View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Working papers series

WP ECON 11.15

Valuing Qalys at the end of life

Jose-Luis Pinto-Prades (U. Pablo de Olavide)

Fernando-Ignacio Sánchez-Martínez (U. de Murcia)

Belen Corbacho (Andalusian Agency for Health Technology Assessment)

JEL Classification numbers: I18, D06, H4

Keywords: QALY weights, end of life, palliative care, life extension

UNIVERSIDAD PABLO DE OLAVIDE SEVILLA

Department of Economics

VALUING QALYS AT THE END OF LIFE

Jose-Luis Pinto-Prades[1]
Fernando-Ignacio Sánchez-Martínez[2]
Belen Corbacho[3]

[1]University Pablo de Olavide, Seville, Spain
[2]University of Murcia, Spain
[3]Andalusian Agency for Health Technology Assessment, Seville, Spain

Correspondence to: Jose-Luis Pinto, Department of Economics, University Pablo de Olavide. Ctra. Utrera, Km. 1 41013-Sevilla (Spain), jlpinto@upo.es

Abstract

There have been changes in the way that NICE evaluates medical treatments for patients who are in the last stages of their lives. If medicines fulfil some criteria to be considered “end of life” NICE considers if QALYs gained under these circumstances should receive an extra weight. In this paper we provide evidence about the social support that this policy may have. We present the result of three surveys conducted in the Spanish general population (n=813). Survey 1 compared increases in life expectancy for patients at the end of their lives with health gains from temporary health problems. Survey 2 compared health gains for temporary health problems with health gains from end of life palliative care. Survey 3 compared increases in life expectancy with palliative care in both cases for end of life patients. Preferences were elicited with Person Trade-Off and Willingness to pay techniques. Our results suggest that QALYs for end of life treatments have a higher social value than for temporary health problems. However, we also find that people discriminate between different ways of health gains within End of Life treatment. People seem to attach a greater weight to palliative care than to life extension.

Key words: QALY weights, end of life, palliative care, life extension

JEL: I18, D6, H4

1. Introduction

There is a renewed interest in the literature in the area of QALY weights. One of the reasons for this is the decision by the UK government to introduce Value-Based Pricing (Department of Health, Medicines, Pharmacy & Industry Group, 2010) for medicines in 2014. It has been stated (Department of Health, Medicines, Pharmacy & Industry Group, 2011) that “QALY measures may not capture all aspects of the value society gains from new treatments…The Government is proposing to calculate ‘QALY weights’” (p.24). Cancer care seems a clear candidate for the introduction of ‘QALY weights’. In fact, in a public consultation about the value of End-of-Life (EoL) medicines NICE (2008) asked if it should “place additional weight on proven survival benefits in patients with terminal illness and short life expectancy”. A significant majority (63%) of those who responded to the consultation document backed this proposal. Following this consultation, NICE (2009a) established that, if medicines fulfil some criteria they will be regarded as “end of life” and the Appraisal Committee will consider “the impact of giving greater weight to QALYs achieved in the later stages of terminal diseases” and “the magnitude of the additional weight”.

There has been much debate about the potential extra weight that QALYs should receive in the case of end of life treatments (Towse, 2009) (Raftery, 2009) (IOM [Institute of Medicine] 2009) (Maynard & Bloor, 2009). Most of the arguments are based on normative judgments. However, there is very little empirical evidence on whether this decision is, or is not, based on social preferences.

There is some evidence about how doctors and health care professionals perceive the value of cancer drugs. Morris & Perez, (2000) interviewed convenient groups of managers, hospital doctors and nurses about the monetary value of a new cancer treatment. They were asked about both their individual (what they would pay for themselves or members of their family) and social (how much the health system should pay) Willingness To Pay (WTP). On average, doctors were willing to pay more than members of the other two groups, in both respects. Nadler, Eckert, & Neumann (2006) interviewed 90 academic medical oncologists. They were asked to identify the minimum survival benefit offered by a new hypothetical medication with a marginal cost of $70,000 per patient at which they would be prepared to prescribe the new medication. Most doctors (60%) thought that increasing life expectancy by 4 months was enough to justify spending $70,000 per patient. Finally, Nadler, Broderick, Zarotsky, & Kim (2009) interviewed 50 health care professionals about their perceptions of the value of three new cancer therapies (bevacizumab, erlonitib, sunitinib). The monetary value of the QALY implied in their responses was approximately $170,000 per QALY gained. All these results apparently show that US health care professionals have a much higher threshold in their minds than the one used by NICE.

The only study that we are aware of that has been conducted among the general population about this issue has been the *EuroVaQ* project (Donaldson, 2010). *EuroVaQ* was funded by the European Commission to derive a monetary value of a QALY based on surveys of the general public. A total of 17,657 subjects responded to an internet based survey. The study was not mainly devised to evaluate end of life issues but it included one question about the value of QALYs at the end of life. The study shows that the monetary value of a QALY is much higher in the context of risk reductions than in the EoL context. However, it is believed that this result is problematic because the context (risk vs. riskless) is a factor of confusion. More specifically, there is evidence in the literature that WTP is not very sensitive to the size of risk reductions. The question about the monetary value of the QALY in the EoL case was asked under certainty, and this may have produced a value of the QALY for EoL that was lower than risk reductions.

This paper offers new evidence on social perceptions about the value of EoL treatments. The objective was to determine if members of the general population want to give greater weight to QALYs achieved in the later stages of terminal diseases. The study attempts to respond to several questions: 1) Do treatments that extend life at the EoL have more social value than treatments that improve quality of life for temporary health problems? 2) Do treatments that improve quality of life at the EoL have more social value than treatments that improve quality of life for temporary health problems? That is, do EoL treatments have more social value than treatments for temporary problems even if they do not extend life? 3) Do people discriminate between different ways of obtaining QALYs (gains in life expectancy vs. gains in quality of life) within EoL treatments? This question deals with the value of palliative vs. life extending care. In fact, in their Supplementary Advice to the 2008 "Guide to the Methods of Technology Appraisal" NICE (2009a) contemplated the possibility of giving weights to QALYs only for EoL life extending treatments but not for EoL palliative care. This is understandable, as life extension is the (potential) main benefits of new cancer drugs. However, if it is shown that it is the presence of death per se (and not life extension at the end of life) that gives EoL QALYs an extra weight then EoL palliative care would also be a candidate for this extra weight.

This paper gives the results of three surveys that attempted to respond to each of the above questions. It was found that members of the general population clearly want to give an extra weight to EoL treatments. It was also found that subjects seem to attach a greater value to palliative care than to life extension at the end of life.

2. Methods

2.1. Surveys

Extensive piloting preceded the surveys. After some informal piloting with colleagues and university staff, a face-to-face pilot study (n=120) was conducted in order to test several issues such as the framing of questions, the bids used in payment cards and the visual aids. After this pilot survey three different surveys were conducted using a face-to-face computer-assisted personal interview (CAPI) method. The random route method was followed in order to choose the households to be interviewed. Interviewers selected subjects according to age and sex quotas representative of the adult Spanish population. The surveys took place in the South of Spain, over a period of six months (May to November, 2010).

Each survey was designed to test a different hypothesis. Survey 1 (n=240) compared the value of a health gain achieved through an improvement in quality of life of a temporary health state (T-QoL) with the value of a health gain achieved through an improvement in life expectancy at the end of life (EoL-LE). Survey 2 (n=232) compared the temporary health gain (T-QoL) used in Survey 1 with a health gain at the end of life that was achieved by palliative care (EoL-QoL). In Survey 2, both treatments only improved quality of life without affecting life expectancy. Survey 3 (n=348) included the two EoL scenarios used in Surveys 2 and 3 in order to directly compare the two EoL treatments (EoL-LE vs. EoL-QoL). It could be thought that these two treatments were already being compared, as in Surveys 1 and 2 they were both compared with T-QoL. However, there is psychological evidence (Loomes, 2010; Shafir, Osherson, & Smith, 1989) showing that comparing two objects A and B indirectly by means of a third object C is not always the same as comparing A and B directly.

2.2. Scenarios

Six different scenarios were used in the survey (Table 1). They corresponded to 3 types of health gains (Temporary, Life Extending, and Palliative) and two different durations of benefits (6 and 18 months). Different durations were used for two reasons. First, to test if the potential weight for end of life treatments changed with the size of the health gain. For example, some people may give a relatively higher value to extending life as the gain increases. The second reason was to test the consistency of the answers, e.g. whether WTP increases with the size of the health gain. Quality of life was defined as a “percentage”. Respondents were told that 100% was normal health for somebody of his/her age and 0% a very bad condition, as bad as death. This way of presenting quality of life has been used in the literature (Dolan & Tsuchiya, 2009) (Baker et al., 2010).

In all the scenarios, individuals were asked to imagine that they had been diagnosed with a certain illness and that a treatment increased the chances of improving their condition by 10%. A low probability of improvement was used for two reasons. One was to reduce the anxiety that a 100% chance of recovery could generate especially in the case of EoL-LE treatments. In piloting, interviewers suggested that some subjects had problems with this question as it sounded like “pay or die”. The second reason was that a 100% chance of success would imply very large health gains (e.g. almost 1 QALY in the case of the 18 months duration) and it was desired to avoid hitting budget restrictions in WTP questions. Budget restrictions produce insensitivities in WTP responses. It can be easily checked that in Scenarios 1, 3 and 5 the health gain was equivalent to 0.025 QALYs while in Scenarios 2, 4 and 6 it was 0.075 QALYs. All the scenarios were illustrated using visual aids such as those in Figures 1 and 2.

[FIGURE 1]

[FIGURE 2]

2.3. Elicitation procedures

Two sorts of questions were asked in order to elicit preferences: Willingness to Pay (WTP) and Person Trade-Off (PTO). These two methods were used in order to obtain preferences from an individual and a social point of view. WTP is a well grounded methodology in Welfare Economics and it is used in several parts of the public sector in order to take regulatory or investment decisions. However, WTP has been criticized (Culyer et al., 2007) as a measure of preference in a public health system. It is argued that since a public health system has a closed budget, the relevant issue is not how much subjects are WTP but the opportunity costs of funding new treatments. Moreover social decisions incorporate an equity element that is absent in WTP. For this reason, PTO has been proposed as an alternative method to elicit preferences in the health care sector (Nord, et al 1999).

In order to elicit WTP, a payment ladder containing a very wide range of sums of money was used (5, 10, 20, 50, 100, 200, 300, 500, 700, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 7000, 10000, 25,000, and more than 25,000 euros). Subjects saw these amounts of money on the computer screen. They had to state if they would pay each of these amounts or not. This produced an interval where WTP was located. The subject was then asked to state their maximum WTP within the interval.

In PTO questions people were asked to choose between two patients (A or B) according to the type of health gain. If the subject chose patient A, the next question was whether they preferred two patients like B or one patient like A. If they still preferred 1A to 2B, the number of B's was increased (3, 4, 5, 10, 15, 20, 30, 50, 100, 5000, 1000, and more than 1000) until the subject changed their choice. Again, an interval was produced where indifference was located. The subject was asked to state the number of patients of one type that was equivalent to 1 patient of their more preferred type.

2.4. Structure of the questionnaires

The three surveys had a similar structure (see Table 2). They started with a short description of the motives behind the study. The following paragraphs tried to explain the concept of quality of life as a percentage. After this introduction, preference questions were structured in three blocks. In each of the three surveys, the first block included two WTP questions, the second block contained the PTO questions, and finally two more WTP questions were asked. At the end of the questionnaire, subjects responded to a series of socio-demographic questions. A final question was asked about the degree of difficulty of the survey (from 1 to 5).

[TABLE 2]

2.5. Analysis of the data

There is not a clear way of aggregating PTO responses. The two methods applied in Baker et al., (2010) were used which had been shown to have desirable theoretical properties. They are dubbed the “ratio of means” and the “median of ratios”:

- Ratio of means

This procedure assigns a value of 1 to the most preferred type of patient and a value of $1/N_i$ to the other patient, where N_i is the number of patients of the least preferred type that are equivalent to 1 patient of the most preferred type. For example, if 1A=10B then $1/N_i=0.1$. The average of these values is estimated for each type of health gain and the ratio of these averages is the aggregate measure of preference. That is, let the relative value of each type of patient be V(.). Assume the sample is split between those (say type i subjects) that prefer 1A to 1B and those (say type j subjects) that have the opposite preferences. For type i subjects $V_i(A)=1$ and $V_i(B)=1/N_B$. For type j subjects $V_j(B)=1$ and $V_j(A)=1/N_A$. Next, the means V(A) and V(B) are calculated [$V(.)\leq 1$]. The measure of central tendency would be:

$$\text{Ratio of means} = \frac{\text{mean V }(A)}{\text{mean V }(B)} \text{ or } \frac{\text{mean V}(B)}{\text{mean V}(A)} \qquad (1)$$

- Median of ratios

In this procedure, firstly one option (A or B) is chosen as the base (the denominator) and the ratio for both options is estimated. That is, assume that

A is the base option. Also assume a subject for whom $1A=N_B$. Then $V_i(A)=N_B$. Assume a subject with opposite preferences. Then $V_j(A)=1/N_A$. The measure of central tendency is the median of V(A). As in the ratio of means procedure, this measure is symmetrical, that is, the median of V(A) is the inverse of the median of V(B).

In the case of WTP the mean is the theoretically correct value. However, it was thought it could be useful to analyse WTP data using “ratio of means” and “median of ratios” methods. In this way WTP and PTO results can be compared more directly. For WTP data these methods were applied as follows:

- Ratio of means

If WTP(A)>WTP(B) then V(A)=1 and $V(B)=\frac{WTP(B)}{WTP(A)}$. If WTP(B)>WTP(A) then V(B)=1 and $V(A)=\frac{WTP(A)}{WTP(B)}$. If WTP(A)=WTP(B) then V(A)=V(B)=1. Once V(A) and V(B) are calculated for each subject, the ratio of means is estimated as with PTO.

- Median of ratios:

Firstly, A or B is chosen as the base. Assume it is A. Then $V(A)=\frac{WTP(B)}{WTP(A)}$. Then the V(A) is estimated for each subject and the median is calculated.

3. Results

3.1. Characteristics of the samples

[TABLE 3]

Socio-demographic characteristics of the samples can be seen in Table 3. Seven subjects were excluded as they were not willing to pay anything in any of the four WTP questions. The reason given was "*the government should provide all these treatments free of charge*". This was interpreted as a protest response. The final number of observations was then 813. Samples were representative of the Spanish adult general population in terms of age and sex. Differences in socio-demographic characteristics between samples are not very relevant in this case since all the hypotheses are tested within each sample.

3.2. WTP results

Results of WTP questions can be seen in Table 4. In all six cases WTP for the 18 months scenario is higher than for the 6 months scenario (differences significant at the 1% level with t-test and Wilcoxon). The main results are as follows:

[TABLE 4]

1. WTP for EoL-LE treatments is higher than for T-QoL treatments.

   However, the percentage of people that were not willing to pay anything was higher for EoL-LE than for T-QoL treatments. The basic reason they gave was that the health gain was insufficient. The fact that average WTP is higher for EoL-LE treatments than for T-QoL treatments, in spite of the existence of more people with zero WTP for EoL-LE, implies that society is split into two very different groups. One group think that a 6 month (even 18 month) increase in LE is not long enough in order to spend any money on it. However, those who give a positive value to life extending treatments are willing to pay substantially more for EoL-LE treatments than for T-QoL treatments.

2. WTP for EoL-P treatments is higher than for T-QoL treatments.

Improving quality of life at the end of life (Palliative care) was considered more important than improving quality of life for Temporary health problems. The issue of zero WTP was much less relevant in Palliative care with respect to Life Extending treatments. If the results of Surveys 1 and 2 are combined, it looks as if EoL-P receives an even higher weight than EoL-LE treatments since differences between each of these options and T-QoL are bigger for EoL-P than for EoL-LE.

3. WTP for EoL-P treatments is higher than for T-QoL treatments.

   The third survey confirmed that EoL-P treatments receive a higher social value than EoL-LE. This is considered to be a highly relevant result as it suggests that NICE End of Life criteria should also be applied to palliative treatments and not only to Life Extending treatments as currently occurs.

3.3. PTO results

The picture that emerges from the analysis of PTO data is very similar to that from WTP, at least at the ordinal level. The summary of these results can be viewed in Table 5.

[TABLE 5]

1. It is confirmed that EoL-LE treatments receive an extra weight in relation to Temporary health problems. More people chose the EoL-LE over the Temporary patient, as was the case with WTP. If WTP and PTO are compared using the same methodology ("ratio of means" and "median of ratios") it can be observed (Tables 4 and 5) that the picture

given by both methods is not very different. For PTO the ratios move between 1.14-1.5 and for WTP between 1.3-2.0. In both methods ratios increase with duration, implying that EoL-LE becomes more attractive *in relation* to Temporary health gains when life expectancy increases.

2. It is confirmed that EoL-P treatments receive extra weight in relation to Temporary health problems. However, the ratios are more extreme with PTO than with WTP. The ratio of means is around 3.30 in PTO and around 1.8 in WTP. The median of ratios with PTO (18.0 and 15.0 for the six months and 18 months durations, respectively) is clearly way out of any ratio produced by WTP. These higher ratios between EoL-P and T-QoL than between EoL-P and T-QoL would suggest (as in WTP) that EoL-P treatments receive a higher social value than EoL-LE treatments.

3. It is confirmed that EoL-P receives higher weight than EoL-LE when both contexts are compared directly. However, the ratios obtained with the direct comparisons are much smaller than what could be expected from the ratios obtained in Surveys 1 and 2. This shows the relevance of comparing options directly.

3.4. Inconsistencies

The picture appears to be (and *is considered to be*) quite consistent. However, this does not imply that the subjects did not face problems in responding to the questions. The analysis of the responses to the last question of the survey revealed that two thirds of the respondents found the questions “difficult” or “very difficult”. This may explain some of the inconsistencies observed at the individual level. For example, in WTP questions many people gave the same answer irrespective of the duration of health gains (6 vs. 18 months). This happened in 22.7% of the valuations in the Temporary health problem scenario, in 41.8% of the subjects valuing the Life Extending treatment and in 34.3% of the cases in the Palliative care context.

[TABLE 6]

Another type of inconsistency emerges when WTP and PTO responses are compared at an individual level. As can be seen in Table 6, there were people who favoured one option in PTO but they were willing to pay less for this option. It could be argued that this does not represent an inconsistency since the perspectives were different (individual vs. social). It is accepted that social preferences can incorporate considerations that are not included in individual WTP questions, and that this can lead to a change in preferences between WTP and PTO. However, it is suggested that another (and complementary) explanation of these “inconsistencies” is imprecision and error. It is not strange to assume that responses to such questions can be subject to some degree of error. This is perfectly possible given that these kinds of questions are complicated and preferences can be imprecise. One of the reasons to think in error as another explanation of the reversals between WTP and PTO in the results is that they were not random. Those subjects who were in the “minority” in PTO or in WTP (Option A in both cases) produced more inconsistencies. For example, in Surveys 1 and 2 the less preferred option is the Temporary health problem. In Survey 3, the less preferred option is EoL-LE. Concentrating on the choices when the duration is 6 months (for 18 months the picture is basically the same), if those who chose Option A in Survey 1 (n=107), in Survey 2 (n=40) and in Survey 3 (n=124) are added, this gives a total of 271 subjects choosing option A (the *minority* option). Of those subjects, a total of 108 (39.9%) reversed their preferences in WTP, that is, they were willing to pay more for Option B. However, among those who chose Option B (the *majority* option) in PTO (n=542) only 14.8% reversed their preferences in WTP. It is the same with WTP. A total of 175 were willing to pay more for Option A and 459 were willing to pay more for Option B. Among those who were prepared to pay more for Option A (the *minority* option), 80 (45.7%) reversed their preferences in PTO. This only happened in 108 (23.5%)

of those subjects who were willing to pay more for Option B (the *majority* option). It is thought that these reversals would have been similar in both groups (majority and minority), if the explanation of the discrepancy between WTP and PTO were some kind of discrepancy between individual and social values. There is no reason to think that individual and social preferences are more different in one group than in the other. However, asymmetry is justified if it is assumed that preferences are stochastic, that is, subject to error/noise, given the difficulty of the task. An illustration of this effect is presented in the Appendix. If the interpretation is correct, the ratios of the relative benefit of one option against the other are biased downwards in all cases.

4. Discussion

The evidence that this paper presents suggests that: a) QALYs gained at the end of life receive a greater weight than QALYs gained from alleviating Temporary health problems, and b) Palliative care receives even greater weight than (short) life extensions at the end of life. It thus seems that the main reason for the extra weight of End of Life treatments is not only that QALYs obtained from quality of life improvements are different from QALYs obtained from increasing life expectancy, as has been suggested in the literature (Mason, Jones-Lee, & Donaldson, 2009). The reason that justifies this extra weight seems to be related to the presence of death itself. Apparently, the message of this paper is that people give an extra value to what can be done (life extending or not) for patients that are in the last stages of their lives.

What are the consequences for health policy? First, the fact that NICE has approved some EoL medicines (NICE, 2009b) with an extra weight of 1.6 for EoL QALYs does not seem unreasonable according to our results. Sunitinib was approved at a cost per QALY of about £50,000 (NICE 2009b) representing an extra weight of 1.6 in relation to the £30,000 per QALY threshold. However, what our data cannot say is whether the threshold itself is reasonable or not. What our data suggest is that the extra weight is reasonable but it does not follow from this that the threshold used (£50,000 per QALY) is acceptable. The reason is that if treatments that give less QALYs per monetary unit are approved, treatments that are at the margin will not receive funding and a lower cost per QALY will be needed (within a fixed budget). This will reduce the threshold and the weight of EoL treatments will have to be estimated in relation to this new threshold.

Another consequence is that NICE may have to change some of its methods in order to evaluate drugs that are used in EoL situations. NICE supplementary advice on EoL states that it is "technically more accurate….to include only the QALYs gained through extension of life and not the QALYs gained through improved quality of life during any extended 'progression free' period". Our study suggests that this is not correct. If anything, QALYs gained through improvements in Quality of Life should receive a higher weight. Cancer drugs could receive a premium if they reduce toxicity even if they do not extend life. If this is the case, quality of life measurement is of paramount importance in the evaluation of cancer drugs. This requires a more stringent set of rules in order to incorporate quality of life into economic evaluations. While NICE has been very demanding in the evidence that relates to life extension, we think that it has accepted weaker evidence in relation to quality of life. In some cases (2010a) the trial did not include quality of life data and the economic evaluation adapts utilities used in patients with a different type of cancer. In other cases (2010b) utilities come from nurses and clinicians. Indirect methods such as "mapping" were also used (2010c). It is also surprising that utilities are allocated to very vague health states such as "progress-free survival", "stable disease" or "progressive disease". In summary, whilst NICE has rejected several drugs based on lack of evidence about the effect of the drug on life expectancy, it seems that quality of life has played no relevant role in these decisions. Apparently this is not exclusive to evaluations conducted by NICE. Tengs (2004) reviewed 110 evaluations of cancer treatments and found that quality of life did not make much difference in resource allocation decisions. This is really surprising since we are dealing with treatments with serious side effects. Garau et al. (2011) have suggested that this result can be explained by limitations in the QALY model (e.g. failure of the constant proportional trade-off assumption) or, at least, in the way that quality of life is measured (e.g. using an instrument, such as the Euroqol, that may not be sensitive enough). Our results suggest that stronger evidence is needed on quality of life when evaluating cancer drugs.

One issue that can be raised is to what extent a survey like this one can produce results that are normative from a social policy point of view. We think that surveys in general, and this one in particular, have limitations from a normative point of view. Some of them are discussed below.

This survey asked questions that are very unfamiliar and emotional. If we put these two things together (emotions and unfamiliarity) responses can be driven by factors (biases) that are not normatively desirable, as the literature on affect has shown (Slovic, 2007) (Finucane, Alhakami, Slovic, & Johnson, 2000). For this reason, it is important to be able to justify empirical results with a normative theory. Cookson (2011) suggests that concern for the worst-off (severity) can provide a normative justification for giving extra-weight to QALYs gained at the end of life. However, this does not mean that subjects' responses are only driven by severity concerns and have normative validity. There may be an emotional component behind the weight for EoL treatments driven by the presence of near death. We are not sure to what extent talking about near death issue raises emotions that may not be normatively desirable to guide social policy. Take, for example, the result that Palliative Care is strongly favoured by the population. Does this imply that alleviating pain at the End of Life is more important than alleviating the same degree of pain after an operation for a temporary health problem? We are not sure to what extent our results can be interpreted in this way.

We also have concerns about the influence of framing effects on our results. The literature has shown that framing effects can dramatically change preferences (Ubel, Baron, & Asch, 2001). For example, we are not sure to what extent using percentages as a way of illustrating quality of life may have affected our results. This method has been used in the literature (Dolan & Tsuchiya, 2009) (Baker et al., 2010) and it has been shown to produce consistent results. However, it would be interesting in future research to see what would happen if quality of life were described in a less abstract manner (e.g. symptoms). Our results should be confirmed by other research using different methods and different framing. However, the fact that we have similar results with two different methods and in three different and independent samples suggests that there is something special about health gains at the end of life that is not captured by the assumption that all QALYs are created equal.

A final issue related to our results is opportunity costs. The paper suggests that people may want to give extra-weight to treatments that benefit patients with a terminal illness. One criticism could be that subjects are not aware of the opportunity costs that this policy involves. More resources for end of life problems mean fewer resources for other health problems. To what extent have the methods used in this survey allowed people to think about this problem? Our survey used two methods that required subjects to take into account opportunity costs in one way or another. In WTP questions people have shown that they prefer to spend their own money (and this has clear opportunity costs) on end of life treatments rather than on temporary health problems. If we go to PTO questions, opportunity costs are very clear. Of course, there could be better ways of showing subjects the consequences of giving extra weight to EoL health gains. For example, people could be asked to choose where to disinvest if more weight is given to EoL health gains. Maybe this framing can help people to perceive opportunity costs more clearly. The role of opportunity costs should be the object of future research.

In summary, we have shown that the decision of giving extra weight to QALYs gained at the end of life may be supported by the general population. Our study also shows that the main reason for this weight is not that QALYs obtained by increasing life expectancy are more valued than those obtained by improving quality of life. We show that quality of life at the end of life can be even more important than life extension. This has consequences for the way that public agencies evaluate medicines.

BIBLIOGRAPHY

Baker, R., Bateman, I., Donaldson, C., Jones-Lee, M., Lancsar, E., Loomes, G., SVQ Research Team. (2010). Weighting and valuing quality-adjusted life-years using stated preference methods: Preliminary results from the social value of a QALY project. *Health Technol Assess*, *14*(27), 1-162.

Brazier, J. (2007). Searching for a threshold, not setting one: The role of the national institute for health and clinical excellence. *Journal of Health Services Research & Policy*, *12*(1), 56.

Cookson R. (2011) "Can the NICE 'end of life' premium be given an ethical as opposed to a political justification?" Paper presented at the Conference Issues on Rationing in Health Care September 19th-20th 2011 London School of Economics.

Culyer A, McCabe C, Briggs A, et al. Searching for a threshold, not setting one: The role of the national institute for health and clinical excellence. *J Health Serv Res Pol* 2007; 12(1): 56-8.

Culyer, A., McCabe, C., Briggs, A., Claxton, K., Buxton, M., Akehurst, R., Department of Health, Medicines, Pharmacy & Industry Group. (2010). *A new value-based approach to the pricing of branded medicines*. London.

Department of Health, Medicines, Pharmacy & Industry Group. (2011). *A new value-based approach to the pricing of branded medicines. Government response to consultation*.

Dolan, P., & Tsuchiya, A. (2009). The social welfare function and individual responsibility: Some theoretical issues and empirical evidence. *Journal of*

Health Economics, *28*(1), 210-20. doi:10.1016/j.jhealeco.2008.10.003

Donaldson, C. (Coordinator). (2010). *European value of a quality adjusted life year*. European Commission. SP5A-CT-2007-044172. Sixth Framework Programme.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*(1), 1-17.

Garau, M., Shah, K. K., Mason, A. R., Wang, Q., Towse, A., & Drummond, M. F. (2011). Using qalys in cancer: A review of the methodological limitations. *Pharmacoeconomics*, *29*(8), 673-685.

IOM (Institute of Medicine). (2009). *Assessing and improving value in cancer care: Workshop summary.* Washington, DC: The National Academies Press.

Loomes, G. (2010). Modeling choice and valuation in decision experiments. *Psychol Rev*, *117*(3), 902-24. doi:10.1037/a0019807

Mason, H., Jones-Lee, M., & Donaldson, C. (2009). Modelling the monetary value of a QALY: A new approach based on UK data. *Health Econ*, *18*(8), 933-50. doi:10.1002/hec.1416

Maynard, A., & Bloor, K. (2009). NICE wobbles. *Journal of the Royal Society of Medicine*, *102*(6), 212. doi:10.1258/jrsm.2009.09k031

Morris, J., & Perez, D. (2000). Willingness to pay for new chemotherapy for advanced ovarian cancer. *N Z Med J*, *113*(1108), 143-6.

Nadler, E., Broderick, W. C., Zarotsky, V., & Kim, J. (2009). How do medical and pharmacy directors perceive the value of new cancer drugs? *Drug Benefit Trends*, *21*(4), 120-130.

Nadler, E., Eckert, B., & Neumann, P. J. (2006). Do oncologists believe new cancer drugs offer good value? *Oncologist*, *11*(2), 90-5. doi:10.1634/theoncologist.11-2-90.

NICE. (2008). Appraising end of life medicines: Consultation document. Retrieved from http://www.nice.org.uk/media/26E/43/Endoflifemedicines.pdf

National Institute for Clinical Excellence (2009a). Update report on the application of the 'End-of- Life' Supplementary Advice in health technology appraisals.

NICE. (2009b). *Sunitinib for the first-line treatment of advanced and/or metastatic renal cell carcinoma*. NICE Technology appraisal guide 169.

National Institute for Clinical Excellence (2010a). Trabectedin for the treatment of advanced soft tissue sarcoma NICE technology appraisal guidance 185.

National Institute for Clinical Excellence (2010b). Pemetrexed for the maintenance treatment of non-small-cell lung cancer. NICE technology appraisal guidance 190.

National Institute for Clinical Excellence (2010c). Sorafenib for the treatment of advanced hepatocellular carcinoma. NICE technology appraisal guidance 189.

Nord, E., Pinto, J. L., Richardson, J., Menzel, P., & Ubel, P. (1999). Incorporating societal concerns for fairness in numerical valuations of health programmes. *Health Economics*, *8*(1).

Raftery, J. (2009). Should NICE's threshold range for cost per QALY be raised? No. *BMJ*, *338*, b185.

Shafir, E. B., Osherson, D. N., & Smith, E. E. (1989). An advantage model of choice. *Journal of Behavioral Decision Making*, *2*(1), 1-23. doi:10.1002/bdm.3960020102

Slovic, P. (2007). “If I look at the mass I will never act” : Psychic numbing and genocide. *Judgment and Decision Making*, *2*(2), 79-95.

Tengs, T. O. (2004). Cost-Effectiveness versus cost--utility analysis of interventions for cancer: Does adjusting for health-related quality of life really matter? *Value in Health*, *7*(1), 70-78.

Towse, A. (2009). Should NICE's threshold range for cost per QALY be raised? Yes. *BMJ*, *338*, b181.

Ubel, P. A., Baron, J., & Asch, D. A. (2001). Preference for equity as a framing effect. *Medical Decision Making*, *21*(3), 180.

Figure 1. Example of a visual aid for WTP questions: Scenario 5 (T-QoL, 6 months).

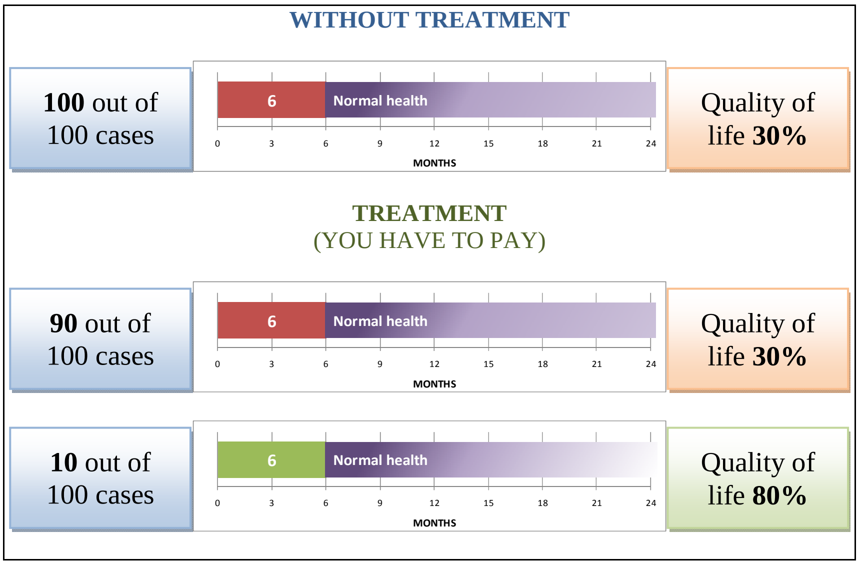
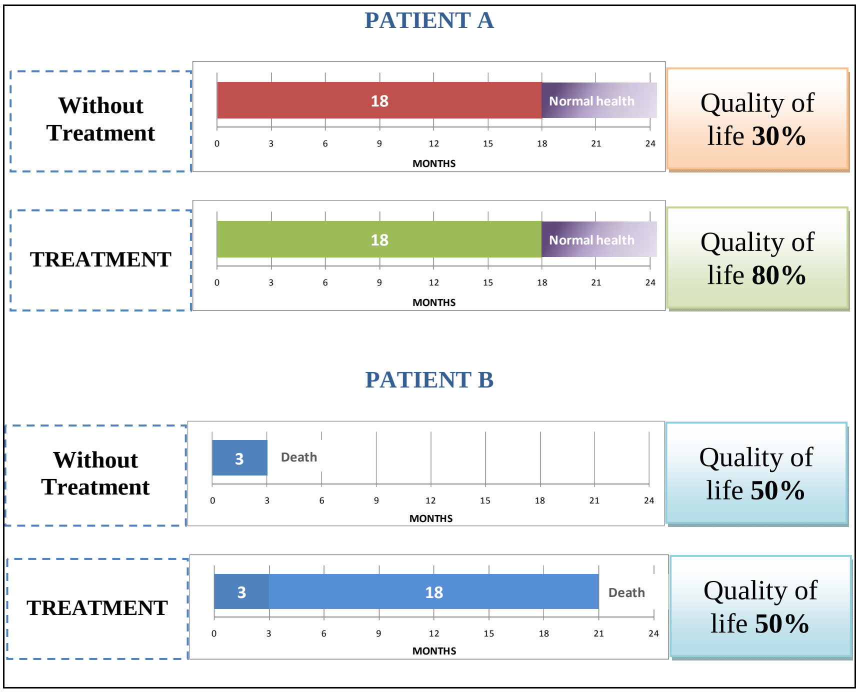


Figure 2. Example of a visual aid for PTO questions:
Scenario 6 (T-QoL, 18 months) vs. Scenario 2 (EoL/LE, 18 months).



TABLES

Table 1. Scenarios used in the surveys.

	Without treatment	With treatment if successful (10% chance of success)	QALY Health gain
Scenario 1. End of life, life extending treatment (**EoL-LE**), 6 months.	Life expectancy: 3 months Quality of life: 50% of normal health	Life expectancy: 9 months Quality of life: 50% of normal health	10% chance of 6 months life extension at 50% (0.025 QALY)
Scenario 2. End of life, life extending treatment (**EoL-LE**), 18 months.	Life expectancy: 3 months Quality of life: 50% of normal health	Life expectancy: 21 months Quality of life: 50% of normal health	10% chance of 6 months life extension at 50% (0.075 QALY)
Scenario 3. End of life, palliative care (**EoL-QoL**), 6 months	Life expectancy: 6 months Quality of life of 30% of normal health	Life expectancy: 6 months Quality of life: 80% of normal health	10% chance of 50% QoL improvement for 6 months (0.025 QALY)
Scenario 4. End of life, palliative care (**EoL-QoL**), 18 months	Life expectancy: 18 months Quality of life: 30% of normal health	Life expectancy: 18 months Quality of life: 80% of normal health	10% chance of 50% QoL improvement for 18 months (0.075 QALY)
Scenario 5. Temporary health benefit (**T-QoL**), 6 months	Life expectancy: not affected Quality of life: 30% of normal health during 6 months; then normal health.	Life expectancy: not affected Quality of life: 80% of normal health during 6 months; then normal health.	10% chance of 50% QoL improvement for 6 months (0.025 QALY)
Scenario 5. Temporary health benefit (**T-QoL**), 18 months	Life expectancy: not affected Quality of life: 30% of normal health during 18 months; then normal health.	Life expectancy: not affected Quality of life: 80% of normal health during 18 months; then normal health.	10% chance of 50% QoL improvement for 6 months (0.075 QALY)

Table 2. Structure of the surveys

Introduction				
		Survey 1	*Survey 2*	*Survey 3*
Part 1 WTP(A)	Question 1	**T-QoL**, 6 months (Scenario 5)	**T-QoL**, 6 months (Scenario 5)	**EoL-P**, 6 months (Scenario 3)
	Question 2	**T-QoL**, 18 months (Scenario 6)	**T-QoL**, 18 months (Scenario 6)	**EoL-P**, 18 months (Scenario 4)
Part 2 (PTO)	Question 3	T-QoL vs. EoL/LE, 6 months (Scenarios 5 vs. 1)	T-QoL vs. EoL-P, 6 months (Scenarios 5 vs. 3)	EoL-P vs. EoL/LE, 6 months (Scenarios 3 vs. 1)
	Question 4	T-QoL vs. EoL/LE, 18 months (Scenarios 6 vs. 2)	T-QoL vs. EoL-P, 18 months (Scenarios 6 vs. 4)	EoL-P vs. EoL/LE, 18 months (Scenarios 4 vs. 2)
Part 3 WTP(B)	Question 5	**EoL/LE**, 6 months (Scenario 1)	**EoL-P**, 6 months (Scenario 3)	**EoL/LE**, 6 months (Scenario 1)
	Question 6	**EoL/LE**, 18 months (Scenario 2)	**EoL-P**, 18 months (Scenario 4)	EoL/LE, 18 months (Scenario 2)
Socio-demographic questions				

Table 3. Socio-demographic characteristics of subjects

	Survey 1 (n=239)	Survey 2 (n=232)	Survey 3 (n=342)	Total (N=813)	Spanish population
Male/Female (%)	51/49	51/49	52/48	51/49	51/49
Mean (SD) age in years	42.6 (15.5)	43.8 (15.2)	44.0 (14.9)	43.5 (15.1)	42.7 (16.9)
Marital status (%)					
Married/Cohabiting	44.8	59,5	59.1	55.0	63.1
Single/Divorced/Widow	55.2	40,5	40.9	45.0	36.9
Education level (%)					
Illiterate/Primary studies	37.7	31.0	49.1	40.6	30.1
Secondary studies	41.4	43.5	28.1	36.4	45.1
University studies	20.9	25.4	22.8	23.0	24.7
Employment status (%)					
Employed	51.0	59.1	56.7	55.7	48.0
Unemployed	17.2	11.2	15.8	14.9	12.0
Inactive	31.8	29.7	27.5	29.4	40.0
Income level (%)					
Up to €1,500	82.8	55.6	63.7	67.0	52.3
€1,501-2,000	10.9	17.7	20.5	16.9	17.2
€2.001-3,000	5.0	14.7	14.0	11.6	19.5
More than €3,000	1.3	12.1	1.8	4.6	11.0

Table 4. Willingness to Pay (WTP) data

	6 months			18 months		
A	**T-QoL**	**T-QoL**	**EoL/LE**	**T-QoL**	**T-QoL**	**EoL/LE**
B	**EoL/LE**	**EoL-P**	**EoL-P**	**EoL/LE**	**EoL-P**	**EoL-P**
Mean WTP(A) (€)	371.1	499.8	976.9	556.6	858.0	1480.0
Mean WTP(B) (€)	647.2	1247.3	1227.9	1239.1	2083.1	1786.9
% WTP(A) > WTP(B)	23.4	11.2	27.2	19.2	12.1	26.9
% WTP(A) = WTP(B)	19.7	14.7	28.7	14.6	9.1	33.3
% WTP(A) < WTP(B)	56.9	74.1	44.2	66.2	78.8	39.8
% WTP(A) = 0	11.3	6.5	17.3	6.7	1.3	15.5
% WTP(B) = 0	23.0	12.5	6.4	12.6	6.5	9.1
Mean ratio A>B	0.62	0.51	0.71	0.54	0.49	0.76
Mean ratio B>A	0.80	0.91	0.86	0.85	0.92	0.89
Ratio of means	1.29	1.78	1.22	1.57	1.86	1.17
Median of ratios (B=1)	1.3	2.0	1.0	2.0	2.1	1.0

Note: differences between mean WTP were always statistically significant at $p<0.01$ with t-test and Wilcoxon text.

Table 5. Relative values from PTO

	6 months			18 months		
A **B**	**T-QoL** **EoL/LE**	**T-QoL** **EoL-P**	**EoL/LE** **EoL-P**	**T-QoL** **EoL/LE**	**T-QoL** **EoL-P**	**EoL/LE** **EoL-P**
%A>B	44.8	17.2	36.3	40.6	17.2	36.5
%A<B	55.2	82.8	63.7	59.4	82.8	63.5
Mean ratio A>B	0.57	0.26	0.65	0.52	0.26	0.66
Mean ratio B>A	0.65	0.87	0.82	0.71	0.87	0.82
Ratio of means	**1.14**	**3.35**	**1.26**	**1.36**	**3.30**	**1.25**
Median of ratios (B=1)	1.5	18.0	1.5	1.5	15.0	1.5

Table 6. Comparison between WTP and PTO

	6 months				18 months			
Chosen option	WTP(A) vs. WTP(B)				WTP(A) vs. WTP(B)			
	A<B	A=B	A>B	Total	A<B	A=B	A>B	Total
A (T-Qo)	**48**	23	36	107	**57**	11	29	97
B (EoL/LE)	88	24	**20**	132	101	24	**17**	142
Total	136	47	56	239	158	35	46	239
A (T-Qo)	**15**	12	13	40	**20**	8	12	40
B (EoL-P)	157	22	**13**	192	163	13	**16**	192
Total	172	34	26	232	183	21	28	232
A (EoL/LE)	**45**	33	46	124	**40**	39	46	125
B (EoL-P)	106	65	**47**	218	96	75	**46**	217
Total	151	98	93	342	136	114	92	342

Strong Preference Reversals in **bold**.

APPENDIX

Let us assume that most subjects (say 80%) think that condition A is worse than condition B. Let us also assume that subjects apply the same preferences in WTP and PTO. That is, those who think that A is worse than B are willing to pay more to avoid A than to avoid B and choose patient A in PTO. However, let us also assume that preferences are noisy (stochastic). This can be justified because preferences are imprecise and the task is complicated. This implies that there is an element of error in responses. Finally, we will also assume that error is bigger in WTP than in PTO since people may find it more difficult to discriminate between options when they are evaluated separately rather than jointly. In this example we will assume that the error rate in WTP is 25% and in PTO is 10%.

We start by supposing that we have a group of 1000 subjects with the above (stochastic) preferences. As 800 think that condition A is worse than condition B, there is a majority of subjects who should pay more for A than for B and who should choose A in PTO. However, out of this 800 there are 200 (25%) who state that WTP(A)<WTP(B) and 600 rightly state that WTP(A)>WTP(B). Of those 600 there are 60 (10%) that make a mistake in PTO and then choose B. If we apply the same reasoning in all cases, we have the following groups:

1000 subjects							
800 think that A is worse than B				**200** think that B is worse than A			
600 say that WTP(A)>WTP(B)		**200** say that WTP(A)<WTP(B)		**50** say that WTP(A)>WTP(B)		**150** say that WTP(A)<WTP(B)	
540 Choose A	**60** Choose B	**180** Choose A	**20** Choose B	**5** Choose A	**45** Choose B	**15** Choose A	**135** Choose B
No mistake	*PTO mistake*	*WTP mistake*	*WTP & PTO mistake*	*WTP & PTO mistake*	*WTP mistake*	*PTO mistake*	*No mistake*

These stochastic preferences would have produced the following table:

	WTP(A)>WTP(B)	WTP(A)<WTP(B)	**Total**	Error rate in PTO
Choose A in PTO	545	195	**740**	26%
Choose B in PTO	105	155	**260**	40%
Total	**650**	**350**		
Error rate in WTP	16%	56%		

So even if even if there is no discrepancy between individual and social values, the stochastic nature of preferences would produce a response pattern similar to that observed in our data. That is:

1. There is discrepancy between WTP and PTO.
2. Inconsistencies are asymmetric: they are bigger for the groups that are in the minority in both WTP and PTO.