HESG/Aberdeen, July 99
version 1; June 2
approx. 4800 words
not for citation
comments welcome

Non-welfarist TTO

Aki TSUCHIYA, at7@york.ac.uk
Centre for Health Economics, University of York

Abstract:

The PTO method for evaluating hypothetical health states is observed to yield higher values than more conventional methods, and this is attributed to a reflection of people's concern for distribution. This paper reports on an empirical study that demonstrates that the TTO method adapted to a societal perspective (i.e. TTO from the perspective of PTO, but without the distributional element) will already yield higher values than conventional TTO, and therefore, distribution is not the sole element causing PTO values to be higher.

Acknowledgements

This study was carried out as a part of a much larger Research Project on Cancer Elimination Strategies (1996-), funded by the Ministry of Health and Welfare, Japan. Many people have offered help at various stages. Thanks are due to: Richard Cookson, Paul Dolan, Toshihiko Hasegawa, Paul Kind, Hidekazu Komatsu, Shuzo Nishimura, Kenji Miyazaki, José-Luis Pinto, Nigel Rice, Anne Spencer, Makoto Tamura, Alan Williams, and Satoshi Yamashita. The author is an Overseas Post-Doctoral Research Fellow (1998-2000) of the Japan Society for the Promotion of Sciences, and a Visiting Research Fellow to the Centre for Health Economics, the University of York. The usual disclaimer applies.

1. Introduction

The so-called PTO (person trade-off) method has been proposed as a way to elicit people's valuations regarding HRQOL (health related quality of life) from a "societal perspective", or people's "social values" (Nord, 1995; Murray, 1996; Pinto Prades, 1997). Nevertheless, there has been concerns raised against PTO, since the values thus elicited will not only reflect social values regarding the health states themselves, but will be confounded by distributional concerns: i.e. given that a fixed amount of health gain was to be distributed, should a large number of people receive a small gain each, or should a large number of people receive a small gain each? Distributional concerns should be in the aggregation phase (if ever), and not in the stage where preferences for different health states themselves are quantified (see for example Dolan, 1998; Tsuchiya, 1999). This paper reports on the possibility of amending the TTO (time trade-off) method to suit the social perspective, and thus to provide a method to elicit social values regarding HRQOL that will not be affected by people's distributional concerns.

The "perspective" of an evaluation relates to whose health or life it is at stake in the hypothetical situation the respondents are asked to consider. Roughly speaking, there are two possible perspectives[1]. One is a *personal* perspective, where the respondent is asked to imagine him/herself in a given health state, and is expected to make contingent decisions as a competent informed consumer in a free competitive market would. This perspective is compatible with the welfarist interpretation of CEA (cost-effectiveness analysis). The other is a *social*, or *impersonal*, perspective, where it is *not* the respondent but somebody else whose health and/or life is at stake, and the respondent is expected to make hypothetical proxy decisions either on behalf of the patient(s) or of the community at large. This perspective is compatible with the non-welfarist interpretation of CEA. Except for when PTO questions are employed, the conventional practice of HRQOL evaluation to date has employed the personal perspective. The central argument supporting this practice comes from the claim that each individual is the best judge for his/her own welfare, and that social welfare should be based on individual welfare as judged by him/herself (*viz.* welfarism).

Nevertheless, in a public health care resource allocation context there are two reasons for which it may be more preferable to divert from this perspective and to employ non-welfarism. The first is that the use of these figures will not be restricted to the respondents of the elicitation exercise him/herself. In fact, on most occasions the figures will be used to decide the treatment of those other than the respondents themselves. For example, under TTO, people who are quite willing to

---

[1] Other possibilities are intermediate perspectives, as the one employed in Richardson, Nord (1997)mentioned below.

trade their own life years off for a better HRQOL, may well be less willing to prescribe the same for others, because the preference of others is less clear, and shortening a life is irreversible once the death occurs. (If so, HRQOL values under the personal perspective may be smaller than those under the impersonal perspective.) Or, conversely, some people may have particular preferences for themselves, and while they may accept a health care policy based on judgements from a social perspective that did not happen to coincide with their own personal preferences, there is less reason for them to accept a policy based on personal preferences of others' which did not coincide with their own.

The second argument for an impersonal perspective, and thus non-welfarism, is that when health care services are not purchased by individual consumers but distributed through some form of non-market collective system, then this perspective can be expected to provide HRQOL values with a higher degree of "context validity". Here, the costs are borne by the public (either through taxes or social insurance premiums, or both), or borne collectively (through managed care arrangements) and the criteria for preferable resource allocation may not coincide with the one dictated under the market-oriented personal perspective. For instance, society may want health care resource allocation to reflect concerns that are not particularly welfarist, such as equal treatment of patients with equal treatment prospects, regardless of their willingness or ability to pay.

These should be no major concern at the practical level should the HRQOL values elicited under the personal perspective not be significantly different from those elicited under the impersonal perspective. Nevertheless, since the effect of perspective has not been studied much to date, and since most HRQOL evaluations are carried out under the personal perspective, it is important to investigate whether in fact there is any difference between the two perspectives.

There have been some studies that compare HRQOL scores obtained from PTO and other methods (see for example Tamura et al., 1996; Ubel et al., 1996; Pinto Prades, 1997). Nevertheless, the disagreement found between these two formats are composed of two elements: the difference that comes from the methods (or in other words, the medium of trade-off: e.g. persons vs. time), and the difference that comes from the perspectives (or in other words, the difference between welfarist values and non-welfarist values). This present paper is based on an empirical study that addressed both these issues the methods and the perspectives, and reports on the latter[2].

If we think about the relationship between PTO and conventional TTO, there are several missing links to bridge these two. Under conventional TTO, the trade-off in question is *intra*-personal: a typical TTO asks the subjects to trade-off his/her own quantity of life for his/her own quality of life.

[2] A separate paper that addresses the comparison between societal TTO and PTO is in preparation.

On the other hand, PTO asks to trade off quantity of life in one group of people for quality of life in another group of people: i.e. the trade-off in question is *inter*-personal. Somewhere in-between these, we can think of four more variants of TTO:

> trading off life-years of *one person* (other than the respondent) for HRQOL of *his/her own*[3],
> trading off life-years of *a group of people* for HRQOL of *their own*,
> trading off life-years of *one person* for HRQOL of *another*[4], and
> trading off life-years of *one group* for HRQOL of *another group*.

This paper reports on a comparison between the last variant and conventional TTO. Note three things. First, what is traded-off in this last variant is quantity of life in terms of time and hence this belongs to the TTO category. Second, it nevertheless pursues trade-offs across different individuals and as such, shares an important characteristic of the PTO: viz. the societal perspective. Third, unlike PTO, since the number of people in the two groups are the same and are kept constant, responses to this variant will *not* reflect people's distributional concerns.

2. Method

2.1. The scenarios

The paper is based on a postal survey that assign people different "scenarios" that asked for the same thing in different ways: i.e. some respondents are asked to value health state X in conventional TTO, while others will be asked to value the same state X in societal TTO. These scenarios will be represented as TTO:C and TTO:S respectively, and in essence will ask:

> <TTO:C>
> Will you prefer to live for 10 years in health state *D* and die, or will you prefer to live in full health for *d* years and die? (*D* is an adverse health state, and $d < 10$.)

> <TTO:S>
> Suppose there are 2 groups with equal numbers of people each. Those in Group 1 will either live for 10 years in adverse health state *D* with treatment and then die, or die shortly without treatment. Those in Group 2 will either live for *d* years in full health with treatment and die, or die shortly without treatment. If there are not enough resources to treat everybody, will you prefer to treat those in Group 1, or will prefer to treat those in Group 2? (*D* is an adverse health state, and $d < 10$.)

Further, apart from the two methods, two different "numeraires" are used[5]. The numeraire of a

[3] This was employed in Richardson, Nord (1997).
[4] A format very close to this was employed in Busschbach et al. (1993).
[5] The entire study defined 10 different scenarios in terms of method (TTO and PTO), perspective (personal and societal), numeraire, and frame (anchored at full health or at death): for more details, see

scenario in this study refers to the number assigned to the prospect of living in the evaluated health state: i.e. the length of time to be lived in the evaluated health state. The choice of numeraire under the TTO method concerns the issue of "constant proportional time trade-off" (Loomes, McKenzie, 1989). The QALY concept itself assumes that the rate of trade-off between quality and quantity (measured in terms of time) is constant, while in practice, preference obtained from the general public is known to violate this[6].

The "basic" numeraire for this study is 10 years' duration, and the "alternative" numeraire is 5 years' duration. Thus we have scenarios: TTO:Cb, TTO:Ca, TTO:Sb, and TTO:Sa. Note that the choice of the alternative numeraire in this study is relatively conservative in the sense that much larger changes may have been used. The main reason for this is that, as is presented below, there are seven hypothetical health states covering a wide range between death and full health, and a choice of a smaller alternative numeraire is likely to have made the trade-off insensitive to the difference between the health states.

2.2. The questionnaire

The ideal protocol to elicit and to compare HRQOL under the two different forms of TTO questions would be to conduct a series of interviews to elicit actual values of HRQOL for each scenario, and to compare them. Unfortunately, due to limited resources, interviews were not feasible and a postal survey was carried out. The extent of reliable information gathered through postal surveys is rather limited compared to that obtained through the most preferred face-to-face interviews, and the so-called "binary choice" technique is often used to amend for this[7].

Chapters 4 and 5 of Tsuchiya (1998).

[6] Buckingham et al. (1996) compares daily-TTOs, annual-TTOs, and lifetime (20 years) -TTOs; and the results show that the 1-year numeraire yields the highest HRQOL values and the 1-day numeraire the lowest. This suggests that, starting from a lifetime and diminishing the numeraire, people are willing to trade-off some quantity (measured in time) for quality of life at the beginning; but as the numeraire becomes smaller, there is a point, before reaching 1 year, beyond which poor quality counts as less and people will not trade quantity off for quality; and as the numeraire continues to diminish, there is another threshold, before reaching 1 day, beyond which people will resume the trade-off. This means that, when life expectancy is long enough, people will be ready to trade-off a portion of it for better health, but when life expectancy is only a couple of years, people will prefer to live in adverse health than to trade if off, and when it is as short as several weeks, being in better health matters more than living some days longer - and it all seems to make good sense. While the first threshold has been discussed in the literature, the second has not received much attention.

[7] Under this technique, each subject is presented with a choice between two options together implying a particular HRQOL: for example "to live in condition *D* for 10 years and die, or in full health for 8 years and die". Subjects are divided into several groups each of which will be given different trade-offs; i.e. some will be offered 7 years and others will be offered 9 years to live in full health, while the prospect of living in condition *D* is kept constant across all subjects. The proportion of those who choose to live with condition *D* will differ depending on the alternative years of life in full health: the shorter (longer) the length of time offered to live in full health, the more (less) likely people will choose condition *D*. Thus, by estimating the cumulative probability distribution across a range of alternatives, the median of the probability density function will be obtained. This will be the estimated HRQOL value of

In this study, nevertheless, the respondents were to be divided up into 10 groups by scenario (out of which 4 are relevant to this paper), and since following the binary choice technique will require a further subdivision with fewer samples in each sub-group, this line was not pursued. Note that the main objective of the study is the different *proportions* depending on perspective, and the actual numerical values are of secondary interest. Instead, "ternary choice" questions, explained below, were employed.

Each questionnaire[8] had an introductory section with an example given under the relevant scenario, followed by the evaluation of the seven hypothetical health states. The introductory section explained that the evaluation questions to follow all dealt with the same kind of situation laid out in the same manner, where only the health states and some of the numbers changed. It was also emphasised that there were no correct or wrong answers.

The following section is for the actual choices. For each health state, there was an preparative segment describing the evaluation problem together with a short description of the state in words, followed by a box summarising the alternatives with their respective outcomes, and finally a list of the alternatives, one of which the respondent was asked to select.

All respondents were presented with the same set of hypothetical health states, in the same order, with "ternary choice" questions implying the same HRQOL value for each health state. The only difference between respondent/scenario groups lies in the ways in which these ternary choice questions were asked: i.e. the scenarios. Ternary choice questions, with an additional option for those who found themselves indifferent between the standard binary alternatives, were employed in order to see the difference in the proportion of those who select the indifferent option across different scenarios.

2.3. The hypotheses

There are two null hypotheses in this study. They are that HRQOL values are:

$H_0 1$: not sensitive to perspective ("C" or "S"), and
$H_0 2$: not sensitive to numeraire ("a" or "b").

Note that the comparison between the personal TTO and impersonal TTO involves a difference in numeraire, and the effect of perspective per se cannot be elicited in this study. Nevertheless, for reasons suggested above, let us set the following alternative hypothesis regarding perspective is:

H1: HRQOL values are larger when the perspective is social than personal:
i.e. TTO:S > TTO:C.

---

condition *D*.

[8] A sample of the actual questionnaire (in Japanese) is available on request.

The expected effect of different numeraire in this range is that HRQOL will be larger when the numeraire is smaller (Buckingham et al., 1996). Thus, the alternative hypothesis for numeraire is:

H2: HRQOL values are larger when the numeraire is smaller: i.e. TTOa > TTOb.

This is because it is expected that, as the numeraire becomes smaller, the difference between full health and adverse health will be thought to count less.

2.4 The hypothetical health states and their HRQOL scores

Seven hypothetical health states, shown in the 2nd column of Table 1, were selected from the 22 "indicator conditions" of the Global Burden of Disease (GBD) version 5 (Murray, 1996). The particular seven states were chosen for their relative ease to imagine, and to cover the whole range of the spectrum between full health and death. Each health state was presented in lay terms, with a short description. The order of appearance, which was decided randomly, is shown in the 1st column of Table 1. The implied HRQOL values, on the last column, were taken from the corresponding "disability weights" (*ibid.*, p.39, Tb.1.3, 2nd column) [9].

In general, the hypothetical health states in HRQOL evaluation exercises can be presented in multi-attribute health status classification systems using, for example, the EuroQol or the HUI (Health Utilities Index): i.e. posing different attributes, or dimensions, of health and specifying a certain level in each. This will enhance the reliability of the responses by eliminating different effects that names of diseases, conditions or disabilities are likely to have on different subjects. It will also enable the elicited values to be added on to the common pool of HRQOL data, for use in practical settings. Nevertheless, multi-attribute descriptions were not used in this study for the following two reasons. First, since the questionnaires were administered in the form of a postal survey, the use of systematic descriptions may have resulted in information overload for the subjects, faced with unfamiliar decision making. Second, the study focuses on the expected

---

[9] There are two deviations from the original exercise. The first deviation concerns whether to fix the number relating to the comparator or the health state in question. The GBD exercise fixes the number of people of the comparator (i.e. those who will regain full health with treatment) at 1000, and the number of people who, with treatment, will be in the condition of question reflects the implied weight. On the other hand, the PTO scenario of the present study fixes the number of people in the health state in question at 1000, so that the implied weight is reflected in the number relating to the comparator. This rearrangement enables parallel scenarios to be formed in TTO: the TTO scenarios fixes the number of years for the prospect involving the sate in question at 10 years, as is done in most standard TTO questions. Second, the duration in the GBD scenario is 1 year, while the duration under the basic scenario of this study is 10 years (and 5 years for the alternative TTO scenario). This is because defining the corresponding TTO questions for the mild states with a 1-year reference seems highly fictitious, and therefore inappropriate for postal surveys. Note further that, since the burden of disease measures the loss of health in death-equivalents, the disability weights are anchored at 0 for full health and 1 for death. The values in Table 1 are equal to 1 – (the disability weight for the relevant health

different proportions between different scenarios, and hence the actual health states themselves are not of central importance. As such, this study does not aim to elicit cardinal values and hence generated data cannot contribute to the common pool of HRQOL data.

2.5 The Analysis

The main body of the obtained data can be demonstrated as stacked bar graphs which indicate, by scenario and/or by health state, the proportions of the ternary responses. For simplicity, the notation "QOL>d" indicates that the respondent chose to treat the patients in group1 (TTO:S), or to live the shorter life with the given condition (TTO:C). On the other hand, "QOL<d" indicates the opposite preference, and "QOL=d" indicates indifference. Two different analyses are carried out on this data, and each are described below.

2.5.1 Different proportions of responses by different scenarios

The first analysis is on the different proportions, and there are two assumptions:

A1: there is an underlying probability distribution for the particular proportion between the ternary alternatives under each scenario, and

A2: these distributions have the same shape for all scenarios, with the same variance.

These two assumptions mean that the underlying distributions for each of the scenarios are, for example, normal or logit distributions with the same variance, the issue here being whether their mean will differ from each other. The assumption of equal variance will be relaxed in the second analysis.

Thus, if HRQOL of the same hypothetical health state is larger under a particular scenario compared to another, this implies that the mean of the distribution underlying this scenario is larger than that of the other, and hence the proportion of respondents choosing QOL>d for a given level of *d* under this scenario will be larger than for the other, and the proportion choosing QOL<d will be smaller. The simplest statistical test for this is to perform a *Z*-test for the equality between two proportions for binomial distributions. Therefore, for each hypothesis, the proportion of QOL>d and QOL<d under one scenario will be compared to the corresponding proportions obtained under another scenario.

While this allows for statistical tests generating *p*-values, there are two major limitations. The first is that A2 may not be appropriate. The second is that, depending on where the given *d* lies relative to the mean of the underlying distribution, the ordering of the scenarios is incomplete. The second analysis attempts to overcome these two limitations.

---

condition).

2.5.2 The underlying probability functions

The conventional binary choice technique offers different values to different respondents without the QOL=d option, and, by using the proportion of replies corresponding to QOL<d under each implied value of *d*, estimates the (often logistic) cumulative frequency distribution underlying the responses. The median/mean of this distribution will be the elicited HRQOL value. For reasons stated above, the present study does not follow this technique, and a unique *d* for each health state is implied across all scenarios. Nevertheless, the use of ternary choice questions (i.e. the addition of the QOL=d option) generates data that, under certain strong assumptions, will allow a calculation of the underlying probability function. The second analysis will explore this approach.

The first assumption above is maintained, the second is dropped, and a third is adopted:

> A3: on average, the percentage of the "allowance" people will give for choosing "QOL=d" is the same across all scenarios.

A3 refers to the width or range in the numbers subjects will see as equivalent or acceptable. For example, if a subject felt that 10 years in condition D was equivalent to *8.5 years* in full health all other things being equal, and if a questionnaire offered a choice between 10 years in D and *8.3 years* in full health: would he/she select the QOL=d option, or the QOL>d option? In other words, is 8.3 close enough to 8.5 to justify a QOL=d choice, or is it significantly smaller? The response will depend on the width of the allowance people will give to the equivalence. If the allowance was 0%, then, since 8.5 > 8.3, the respondent will prefer the 10 years in condition D and will choose the QOL>d option. But if it was 5% to each side, then since the difference between 8.3 and 8.5 is less than 3%, the given trade-off will fall within the range of equivalence and QOL=d will be chosen. A3 assumes that whatever magnitude this allowance will be, on average, it is the same across all scenario-groups.

Relaxing A2 and adopting A3 enables a tentative calibration of the mean and variance of the underlying probability density function, by employing the logit model. For example, fixing the allowance arbitrarily at 90% (i.e. $\pm$5%) implies that the proportion of QOL<d will represent the cumulative distribution for $d \times 0.95$, and the proportion of QOL$\leqq$d will correspond to the cumulative distribution for $d \times 1.05$, and thus, the data set can be interpreted as one generated from a binary choice procedure (see Figure 1). Clearly, extrapolating to the entire probability distribution from two points of observation under strong and arbitrary conditions is contentious[10].

[10] The rank order of median and variance of scenarios is indifferent to the choice of the allowance (here, $\pm$5%). Thus, note that the arbitrariness does not refer to this choice, but to the assumptions regarding the nature of this allowance, and the distribution being symmetric. There is evidence that the underlying probability distribution of HRQO evaluation may not be symmetric, especially close to the

Therefore, the results are highly tentative, and though some inference will be made regarding the *ordering* of the average of the distribution function by scenarios, *cardinal* values are not addressed. Also note that, since the logistic curves are calculated from two points, the procedure cannot be referred to as an "estimation", and statistics concerning goodness of fit are irrelevant.

The averages and variances of the distribution thus derived are affected by several elements. Firstly, they are both affected by the choice of particular transformations (i.e. logit or probit). Secondly, the variances depend on the magnitude of the allowance. These two elements will affect all scenarios proportionally, and thus will not prevent ordinal comparison of averages and variances across them. Thirdly, given identical proportions of "QOL$\leqq$d" and "QOL<d" responses, variance of severe health states will be smaller relative to variance of mild health states. This is illustrated in Figure 1: the distance between $0.95d$ and $1.05d$ is relatively smaller (larger) for severe (mild) states, and thus for a given distribution of respondents, the cumulative function will be more (less) steep for severe (mild) states, implying smaller (larger) variance of the probability distribution. Due to these elements, reference to variance is limited. The treatment of averages will be with reference to the analysis of proportion under A1 and A2.

3. Results

3.1 Samples and respondents

Over 1100 names of those over 18 years of age were drawn randomly from the residents registry of Sakyo-ku, Kyoto. They were randomly divided into 10 groups, out of which 4 are related to this present study. After the questionnaires were sent by post, more than a dozen were sent back due to wrong addresses and unknown addressees, resulting in 411 questionnaires sent amongst the 4 groups related to this study. 10 days after the dispatch of the questionnaires, a reminder post card was sent, specifying a deadline. In the end, a total of 206 responses were obtained from these 4 groups, resulting in a response rate of 50.1 %.

The 2nd column of Table 2 indicates the proportion of different age/sex groups of the registry, the 3rd column is the same for all respondents, and the 4th column onwards is for the respondents of each of the 4 scenarios. χ^2-tests indicate that while respondents for scenario TTO:Sa, TTO:Sb, TTO:Ca are representative of both the residents registry and the respondents as a whole (p values between 0.22 and 0.823), respondents for scenario TTO:Cb are a poor representative of the registry ($p<0.01$) and of the respondents as a whole ($p=0.14$).

---

two anchoring points, death and full health.

3.2. Different proportions by scenario

3.2.1. Results by scenario, pooling across different health states

Figure 2 (1) shows the distribution of different responses for each scenario[11]. This is derived by averaging across the seven hypothetical health states. As can easily be seen, there are different patterns depending on the scenario. *Z*-tests for the equality of proportions in scenario pairs were carried out, and the results are shown in Table 3. The third through the fifth columns indicate the *Z*-values representing the degree of equality between the proportion of respondents with particular responses. Thus, the first number on the third column (0.578) represents the *Z*-value concerning the equality of the proportion of those who replied "QOL>d" under scenario TTO:Sa and under scenario TTO:Ca. The negative *Z*-value indicates that the proportion under the first scenario (TTO:Sa) is smaller than that under the second (TTO:Ca). The sixth through the eighth columns indicate one-sided *p*-values corresponding to the *Z*-values. Thus, given the proportions of those who replied "QOL>d" under TTO:Sa, the chances that the statement on the equality of these two proportions is correct is 56% (=0.282×2).

The ninth column is marked "< (>)" when the third column is negative (positive) *and* the fifth is positive (negative). This is because, under assumptions A1 and A2, if the proportion of those replying QOL>d under one scenario is smaller (larger) than under the other *and* the proportion of those replying QOL<d under the former is larger (smaller) under the latter, then the implied value of QOL under the former is smaller than under the latter. The two rows marked "?" implies inconclusiveness, and suggest that, at least for these scenario pairs, assumption A2 (equality of variance of underlying distribution) may not have been appropriate.

3.2.2. Breaking down to individual health states

Figure 2 (2) to (8) show the different proportions by each of the seven hypothetical health states. Table 4 illustrates the results of *Z*-tests and *p*-values on these: since the entire result (four scenario pairs for each of the seven health states, under each of the ternary choices) is too large to present, this offers a somewhat arbitrary summary. It indicates what, by the same criterion as in Table 3, the symbol corresponding to the ninth column there is, and whether this is accompanied by a *p*-value less than 0.1.

3.3. The underlying probability functions

3.3.1. By scenario, pooling across health states

---

[11] The percentages are the proportion of actual responses, and not necessarily of the total number of respondents. Most questions had a small percentage of blank responses.

The median of the underlying probability distribution for each scenario is calculated by logit transformations of the proportion of replies QOL<d and QOL≦d, assuming a width of allowance of 5% to both sides. For this procedure, A2 is dropped and A3 (equality of allowance for "QOL=d") is adopted. The results are illustrated in Figures 3 and 4. There are two things to note. Firstly, the ranking based on the implied weights is completely consistent with Table 3 (and with the actual observed distribution shown in Figure 2), and further enables additional indications regarding the scenario pair (TTO:Sb and TTO:Cb) where the orderings could not be discerned by the *Z*-tests. Secondly, the figures shows A2 is likely to have been inappropriate for some of the scenario pairs.

3.3.2. By health states, within scenarios

The underlying probability functions were calculated in the same way for each health state within each scenario (graphs not shown). Then, these were compared for particular scenario pairs. The third through the ninth column of Table 4 shows the difference in the medians thus calculated under each health state, where the medians are rounded to the closest second decimal point. The tenth column indicates the results of a sign test in terms of one-sided *p*-values calculated from the cumulative percentage of the binomial distribution, $r \sim B(N, p)$, where r is the number of negative signs observed, N is the number of health states, and $p = 0.5$. Thus for example, the number "0.063" on the tenth column for the first scenario pair demonstrates that, supposing that the two scenarios have the same mean, the probability of actually observing six negatives out of the seven health state pairs is 6 %: i.e. this corresponds to the chances of the statement rejecting the null hypothesis $H_0 1$ being incorrect. While the *Z*-test and the logistic curves show the *magnitude* of the possible difference between a given scenario pair, this sign test will indicate how *persistent* the difference is across various health states.

3.4. Results concerning the first hypothesis

The first hypothesis is on the relevance of perspective. This is examined by comparing:

TTO:Sa and TTO:Ca, and
TTO:Sb and TTO:Cb.

As regards numeraire *a* (duration = 5 years), the analysis on proportions indicate TTO:Sa > TTO:Ca, both when data are pooled across the seven hypothetical health states (Table 3), and when they are analysed separately (Table 4). This is supported by the logistic curves when all states are pooled (Figure 3-1), and when states are broken down (Table 4).

The results are less clear for numeraire *b* (duration = 10 years). The results of the *Z*-test for pooled data are inconclusive, and those for individual health states largely suggest the same. The logistic curves (Figure 3-2) are similar to those under numeraire *a*, and the sign test indicate TTO:Sb > TTO:Cb, but statistical significance is weak.

The observed results are that, when the duration is 5 years, societal TTO yields larger values compared to conventional TTO, and the same may also be said of a duration of 10 years, though this is less certain. The variance of the responses differ between societal and conventional TTO, and somewhat surprisingly, societal TTO has a smaller distribution.

3.5. Results concerning the second hypothesis

The second hypothesis concerns the effect of numeraire, and is examined by comparing:

TTO:Sa and TTO:Sb, and
TTO:Ca and TTO:Cb.

The social perspective indicates the anticipated difference (numeraire *a* > numeraire *b*) in all four analyses (Tables 3 and 4, Figure 4-1, and Table 5). Statistical significance varies across different health states and analysis: Table 4 indicates that there are some health states where the relationship is clear, but the results of the signs test is poor. The logistic curves suggest that A2 (equal variance) is appropriate, and the difference is likely to be small.

Regarding the personal perspective, the analyses of proportions and the sign test on the logistic curves on individual health states suggest the opposite relationship (numeraire *a* < numeraire *b*, Tables 3, 4 and 5), but statistical significance is poor. Again, the logistic curves suggest that A2 (equal variance) is appropriate, and the difference is likely to be small.

To summarise, while there are some indications that a difference in the duration of 5 years and 10 years may affect TTO valuations of hypothetical health states, the difference detected in this study is small.

4. Conclusion

The main conclusion to be drawn from this study is that in valuation of hypothetical health states, perspective matters, and the magnitude of different perspectives is larger than the magnitude of different durations of 5 years and 10 years. It has been observed that PTO values yield high values relative to conventional TTO from the personal perspective, and the difference has been attributed to the distributional element expected to be present in PTO. It is controversial whether a valuation method should reflect people's concern for distribution. However, this study has shown that the distribution element is not the sole element causing PTO values to be higher.

Values obtained from societal TTO, which shares the impersonal perspective of PTO but does *not* share the distributional element, are already larger than those obtained from conventional personal TTO.

References

Buckingham, J.K., Birdsall, J., Couglas, J.G. (1996), Comparing three versions of the time tradeoff: Time for a change? *Medical Decision Making, 16*: 335-347

Busschbach, J.J.v., Hessing, D.J., de Charro, F.T. (1993), The utility of health at different stages in life: A quantitative approach. *Social Science and Medicine, 37*(2): 153-158

Dolan, P. (1998), The measurement of individual utility and social welfare. *Journal of Health Economics, 17*: 39-52

Loomes, G., McKenzie, L. (1989), The use of QALYs in health care decision making. *Social Science and Medicine, 28*(4): 299-308

Murray, C.J.L. (1996), Rethinking DALYs. In Murray, C.J.L. and Lopez, A.D. (eds.), *The Global Burden of Disease* : Harvard University Press

Nord, E. (1995), The Person Trade-Off approach to valuing health care programes. *Medical Decision Making, 15*(3): 201-208

Pinto Prades, J.L. (1997), Is the person trade-off a valid method for allocating health care resources? *Health Economics, 7*: 71-81

Richardson, J., Nord, E. (1997), The importance of perspcetive in the measurement of quality-adjusted life years. *Medical Decision Making, 17*(1): 33-41

Tamura, M., Nozaki, M., Fukuda, T. (1996), An empirical study on utility assessment methods for quality adjusted life years. *Japanese Journal of Health Economics and Policy, 3*: 87-103 (in Japanese)

Tsuchiya, A. (1998), *Economics and social value of health: An exploration of the non-welfarist approach.* Unpublished PhD thesis, Kyoto University

Tsuchiya, A. (1999), *Does CBA really have a better welfare economic foundation? - Going back to basics, and re-examining welfarism.* Paper presented at the HESG, Birmingham

Ubel, P.A., Loewenstein, G., Scanlon, D., Kamlet, M. (1996), Individual utilities are inconsistent with rationing choices: A partial explanation of why Oregon's cost-effectiveness list failed. *Medical Decision Making, 16*(3): 108-116

Tables

Table 1: The 7 health states, their HRQOL values and their order of presentation

order	health state (D)	abbreviations	QOL (d)
4	vitiligo on one third of face, no itching or pain	Vitiligo	0.980
6	severe sore throat, difficulty swallowing	SSTh	0.923
1	below-the-knee-amputation, with crutches	BKA	0.719
5	deafness	Dfness	0.667
3	blindness	Bldness	0.376
2	severe migraine, no temperature or nausea	Migrn	0.262
7	quadriplegia, no pain	Qdplg	0.105

Table 2: The respondents: share of age/sex groups (%)

	registry	all respnds	TTO:Sa	TTO:Sb	TTO:Ca	TTO:Cb
m <30	24.2	16.5	17.0	15.7	15.6	18.2
f <30	19.1	16.5	20.8	9.8	20.3	13.6
m 30 to 69	20.3	28.8	28.3	29.4	29.7	27.3
f 30 to 69	22.7	20.8	20.8	27.5	23.4	9.1
m 70+	5.0	6.1	1.9	5.9	6.3	11.4
f 70+	8.8	11.3	11.3	11.8	4.7	20.5
total	100	100	100	100	100	100
n	146687	212	53	51	64	44

Table 3: *Z*- and *p*-values (one-sided) of the equality of response distributions

hypo-thesis	scenario pair		*Z*-values ‡			*p*-values			ord ¶
			QOL>d	QOL=d	QOL<d	QOL>d	QOL=d	QOL<d	
$H_0$1 †	TTO:Sa	TTO:Ca	0.578	1.609	-1.697	0.282	0.054	0.045	>
	TTO:Sb	TTO:Cb	-0.409	0.879	-0.317	0.341	0.190	0.376	?
$H_0$2 †	TTO:Sa	TTO:Sb	0.712	0.628	-1.079	0.238	0.265	0.140	>
	TTO:Ca	TTO:Cb	-0.261	0.041	0.194	0.397	0.484	0.423	<

† $H_0$1: HRQOL values are independent of perspective.
$H_0$2: HRQOL values are independent of base-line numeraire.

‡ A positive (negative) *Z*-values implies that the share of those replying, for example, "QOL>d" is larger (smaller) under the first of the two scenarios than under the second.

¶ ord: the relative size of the medians of the two scenarios.
>: the first of the two scenarios has a larger median than the second.
<: the first of the two scenarios has a smaller median than the second.

Table 4: *Z*- and *p*-tests by individual health states

		Vitiligo	SSTh	BKA	Dfness	Bldness	Migrn	Qdplg
TTO:Sa	TTO:Ca	e f	a f	e f	e	a	a	a f
TTO:Sb	TTO:Cb	b	e	e	e f	e f	e	c
TTO:Sa	TTO:Sb	a f	a	bf	e f	b	a	a
TTO:Ca	TTO:Cb	e	d	e	c f	d	d f	d

legends:

a) the *Z*-test suggests that first of the scenario pair has a larger value, and at least one relevant *p*-values is smaller than 0.1
b) the *Z*-test suggests that first of the scenario pair has a larger value, but statistical significance is low
c) the *Z*-test suggests that first of the scenario pair has a smaller value, and at least one relevant *p*-values is smaller than 0.1
d) the *Z*-test suggests that first of the scenario pair has a smaller value, but statistical significance is low
e) the results of the *Z*-test are inconclusive
f) QOL=d has a *p*-value smaller than 0.1

Table 5: The differences in the mean HRQOL between scenario pairs, by health state † ‡

hypo-thesis	scenario pair		Vitiligo	SSTh	BKA	Dfness	Bldness	Migrn	Qdplg	sign test §
$H_0 1$ ¶	TTO:Sa	TTO:Ca	0.10	0.26	0.25	0.01	0.05	-0.04	0.01	0.063
	TTO:Sb	TTO:Cb	0.05	0.07	0.07	-0.01	0.06	0.01	0.00	0.109
$H_0 2$ ¶	TTO:Sa	TTO:Sb	0.08	-0.04	0.04	0.00	0.01	-0.07	0.01	0.344
	TTO:Ca	TTO:Cb	0.03	-0.23	-0.14	-0.02	0.02	-0.02	0.00	0.109

† The calculations are based on medians of the logistic curves.

‡ The negative sign indicates that the median of the first scenario is smaller than of the second.

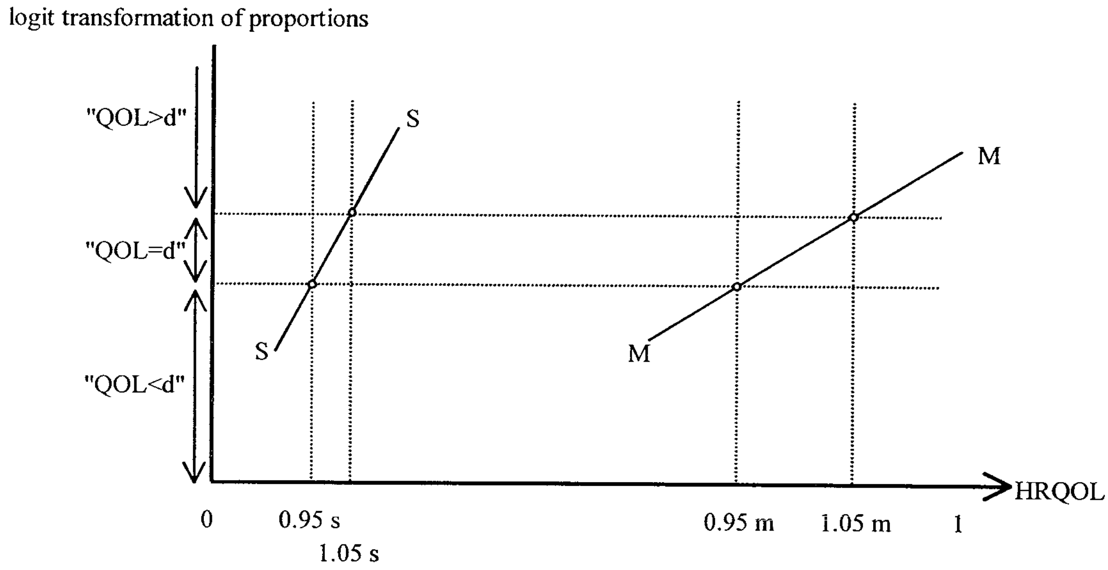
¶ $H_0 1$: HRQOL values are independent of perspective.

$H_0 2$: HRQOL values are independent of base-line numeraire.

§ The sign test: the chances of the same number of negative sings as the observed actually occurring, given the null hypothesis that the two scenarios have the same mean. This is calculated as: min {cumulative of binomial distribution, 1-(cumulative of binomial distribution)}.

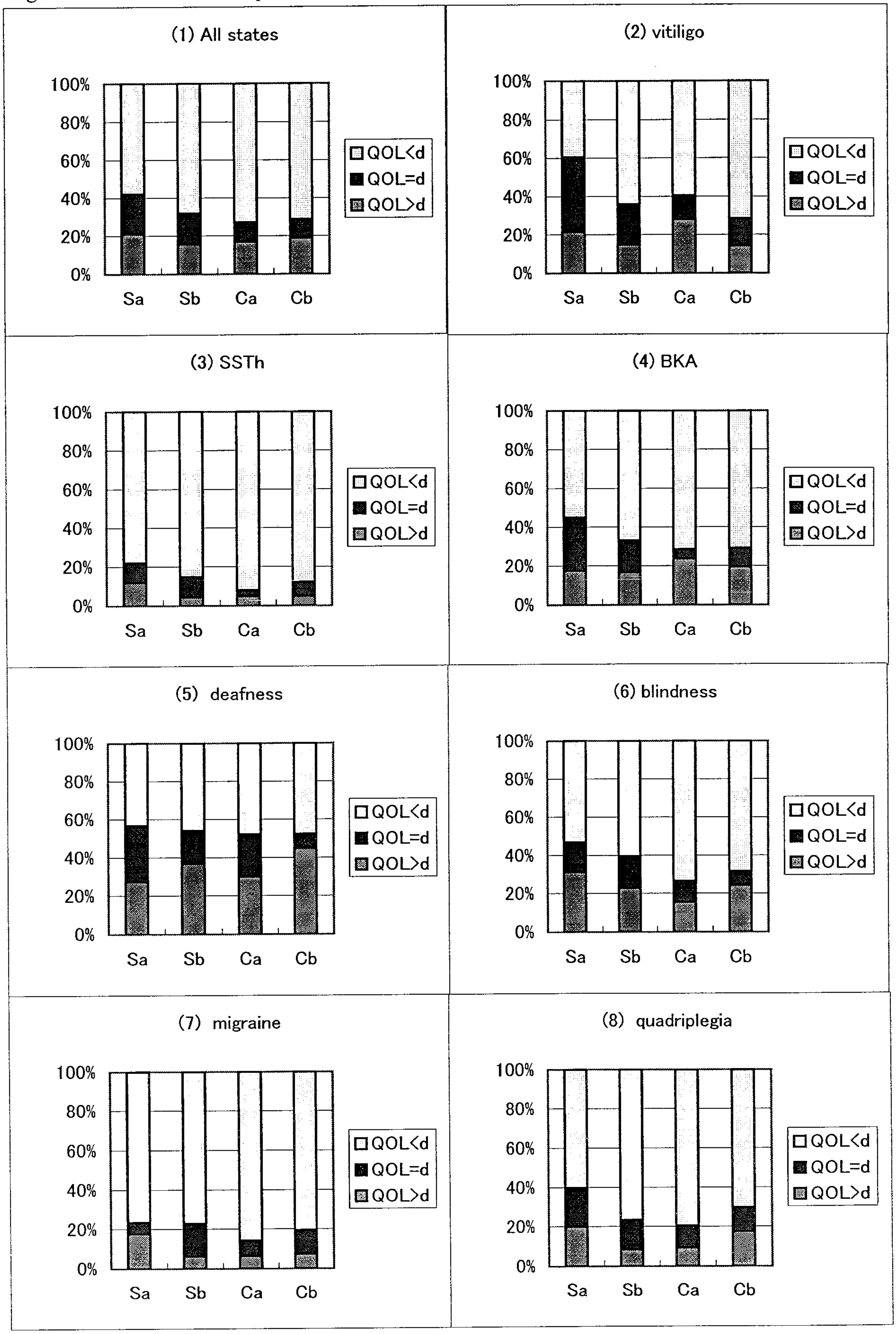
Figures

Figure 1: Calibration of cumulative frequency functions by logit transformation



s: implied weight for a severe health state
m: implied weight for a mild health state
s < m, and therefore (1.05s - 0.95s) < (1.05m - 0.95m)
SS, MM: cummulative frequency function for a severe and a mild health state

Figure 2: Distribution of responses



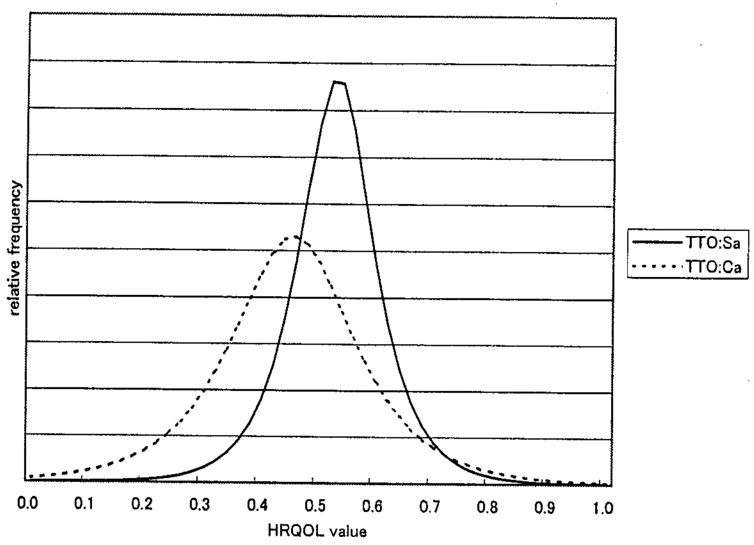


Figure 3(1): Hytpothesis 1(a)

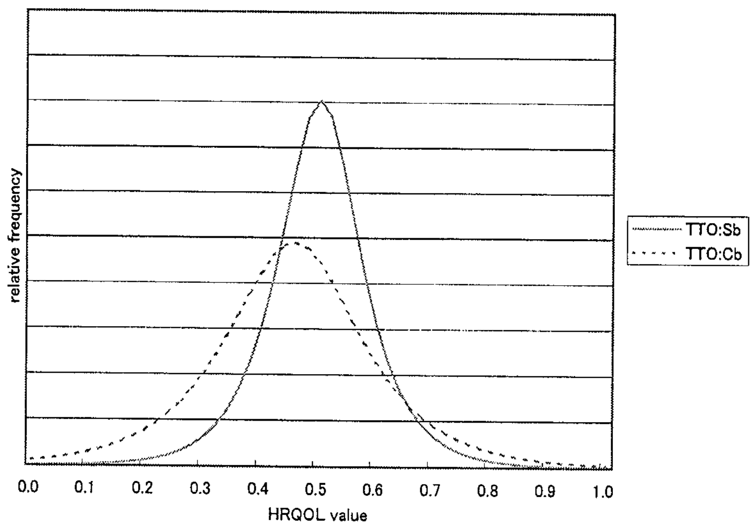


Figure 3(2): Hypothesis 1(b)

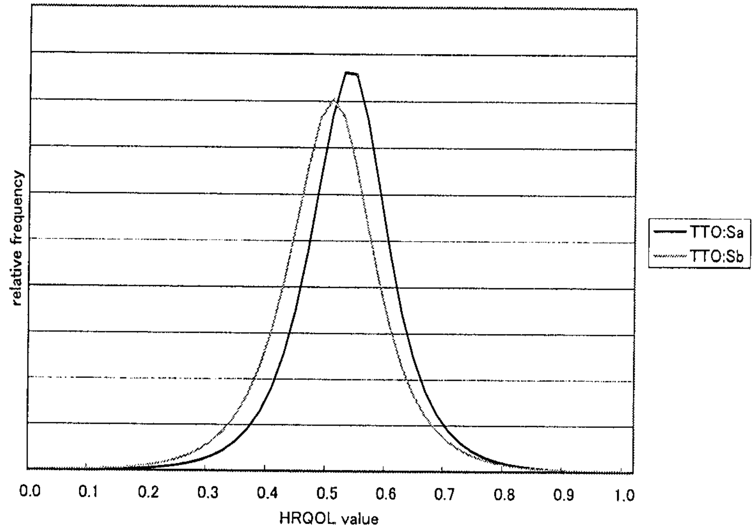


Figure 4(1): Hypothesis 2(S)

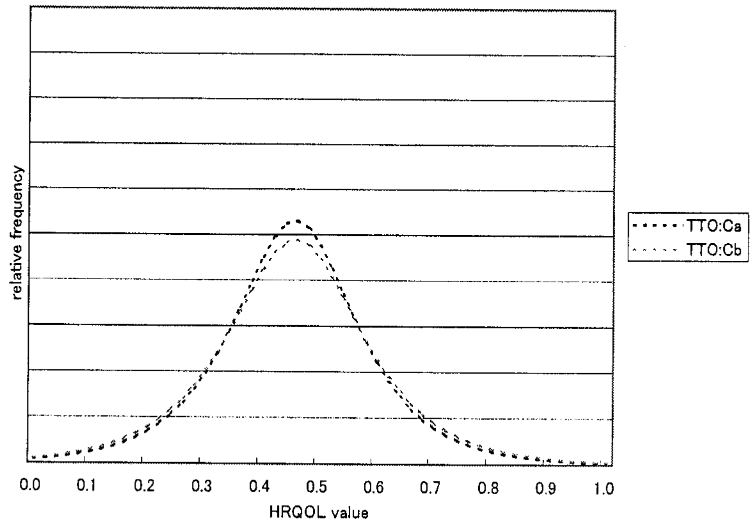


Figure 4(2): Hypothesis 2(C)