

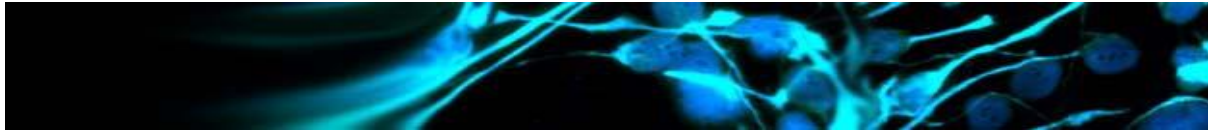
Research Paper 10/01

Using QALYs in Cancer: A Review of the Methodological Limitations

Martina Garau, Koonal K Shah, Anne R Mason, Qing
Wang, Adrian Towse, Michael F Drummond

26 October 2010

A revised version of this research paper has been published as Garau, M., Shah, K.K., Mason, A., Wang, Q., Towse, A. and Drummond, M. (2011) Using QALYs in cancer: A review of the methodological limitations. *Pharmacoeconomics*. 29(8), 673-685.



Using QALYs in cancer: a review of the methodological limitations

Office of Health Economics
Research Paper 10/01
26 October 2010



Martina Garau, Office of Health Economics
Koonal K Shah, Office of Health Economics¹
Anne R Mason, Centre for Health Economics, University of York
Qing Wang, Bristol-Myers Squibb
Adrian Towse, Office of Health Economics
Michael F Drummond, Centre for Health Economics, University of York

Corresponding author:

¹12 Whitehall, London, SW1A 2DY; Email: kshah@ohe.org; Telephone: 020 7747 8856

Acknowledgements

The authors are grateful for the contributions of Claire Devaney, Nancy Devlin, the anonymous reviewers, and participants at the Office of Health Economics oncology workshop in September 2009.

Sources of support

This paper is based on part of a consulting project commissioned and funded by the Pharmaceutical Oncology Initiative (POI) group. The material presented in this paper is independent of the funders. There are no conflicts of interest to declare.

OHE Research Papers are intended to provide information on and encourage discussion about a topic in advance of formal publication. Any views expressed are those of the authors, and do not necessarily reflect the views or approval of OHE, its Editorial Board, or of its sponsors. Once a version of the Research Paper's content is published in a peer-reviewed journal, this typically supersedes the Research Paper and readers are invited to cite the published version in preference to the original version.

Table of Contents

Abstract.....	3
1. Introduction.....	4
2. The QALY framework.....	6
3. Methods for the literature search.....	7
4. Findings from the literature review.....	8
4.1 How to describe health states: are generic measures, in particular the EQ-5D, sensitive enough?	8
4.2 How to value health states: is the time trade-off and its assumption of constant proportional trade-off valid in the context of cancer treatments?.....	11
4.3 Whose values to use: should patients' health status be valued by the general population or by the patients themselves?.....	14
5 Summary and policy implications.....	17
Reference list	19
Table 1: Classification of studies relating to 'whose values to use' (7 empirical studies + 2 reviews).....	24

Abstract

Objective: The objective of this paper is to examine how well the QALY captures the health gains generated by cancer treatments, with particular focus on the methods for constructing QALYs preferred by NICE.

Methods: Data were obtained using a keyword search of the MEDLINE database and a hand search of articles written by leading researchers in the subject area (with follow-up of the references in these articles). Key arguments were discussed and developed at the Office of Health Economics oncology workshop in September 2009.

Results: Three key issues emerge. First, the EQ-5D, NICE's preferred measure of health-related quality of life in adults, has been found to be relatively insensitive to changes in health status of cancer patients. Second, the time trade-off, NICE's preferred technique for estimating the values of health states, involves making assumptions that are likely to be violated in end-of-life scenarios. Third, the practice of using valuations of members of the general population, as recommended by NICE, is problematic because such individuals typically display a misunderstanding of what it is really like for patients to live with cancer.

Conclusions: Because of the way in which it is constructed, the QALY shows important limitations in terms of its ability to accurately capture the value of the health gains deemed important by cancer patients. The paper concludes by proposing a research agenda for addressing these limitations.

1. Introduction

Economic evaluations of health care technologies are a critical element in informing decisions about resource allocation in a number of jurisdictions. For example, cost-effectiveness has informed decisions on new drug reimbursement in the Netherlands since 2005,^[1] and the introduction of ‘comparative effectiveness’ evaluations in the United States is viewed as “a cornerstone in controlling runaway health care costs”.^[2] In England and Wales, the National Institute for Health and Clinical Excellence (NICE) produces guidance on the appropriate use of selected health technologies in the National Health Service (NHS). Its guidelines endorse the use of the quality-adjusted life year (QALY) to measure health.^[3] The QALY combines length of life with health-related quality of life (HRQL) to form a single, generic measure of health improvement. It acts as a ‘common currency’ for assessing the value-for-money generated by treatments both across and within diseases areas. In order to ensure consistency across evaluations, NICE specifies a prescribed set of methods for constructing QALYs in the reference case economic analysis. However, these methods have been found to be inappropriate for certain medical conditions,^[4] which has led to concerns that the QALY may not be sensitive to some changes in health that are relevant for specific diseases or interventions.

A large proportion – nearly 25% – of the Technology Appraisals produced by NICE has focused on cancer interventions. This reflects the increasing number of new medicines developed to treat different types of cancer, including rare forms where no alternative therapy is available. The focus on cancer is also due to the decision of the Department of Health, as part of the Cancer Reform Strategy, to refer all new cancer medicines and significant new licensed indications to NICE.^[5] However, decisions by NICE not to recommend the use of certain medicines aimed at extending the life of advanced cancer patients on cost-effectiveness grounds^[6] have resulted in criticism from charities, campaigners and health professionals.^[7]

In early 2009, NICE undertook a review of its approach to appraising life-extending, end-of-life treatments. Based on this review, it issued supplementary advice which indicated that under certain circumstances, end-of-life treatments showing survival improvements may be recommended for use even if their incremental cost-effectiveness ratios exceed the upper end of the range normally approved by the Appraisal Committee.^[8] During the first four months following its implementation, the supplementary advice was applied to the appraisals of a number of treatments for various cancers (renal cell carcinoma, hepatocellular carcinoma and multiple myeloma).^[9]

There are likely to be a number of factors contributing to the negative decisions made by NICE in its assessment of cancer medicines, such as the cost of the products, the lack of availability of trial data relating to reliable clinical endpoints, and the inadequacy of the QALY in capturing changes in health that are relevant for cancer. The latter is the subject of this paper, which reports the results of a literature review of the estimation of QALY gains and the use of the QALY in the economic evaluation of cancer treatments. The review seeks to investigate how well the QALY captures the health gains generated by cancer treatments, to identify the potential limitations of the existing QALY framework, and to examine whether the published literature provides any suggestions as to

how to address these limitations. The paper focuses in particular on the instruments and features of the QALY approach currently embraced by NICE. In light of the supplementary advice noted above, and the fact that new oncology treatments are increasingly targeted at advanced disease, the main emphasis of the paper is on end-of-life cancer. Its findings may therefore be relevant for the evaluation of other life-threatening conditions. Conversely, the fact that many of the relevant treatments intend or expect only to generate improvements in survival of a few weeks or months, the patients affected by these types of disease are likely to differ considerably from other populations whose health gains are also assessed using the QALY methodology.

In the following paper, Section 2 sets out a simple version of NICE's framework to show how QALY estimates are produced. Section 3 presents the methodology employed to conduct the literature review. Section 4 discusses the results of the review, and Section 5 summarises the findings of the paper and explores some implications for policy and future research.

2. The QALY framework

NICE's approach to generating QALY estimates can be described using a three-step framework. The first step is to describe the health states experienced by patients. This is usually done using generic preference-based measures which can be employed across different disease areas. NICE's preferred measurement measure, the EQ-5D,^[10] is the output of a questionnaire that is administered directly to respondents either in the context of clinical trials or in clinical practice.

The second step is to convert each resulting EQ-5D health state into a value, or coefficient, based on the preferences of a sample of the general population. Techniques used to elicit individual preferences include choice-based methods such as the time trade-off (TTO), where respondents are asked to make a trade off between quality of life and length of life, and the standard gamble, where respondents are asked to choose between a certain outcome and a gamble involving both a positive and a negative outcome (usually perfect health and death). Rating scales, such as the visual analogue scale (VAS) may also be used, but these are generally considered inferior to choice-based methods.^[11] NICE recommends the use of a set of tariff values estimated using the TTO for EQ-5D health states based on a study involving over 3,000 members of the UK population.^[12]

The final step is to calculate the QALY gain associated with an intervention. This involves multiplying the duration of each health state experienced by patients in the treatment pathway by the corresponding HRQL value for that health state. The resulting values are then summed according to the likely sequence of health states (with a discount rate applied to health states occurring in the future), estimated from primary data or by modelling the long-term benefits of treatment by extrapolating from short-term data. This paper focuses on the methods used to assess HRQL (i.e. the first two steps).

3. Methods for the literature search

Two methods were used to obtain data for the review. The first method was an electronic search of the MEDLINE database using logical combinations of keyword terms related to health state valuation, QALYs, EQ-5D and cancer. 155 potentially relevant English language records were identified using the keyword search.

The second method was a hand search of published articles written by key authors in the field of health state valuation, as identified through an inspection of the most relevant records found using the keyword search and through recommendations by colleagues. The personal web pages and academic curriculum vitae of these authors were used to obtain additional articles that had not been identified using the keyword search. A further 191 potentially relevant records were identified using the key author search.

Records were selected for the review if they: (i) focused on cancer or on end-of-life diseases with similar characteristics to cancer; or (ii) discussed measures, instruments and techniques related to the construction of QALYs. We considered empirical studies, systematic reviews, and theoretical papers. Reference lists from included papers were checked to identify further relevant studies. Based on these criteria, 61 records were included in the review. We categorised these into three broad methodological areas: 1) how to describe health states; 2) how to value health states; and 3) whose values to use for the valuation of health states.^[13]

4. Findings from the literature review

4.1 How to describe health states: are generic measures, in particular the EQ-5D, sensitive enough?

The first step is to describe the changes in HRQL experienced by patients. This is typically done using the health status classification system of a generic multi-dimension instrument such as the EuroQol Group's EQ-5D.^[10] NICE recommends the use of the EQ-5D except when data are unavailable or inappropriate for the treatment being evaluated.^[3] There are a number of other candidate generic instruments, such as the Short Form 6D (SF-6D) and the Health Utilities Index (HUI);^[14] and also condition-specific measures (CSMs) which are used alongside or instead of generic instruments in order to capture changes in symptoms and side effects that are particular to the disease in question.

The EQ-5D questionnaire comprises five dimensions: mobility, self-care, usual activity, pain/discomfort and anxiety/depression. Three severity levels are defined for each dimension: no problems, some problems, and extreme problems. This gives a total of 243 unique health states into which individuals may classify themselves. The EQ-5D can be said to lack descriptive richness as it comprises a relatively small number of dimensions and levels. The measure was designed to be simple and versatile, and is therefore intentionally brief.^[15] However, this means that it may not be sensitive enough to capture the often subtle changes in health brought about by some cancer treatments.

Compared to other HRQL measurement instruments such as the SF-6D, the EQ-5D has been found to be less sensitive to health status changes in situations where the degree of 'vitality' is an important element.^[16] The EQ-5D does not include a dimension incorporating vitality, energy or fatigue, as these aspects were previously considered to have only a 'trivial impact' on health state values (the choice of dimensions was informed by a 'survey of lay concepts of health', but our literature search did not identify any information about the methods or results of this survey).^[15] Many studies have since reported that vitality has a significant impact on the HRQL of cancer patients both with and without treatment.^[17-21] We did not, however, identify any empirical studies that have specifically tested the effect of expanding the EQ-5D to incorporate a vitality dimension on cancer patients' HRQL scores. As well as vitality, the EQ-5D excludes a number of other aspects that may be of importance when assessing the health of cancer patients, such as 'sense of coherence'.^[22]

In terms of the number of levels within each dimension, the EQ-5D is restricted in its ability to capture small changes in health because of its relative crudeness. Feeny shows that this problem is particularly applicable to cancer patients, whose health gains from treatment tend to be small.^[23] In some cases, a 'ceiling effect' occurs whereby an individual in fairly good health cannot be distinguished from one in perfect health because both would classify themselves using the least severe level.^[24,25]

Two main ways of addressing concerns about the EQ-5D's lack of sensitivity have been explored in the literature. The first is to refine the system by increasing the number of dimensions and levels. Adding a vitality dimension, for example, would certainly help to capture improvements in an individual's power and vigour that cannot easily be described in terms of the five existing dimensions. However, as noted above we identified no research examining the practicability of doing this. Furthermore, it may be argued that changes in energy levels and tiredness are to some extent already captured indirectly through the 'usual activities' dimension. Thus, the evidence is not yet strong enough to support the inclusion of this extra dimension.

Increasing the number of levels from three to five (for instance) would improve the sensitivity of the EQ-5D – indeed, early prototypes of the EQ-5D included a five-level version^[26] – but would also generate a considerably greater number of unique health states: 3,125 states (5^5) compared to the current 243 (3^5). As well as increasing the burden associated with the administration of the questionnaire, this will likely make the task of calculating weights for all possible health states – whether done directly using general population values or using regression analysis to impute values – considerably more complex and time-consuming. Given the fact that other generic instruments such as the HUI already comprise several hundred thousand health states,^[27] this may well be manageable in terms of the burden of administering the questionnaire. The EuroQoL Group has now developed a new five-level version, the EQ-5D-5L, which appears a valid and reliable extension of the three-level system.^[28-31] The workload associated with the task of calculating values now needs to be carefully weighed up against the benefits of improving the instrument's descriptive power.

The second option is to use CSMs rather than generic measures on the basis that they have been found to be more sensitive to changes in cancer patients' health status.^[32,33] NICE's endorsement of generic measures is based on the fact that they facilitate comparability across different treatments and conditions. In practice, however, despite the fact that researchers have long advocated the use of at least one generic measure if utility values are thought to have an impact on study results,^[34] many pivotal trials rely entirely on CSMs because they are considered to focus on the aspects of health that are most important and relevant to patients. However, unless they are preference-based, CSMs are only of limited use in economic evaluations because they do not provide a valuation of health, and therefore cannot be used to calculate QALYs.

It is, in theory, possible to use mapping (or 'cross-walking') to transform condition-specific, non-preference-based scores into scores for generic instruments. However, a recent review of the evidence on mapping from CSMs, which covered a variety of disease areas including cancer, concluded the use of mapping is always inferior to employing generic measures directly.^[35] Furthermore, mapping is viable only if the generic target measure covers all of the important aspects of health covered by the CSM – it does not overcome any inadequacy in the descriptive system of the generic measure.^[36]

An alternative approach is to develop a preference-based measure using an existing CSM. This approach has already been used, for example, to enable the use of a CSM for urinary incontinence in economic evaluation,^[37] and a recent study has examined the feasibility of deriving a preference-based measure from the cancer-specific EORTC QLQ-C30 instrument.^[38] There are concerns that

CSMs may not be comparable to generic measures, and in particular that they may fail to capture important information about co-morbidities.^[36] Comparability between different descriptive systems can be improved by ensuring consistency in the methodology and sample type used to obtain values.^[4]

4.2 How to value health states: is the time trade-off and its assumption of constant proportional trade-off valid in the context of cancer treatments?

The second step of the QALY calculation is to value each health state using the preferences of a sample of the general population. One of the methods used to elicit these preferences is the TTO, a choice-based means of deriving weights for different health states. NICE's most recent guidelines recommend the use of the TTO technique for estimating the values of health states defined by the EQ-5D or other validated HRQL measures.^[3]

The TTO method involves asking participants to choose between two certain scenarios: 1) ill health state i for time t followed by death; and 2) perfect health for time $x < t$ followed by death. The respondent's indifference point is found by varying time x – at this point the required preference score for state i is given by $h_i = x/t$.^[39]

One of the key assumptions underlying the use of TTO in the traditional QALY model is that of 'constant proportional trade-off' (CPT). This states that individuals are willing to trade a constant proportion of their life expectancy to obtain a proportional improvement in HRQL, regardless of the number of the life years remaining.^[40] This implies that the proportional amount of time which is traded off is independent of the duration, or magnitude, of t (i.e. the value h_i should hold whether the state i lasts 5, 10 or any other number of years).

In the development of weights for EQ-5D health states in 1995, general population samples were given a 10-year TTO framework^[41] – a scenario which does not reflect the typically much shorter duration of health states in cancer, particularly in late-stage cancer where remaining life expectancy may be less than one year. Empirical studies of time-related preferences have identified important challenges to CPT that may be of particular relevance for late stage cancer patients.

One challenge relates to the concept of people having a 'maximal endurable time' (MET) beyond which attitudes towards survival become more negative.^[42] Dolan and Stalmeier observed that, when asked to consider a relatively severe health state, approximately half of all respondents indicated that they preferred less time (10 years) to more time (20 years) in that state, thereby expressing MET preferences.^[37] In other words, these individuals become willing to sacrifice a higher proportion of length of life to achieve a smaller gain in quality of life (or even no gain in quality of life at all), in contradiction to the CPT assumption. A number of other studies have similarly found evidence of MET preferences.^[43-46]

Another challenge relates to the change in people's attitudes when their remaining expectancy becomes very short. Miyamoto and Eraker reported that people with less than one year of remaining life expectancy were unwilling to give up any of that time to improve their quality of life, a phenomenon termed as 'indifference to health quality at short duration'.^[47] This type of preference was not observed in situations where remaining life expectancy was longer than one year. Attema and Brouwer's recent review found a number of studies that reported violations of CPT in this direction.^[48] Figure 1 provides a simple illustration of these challenges to CPT.

Figure 1: Illustration of challenges to CPT

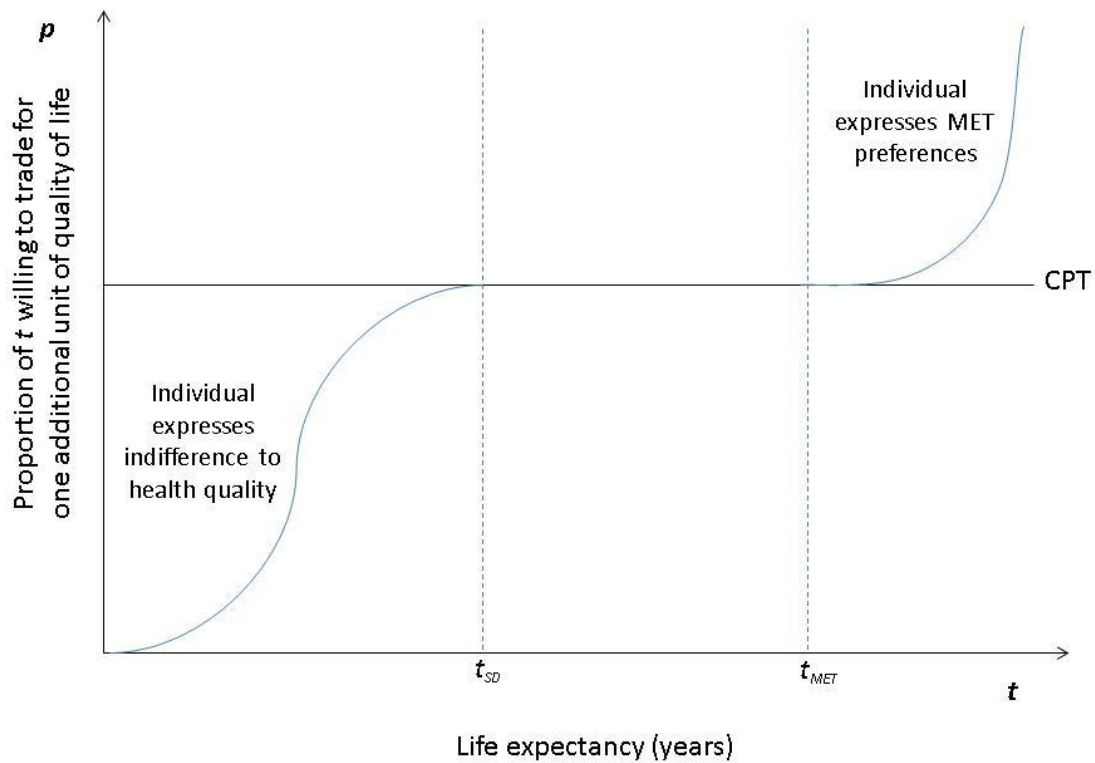


Figure 1 provides a simple illustration of the challenges to CPT. The x-axis denotes t , or life expectancy (years of life remaining). The y-axis denotes the proportion of t that one would trade off for a specified improvement in HRQL. If the CPT assumption is valid, this proportion should remain the same regardless of the value of t , as indicated by the horizontal line labelled 'CPT'. When t is very short (less than t_{SD}), individuals may become less willing to give up life expectancy to achieve a gain in HRQL. This is the 'indifference to health quality at short duration' challenge. Conversely, when t becomes sufficiently long (greater than t_{MET}), individuals may become willing to give up an increasingly large proportion of their remaining life expectancy in order to achieve a gain in HRQL. This is the MET challenge.

The literature provides a few suggestions of ways to overcome the problems associated with CPT violation, including the varying of preference weights according to duration, the use of subjective life expectancy (SLE), and the development of a duration-utility value function. Dolan argued that preference weights for health status should be varied according to different time durations.^[49] Having demonstrated the feasibility of using the VAS to elicit valuations for both long and short durations, he proposed the use of a two-stage weighting process: first, to generate a TTO-VAS mapping function based on TTO and VAS valuations for a long duration, and then to elicit VAS valuations for a shorter durations which can then be converted into short-duration TTO scores using the mapping function estimated in the first stage.

Stiggelbout et al. proposed the use of SLE – a life expectancy that is realistic for patients – as the anchor timeframe in TTO exercises, rather than the artificially standardised 10-year duration.^[50] TTO values could then be generated for different SLEs, although the authors did not explore in detail how the TTO protocol would need to be adjusted to achieve this. Also, their study did not examine individuals' preferences when remaining life expectancy was shorter than one year. There is some recent evidence confirming that the use of SLE does have some impact on one's willingness to trade – specifically, the fewer the number of subjective expected life years, the smaller the proportion of time traded off.^[51] Nevertheless, some issues around the practicability of applying the SLE approach remain unresolved.

Finally, Buckingham and Devlin suggested a number of ways of dealing with CPT-related issues,^[52] including the development of a value function describing the relationship between duration and utility, and the provision of multiple tariffs for a range of durations. The value of the HRQL component part of the QALY would then depend on how long the patient was going to live for. Sharma and Stano have similarly called for the development of a two-dimensional algorithm based on health status and its duration.^[53]

A point relevant to both sections 4.1 and 4.2 is that as patients progress to advanced stage disease, it is not only their time trading preferences that change, but also their preferences about the relative importance of different dimensions of health and wellbeing. However, we did not find a substantial literature on this topic and it is therefore not discussed further in this paper.

4.3 Whose values to use: should patients' health status be valued by the general population or by the patients themselves?

Determining a value for each health state requires a decision to be made as to whether these values should be elicited from members of the general population (acting as 'hypothetical patients') or from individuals with experience of those states (either as past or present patients). The argument for using patient valuations rests on the belief that patients themselves should be best judges of the relative desirability of their own health states. However, the US Public Health Service Panel on Cost-Effectiveness in Health and Medicine recommended that a representative sample of the general population should be used for health state valuations, as long as the judgements supporting these valuations are "informed, unbiased, and competent".^[54] This recommendation is mirrored in NICE's guidelines,^[3] and is based on the argument that in publicly funded health care systems, the aim of economic evaluation is not to make decisions at the individual patient level, but to guide policies that fulfil the interests of society as a whole. A further argument is that unlike patients, the general population tends not to have a vested interest in getting access to treatment and is therefore more likely to give an unbiased view of the value of the health gain it generates.

However, the practice of using valuations of the general population becomes problematic if they differ substantially from patient valuations because the "informed, unbiased, and competent" requirement has not been met – for example, because respondents have misunderstood what is really like for patients to live with illness. There are a number of factors that may contribute to these discrepancies (see Ubel et al.^[55] for a more comprehensive discussion). Some of these relate to the limited scope of any approach being used to describe health states to a general population sample. As demonstrated by Insinga and Fryback, differences may occur because the descriptive system omits a relevant dimension of health, or because within a given dimension of health patients rate themselves as being between the specified severity levels.^[56] Thus, if a dimension such as vitality plays a major part in determining the HRQL of cancer patients but is not included in the descriptive system, then the general population sample will not be basing its valuations on the health state actually being experienced by patients. A related issue is the observation that general population respondents typically focus on the negative aspects of ill-health whilst ignoring unaffected life domains that the descriptive system does not bring to their attention (this is often referred to as the 'focusing illusion').^[57]

Another issue is the fact people's evaluations of a particular health state tend to depend on their current level of health.^[58] A very slight improvement in mobility may seem insignificant to a fully-functioning, healthy individual, but might be very important and meaningful for (and therefore appear much larger to) a late-stage cancer patient who is mostly bed-ridden. There are also important differences between the way in which patients adapt to states of ill-health over time (for example, by adjusting their activities in order to lessen the impact of their disability) and healthy individuals' perception of their ability to adapt. There is substantial evidence showing that people are poor predictors of their ability to adapt to ill-health.^[59] Some researchers argue that certain elements of adaptation (such as the lowering of expectations) are regrettable and that it would be inappropriate for these factors to influence health care prioritisation decisions.^[60] Moreover, the

patient adaptation phenomenon may not apply to end-of-life cancer patients who are not afforded the time to adapt to their new, temporary health state.

The most comprehensive review of empirical studies in this area was conducted by de Witt et al.,^[61] who analysed the results of 38 studies comparing patient and non-patient valuations (12 of which explicitly focused on cancer patients). They reported that of the 27 studies indicating at least some difference between the two groups, 22 reported higher (better) patient values; two reported lower patient values; and three showed contradictory results. These results are further supported by the findings of more recent empirical studies^[56,62-65] (see Table I for a summary of the empirical data considered in this section). Hence, the majority of evidence indicates that on average patients tend to value a given health state more highly than do non-patients (although it should be noted that a number of researchers have pointed out important caveats and exceptions to this rule^[65-67]). This means that for interventions aimed at, say, restoring low HRQL individuals to full health, the use of public valuations will generate larger health gains than if patient valuations were to be used. However, this also implies that interventions aimed at extending the lives of such individuals will be valued lower by general population samples than by patients due to their perceptions about the level of HRQL being maintained.^[68] Thus the use of general population values rather than patient values will tend to favour interventions aimed at achieving perfect functioning whilst disfavouring life-prolonging interventions.^[69]

With specific regard to cancer, Slevin et al. found that when responding to a questionnaire describing hypothetical chemotherapy scenarios, cancer patients were more likely than general population controls to accept intensive treatments for small potential benefits.^[70] The results indicate that individuals' preferences change considerably when they are faced with the genuine possibility of death, and Slevin et al. suggest that perhaps only end-of-life patients themselves can appropriately evaluate preferences relating to life and death. This is consistent with findings discussed in section 4.2 which indicate that end-of-life patients become increasingly reluctant to sacrifice life expectancy for improved quality of life.

The question of whether improved quality of life is preferable to an increase in life expectancy is a matter of value judgement, and it is clear that patients and the general population differ in the way in which they approach this trade-off. If we accept the normative argument for using general population values, then it is important that respondents are well-informed about what the patient experience is actually like. For example, health state descriptions could be made more realistic by adding extra dimensions or by providing respondents with testimonials from patients with the illness in question.^[69] Steps could also be taken to lessen psychological phenomena such as the focusing illusion – for example, by providing reminders of the life domains that remain unaffected by the illness. However, there will be a trade-off between providing sufficient detail and overburdening respondents. There is also some evidence that the provision of clinical information can induce an emotional and unconsidered response in general population respondents,^[71] which suggests that any hint that the health state relates to cancer (or some other high-profile disease) is likely to introduce bias. Evidence on the correlation between cancer patients' health state valuations and those of their family caregivers (who have direct opportunities to observe the patients' problems over time and might therefore be described as being relatively well-informed

about their preferences) is mixed.^[72,73] These factors indicate that the task of informing general population respondents about the patient experience is far from straightforward.

Dolan and colleagues go further, arguing that even patients' valuations will not correctly predict the degree to which health states will actually affect them, and that preferences are problematic from whomever they are elicited.^[74] They propose evaluating health technologies in terms of their effect on subjective wellbeing, which they suggest offers a more accurate way of assessing the actual impact of health states on people's lives.

5 Summary and policy implications

A review of 110 evaluations of cancer treatments found that in the majority of cases, the use of the QALY yielded resource allocation decisions that were no different than if no quality of life adjustment had been made.^[75] Although this might indicate that most of the benefit from cancer treatment comes from improvements in length of life rather than quality of life, it equally could mean that the QALY construction methodology is inadequate, insofar as it is failing to capture important quality of life issues.

This review has shown that the QALY, because of the way in which it is constructed, may indeed fail to accurately capture the value of the health gains that are deemed important by cancer patients. The evidence suggests that the main issues with using the QALY in the context of cancer relate to the techniques used to describe and value health states, and the most appropriate source of these valuations.

The EQ-5D descriptive system has been found to lack sensitivity in measuring changes in the health states of cancer patients. One possible solution to this problem is to increase the number of dimensions and levels in the system in order to improve sensitivity. Whilst research in this area is ongoing, it may be necessary to examine the practical implications of adding extra dimensions/levels specifically in the context of cancer. The benefits of improved sensitivity will need to be balanced against the increased complexity arising from the larger number of health states. A second possible solution is to develop cancer-specific instruments that are preference-based and amenable to valuation. Given the availability of other generic instruments such as SF-6D, it also seems advisable to further explore how they perform in cancer to see if they represent a more valid alternative to EQ-5D.

When valuing health states, the use of the TTO technique requires making the assumption of CPT. This assumption has been challenged in the context of cancer where patients typically have a short remaining life expectancy and/or may experience a maximal endurable time beyond which they are unwilling to extend their life. Under these circumstances, a weighting system based on CPT may be invalid. Different timeframes have been proposed to resolve these non-CPT issues, but none have been widely accepted in routine practice. More research is needed to improve the understanding of the relationship between length of the health states and corresponding value. Some of the studies commissioned by the EuroQoL group are currently exploring new approaches to tackle this issue [<http://www.euroqol.org/home.html>].

With regard to the third issue of whose valuations should be used, a growing body of empirical evidence indicates that patients tend to value a given health state more highly than non-patients.^[56,61-65] This has important implications in terms of resource allocation if it is deemed appropriate for health state valuations to be elicited from the general population, as is currently the case in the UK and elsewhere. A key concern here is that these valuations will be biased if they reflect a misunderstanding on the part of general population respondents about the true nature of the illness and its impact on one's health. Future research should therefore focus on ways of making health state descriptions more comprehensive and/or realistic in order to obtain better-

informed valuations. The main challenge will then be to avoid contaminating respondents' preferences by providing too much detail – this is particularly important for prominent diseases such as cancer, given that the public may have misguided preconceptions about conditions that have received a large amount of media exposure. There is also need for further debate about the role that subjective wellbeing can play in assessing the impact of health states on people's lives.

It is clear, therefore, that the methods currently being used to assess health benefit are far from perfect when evaluating health care technologies for cancer and other end-of-life diseases. Given the increasing number of medicines being developed to treat these diseases, it is important that the methods are improved in order that health care budgets deliver value for money. The long-term research agenda should address the problems raised in this paper by exploring ways of modifying the QALY and the way in which it is generated, along the lines proposed above. The agenda is developed further elsewhere.^[76] To some extent this research is already underway – as part of the work being commissioned by the EuroQol group, for example. In the meantime, where there is an absence of agreement about what these modifications should entail, it is important for health technology assessment bodies such as NICE to recognise explicitly the limitations of the QALY when appraising cancer treatments. In cases where the QALY is likely to undervalue (or indeed, to overvalue) the actual health benefit accruing to cancer patients, this consideration should inform resource allocation decisions – for example, by means of a deliberative decision-making process where the implications of adopting alternative methods for constructing QALYs are considered alongside the clinical- and cost-effectiveness evidence.

Reference list

1. Boersma C, Broere A, Postma MJ. Quantification of the potential impact of cost-effectiveness thresholds on Dutch drug expenditures using retrospective analysis. *Value Health* 2010; 13: 853-6
2. Weinstein MC, Skinner JA. Comparative effectiveness and health care spending — implications for reform. *New Eng J Med* 2010; 362: 460-5.
3. NICE. Guide to the methods of technology appraisal; 2008. Available from URL: <http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf> [Accessed 2009 Nov 16]
4. Brazier J, Ratcliffe J, Salomon J, et al. Measuring and valuing health benefits for economics evaluation. Oxford: Oxford University Press, 2007
5. Department of Health. Cancer Reform Strategy; 2007. Available from URL: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_081013 [Accessed 2009 Nov 16]
6. NICE. Technology appraisal guidance 178: Bevacizumab (first-line), sorafenib (first- and second-line), sunitinib (second-line) and temsirolimus (first-line) for the treatment of advanced and/or metastatic renal cell carcinoma; 2009. Available from URL: <http://guidance.nice.org.uk/TA178/Guidance/pdf/English> [Accessed 2009 Nov 16]
7. Hawkes N. The Times. 2008 Aug 7. Available from URL: http://www.timesonline.co.uk/tol/life_and_style/health/article4474425.ece [Accessed 2009 Nov 16]
8. NICE. Appraising life-extending, end of life treatments; 2009. Available from URL: <http://www.nice.org.uk/aboutnice/howwework/devnicetech/endoflifetreatments.jsp?domedia=1&mid=88ACDAE5-19B9-E0B5-D422589714A8EC6D> [Accessed 2009 Nov 16]
9. NICE. Update report on the application of the 'end-of-life' supplementary advice in health technology appraisals; 2009. Available from URL: www.nice.org.uk/media/835/8E/ITEM7EndOfLifeTreatments.pdf [Accessed 2009 Nov 16]
10. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001; 33: 337-43
11. Brazier J, Deverill M, Green C, et al. A review of the use of health status measures in economic evaluation. *Health Technol Assess* 1999; 13: 9
12. Dolan P. Modelling valuations for EuroQol health states. *Med Care* 1997; 35: 1095-108
13. Dolan P. The measurement of health related quality of life for use in resource allocation in health care. In: Culyer AJ, Newhouse JP, editors. *Handbook of health economics*. Vol. 1. Amsterdam: Elsevier Science, 2000

14. Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the economic evaluation of health care programmes. 3rd ed. New York: Oxford University Press, 2005
15. Williams A. The role of the EuroQoL Instrument in QALY calculations. CHE Discussion Paper 130, 1995
16. Grieve R, Grishchenko M, Cairns J. SF-6D versus EQ-5D: reasons for differences in utility scores and impact on reported cost-utility. *Eur J Health Econ* 2009; 10: 15-23.
17. Broeckel A, Jacobsen PB, Horton J, et al. Characteristics and correlates of fatigue after adjuvant chemotherapy for breast cancer. *J Clin Oncol* 1998; 16: 1689-96
18. Stone P, Ream E, Richardson A, et al. Cancer-related fatigue – a different of opinion? *Eur J Cancer Care* 2003; 12: 20-7
19. Esbensen BA, Osterlind K, Roer O, et al. Quality of life of elderly persons with newly diagnosed cancer. *Eur J Cancer Care* 2004; 13: 443–53
20. Hofman M, Ryan JL, Figueroa-Moseley CD, et al. Cancer-related fatigue: the scale of the problem. *Oncologist* 2007; 12: 4-10
21. Dobrez D, Cella D, Pickard S, et al. Estimation of patient preference – based utility weights from the functional assessment of cancer therapy-general. *Value Health* 2007; 10: 266-72
22. Kröz M, Büssing A, von Laue HB, et al. Reliability and validity of a new scale on internal coherence (ICS) of cancer patients. *Health Qual Life Outcomes* 2009; 7:59
23. Feeny DH. The roles for preference-based measures in support of cancer research and policy. In: Lipscomb J, Gotay CC, Snyder C, editors. *Outcomes assessment in cancer: measures, methods and applications*. New York: Cambridge University Press, 2005
24. Bharmal M, Thomas J. Comparing the EQ-5D and the SF-6D Descriptive Systems to Assess Their Ceiling Effects in the US General Population. *Value Health* 2006; 9: 262-71
25. Barton G, Sach T, Doherty M, et al. An assessment of the discriminative ability of the EQ-5D, SF-6D, and EQ VAS, using sociodemographic factors and clinical conditions. *Eur J Health Econ* 2008; 9: 237-49
26. van Agt H, Bonsel G. The number of levels in the descriptive system. In: Kind P, Brooks R, Rabin R, eds., *EQ-5D concepts and methods: a developmental history*. Netherlands: Springer, 2005
27. Horsman J, Furlong W, Feeny D, et al. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes* 2003; 1. Available from URL: <http://www.hqlo.com/content/1/1/54> [Accessed 2009 Nov 16]
28. Pickard AS, De Leon MC, Kohlmann T, et al. Psychometric Comparison of the Standard EQ-5D to a 5 Level Version in Cancer Patients. *Med Care* 2007; 45: 259-63

29. Pickard AS, Kohlmann T, Janssen MF, et al. Evaluating equivalency between response systems: application of the Rasch Model to a 3-level and 5-level EQ-5D. *Med Care* 2007; 45: 812-9
30. Kind P. Size Matters: EQ-5D in Transition. *Med Care* 2007; 45: 809-11
31. Janssen MF, Birnie E, Haagsma, JA, et al. Comparing the Standard EQ-5D Three-Level System with a Five-Level Version. *Value Health* 2008; 11: 275-84
32. Kind P, Macran S. Eliciting social preferences weights for functional assessment of cancer therapy-lung health states. *Pharmacoeconomics* 2005; 23: 1143-53
33. Krahn M, Bremner KE, Tomlinson G, et al. Responsiveness of disease-specific and generic utility instruments in prostate cancer patients. *Quality Life Res* 2007; 16: 509-22
34. Neyt M. Towards more consistent use of generic quality-of-life instruments. *Pharmacoeconomics* 2010; 28: 345-6
35. Brazier J, Yang Y, Tsuchiya A. Review of methods for mapping between condition specific measures onto generic measures of health. In: Office of Health Economics. Report of the Office of Health Economics Commission on NHS outcomes, performance and productivity, 2008. Available from URL: <http://www.ohe.org/page/Commissionreport.cfm> [Accessed 2009 Nov 16]
36. Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability? *Health Econ* 2010; 19: 125-9
37. Brazier J, Czoski-Murray C, Roberts J, et al. Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire. *Med Decis Making* 2008; 28: 113-26
38. Rowen D, Brazier JE, Young TA, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. HEDS Discussion Paper 10/01, 2010. Available from URL: <http://eprints.whiterose.ac.uk/10872/> [Accessed 2010 Jul 23]
39. Torrance GW, Thomas W, Sackett D. A utility maximization model for evaluation of health care programs. *Health Serv Res* 1972; 7: 118-33
40. Dolan P, Stalmeier P. The validity of time trade-off values in calculating QALYs: constant proportional time trade-off versus the proportional heuristic. *J Health Econ* 2003; 22: 445-58
41. Gudex C. Time trade-off user manual: props and self-completion methods. Centre for Health Economics Working Paper No 020cheop, 1994
42. Sutherland HJ, Llewelyn-Thomas H, Boyd NF, et al. Attitudes towards quality of survival: the concept of "maximum endurable time". *Med Decis Making* 1982; 2: 299-309
43. Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Econ* 2006; 15: 393-402
44. Stalmeier P, Bezembinder T, Unic I. Proportional heuristics in time tradeoff and conjoint measurement. *Med Decis Making* 1996; 16: 36-44

45. Stalmeier PFM, Chapman GB, de Boer AGM, et al. A fallacy of the multiplicative QALY model for low quality weights in students and patients judging hypothetical health states. *Int J Technol Assess Health Care* 2001; 17: 488-96
46. Stalmeier PFM, Lamers LM, Busschbach JJV, et al. On the assessment of preferences for health and duration: maximal endurable time and better than dead preferences. *Med Care* 2007; 45: 835-41
47. Miyamoto JM, Eraker SA. A multiplicative model of the utility of survival duration and health quality. *J Exp Psychol* 1988; 117: 3-20
48. Attema AE, Brouwer WBF. On the (not so) constant proportional trade-off in TTO. *Qual Life Res* 2010; 19: 489-97
49. Dolan P. Modelling valuations for health states: the effect of duration. *Health Policy* 1996; 38: 189-203
50. Stiggelbout AM, Kiebert GM, Kievit J, et al. The 'utility' of the time trade-off method in cancer patients: feasibility and proportional trade-off. *J Clin Epidemiol* 1995; 48: 1207-1
51. van Nooten FE, Koolman X, Brouwer WBF. The influence of subjective life expectancy on health state valuations using a 10 year TTO. *Health Econ* 2008; 18: 548-558
52. Buckingham K, Devlin N. A note on the nature of utility in time and health and implications for cost utility analysis. *Soc Sci Med* 2009; 68: 362-7
53. Sharma R, Stano M. Implications of an economic model of health states worse than dead. *Journal of Health Econ* 2010; 29: 536-40
54. Gold MR, Siegel JE, Russell LB, et al. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996.
55. Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Qual Life Res* 2003; 12: 599-607
56. Insinga RP, Fryback DG. Understanding differences between self-ratings and population ratings for health in the EuroQOL. *Qual Life Res* 2003; 12: 611-9
57. Schkade DA, Kahneman D. Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychol Sci* 1998; 9: 340-6
58. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica* 1979; 47: 263-91
59. Dolan P, Kahneman D. Interpretations of utility and their implications for the valuation of health. *Econ J* 2008; 118: 215-34
60. Menzel P, Dolan P, Richardson J, Olsen JA. The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Soc Sci Med* 2002; 55: 2149-58

61. de Witt GA, Busschbach JJV, de Charro F. Sensitivity and perspective in the valuation of health status: whose values count? *Health Econ* 2000; 9: 109-26
62. Giesinger JM, Golser M, Erharter A, et al. Do neurooncological patients and their significant others agree on quality of life ratings? *Health Qual Life Outcomes* 2009; 7: 87
63. Polsky D, Willke RJ, Scott K, et al. A comparison of scoring weights for the EuroQol derived from patients and the general public. *Health Econ* 2001; 10: 27-37
64. Ratcliffe J, Brazier J, Palfreyman S, et al. A comparison of patient and population values for health states in varicose veins patients. *Health Econ* 2007; 16: 395-405
65. Mann R, Brazier J, Tsuchiya A. A comparison of patient and general population weightings of EQ-5D dimensions. *Health Econ* 2008; 18: 363-72
66. Ditto PH, Hawkins NA, Pizarro DA. Imagining the end of life: on the psychology of advance decision making. *Motiv Emot* 2005; 29: 481-502
67. Lacey HP, Fagerlin A, Loewenstein G, et al. It must be awful for them: perspective and task context affects ratings for health conditions. *Judgm Decis Mak* 2006; 1: 146-52
68. Harris J. QALYfying the value of human life. *J Med Ethics* 1987; 13: 117-23
69. Brazier J, Akehurst R, Brennan A, et al. Should patients have a greater role in valuing health states? *Appl Health Econ Health Policy* 2005; 4: 201-8
70. Slevin ML, Stubbs L, Plant JJ, et al. Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses, and general public. *BMJ* 1990; 300: 1458-60
71. Roberts T, Bryan S, Heginbotham C, et al. Public involvement in health care priority setting: an economic perspective. *Health Expect* 1999; 2: 235-44
72. O'Leary JF, Fairclough DL, Jankowski MK, et al. Comparison of time-tradeoff utilities and rating scale values of cancer patients and their relatives: evidence for a possible plateau relationship. *Med Decis Making* 1995; 15: 132-7
73. Tang ST, McCorkle R. Use of family proxies in quality of life research for cancer patients at the end of life: a literature review. *Cancer Investig* 2002; 20: 1086-104
74. Dolan P, Lee H, King D, et al. How does NICE value health? *BMJ* 2009; 339: 371-3
75. Tengs TO. Cost-effectiveness versus cost-utility analysis of interventions for cancer: does adjusting for health-related quality of life really matter? *Value Health* 2004; 1: 70-8
76. Shah KK, ed. Assessing and appraising oncology medicines: what are the key areas of methodological research to pursue? OHE workshop summary; 2009. Available from URL: <http://www.ohe.org/lib/liDownload/663/Oncology%20Workshop.pdf?CFID=3043207&CFTOKEN=46888688> [Accessed 2010 Sep 29]

Table 1: Classification of studies relating to ‘whose values to use’ (7 empirical studies + 2 reviews)

Study	Country	Design	Sample size	Cancer-specific?	Patient versus non-patient valuations
de Wit et al. (review) [58]	Various	1	8 studies	4 – yes 4 – no	2 > 4 ≥ 1 = 1 ><
		2	4 studies	0 – yes 4 – no	1 > 1 = 2 ><
		3	4 studies	2 – yes 2 – no	3 > 1 ≤
		4	1 study	1 – yes 1 – no	1 =
		5	5 studies	0 – yes 5 – no	2 > 1 ≥ 2 =
		6	7 studies	0 – yes 7 – no	3 > 1 ≥ 3 =
		7	3 studies	3 – yes 0 – no	3 >
		8	3 studies	2 – yes 1 – no	2 = 1 ≤
		9	3 studies	0 – yes 3 – no	1 > 1 ≥ 1 =
Giesinger et al. [59]	AUT	6	1-99	yes	≥
Insinga and Fryback [53]	UK	9	1,000+	no	><
Ratcliffe et al. [61]	UK	9	100-499	no	>
Mann et al. [62]	UK	9	1,000+	no	><
Lacey et al. [64]	USA	1	100-499	no	><
Polsky et al. [60]	USA, CAN	9	500-999	no	>
O’Leary et al. [69]	USA	1, 6	100-499	yes	≥
Tang and McCorkle (review) [70]	Various	6	25 studies	25 – yes 0 – no	9 > 8 =

Table 1 legend

Design (classification of designs adapted from the review by de Witt et al.):

1. Patient and non-patient groups value hypothetical health states related to the actual health state of the patient.
2. Patient and non-patient groups value hypothetical health states unrelated to the actual health state of the patient.
3. Different patient groups (i.e. with different stages of disease) value hypothetical health states related to the actual health state of the patient.
4. Different patient groups value hypothetical health states unrelated to the actual health state of the patient.
5. General population samples value hypothetical health states. The analysis entails comparing the values of respondents in normal health with the values of respondents in dysfunctional health states.
6. Patient and proxy groups value the actual health state of the patient.
7. Patient and non-patient groups choose between hypothetical treatment choices (which are linked to hypothetical health states).
8. Values for health states are elicited from the patients before they enter that health state, and again from the same patients after they have obtained experience with the health state.
9. Patients describe and value their actual health state based on a classification system or profile. The patient value is compared with a population value corresponding to that health state.

Sample size: For reviews, this column indicates the total number of empirical studies included in the review.

Cancer-specific: This column indicates whether or not the study explicitly focused on cancer; or for reviews, the number of studies in the review that did so.

Patient versus non-patient valuations: > patient values higher than non-patient values; \geq patient values higher than or equal to non-patient values; = patient values equal to non-patient values; \leq patient values lower than or equal to non-patient values; >< contradictory results.

Table 2: List of acronyms used in this paper

NICE	National Institute for Health and Clinical Excellence
NHS	National Health Service
QALY	Quality-adjusted life year
HRQL	Health-related quality of life
TTO	Time trade-off
VAS	Visual analogue scale
CSM	Condition-specific measure
SLE	Subjective life expectancy
MET	Maximal endurable time
CPT	Constant proportional trade-off