ASSESSING TECHNOLOGIES AT THE END OF LIFE: A REVIEW OF EMPIRICAL EVIDENCE.

REPORT BY THE DECISION SUPPORT UNIT

Simon Dixon, Louise Longworth, Allan Wailoo
School of Health and Related Research, University of Sheffield

26 November 2009

NICE Decision Support Unit
ScHARR, University of Sheffield
Email: dsuadmin@sheffield.ac.uk
Tel: 0114 222 0736

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) is a collaboration between the Universities of Sheffield, York and Leicester. We also have members at the University of Bristol, London School of Hygiene and Tropical Medicine and Brunel University. The DSU is commissioned by The National Institute for Health and Clinical Excellence (NICE) to provide a research and training resource to support the Institute's Technology Appraisal Programme. Please see our website for further information www.nicedsu.org.uk

The production of this document was funded by the National Institute for Health and Clinical Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

CONTENTS

1. INTRODUCTION

In January 2009, supplementary advice was issued to the NICE Technology Appraisals Committees in relation to treatments that extend life for those with short life expectancy. The advice came as the Institute's response to the recommendations made in the Richards Report (Richards, 2008) into patients that choose to pay for drugs not recommended for use in the NHS. Specifically, it was recommended that the Department of Health should work with NICE to find a way of making drugs that would not be considered cost effective available in the NHS (Recommendation 5). This was not a recommendation based on scientific evidence but "a common perception that the value that society places on supporting patients nearing the end of their life is not sufficiently reflected in assessing the cost-effectiveness of new drugs." (section 5.21). In particular, the views sought to inform the review were highly selective both in terms of the methods for identifying and including stakeholders and the method for eliciting their views.

The NICE supplementary advice to the Appraisal Committees, updated in July 2009, allows greater flexibility in recommendations for treatments with incremental cost effectiveness ratios (ICERs) in excess of £30,000 per QALY, the upper end of the threshold range operated by NICE. To be eligible to be considered under this advice, treatments must meet three criteria:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months and;
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared to current NHS treatment, and;
- The treatment is licensed or otherwise indicated, for small patient populations.

In addition, when taking these criteria into account, the Committee must be persuaded that the estimates of the extension to life are robust and that the assumptions used in the reference case economic modelling are plausible, objective and robust.

Whilst many think of this supplementary advice as increasing the acceptable threshold ICER for such treatments, the advice indicates that this is not so and instead focuses

on the possibility for giving greater weight to the benefits of the intervention than are reflected in the standard approach. Indeed, since the threshold represents the opportunity cost of implementation i.e. the value of what is displaced in the NHS, a higher threshold is not logically consistent. There are two ways in which the benefits generated by treatments meeting the criteria may be higher than is reflected by current estimation methods. Firstly, at the individual patient level, it may be the case that quantity of life is more important relative to quality of life than is the case in other situations. In this situation, no departure from the principle of QALY maximisation embodied in NICE's current methods is required. The only claim is that in the end of life (EOL) situation there may be reasons to think that current methods do not measure QALYs accurately. This efficiency element to EOL requires a more accurate means of assessing health state utilities than is currently the case through the tariffs associated with instruments like EQ5D.

At the societal level, it may be the case that the general public would prefer to assign greater weight to the (accurately measured) health benefits to EOL patients than for equivalent health benefits in other patient groups. This is typically thought of as an equity argument. This entails a departure from the principle of QALY maximisation although it should be recognised that the reasons society may wish to attach greater weight to certain groups of patients or to particular treatments may not be for equity reasons. For example, society may prefer to give greater weight to those of working age because of the economic benefits this may generate for the wider economy.

The aim of this document is to review evidence relating to these two issues of benefit measurement. We focus on the criteria relating to short life expectancy and life extension. Substantial evidence already exists in relation to the issues associated with rarity (McCabe et al. 2005, 2006, 2007), including work undertaken by NICE (NICE Citizens Council 2004).

We also recognise that there are other means by which greater benefits could be assigned to end of life treatments. For example, there are potentially positive externalities transmitted to friends and relatives associated with prolonging the life of a terminally ill person. Just as it is suggested that the patient 'savours' or 'craves' life more, the friends and relatives may place greater value on the extended life. Whilst this is plausible, an understanding of how to measure and value these aspects of

benefits beyond the patient is not well developed and is distant from current NICE methods in relation to assessment of all interventions, not just those that fall under the end of life criteria.

We have reviewed empirical evidence in relation to both individual and societal valuations and considered the appropriateness of existing methods to the specific situation outlined by the end of life criteria.

2. INDIVIDUAL UTILITY

2.1 VALUATION METHODS

There are several methods that may be used to elicit valuations of health states from individuals. Of these methods, two choice based approaches are most widely used: the Time Trade-Off (TTO) and Standard Gamble (SG). The NICE methods guide explicitly recommends the EQ5D instrument where available for which TTO based preference weights exist, and direct elicitation by TTO where EQ5D is not available.

The TTO obtains health state valuations by asking an individual to choose between a period of time (*y*) in the health state to be valued (*i*) and a shorter duration (*x*) in a better health state (usually full health) followed by death. The shorter period is altered until, for states considered better than death, a point of indifference is reached. At this point the utility for health state *i* is simply the ratio *x/y* (provided the reference is full health).

The SG method obtains valuations by asking respondents to consider the choice between a certain option (a set period *x* in health state *i*) and a risky one. The risky choice has a probability (*p*) of full health for the same period *x* and an inverse probability (*1-p*) of immediate death. The probability *p* is altered until a point of indifference is found at which point the value of health state *i* is equal to *p* (for states considered preferable to death).

Given the context of the NICE end of life supplementary advice, we focus our review on studies that have focussed on the examination of two assumptions employed when calculating quality adjusted life years; the constant proportional trade-off (CPT)

assumption and the utility independence (UI) assumption. CPT is relevant in the context of using time trade-off values, and assumes that the rate at which an individual trades off length of life for improved health is constant, irrespective of the duration of life in the health state of interest. In other words, the ratio *x/y* is constant irrespective of the value of *y*. UI is relevant in the context of using standard gamble values, and assumes that the utility value derived from the gamble (*p*) is independent of the duration used for the gamble (*x*).

It is important to recognise that the examination of the CPT and UI literature has not been undertaken as a test of the CPT or UI assumption per se, but as a test of proportionality or independence in the presence of death. However, the most common tests of CPT and UI to date have been undertaken by trading off the duration of ill-health that leads to death[1] . As such, we would argue that any deviation from proportionality may be attributable to either violation of CPT or the contextual effect of death. Indeed, this contextual effect is cited in one of the studies reviewed as a reason for examining the use of alternative durations within the TTO exercise. When examining the validity of using a 12 week duration leading to death which was indicated by their study design, Cook and colleagues (1994) argued that:

> "…rating such an option would be so seriously distorted by the imminence of death that it is unlikely that answers would bear any relation to the utility of the health state revealed in another context. The pragmatic solution to this problem was to measure the chronic health state for two substantially different periods before death and to determine empirically whether there was a quantitatively significant problem when these dissimilar time periods were used to evaluate the health state"
> (Cook et al, p159).

2.2 REVIEW METHODS

The review focuses on empirical evidence that utility values change near the end of life when either SG or TTO methods are used, concentrating on CPT and UI. The literature search is based on a previous review of the CPT and UI literature undertaken by Tsuchiya and Dolan (2005). They identified 11 studies that examined

[1] Other formulations of the TTO are possible that do not vary the time leading up to death as the basis for calculating utilities. For example, Buckingham et al (1996) examined the possibility of giving up time within each day for remaining life or within each year of remaining life as opposed to reducing the duration of remaining life.

these two assumptions, and so we have updated this search by identifying all studies that have cited either the Tsuchiya and Dolan paper or any of the CPT/UI papers that they themselves identified. This search identified 1179 potentially relevant papers.

The inclusion criteria used to select papers were:

- Empirical investigation of health state values
- Use of either SG or TTO methods to value health
- If TTO is used, the time to death is traded
- If SG is used, the time period over which the health state is valued ends in death

We extracted descriptions of the studies from relevant papers and, where possible, summary statistics relating to the health state valuations for the different scenarios.

2.3 REVIEW RESULTS

Applying the inclusion criteria removes two papers from the Tsuchiya and Dolan review; Buckingham et al (1996) do not trade time to death, and Dolan (1996) uses visual analogue scales to rate health states.

Applying the criteria to the identified abstracts by two of the authors (AW and SD) yielded approximately 50 papers that were ordered, and from this 8 papers have been added to the reduced Tsuchiya and Dolan review. This produces an evidence base of 17 papers (see Table 1).

Most studies were undertaken in small convenience samples of less than 100 participants. Seven studies examined durations less than two years (Cook et al 1994, Sackett et al 1978, Schulz 2003, Bala 1999, Franic 2003, Shumway 2005, Stigglebout 1995). It is also worth noting that only one study used between sample testing (Bleichrodt et al 1996), with all others only testing within a single sample with the same participants giving responses to shorter and longer durations.

The numerical values for the different time horizons presented in the papers are given in Table 2. For 5 studies, the results were not presented numerically and so can not be presented. Given the heterogeneity of the studies in terms of health states, durations, participants and the procedure relating to the actual elicitation, a quantitative synthesis of results can not be undertaken.

In general, the results tended to show higher mean utility values for short durations. Whilst the study sample sizes are generally small it should be noted, however, that not all of these differences were statistically significant with some studies finding lower mean values as statistically significant.

In qualitative terms we would consider three points to be of note. Firstly, the differences in utilities between longer and shorter durations are generally small (irrespective of statistical significance of these differences) with most mean differences being less than 0.03. Secondly, it appears that more severe health states produce greater differences. This is partly informed by within study differences (e.g. Kirsch 2000 and Franic 2003), however, the differences between studies are confounded by the fact that several are among the few with durations less than 2 years. Thirdly, the study by Torrance appears to be an outlier, showing much greater differences than all other studies. However, all of these observations need to be treated with caution as they are clearly not a representative sample of all studies identified, and summarised in Table 1.

2.4 DISCUSSION

Overall, the studies show an inconsistent set of results, with some reporting higher utilities for shorter durations and other reporting the opposite. Those which report the results in a manner that allows summary statistics to be reported in Table 2 show a more consistent pattern of higher values where shorter durations are used. However, this is a biased sub-sample of all the studies identified as 3 of the 5 studies that did not present results in a way that could be summarised in Table 2 found lower utilities for shorter durations.

A formal meta-analysis of findings is not possible due to the differences in study designs. We would also argue that it would be imprudent to synthesise the evidence

qualitatively and attempt to suggest that the balance of evidence favours one conclusion or another. However, we would venture to say that if one were to consider the evidence to be supportive of higher utility values for shorter durations, the effect of shorter durations is small – typically, less than 0.03.

Comparison with Tsuchiya and Dolan

The review has focused on TTO/SG valuations where the health state whose duration is adjusted leads to death. This is because the elicitation technique and the context most closely replicates the decision making situation under investigation. This excludes some studies that were included in the Dolan and Tsuchiya. Of these, the paper by Dolan (1996) is worth noting as it does fit the context of end of life (but does not utilise a choice-based method) and has several important features Firstly, it is based on EQ5D health states (and is in fact based on a subset of participants in the original EQ5D valuation study). Secondly, it estimates a full tariff rather than restricting itself to a small set of health states and, thirdly, it includes durations of one year and one month (which puts it within the limits of the end of life guidance). Both the mean valuations of the health states included and the full tariff, suggest that that utility values are a decreasing function of both severity and duration of the health state. When considering the mean valuations of the health states, the 10-year values are about 0.05 below the 1-month values for less severe states, and 0.15 below for more severe states. The finding that larger discrepancies were observable for more severe health states is in line with the studies shown in Table 2.

One other point of note relating to the Dolan study is the interpretation of the observed relationship between value and duration. In common with other authors, the relationship is interpreted as showing some individuals becoming intolerant to severe health states. In the extreme, some studies identify a 'maximal endurable time', such that for severe ill health, longer durations of survival are rated lower than shorter durations. If such a situation were encountered at the end of life, it is possible that extending life with a new treatment would be valued lower than the use of the conventional treatment. However, the value-duration relationship identified by Dolan is also present, to a lesser degree, with health states that would not be considered as becoming intolerable (e.g. values of around 0.8). Consequently, it seems reasonable

to assume that an alternative factor is generating the systematic discrepancies in values.

Finally, in comparison with Tsuchiya and Dolan we have included 8 additional studies, and tend to be more supportive of the notion that higher values are given to health states for shorter durations of life. However, we too note the heterogeneity of studies and results preclude any firm conclusions being drawn.

Explanations for non-CPT

Most studies have examined the potential variation of utilities with duration in TTO and SG experiments purely for the purpose of testing the assumptions of CPT and UI, respectively. In some circumstances, reduced desire to trade time in TTO questions has been described in different terms such as short-term indifference or an example of lexicographic preferences (whereby longer duration is always preferred in any choice presented). However, only a small number of studies have tried to explain the reasons behind the results.

Loss aversion, where the reduction in value from a loss is proportionally greater than the rise in value from an equivalent gain, is highlighted by several authors as explaining higher short-term utilities, or differences between conventional TTO values and alternative valuation methods. Bleichrodt and colleagues (2003) highlight two possible viewpoints regarding the role of loss aversion. Firstly, that loss aversion is 'valid' and therefore decision analysis needs to incorporate it, or secondly, that it is a bias that causes preferences to violate procedure invariance and should be excluded by basing utilities on scenarios where the effects of loss aversion are minimised. These two viewpoints clearly lead to quite different solutions for developing utility assessment in the future; which position to adopt is open to debate.

It should be noted, however, that the current EQ5D tariff is based on TTO values generated from a valuation exercise that used 10-year health state durations. Bleichrodt and colleagues consider 10 years to be a short enough duration for loss aversion to influence utility values. Consequently, if it is felt that loss aversion is a legitimate influence on health state valuations, then the current EQ5D tariff probably incorporates some degree of this influence (although, possibly not its full extent for

contexts where duration is considerably less than 10 years). If it is not considered legitimate, then the valuation tariff potentially overestimates utility values, all other things being equal. Also of note is that SG values are immune from loss aversion as the trade-off is not in terms of length of life, but risk.

Other authors highlight the possibility of some individuals possessing a maximal endurable time for living in severe health states. When such preferences are held by individuals within a sample, this will result in the mean value for the sample being higher for shorter durations.

Other issues

Another study that does not meet the inclusion criteria of this review but is of note was undertaken by Tsevat and colleagues (1995). Within this study, terminally ill patients (defined as having a less than 50% chance of survival at 6 months) valued their own health using the TTO technique at multiple time points. For those patients valuing their health at baseline and 6 months, utilities increased on average by 0.08 ($p<0.0001$). Whilst this is consistent with the notion of life becoming 'more valuable' it may also reflect a selection bias in that these patients are by definition 'survivors' and consequently may be moving into remission from their serious illness. Interpreting the results of this study, therefore, is fraught with difficulties.

Within this review, we have identified loss aversion as one possible explanation for higher than expected utility values at shorter durations. However, there are other potential problems with the calculation of utility values that may have an influence on 'end of life' appraisals such as the discounting of utilities within the EQ-5D tariff.[2] There are also other more fundamental reasons why stated preference questions may not match expectations (Baron, 1997), and other reasons still why QALYs may not match preferences (Gafni and Birch, 1995).

If we feel that the utility values recommended within the reference case, and the associated QALY model need refining to incorporate specific treatment contexts, we

[2] When utilities are calculated from TTO data no account is taken of an individual's marginal rate of time preference and hence the unequal value of the years traded.

should consider all possible improvements to our methods. A series of ad hoc adjustments is likely to result in a confusing set of methods that are applied in different circumstances, which in turn will lead to problems of interpretation and inconsistency. An overview of these various issues would help produce a clearer picture of where improvements can be made, and where these improvements are or particular importance to the decision making context within which NICE operates.

To a certain extent, this dilemma is produced by the adoption of an extra-welfarist perspective within cost-utility analysis that focuses on the valuation of 'contextless health', and the desire for decision makers to incorporate individual and/or societal preferences more fully which requires valuation of 'health within a specific context'. A clearer position in this regard is needed for this dilemma to be resolved.

Future research

If the evidence to date is not sufficient to identify whether being at the end of life changes the value placed on health, can we be at least confident that an appropriately designed study of the kind reviewed here, is a valid approach to identify any end of life effects? In principal, valuation exercises of this kind do conform with reference costs methods and incorporate proximity to death into the valuation task. However, several problems have been identified in the studies undertaken to date. Firstly, several studies have identified the possibility of a proportional heuristic being used by respondents when asked to value the same health state based on different durations. In other words, respondents appear to use proportionality of trade-off as a cognitive short-cut to answering the valuation exercises, even when they state clear rankings of health states and durations.

Secondly, several of the studies have identified procedural invariance with respect of TTO relating to loss aversion. When the valuation exercise is framed in an alternative manner, for example, increasing the duration of survival but in a less desirable health state within a TTO exercise, systematically different valuations are given. This observation is most noticeable with tasks using short durations of life and has been attributed to loss aversion. This raises the question of which variant of the valuation task is of most relevance to the end of life as defined by the NICE supplementary guidance? It could be argued that the situation of trading off an extension of life for

worse health is more akin to the situation of life extending therapies; this would indicate that the 'normal' TTO variant is not relevant to the end of life context.

This point is in itself dependent on the idea that the reasoning behind the NICE end of life supplementary advice relates to current utility values not capturing all important aspects of health value. Alternatively, the guidance may come from a view that individuals in this position are more deserving of health care resources, in other words, it is an equity issue. Ultimately, until we can be certain of which of these reasons predominates, we are unable to identify methods by which we can quantify a possible solution. Clarity on this point is essential before the research agenda can move forward.

As mentioned previously, there are several other potential problems with utility values and the conventional QALY model recommended in the reference case. The end of life debate has brought to a head one possible problem with the reference case approach, but many others remain. An overview of these problems and their relevance to NICE should be considered in order to focus methodological research onto the issues that of greatest importance to the Institute. Such an endeavour would be helped greatly by an explicit statement of the Institute's vision of benefit valuation within technology appraisal. This process has started through the latest revision of the reference case, the methods priority setting exercise and its related MRC research programme. However, the Institute should consider taking a greater lead in future work to ensure the research agenda meets its needs.

There are other further gaps between the general concepts of CPT/UI and the tightly defined NICE end of life supplementary advice. In particular, there may be reasons to believe that the willingness to trade length of life for quality of life is not simply a function of remaining life expectancy as is assumed in the existing policy. Indeed, much of the literature pertaining to preferences of those approaching death indicates that quality of life retains critical importance, albeit that some of the domains of standard measurement tools may not be relevant in this situation (see Coast and Lavender 1998 for a review). One factor that may be of particular relevance in this context is the time from terminal diagnosis rather than life expectancy per se. It may be the case that the value of life extension is greater not amongst those with a short

life expectancy per se, but a subgroup for who the time from terminal diagnosis is short. It is commonly cited that the ability to "put one's house in order" at the end of life is of critical importance, but this indicates that it is life extension in the face of a sudden terminal diagnosis that is most valuable and the focus should therefore not be on life expectancy in defining this patient group. This change in focus would also clearly distinguish a patient group in terms of a characteristic that is common to all conditions: everybody reaches a point where life expectancy is less than 2 years at some point in their life.

3. SOCIAL VALUES

This section examines the evidence on the relative values or weights society places on increasing life expectancy or health for people at the end of life.

Most previous work conducted in this area has focussed on whether the principle of QALY maximisation (a QALY is equal regardless of who receives it) reflects social preferences. However these studies have generally focussed on attributes other than extending life at the end of life. Attributes that have been examined include characteristics of populations such as: health gain; whether other treatments are available; previous health; lifestyle 'choices'; the impact on others; claims based on compensation or reward and; time spent waiting for treatment (Dolan and Shaw 2003).

The aim of this part of the review is to identify evidence that quantifies the relative social value placed on increasing the health of people at the end of life compared with other groups. A secondary aim is to examine possible methods for obtaining such evidence.

3.1 REVIEW OF EVIDENCE ON SOCIETAL PREFERENCES, METHODS.

A literature search was conducted. The following databases were searched Medline, Embase, Econlit, NHS EED and the Cochrane Library. Search terms which covered

"end of life" or related terms such as "terminal" were combined with those relating to QALYs. The search terms are provided in Appendix 1.

In addition a review of methods used to assess societal preferences for different groups of individuals was conducted by reviewing two recent reports (Baker et al, 2008; Dolan et al 2008) on the 'social value of a QALY' commissioned by NICE and a published review (Dolan et al, 2005) of whether people's preferences are consistent with QALY maximisation. These documents were also reviewed for relevant articles relating to end-of-life and QALY maximisation.

The inclusion criteria for the literature search and the review of the existing review and reports were:

- Papers quantifying the preferences of the general population, and;
- Papers reporting preference weights towards treating people at the end of life or with terminal diseases.

3.2 REVIEW OF EVIDENCE ON SOCIETAL PREFERENCES, RESULTS

The search identified 247 possible papers. The titles and abstracts of these papers were reviewed (LL). Following this initial sift of papers only two papers potentially met the inclusion criteria and were reviewed in full (Winter and Parker, 2007; Reese et al, 2005). The majority of papers discarded at this stage focussed on the aspects of life towards the end-of-life considered important by patients (for example being able to get personal affairs in order before death, having a pain-free death and not being a burden to loved ones) or reported the estimated prognosis of patients with specific conditions or treatments and therefore did not meet the inclusion criteria.

Of the two papers fully reviewed, one did not meet the inclusion criteria. It considered the use of prospect theory and whether less healthy people would express stronger preferences for life-prolonging treatments compared to healthier people. As the paper focuses on individual preferences it was not included in this part of review (Winter and Parker, 2007). The main focus of the other paper was the personal perspective however participants were also asked to consider their preferences for others at the end of life and so has therefore been included in the review (Reese et al. 2005).

Dolan et al (2005) conducted a review to assess whether people's preferences differ from the assumption that a QALY is valued equally regardless of who receives it which underpins the rationale for QALY maximisation Dolan et al (2005). They included papers that allowed inferences to be drawn regarding the health gain of one group compared to another. The authors identified 64 studies with empirical data looking at QALY maximisation and report some details of the methodology used by all of the studies in their review. The majority of studies are stated as being based on self-report questionnaires. Other designs included: postal questionnaire (it is unclear from the paper how this differs from self-report questionnaire); structured interview; experimental; focus group; telephone survey and; ethnography. The studies used random samples of the general population, convenience samples of the general population, students, patients, health professionals or academic staff. Of the 64 studies included in the review, 18 examined whether societal preferences are a linear function of changes in quality and length of life. The abstracts of these 18 papers were reviewed to assess whether they included information on societal preferences for weighting the QALYs of people at the end of life. Although not the sole focus of the papers, four articles included potentially relevant information.

One of the reports commissioned by NICE (Baker et al. 2008) identified a further two papers for inclusion in the review. In total seven journal articles and the two reports produced for NICE were included in the review and are summarised below.

Baker et al (2008)

Baker and colleagues undertook qualitative research in members of the general public in Norwich and the North East of England. The aim of this was to identify the characteristics of treatment beneficiaries that were considered important to respondents. It included open ended discussions, simple ranking procedures and more a complex ranking task (Q methodology). A discrete choice experiment and a person trade-off approach (also referred to as 'matching' in the report) were then used to estimate the relative weights attached to these characteristics by respondents. A feasibility study (including willingness to pay and standard gamble type questions) was also undertaken to assess whether it is feasible to estimate the willingness to pay of a QALY.

The results of the qualitative component of the study found that age and severity of illness were important characteristics. The discussion of severity appears to have focussed on quality of life rather than survival, in particular the quality of life without treatment and how much could be gained by treatment. The ranking procedure found that the life expectancy of patients before treatment was ranked third most important; such that those with shorter life expectancies were ranked higher. The most important attribute was the quality of life of patients before treatment and secondly, whether no other treatment was available.

For both the discrete choice and PTO questions, respondents received an introductory explanation of the task assisted by the presentation of the options using diagrams. The discrete choice study included attributes relating to age at onset, age at death, gain in life expectancy, quality of life if untreated and gain in quality of life with treatment. The attribute for 'gain in life expectancy' included response categories ranging from 0 to 79 years. Respondents were asked to choose their preference for one of two options. An example of the text of one of the questions is given below.

> *Now, would you give priority to treating the same number of people with:*
> *Condition A: Life expectancy extends from 10 to 20 at 90% of health OR*
> *Condition B: Health improves from 30% to 70%. Life extends from 60 to 70?*

The PTO (or 'matching') approach used similar diagrams and questions but varied the number of people treated in each condition. Depending upon the respondents' prior responses the number in each group would change until the respondent switched their preference to the alternative condition (thereby inferring a point of indifference between the two options). Examples of the questions are noted below.

Question 1: Now, would you give priority to treating:

> *100 people with condition A: Health improves from 60% to 80% between the ages of 20 to 40 OR*
> *100 people with condition B: Health improves from 60% to 80% between the ages of 40 to 60*

Question 2: What if only 40 people with condition A could be treated? Would you give priority to treating 40 people with condition A or would you give priority to treating 100 people with condition B?

Relative weights were estimated by age of onset, age of death and severity (quality of life without treatment). Two alternative approaches were used to estimate the weights from the discrete choice data. The first was a 'probability of choice' method, in which the number of QALYs in a scenario is varied until the estimated probability of choosing that scenario is equivalent to that for a specified basecase scenario. The second approach was a 'compensating variation' approach using the utility of QALY gain as the numeraire and using this to value changes from the specified basecase. The estimated weights for both methods were similar but a narrower range for compensating variation approach: range 0.38 to 1.61 for probability of choice; and range 0.82 to 1.1 to the compensating variation approach. In the quantitative analyses 'gain in life expectancy' was combined with 'gain in quality of life' into a QALY variable. Therefore it is not possible to differentiate the contribution of survival gain in the results or to determine a weight for life expectancy without treatment. The compensating variation approach has an advantage over the 'probability of choice' method in that it can calculate weights for each individual attribute, whereas the weights from the 'probability of choice' method produce weights for scenarios.

For the PTO data, the ratio of the responses to condition A vs B were calculated to estimate the relative weights. The authors report methodological choices here in terms of whether to calculate the ratio of mean values or whether to calculate the average (median or mean) ratio. The severity weights ranged from 0.24 to 1.00 depending on age and baseline HRQL with the mean ratio approach, and from 0.36 to 1.00 using the ratio of means method of calculation.

The willingness-to-pay feasibility study focuses on treatments affecting quality of life (using examples of stomach/bowel problems and head pain) rather than treatments increasing survival and the empirical results are therefore not relevant to this review. This part of the study was not aimed at deriving weights for different populations. The authors also state that the results from this study should not be used to inform policy

decisions as they were not obtained from a representative sample of the general population.

In terms of the methods used the authors note that caution should be taken if undertaking studies similar to the willingness-to-pay and standard gamble exercise, particularly around the presentation of probabilities and the extent to which answers are sensitive to key dimensions (e.g. duration of illness). Disagreement about the preferred method of the two approaches to estimating weights was noted. Some of the authors considered the person trade off to be preferable whereas others considered that both approaches had their merits. Concerns expressed about the results from the discrete choice method included that the design may have been compromised, the functional form of the statistical model and the potential impact of multicollinearity between the age and severity variables, and difficulties in expressing the severity characteristic. Concerns with the person-trade off estimates included whether the underlying assumption that the number of beneficiaries can be multiplied by the number of QALYs, whether the results can be generalised to smaller QALY gains and that the magnitude of weights had not been established. In general the authors found that respondents were largely insensitive to the size of risk eliminates in the questions and that many respondents found probabilities problematic.

Baker and colleagues also identified two additional papers regarding QALY preferences published since the Dolan et al (2005) review. Both of these papers are reviewed below (Ratcliffe 2000 and Schwappach 2003).

Dolan et al (2008)

Dolan and colleagues sought to elicit preferences from the general public that would allow QALYs to be weighted according to preferences for certain characteristics. The characteristics explored in this study are not directly relevant to the end of life context that is the focus of this review; however the methods employed may offer insight into future studies assessing preferences towards increasing survival at the end of life. The project included several sub-studies. Of these, the most relevant to this review was a set of pairwise choice questions put to 29 participants. The questions were designed to assess how people would prioritise treatment for groups based on a variety of factors; financial situation; social class; quality of life; timing of the condition; time spent on a

waiting list for treatment; age; marital status and; number of dependents. In response to the question regarding quality of life, the majority of participants (59%) chose to prioritise those in poor quality of life (10% HRQL) when the difference in the total QALYs gained by the two groups was 100 QALYs, but the majority of participants prioritise those in better health (the QALY maximisation option) when the difference in QALYs gained between the two groups was large (1000 QALYs vs 500 QALYs).

The characteristics used in the main study included the timing of ill health in a person's lifetime, the severity of ill health (expressed in terms of quality of life) and NHS or personal responsibility. Respondents were presented with multiple pairwise choices asking them which of two groups they thought the NHS should give priority to when treating. The ordinal responses were then transformed into cardinal measures using Thurstone scaling, which were then used as inputs into a social welfare function. The authors also explored the stability of preferences with and without prior group discussion. Of all the attributes the lowest QALY weight was 0.77 for those with a quality of life of between 50% and 100%. The highest QALY weight of 1.828 was for children under 18 years.

The authors conclude that their social welfare function approach is a feasible method for estimating weights. They note that ideally they would have preferred to include a larger number of pairwise choices and a larger number of levels, particularly for the severity domain, but that it was not practical to do so. They also conclude that a 'resource intensive' method of eliciting preferences including prior discussion prior to the interview was not required.

Cookson and Dolan (1999) and Dolan and Cookson (2000)

Dolan and Cookson convened small focus groups with 60 people from the Yorkshire area to discuss a range of health issues. Two experiments conducted within these focus groups are potentially relevant to this review.

Firstly, people were asked to consider 4 individuals with different health needs and decide which one should have their treatment funded and to discuss their choices (Cookson and Dolan, 1999). Participants were presented with photographs of four individuals accompanied by some text describing their age (8, 18, 42 or 65 years),

condition (facial scaring and psychological problems, hip replacement, HIV positive, cancer), the likely impact of treatment (correct scarring, be able to live independently, increase survival and minimise symptoms, treatment success undefined) and how they acquired the condition (car accident, prior drug misuse, not relevant). Specifically respondents were asked

> *You've got £4000 which will fund one of the scenarios described – what would you spend the money on and why?*

There was variation in the priorities for the different individuals, but on average people ranked the treatment of a child with life threatening cancer first, a hip replacement for a 65 year old woman second, plastic surgery for a 18 year old man was ranked third and lastly treatment to improve symptoms and extend the life of a 42 year old woman who contracted HIV as a result of drug taking. Participants were presented with additional information, including names and photographs, and therefore other characteristics could have influenced results. Consequently, it is difficult to draw conclusions about these results as respondents may have focussed on attributes other than health gain or likely prognosis without treatment (for example age, gender and ethnicity). When participants were encouraged to discuss their choices, the most frequently cited principle was giving a lower priority for self-inflicted illnesses, followed by priority for 'life-threatening' illnesses and prioritising for larger gains in length of life.

Secondly participants were asked to prioritise health resources for two hypothetical groups of patients (Dolan and Cookson 2000). The groups differed in terms of the extent to which their health could be improved in terms of quality of life and length of life. In addition, the baseline prognosis (both in terms of health-related quality of life and length of life) was varied between the groups. Participants were encouraged to discuss what principles they considered when choosing between the two groups. In five of the groups, people were asked to imagine they were a decision-maker who had to choose between the two groups of patients. The other five groups were told that they would personally be in one of the groups but that they did not know which one (the 'veil of ignorance' condition).

The majority of participants chose to give the same priority to both groups of patients (between 57% and 63% depending upon the scenario), although a sizable proportion of people chose to give priority to those who would gain most in quality or length of life (37% to 43%). Participants were more likely to choose between the two groups rather than give the same priority when the difference in end-points went above a certain threshold. The most frequently cited principle by participants was that the groups should have the same priority regardless of benefit. The researchers also recorded various principles cited by respondents to explain their decisions regarding the prioritisation task. The principle of prioritising people with 'life-threatening conditions' was mentioned in discussions, but only twice with all other principles that were noted arising more frequently. The authors state that when respondents were requested to employ the veil of ignorance condition, they tended to consider the question from a societal decision maker perspective.

Ratcliffe (2000)

Ratcliffe surveyed a random sample of university workers to gain preferences for the allocation of a fixed resource (donor livers) using a conjoint analysis method. Respondents were asked to consider how they would allocate a fixed number of livers (100) between two groups of potential liver transplant recipients. The characteristics of the two groups varied in terms of: the expected life years gain from treatment (5, 10 or 20 years); whether the cause of liver disease was alcohol related; the time spent on the transplant waiting list (3, 6 or 12 months); age (40, 50, or 60 years) and; whether the patient had previously undergone a transplant or not. She found that all characteristics were significant in people's preferences, but the factor that caused the largest absolute difference in responses was the cause of liver disease (respondents preferred to give fewer livers to those whose disease was alcohol-related).

Reese et al (2005)

The paper by Reese and colleagues was a study of preferences towards end of life care and whether there are differences based on socio-demographic variables. The study was conducted in the USA and included sample of social work students (n=58), members of the general population (n=153) and medical students (n=82). Participants were asked several questions relating to end-of-life care (e.g. preference for place of death). Participants were also asked to whether they would prefer treatment to

improve symptoms or extending life as much as possible if they had a terminal disease. They were asked to consider this from three perspectives: their own life, the life of a family member and for a terminally ill person in general. Specifically the question put to respondents was:

> *'If you had a very serious medical condition and were told that you had only a one chance in four of living -that is 25%-, which of these two plans of care would you prefer? Relieving pain and discomfort as much as possible, even if that meant not living as long, or extending life as much as possible, even if that meant more pain and discomfort. Which of these same two choices would you select if it was a member of your family that was dying? Which of the same two choices do you think should be emphasized in medical care for the terminal patient in general?'*

The majority of responses in all groups and for all perspectives stated preferences for relieving pain and discomfort (the palliative care option) compared to the life extending option. The group most relevant to this review is the general population sample; however there were no statistically significant differences in the responses to the question between the three groups. For the general population responses, 70% of respondents preferred the palliative care option when considering their own health, 85% preferred the palliative care option when considering the health of a family member and 87% preferred the palliative care option when considering terminally ill patients in general.

Roberts et al (1999)

Roberts and colleagues undertook a discrete choice exercise to explore societal preferences for the following attributes in health resource prioritisation: the number of patients treated, survival without treatment, survival following treatment, quality of life following treatment and the chance of successful treatment. Respondents were presented with two treatment programmes and asked to say which they would support if only one could be provided by a health authority. Details of the clinical condition were provided in some questions and not in others to assess the impact of the additional information. An example of a question is provided below.

A Health Authority is faced with many choices on how to allocate its budget. Imagine a choice involving the allocation of £20 000 where two options for using these funds have been identified. Given that only one of the options can receive funding, which option would you support?

Option D: Ten people within a health authority have a major blood vessel in the heart that may burst at any time. Treatment for this condition has a 1 in 2 chance of being successful (so 1 in every 2 treatments will fail). If it is successful, people would be expected to live for 1 year. However, they will be unable to perform their usual activities and will be moderately anxious or depressed. Without treatment, or if treatment is not successful, people are not expected to live more than a few weeks.

Option K: One hundred people within a health authority are at a high risk of multiple strokes. Treatment for this condition has a 1 in 10 chance of being successful (so 9 in every 10 treatments will fail). If it is successful, people would be expected to live for 5 years. However, they will have some problems with performing their usual activities and will be moderately anxious or depressed. Without treatment, or if treatment is not successful, people are not expected to live more than a few weeks.

The results are not summarised into the overall impact of each attribute. However, there was considerable variation in responses depending upon the pair-wise choice with between 23% and 93% choosing the QALY maximising option. Most respondents did not respond in line with the principle of QALY maximisation. In general, respondents were reluctant to support programmes that left people in a poor state of health following treatment, however it was not possible to determine any preferences towards extending survival at the end of life. The authors state that the results do not support the provision of additional clinical information in future studies although they suggest that further qualitative evidence would be beneficial.

Schwappach (2003)

Schwappach conducted a discrete choice experiment in a convenience sample of German undergraduate students. Participants were asked to consider that they were responsible for a health authority budget that was not sufficient to finance necessary treatment for all patients. They were presented with two groups of patients and asked

'What proportion of the budget for life saving treatments would you allocate to each group?' The characteristics of the two groups varied in terms of healthy lifestyle, social class, age, life expectancy after treatment, quality of life after treatment, and whether they had been prior recipients of intensive care. Participants were also told that without treatment patients in both groups would die within the next three months.

Given that all hypothetical patients in this experiment can be considered to be at the end of life, the results are not informative for estimate preferences towards extending survival at the end of life. However, overall respondents preferred to prioritise groups that would gain most in terms of life expectancy and quality of life, younger people, those who have healthy lifestyles, people in lower social classes and those who had not been prior recipients of intensive care. In terms of the methods, the author notes that the use of a web-based design may not have sufficiently encouraged reflection on some of the questions and that some of the results appear to conflict with an earlier simpler ranking exercise. In addition, it is noted that the sample of respondents was a convenience sample of young, well educated students and it is unclear if the web-based method would be appropriate for a wider group of the general public.

Shmueli (1999)

The study that comes closest to addressing whether people would prioritise treating people with life extending treatments at the end of life is a study by Shmueli. Shmueli conducted face-to-face interviews with a large sample of Jewish Israeli people aged 45-75 to discuss health-related issues including the prioritisation of resources. Respondents were asked to decide which of two individuals they would choose to treat: one individual had an immediately life threatening condition and a limited life span (1 month, 1 year and 5 years) and; the other individual was paralysed but could receive treatment aimed at improving his quality of life.

> *"Two male road-accident injured are brought into the emergency room. Both are around fifty years old, married with children and of fair economic status. The first will die if not treated immediately. However, even if treated, his chances of surviving a month — during which time he will be confined to a hospital bed — are low. The second injured has no life threatening condition,*

but will remain paralyzed in both legs if not treated immediately. If only one can be treated at a time, which of the two should be treated first?"

More people preferred to improve the quality of life of the second individual rather than extend the life of the first individual when the expected extension to survival was relatively small: 1 month (63% vs 27%) and 1 year (48% vs 40%). However the majority preferred to extend the survival of the first individual when the expected survival gain was 5 years rather than improve the quality of life of the second. The results suggest that people preferred to only favour increasing life expectancy when the increase was considered to be substantial.

3.3 DISCUSSION

None of the studies identified within this review have directly estimated societal preference weights that could be applied to QALYs at the end of life. The paper by Shmueli (1999) is perhaps the most relevant to this question. This study found that people only gave more weight to increasing survival at the end of life when the magnitude of the increase was over a sizable threshold. However the results of this study may not be generalisable to the UK setting and the author noted that the results were sensitive to age and religiosity. In addition the paper focussed on two individuals and different results may have been obtained if the questions were posed in terms of treating populations, which is arguably the context most relevant to NICE. In addition specific weights were not calculated.

The two papers by Dolan and Cookson (1999, 2000) suggest some people do wish to prioritise treatments that can extend survival at the end of life; however there was also some support for giving everyone an equal chance of treatment and prioritising those who are likely to have the largest health gains.

In terms of designing future studies, there should be a specific focus on estimating relative QALY weights for extensions to survival at the end of life in order to help inform NICE's current supplementary advice to its Appraisal Committees. Studies should explore preferences for small increments in survival gain (Shmueli, 1999) but also consider small increments in health-related quality of life. In addition, the

sensitivity of results to differences in prognosis without treatment should also be considered (Dolan and Cookson, 2000).

There have been differences in the studies as to whether they focus on the individual or a sub-population/treatment programme. The NICE remit is to decide whether to recommend specific treatments or interventions to the NHS rather than recommend which specific individuals should get treatment. Therefore studies focussing on sub-populations or treatment programmes better reflect the type of decisions made by NICE.

A variety of methods have been used by the papers. Only a couple have explicitly estimated weights associated with characteristics of individuals or treatment (Baker et al., 2008 and Dolan et al., 2008). The majority of studies have used discrete choice methods, different forms of person trade-offs or a combination of the two. Limitations were noted with all of the methods however the authors of the reports considered the person trade-off (Baker et al, 2008) and discrete choice with Thurstone scaling (Dolan et al., 2008) methods to be feasible for undertaking this kind of research. Opinions were divided regarding the results from the standard discrete choice experiment although it is possible that an alternative study design could have improved results (Baker et al 2008).

4. SUMMARY AND CONCLUSIONS

The NICE end of life supplementary advice suggests that the value of health gains generated by life extension for those with a life expectancy without treatment below 2 years should be valued more highly than standard approaches would suggest. We suggest there are two broad levels at which such an approach may reflect preferences – the individual and societal levels. We conducted a review of relevant literature in both areas.

Empirical tests of CPT and UI were reviewed in order to identify whether there is proportionality in health state valuations in the presence of death. We found a paucity of strong evidence amongst 17 papers to support the hypothesis that health state

valuations are greater when there is a short life expectancy. Where evidence does exist it generally indicates that the effect is likely to be small.

We suggest that there is unlikely to be substantial value from undertaking further empirical work to test proportionality. There could be value in further investigation of how individual valuations change in the presence or absence of factors such as time from terminal diagnosis and existing methods are easily adapted to achieve this aim. Of course, additional questions are raised by such an approach such as how to incorporate the fact that for many diseases including cancer, the initial diagnosis is of a condition that has a probability of being fatal. Furthermore, consideration should be given as to whether greater refinement in the accuracy of utility measurement in this area will ultimately lead to changes in decision making. One of the sensitivity analyses currently conducted where technologies are deemed to meet the end of life criteria is to assign full health to the additional survival time, i.e. the maximum feasible value. Our limited experience of these results is that they make little substantial difference to cost effectiveness.

No direct evidence of higher social values for the NICE defined end of life setting was identified, although there is a large literature to suggest that the general population may wish to diverge from QALY maximisation when considering other characteristics. Studies which considered characteristics of the recipients of health benefits that may be relevant to this setting included those which considered severity of untreated illness as a characteristic. One study did consider the attribute of a short life expectancy but was of questionable relevance to the UK setting.

Our review also suggests that existing methods that have been used to estimate the relative importance of characteristics other than those embodied in QALY maximisation are suitable to the investigation of societal preferences in relation to end of life. In particular, the discrete choice experiment framework has been extensively applied and offers a feasible method.

Ultimately, this research agenda requires greater clarity from the Institute relating to their overall vision of benefit valuation. For example, what are the issues that should be included within utility values, and to what extent is the end of life advice based on

a perception of inadequacy of standard approaches with respect to individual utility measurement versus societal valuation? Following on from this, a review of methodological issues relating to utility measurement set within the context of the Institute's vision would help focus research further. The end of life debate has highlighted more clearly than ever before, the central importance of health values to NICE's work and the public's engagement with its processes.

5. References

Baker R, Bateman I, Donaldson C et al (2008). Weighting and valuing quality adjusted life years: preliminary results from the Social Value of QALY Project. Report to NICE

Baron J. (1997) "Biases in the quantitative measurement of values for public decisions", Psychological Bulletin, Vol.122(1):72-88.

Bleichrodt H, Johannesson M. The validity of QALYs: an experimental test of constant proportional tradeoff and utility independence. Medical Decision making 1997;17(1):21-32.

Buckingham JK, Birdsall J, Douglas JG. Comparing three versions of the time tradeoff: time for a change? Medical Decision making 1996;16(4):335-47.

Cook J, Richardson J, Street A. A cost-utility analysis of treatment options for gallstones disease – methodological issues and results. Health Economics 1994;3(3):157-68.

Cookson R and Dolan P (1999). Public views on health care rationing: a group discussion study. Health Policy 49: 63-74.

Dolan P. Modelling valuations for health states: the effect of duration. Health Policy 1996;38(3):189-203.

Dolan P and Cookson R (2000). A qualitative study of the extent to which health gain matters when choosing between groups of patients. Health Policy 51: 19-30.

Dolan P, Edlin R and Tsuchiya A (2008). The relative value of health gains to different beneficiaries. Report to NICE.

Dolan P and Shaw R (2003). A note on the relative importance that people attach to different factors when setting priorities in health care. Health Expectations 6 (1): 53-59.

Dolan P, Shaw R, Tsuchiya A et al (2005). QALY maximisation and people's preferences: a methodological review of the literature. Health Economics 14: 197-208.

Gafni A, and Birch S. (1995) "Preferences for outcomes in economic evaluation: an economic approach to addressing economic problems", Social Science and Medicine, Vol.40:767-76.

McCabe,C., Claxton, K., Tsuchiya, A. (2005) Orphan drugs and the NHS: should we value rarity? BMJ, 331:1016-1019

McCabe, C., Tsuchiya, A., Claxton, K., Raftery, J. (2006). Orphan drugs revisited. *QJM* 99: 341-345

McCabe C, Tsuchiya A, Claxton K, Raftery J. (2007) Assessing the economic challenges posed by orphan drugs: A comment on Drummond et al. International Journal of Technology Assessment in Health Care 2007 Jul;23(3):397-401.

McNeil BJ, Weischselbaum R, Pauker SG. Speech and survival: trade offs between quality and quantity of life in laryngeal cancer. New England Journal of Medicine 1981;305:982-7.

Miyamato JM, Eraker SA. A multiplicative model of the utility of survival duration and health quality. J Exp Psychol Gen 1988:117(1):3-20.

NICE Citizens Council Report (2004) Ultra Orphan Drugs. London 2004

Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. Operations Research 1980;28(1):206-24.

Ratcliffe J (2000). Public preferences for the allocation of donor liver grafts for transplantation. Health Economics 9:137-148.

Reese DJ, Chan CLW, Perry DC *et* al (2005). Beliefs, death anxiety, denial and treatment preferences in end-of-life care: a comparison of social work students, community residents and medical students. Journal of Social Work in End-of-Life and Palliative Care 1(1):23-47.

Richards M. (1998) Improving access to medicines for NHS patients, Crown Copyright.

Roberts T, Bryan S, Heginbotham et al (1999). Public involvement in health care priority setting: an economic perspective. Health Expectations 2: 235-244.

Sackett D, Torrance G. The utility of different health states as perceived by the general public. Journal of Chronic Disease 1978;31:697-704.

Schwappach DLB (2003). Does it matter who you are or what you gain? An experimental study of preferences for resource allocation. Health Economics 12: 255-267.

Shmueli A (1999). Survival vs. quality of life: a study of the Israeli public priorities in medical care. Social Science and medicine 49: 297-302.

Shumway M, Chouljian TL, Battle CL. Measuring preferences for schizophrenia outcomes with the time tradeoff method. The Journal of Behavioral Health Services Research 2005;32(1):14-26.

Stalmeier P, Wakker P, Bezembindr T. Preference reversals: violations of unidimensional procedure invariance. J Exp Psychol Hum Percept perform 1997;23(94):1196-205.

Stiggelbout AM, Kiebert GM, Kievit J, Leer JWH, Habbema JDF, de Haes JCJM. The "Utility" of the time trade-off method in cancer patients: feasibility and proportional trade-off. Journal of Clinical Epidemiology 1995; 48(10):1207-1214.

Tsuchiya A, Dolan P. The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. Medical Decision Making 2005;25:460-467.

Tvesat J, Cook F, Green ML, Matchar DB, Dawson NV, Broste SK, et al. Health values of the seriously ill. Annals of Internal Medicine 1995;122(7):514-520.

Unic I, Stalmeier P, Vefhoef LvDW. Assessment of the time-trade-off values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer. Medical Decision Making 1998;18(3):269-77.

Winter L and Parker B (2007). Current health and preferences for life-prolonging treatments: an application of prospect theory to end-of-life decision making. Social Science and Medicine 65: 1695-1707.

Table 1: Summary of studies examining health state values immediately preceding death

	Previous review	Author*	Year *	Sample size	Duration 1 (shortest)	Duration 2	Duration 3 (longest)	Assumption tested	Utility values reported**	Summary of results
1	Y	Cook et al	1994	96	12 weeks	12 months	-	CPT	Yes	Shorter durations produced higher utility values for 5 out of 7 health states valued. Identical means were produced for the remaining 2 health states. Individual differences not statistically significant.
2	Y	McNeil et al	1981	37	Not specified	Not specified	Not specified	CPT	No	Mean values and differences not reported in any way. Authors report that 'virtually none' of the participants would trade survival when the duration was less than 2 years.
3	Y	Sackett et al	1978	246 general pop., 29 patients	3 months	8 years	Lifetime	CPT	Yes	Shorter durations produced higher utility estimates for both of the 2 health states valued. This was true for both patient and population samples. No statistical tests were undertaken.
4	Y	Pliskin et al	1980	10	5 years	15 years	-	CPT	Yes (can be derived from data presented)	Shorter duration produced higher utilities for all 3 comparisons. These estimates were derived from the data in the paper, and showed no statistically significant differences.
5	Y	Stalmeier et al	1997	39	5 years	10 years	50 years	CPT	No	Shortest duration produced lower utility relative to 10 and 50 years. This difference was statistically significant. No significant difference between 10 and 50 year durations.
				28	5 years	10 years	20 years	CPT	No	Shortest duration produced lower utility relative to 10 and 20 years. This difference was statistically significant. No significant difference between 10 and 20 year durations.
6	Y	Bleichrodt et al	1997	87	10 years	30 years		CPT	Yes	Small inconsistent differences that were not statistically significant for either the within or between samples comparisons.
				85	10 years	30 years		UI	Yes	Shorter durations produced higher mean utility values. These were statistically significant differences for both within and between samples comparisons.

	Previous review	Author*	Year *	Sample size	Duration 1 (shortest)	Duration 2	Duration 3 (longest)	Assumption tested	Utility values reported**	Summary of results
7	Y	Unic	1998	54	5 years	10 years	15-60 years	CPT	No	Shortest durations produced a lower utility compared with all other durations. Differences were statistically significant.
8	Y	Miyamato et al	1988	60	12-15 years	Various	20-24 years	CPT	No	Mean values and differences not reported in any way. Authors report that some subjects were not willing to trade-off length of life at short durations.
9	N	Dolan	2003	91	10 years	20 years	-	CPT	Yes	Shorter duration produced higher mean utility value for the one health state valued for full sample. This difference was statistically significant.
10	N	Schulz	2003	29	1 year	10 years	-	CPT	Yes	Shorter durations produced higher utility estimates for both of the 2 health sates valued. No statistical tests were undertaken.
11	Y	Bala	1999	114	1 year	20 years		UI	Yes	Shorter duration produced higher utility estimates for both of the only health sates valued. No statistical test was undertaken.
12	N	Kirsch	2000	64	2 years	10 years		CPT	Yes	Shorter durations produced higher utility estimates for 4 out of the 6 health states valued, and identical utility estimates for the remaining two states. Differences were statistically significant for 2 of the health states.
13	N	Martin	2000	199	5 years	10 years	15 years	CPT	Yes	Shorter durations produced higher utility estimates for all but one of the comparisons undertaken. All differences were statistically significant.
				199	5 years	10 years	15 years	UI	Yes	
14	N	Franic	2003	18	3 days	Rest of life		UI	Yes	Shorter durations produced higher utility estimates for all 4 of the health states valued, Differences were statistically significant for 3 of the health states.
15	N	Bleichrodt	2003	51	13 years	24 years	38 years	CPT	No	In general, shorter durations produced lower utilities. CPT was rejected for 3 out of the 3 health states valued (although there is some discrepancy between alternative tests) and for two different TTO procedures.
				51	13 years	24 years	38 years	UI	No	In general, shorter durations produced lower utilities. UI was rejected for both of the 2 health states valued.

	Previous review	Author*	Year *	Sample size	Duration 1 (shortest)	Duration 2	Duration 3 (longest)	Assumption tested	Utility values reported**	Summary of results
16	N	Shumway	2005	40	12 months	Rest of life		CPT	No	No statistically significant difference between time frames.
17	N	Stigglebout	1995	54	5 years	20 years	75 years	CPT	Yes	In general, shorter durations produced higher utilities. Few differences were statistically significant.
				63	3	10	Rest of life	CPT	Yes	
				23	3 years	5 years	10 years	CPT	Yes	

*Where author and years are missing, this indicates that the results belong to the preceding reference. This occurs when the same paper publishes the results of more than one experiment or subgroup of respondents.

**This indicates whether the actual values were reported as opposed to graphical presentation of results. Where values are reported, these are given in Table 2.

Table 2: Values elicited in the studies identified

Study	Assumption (sample)	Health state	Shorter duration	Longer duration	Longest duration	Mean utility (shorter duration)	Mean utility (longer duration)	Mean utility (longest duration)
Cook	CPT	Laparoscopic cholecystectomy	12 weeks	12 months		0.90	0.90	
		Moderate pain	12 weeks	12 months		0.90	0.89	
		Severe pain	12 weeks	12 months		0.88	0.87	
		Severe diarrhoea	12 weeks	12 months		0.81	0.81	
		Open cholecystectomy	12 weeks	12 months		0.81	0.80	
		Moderate pain/severe diarrhoea	12 weeks	12 months		0.68	0.66	
		Severe pain/severe diarrhoea/nausea	12 weeks	12 months		0.47	0.44	
Torrance	CPT (GP)	Hospital dialysis	3 months	8 years	Life	0.62	0.56	0.32
	CPT (Patients)	Home dialysis		8 years	Life		0.65	0.39
	CPT (GP)	Hospital dialysis	3 months	8 years	Life	0.81	0.59	0.52
	CPT (Patients)	Home dialysis		8 years	Life		0.72	0.56
Pliskin	CPT	Severe angina	5years	15 years		0.69	0.68	
		Moderate angina	5 years	15 years		0.90	0.89	
		Severe angina (chained)	5 years	15 years		0.72	0.69	
Bleichrodt	CPT (WS)	Generic description 1	10 years	30 years		0.59	0.59	
	CPT (BS)	Generic description 1	10 years	30 years		0.59	0.57	
	CPT (WS)	Generic description 2	10 years	30 years		0.79	0.78	
	CPT (BS)	Generic description 2	10 years	30 years		0.79	0.80	
	UI (WS)	Generic description 1	10 years	30 years		0.70	0.68**	
	UI (BS)	Generic description 1	10 years	30 years		0.70	0.66**	
	UI (WS)	Generic description 2	10 years	30 years		0.90	0.85*	
	UI (BS)	Generic description 2	10 years	30 years		0.90	0.87*	
Dolan	CPT	EQ5D 21223	10 years	20 years		0.47*	0.43*	
Schulz	CPT	Own health	1 year	10 years		0.86	0.80	
		Worst state benign prostatic	1 year	10 years		0.54	0.48	

Study	Assumption (sample)	Health state	Shorter duration	Longer duration	Longest duration	Mean utility (shorter duration)	Mean utility (longer duration)	Mean utility (longest duration)
		hyperplasia						
Bala	UI	Shingles	1 year	20 years		0.48	0.47	
Kirsch	CPT	NYHA I	2 years	10 years		0.93	0.93	
		NYHA II	2 years	10 years		0.78	0.77	
		NYHA III	2 years	10 years		0.55	0.51*	
		NYHA IV	2 years	10 years		0.37	0.28*	
		EQ5D 11122	2 years	10 years		0.77	0.77	
		EQ5D 21232	2 years	10 years		0.41	0.37	
Martin	CPT	Own health	5 years	10 years	15 years	0.90**	0.88**	0.86**
	UI	Own health	5 years	10 years	15 years	0.92*	0.92*	0.91*
Franic	UI	Own health	3 days	Rest of life (ROL)		0.99	0.99	
		Complete alleviation from post chemotherapy nausea and vomiting (PCNV)	3 days	Rest of life (ROL)		0.97	0.93*	
		Partial alleviation from PCNV	3 days	Rest of life (ROL)		0.94	0.81*	
		No alleviation from PCNV	3 days	Rest of life (ROL)		0.87	0.64*	
Stigglebout	CPT	Own health – testicular cancer patients with good prognosis	5 yrs	20 yrs	Rest of Life (ROL) 75yrs	0.95	0.97	0.93^
		Own health – colorectal cancer (treated with curative intent)	3 yrs	10 yrs	Rest of Life (ROL)	0.98	0.96	0.91^^
		Own health - colorectal cancer (incurable)	3 yrs	5 yrs	Rest of Life (ROL) 10yrs	0.95	0.90	0.80^^^

Key
+ p>0.05 * p<0.05 ** p<0.01 ^ test of difference in medians p>0.05 ^^ p<0.05 ^^^ p<0.01
No superscript means that statistical tests were not presented.
CPT = constant proportional trade-off, UI = utility independence, GP = general population, WS = within sample test, BS = between samples test, NYHA = New York Heart Association

Appendix 1: Search strategy

1 end of life.ti.
2 life expectanc*.ti.
3 ((sever* or terminal*) adj ill*).ti.
4 (life adj exten*).ti.
5 prognosis.ti.
6 1 or 2 or 3 or 4 or 5
7 (QALY or quality adjusted life year*).ti.
8 (equit* adj5 economic*).ti.
9 priorit*.ti.
10 age.ti.
11 (utility or utilities).ti.
12 preference*.ti.
13 ((social or community or public) adj2 (value* or valuation)).ti.
14 8 or 11 or 7 or 13 or 10 or 9 or 12
15 6 and 14
16 limit 15 to (english language and humans and yr="2002 -Current")

Table 3: Characteristics of the studies included in the review of societal preferences

Study	Country	Sample	Sample size	Individual or population perspective	Method	Variable considered	Weights calculated for attributes?	Summary of results
Baker et al (2008)	UK	General population	587	Population	Qualitative (group discussion, ranking and Q methodology). Quantitative: Discrete and Person Trade off (WTP also used in another aspect of the study).	Age at onset, age at death, gain in life expectancy, quality of life if untreated and gain in quality of life with treatment.	Yes using the DCE method. Weights for scenarios only using the PTO method.	Respondents generally preferred to give priority to the young and those in worse HRQL. However when the PTO method was used, this appeared to be subject to a threshold effect (such that preferences were not given to the youngest age group or those with the worst HRQL).
Cookson and Dolan (1999)	UK (York area)	General population	60	Individual	Group discussion. Participants asked to rank the treatment of 4 people. Qualitative analysis also conducted.	Various. Health status of the individuals differs in terms of type of problem and effect of treatment. However there also many other differences in terms of age, cause of illness, appearance, gender and ethnicity.	N	A child with life threatening cancer was given highest priority on average, however there were many different variables included making impact of treatment and prognosis difficult to unpick from the results.
Dolan and Cookson (2000)	UK (York area)	General population	60	Population and variant of individual (veil of ignorance)	Discrete choice questionnaire (self-complete) followed by group discussion.	Improvement in survival and HRQL and baseline prognosis (in terms of survival and HRQL).	N	Most participants gave the same priority to both groups of patients but were more likely to choose between the two groups when the difference in end-points went above a certain threshold.

Dolan et al (2008)	UK (representative sample)	General population	559 (main study) plus additional 129	Population	Discrete choice (interview).	Age, quality of life, social class, condition cause and rarity.	Y	Weights ranged from 0.77 (good quality of life) to 1.828 (children)
Ratcliffe (2000)	UK	University workers	303	Population	Discrete choice questionnaire (self-complete).	Life years gained, cause of disease, time spent on the waiting list, age and re-transplantation.	Y (DCE weights for attributes)	All characteristics were important. Weights depend on level of characteristic, but cause of disease has the most absolute impact.
Reese et al (2005)	USA	Social work students Community residents Medical students	58 153 82	Both (own health, health of family member, terminal patients generally)	Self complete questionnaire.	'Reduce pain and discomfort' Nor 'extending life as much as possible' when faced with 75% chance of death.	N	Majority of preferences were for reducing pain and discomfort in all groups and for all perspectives. Preferences for palliative care increased when considering others.
Roberts et al (1999)	UK	General population	91	Population	Face-to-face interviews. Discrete choice questions.	Chance of success of treatment, Number of people receiving treatment, survival without treatment, survival and HRQL with treatment.	N	Variation in responses, although most respondents did not respond in line with the principle of QALY maximisation. Respondents were reluctant to support programmes that left people in a poor state of health following treatment.

Schwappach (2003)	Germany	Undergraduate students	154	Population	Web-based questionnaire	Healthy lifestyle, social class, age, life expectancy after treatment, quality of life after treatment, prioir recipients of intensive care.	N	Preference for larger gain in life expectancy, larger gain in quality of life, younger people, those who have healthy lifestyles, people in lower social classes and those who had not been prior recipients of intensive care.
Shmeuli (1999)	Israel (Jewish Israelis)	General population	2030	Individual	Face-to-face interviews. Asked to consider which of two individuals to treat..	Survival (varying lengths) vs HRQL (varying degrees).	N	On average people favoured increasing HRQL over extending life expectancy, unless the gain in life expectancy was large.