sized change from a better baseline utility of health state. The results from both VAS­S and HUI­­S are clear and we cannot reject that these respondents prefer a change from a poorer health state. The results from Sample II are more ambiguous. In VAS­P, when the utility Gain X is 0.5 or above (Q4-Q6) we can draw the same conclusions. However, when the utility Gain X is 0.25 (Q1-Q3), we cannot reject based on a one-sided test that the subjects were indifferent between the changes. This pattern is not mirrored in EQ5D­­P­ where the only situation where subjects strictly preferred Gain X was from the worst utility state given in Q6.

In RQ1b, the one-sided alternative is that participants prefer a change from a good baseline utility to an equal change from a poor baseline utility state. The results from both VAS­S, VAS­­­P and HUI­­S are clear; this change is not preferred.However in the EQ5D­­P­ results we can identify that Gain X is strictly preferred or the subjects are indifferent between the health changes for 5 of the 6 questions, only in Q6 is Gain X not preferred.

Table 5 reports the Stage 2 results. For a Gain X of 0.25 (Q1-Q3), we see that this is perceived by individuals to be equivalent to a Gain Y\* ranging from 0.1 to 0.43 depending on sample, baseline and elicitation method. A Gain X of 0.5 is equally preferred to a Gain Y\* ranging from 0.24-0.54 whereas a Gain X of 0.75 is perceived to be equivalent to a Gain Y\* in the range of 0.29 to 0.55. In general, we can see that that Gain Y\* is smaller than the corresponding Gain X in almost all situations, but given the small sample sizes and different elicitation methods utilized, we will only draw broad qualitative conclusions and not apply statistical inference at this time.

**TABLE 5 AROUND HERE**

Overall, based on this, we can reject that subjects in this data set comply with the intra-person interval property. Based on the results of RQ1a-b, we can infer that the main reason for this is explained by RQ1a i.e. individuals prefer an “equal” increment from a lower baseline utility state. However, in EQ5D­­P some subjects express preference for a gain when the baseline health state is associated with higher utility

**4. Discussion and conclusions**

In this paper we set out to develop a method that would allow us to measure individual-based preferences over different utility increments. Our results show that in Sample I, there was a clear tendency to prefer an equal utility increment from an initially more severe utility of health state compared to a better baseline state. Overall, the VAS results were similar in Sample II, although the tendency to prefer incremental changes from a more severe baseline was not significant when the utility score of the baseline health state in Gain X was high. These results are similar to previous findings in other decision contexts where “social” values have been elicited. When potential recipients are different individuals and the questions relate more to equity concerns, it has been found that people prefer to give priority to severely ill patients (Nord, 1995; Ubel et al., 1999. In contrast, for the EQ5D­­­P results there was a tendency to prefer equal utility increments from the better baseline utility to incremental changes from an initially more inferior baseline. By varying one gain until the individual identifies indifference between two changes in individual’s utility of health from different proposed baselines, in Stage 2 we also illustrate how not only individual rankings but also strength of individual preferences can be elicited.

On the whole, we find that gains of equal utility increments are not found to be equally preferable by the individual. By violating the intra-person interval property, our data raises some doubt regarding the subtraction method which could lead to sub-optimal resource allocations. This suggest that a new approach which values changes in utility of health directly would better reflect individual preferences. Our study has laid the foundations for a method that could address this question and based on the Stage 2 results potentially determine a multiplier for changes in utility of health starting from a lower baseline level. To illustrate, imagine a comparison of two treatments which both deliver an increment in utility of 0.25. One treatment would deliver an increment from 0.75 to 1 for the individual, whereas the other would provide an individual increment from 0.5 to 0.75. Our results (as reported in Table 5, VAS­­S, Q10.25) would suggest multiplying the latter by the ratio between Gain X and Gain Y\* (i.e. (0.25/0.16 = 1.6)). In a cost-utility analyses the gain from the low baseline state would thereby be inflated to better reflect individual preferences. Hence our approach provides a non-monetary measure which could be used within the cost utility framework. Moving towards eliciting willingness-to-pay for a change in health instead could be another way of estimating the benefit of a change in health directly which could also take into account the results found in this paper.

In the present paper, our method was designed for a world of certainty, for reasons explained in the Introduction. However, our method can in principle be extended to an uncertain world, although this would undoubtedly be more cognitively challenging for respondents. In this case, respondents would be asked to choose between health gains in which the initial baseline health state and the new health states were described in terms of pairs of gambles, potentially resulting in a multiplier to be used in conjunction with EQ-5D scores elicited using the SG.

A possible explanation for the difference in significance and results found between the two samples could be that the VAS method was better understood by participants and, therefore, led to more consistent results. However, as a direct valuation technique, the VAS method has been criticised for lack of theoretical foundation as it involves no choices and thereby no trade-offs (Johannesson et al., 1996; Parkin and Devlin (2006)). We refrain from combining results from the different samples as one limitation of the current study is that population mean utility scores in EQ5D­­P and HUIS were assumed to represent the utility score for the individual participant whereas an individual assessment of the health changes was elicited using the VAS method. Preceding the elicitation task described here with an individual elicitation of utility scores would strictly speaking make the results of the different samples more comparable and is a topic for further empirical investigation. Our findings come with the caveat that they might be at least partially driven by biases in the heath state measurement method itself as well as by measurement error introduced by the method for comparing increments in particular when eliciting strength of preferences. It is a question for future research whether standard elicitation methods can be adapted instead to minimise measurement bias and directly take into account the findings in this paper and reflect them in the in the utility scores given to specific health states.

Based on our results generated from current health state elicitation methods, our findings might call into question the use of the subtraction method for assigning values to any change in health states delivered by an intervention, particularly for the range of health states within the study. Whether this holds for the entire range of utility improvements that could be delivered and/or is generalizable to the wider population is an open question, again one for further research. Likewise, a larger data set would be needed to test whether strength of preferences are significantly different across different baseline and measurement methods. Nevertheless, our study has laid the foundations for a method that could address this question empirically

In summary, in the QALY approach, a gain in health is only indirectly valued through *ex ante* preference elicitation on health states and subsequent subtraction of average health state values from one another. However, our results raised some doubt regarding this subtraction method. Overall, we find that gains of equal utility increments from different baselines are not found to be equally preferable by the individual. Hence our data violate an assumption underlying the subtraction method. In this case, a direct method would be preferable although, up to now, one has not been available. This paper provides the foundations for such an approach i.e. one that allows us to measure individual-based preferences over different utility changes directly. Adopting this method in policy making would therefore have the potential to make cost utility analyses of health care initiatives more representative of underlying preferences than current practice.

**References**

# Drummond, M., Sculpher, MJ., Torrance, GW., O’Brien, BJ., Stoddart, GL. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 2005.

1. Nord E., Enge AU., Gundersen, V. Is the value of treatment proportional to the size of the health gain? Health Economics 2010: 19; 596-607.

# Torrance, GW. Utility Measurement in Healthcare. The Things I Never Got To. Pharmacoeconomics 2006: 24(11); 1069-1078.

1. Dolan, P., Shaw, R., Tsuchiya, A., Williams, A. QALY maximisation and people’s preferences: a methodological review of the literature. Health Economics 2005; 14: 197-208.
2. Gyrd-Hansen, D. Investigating the social value of health changes. Journal of Health Economics 2004: 23; 1101-1116.
3. Lancsar, E., Wildman, J., Donaldson, C., Ryan, M., Baker, R. Deriving distributional weights for QALYs through discrete choice experiments. Journal of Health Economics 2011: 30; 466-478.
4. Exel, J.v., Baker, R., Mason, H., Donaldson, C., Brouwer, W. Public views on principles for health care priority setting: Findings of a European cross-country study using Q methodology. Social Science and Medicine 2015: 126; 128-137.

Horsman, K., Furlong, W., Feeny, D., Torrance G. The Health Utilities Index (HUI®): concepts, measurement properties and applications. Health and Quality of Life Outcomes 2003: 1:54.

EuroQol Group. EQ-5D. <http://www.euroqol.org/>

1. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. Health Economics 2002; 11: 447-56.
2. Nord, E. The person-trade-off approach to valuing health care programs. Medical Decision Making 1995: 15 (3); 201–208.
3. Ubel PA. How stable are people’s preferences for giving priority to severely ill patients? Social Science and Medicine 1999: 49(7); 1049-1055.
4. Johannesson, M, Jonsson, B and Karlsson, G. Outcome measurement in economic evaluation. Health Economics 1996: 5; 279-96.
5. Parkin D, Devlin N. Is there a case for using visual analogue scale valuations in cost-utility analysis? Health Economics 2006: 15; 653–664.