WP ECON 08.09

Trying to Estimate a Monetary Value for the QALY

Jose Luis Pinto Prades (U. Pablo de Olavide)

Graham Loomes (University of East Anglia)

Raul Brey (U. Pablo de Olavide)

Department of Economics

TRYING TO ESTIMATE A MONETARY VALUE FOR THE QALY

April 2008

Jose Luis Pinto Prades*
University Pablo de Olavide,
Carretera de Utrera km 1, 41013 Sevilla

Graham Loomes
University of East Anglia, NR4 7TJ, Norwich, UK

Raul Brey
University Pablo de Olavide
Carretera de Utrera s/n 41013 Sevilla

Abstract

In this paper we study the possibility of estimating a monetary value for the QALY. Using two different surveys of the Spanish population (n=900), we try to establish whether willingness to pay (WTP) is (almost) proportional to the health gains measured in QALYs. We also explore whether subjects' responses are prone to any biases. We find that the monetary value of the QALY is higher the smaller the health gain, pointing to insensitivity in WTP. We also find two clear biases. One is the existence of sequencing effects. The other is the insensitivity of WTP to the duration of the period of payment. All these effects translate into a large variation in estimates of the monetary value of the QALY. We conclude that in order to be able to obtain consistent and stable estimates, we should try to understand better the causes of these problems with a view to developing ways of mitigating them.

*Corresponding author:
Postal address: Department of Economics, University Pablo de Olavide, Carretera de Utrera s/n 41013 Sevilla, Spain.
Telephone: +34 9549 77608
Fax: + 34 954 34 9339
Email: jlpinto@upo.es

1

1. Introduction

In recent years there has been a renewed interest in the issue of the monetary value of Quality Adjusted Life Years (MV-QALY). In fact, the National Institute for Health and Clinical Excellence (NICE) has commissioned one study in order to explore the possibility of estimating this figure. It is not surprising that NICE is interested in this research since, as Donaldson et al (2002) argue, cost-effectiveness analysis is inadequate in cases where a technology is both more costly and more effective. Unfortunately, as Neuman (2000) has shown, this is very often the case, since 80% of new medical technologies involve spending more money in order to achieve better health outcomes. Johannesson (1995) has argued that "in order to reach a decision based on cost-effectiveness analysis, information is then needed about the willingness to pay for QALYs gained" (p. 485) and Johanesson and Meltzer (1998) considered that obtaining this information "should be a research priority" (p. 4).

These authors identified two possible strategies for deriving the Willingness to Pay (WTP) per QALY gained. One is based on WTP for a marginal health change. The second strategy is to derive the WTP per QALY gained using estimates of the value of statistical life in the literature.

Using this latter approach, Hirth (2000) estimated a WTP per QALY ranging from $30,000 to $430,000. More recently, Byrne et al (2005) and King et al (2005) have estimated much lower values, ranging from $1,200 to $32,000. Gyrd-Hansen(2003) estimated the MV-QALY to be 88.000 DKK (about $15,000). These three studies are the only ones, as far as we are aware, that had the specific objective of estimating the monetary value of the QALY. There have been other studies that, although they did not have it as their main objective to estimate WTP per QALY, have produced such a figure. For example, Zehraeus (1998) divided the annual WTP for hormone replacement therapy by the estimated QALY gain to infer a WTP per QALY of 120,000SEK (about $16,000). There have been other studies (Bala et al., 1998; Johnson et al, 1997) that have analyzed the relation between WTP and QALYs but have not provided such a figure.

The study reported in this paper set out to explore the possibility of eliciting the MV-QALY using the first strategy proposed by Johannesson and Meltzer (1998), namely, estimating WTP for a marginal health change.

In that respect, it departs from Byrne et al and King et al who use health gains that cannot be considered marginal. Byrne et al focus on osteoarthritis. Although they suppose a relatively small drop in health status, the condition is chronic and the duration of the state means it cannot be considered to be a marginal health problem. Similarly, Gyrd-Hansen's study involved avoiding (or improving) a chronic health problem and WTP was elicited as a monthly payment for the rest of their lives. Since 6 of the 11 bids used entailed a monthly payment of more than 200€, this cannot be considered as a marginal payment. King et al revolved around two health conditions (cervical spondylotic myelopathy and cerebral aneurysms) that cannot be considered marginal – as can be seen from the fact that many patients said they would accept a 25% risk of death in order to avoid those

conditions. Also, their WTP was around $100,000, which can hardly be considered as marginal. So in this paper we want to explore the possibility of eliciting WTP for a QALY using marginal health gains.

We acknowledge the possibility that estimating a *unique* MV-QALY may not be feasible. Several papers(Bleichrodt and Quiggin, 1999; Dolan and Edlin, 2002) have shown that the conditions for a unique MV-QALY to exist are quite restrictive and are unlikely to hold. However, our objective is more limited. As suggested by Gyrd-Hansen (2005) "one may take a more pragmatic stance and advocate for a 'rule of thumb' in determining a threshold for CEAs. Hence, seeking to apply a unique WTP for a QALY should not be seen as defining the theoretical link between CEA and CBA, but rather as an aid to decision-makers" (p. 428). Estimating the willingness to pay for a QALY using marginal health gains could be useful for decision makers if it can provide some empirical support to a potential cost per QALY threshold for use in regulatory decisions. Our aim is to explore how robust such an estimate might be.

Estimating an MV-QALY using this approach does not prevent us from incorporating elements like a decreasing marginal valuation of health, as Johannesson and O'Connor (1997) suggest. For example, suppose an individual is willing to pay 2000€ for a quality of life gain from 0.9 to 1.0 with a duration of 1 year (0.1 QALY gain), implying an MV-QALY of 20,000€. This does not necessarily mean that if duration is 10 years the WTP should be 10 times more. Abellan-Perpiñan et al (2006) or Bleichrodt and Pinto (2005) have suggested that a non-linear QALY model may be better than the linear model. A non-linear QALY model with a power coefficient for life years of 0.65 (as suggested in Abellan-Perpiñan et al, 2006) would imply that U(0.1, 10years) would be around 0.45 QALYs so that WTP (setting any budget constraints aside) would be about 9.000€.

Also, if one believes, as Nord et al suggest (1999) that the value of health increases with severity, we could estimate the relative value of health gains according to this principle. So if WTP is 2000€ for a one-year health gain from 0.9 to 1.0 and if the health gain from 0.4 to 0.5 is valued twice as much, then WTP for this health gain would be 4000€, implying an MV-QALY of 40,000€. This approach is similar to the approach taken by Karnon et al (2005) in the sense that they use a "peg event for which a reliable valuation exists (road deaths)" in order to derive the value of life in other contexts. Our paper should be seen as an attempt to derive MV-QALY through a "peg event" and to explore if such a value is reliable or not.

2. Definitions

2.1 Marginal health gains

There are several ways of putting into practice the concept of marginal in relation to health gains. Basically, one can combine four factors:

1. Improvement in quality of life.
2. Duration of the quality of life improvement.

3

3. Gain in length of life.
4. Change in risk.

In this paper we will operationalize the concept of marginal health gains in two ways:

1. By considering health problems under certainty that, if untreated, are of short duration and involving a quality of life that is not too low. Of course, the concept of "short" and "low" is quite subjective. In general, the health gains that we will use will not imply more than 0.3 QALYs in the full life of a subject. In Gyrd-Hansen (2003) there are differences in health states that are even smaller but as they applied to chronic health problems, the overall health gain was much larger than ours.
2. By considering chronic health problems under risk: that is, by asking WTP questions about treatments that reduce the (small) risk of such a health problem.

2.2. Deriving Aggregate Monetary Values for a QALY

Given the two measures above – WTP for small health gains under certainty, and WTP to reduce the risk of a large health impact – how should one proceed to estimate the social monetary value of a QALY?

It might seem that one possible route is to estimate a value of a QALY for each individual respondent and then take the population average. Although we shall argue that this is NOT the best way to proceed, let us consider how this approach would be implemented.

First, define the MV-QALY of respondent "i" as follows:

$$\text{MV-QALY}_i = \frac{\text{WTP}_i(\Delta Q_i, t')}{[\Delta Q_i \text{ x } t'] \text{ x } p^*} \qquad [1]$$

Where:

ΔQ: improvement in quality of life due to the medical treatment
t': duration of the effect
p^*: probability of receiving the health improvement
$\text{WTP}_i(\Delta Q, t')$ = willingness to pay in order to achieve health gain for subject i.
MV-QALY_i = monetary value of the QALY for respondent i.

When considering WTP for small gains under certainty, p^* is set equal to 1. Thus for the health improvement to be marginal, we must have either ΔQ or t' or both (very) small. So, for example, if ΔQ x t' = 0.001 of a QALY and if individual i were willing to pay 30€ to achieve that health gain, the supposition would be that 1 QALY is worth 30,000€ to that individual. But this relies on one (or both) of two assumptions: either that WTP is strictly proportional to the duration of a given health state; or that WTP is strictly proportional to (1-h) where h is the health state index; or both. However, neither

4

assumption can be said to be obviously compelling. And indeed, as we shall see, neither assumption is supported by the data from the study reported below.

Alternatively, when considering WTP for a marginal reduction in the risk of a more substantial health effect, p* is small. To take the case where the health effect is death, research seeking to establish a *value of statistical life* (VSL) – sometimes also known as the *value of preventing a (statistical) fatality* (VPF) – has typically asked about WTP for risk reductions of the order of 10^{-5}. However, what is being measured here is the *marginal rate of substitution of wealth for risk* of death, and except under some very special assumptions (generally regarded as implausible) about the linearity of people's utility function, it is not possible to infer the individual's value of preventing his own certain death simply by multiplying that WTP response by 10^{5}, as Expression 1 might seem to imply.

Besides the lack of theoretical justification for computing each individual's MV-QALY according to Expression 1, there are also practical reasons for being wary of such a procedure. The very fact that the change being valued is small allows the possibility that responses may be susceptible to errors that may be modest in *absolute* terms but may be *relatively* large: for example, suppose that an individual's 'true' value of reducing the risk of losing 1 QALY by 0.001 is actually 30€, but that uncertainty about her preferences means that she might on some occasions give a response as low as 10€ while on other occasions her response might be 50€. Multiplying these responses by the inverse of the risk reduction – i.e. multiplying by 1,000 – greatly magnifies the noise in the original response, leading to a range between 10,000€ and 50,000€.

Given what we know about the difficulties people have in handling small probabilities, and given the greater scope for upward than downward biases (someone whose true value is 30€ is constrained by zero to never underestimate by more than 30€, but there is no such tight constraint on the potential for giving an overestimate), there is a very real danger that a small number of high WTP responses to the risk reduction question might translate into a long thin tail of outliers that may greatly inflate the mean MV-QALY. In the case of MV-QALYs, the danger is further aggravated by the possibility that the right-hand tail of WTP estimates might interact with underestimates of the size of QALY gain afforded by a particular treatment. And as we shall see, mean values of a QALY estimated in this way do indeed result in much higher and more volatile estimates than are generated by an alternative aggregation and estimation procedure.

That alternative procedure can be understood/motivated as follows. A population faces some chance of suffering condition X, but that chance can be reduced by treatment B. Providing some particular level of treatment B would, let us say, prevent/cure 10 cases of X. However, at the time of provision there is no way of knowing which 10 members of the population will benefit. All that we can work with is the overall *expected* size of this benefit in QALY terms – which we arrive at by eliciting each individual's estimate of the QALY loss involved in suffering a case of X (equivalently, the QALY gain entailed by preventing a case of X) and taking the population mean of these estimates. If, for example, a representative sample of the population were, on average, to consider that a

case of X constituted a loss of 2 QALYs, the total expected QALY benefit from providing treatment B to 10 randomly-selected members of the population would be 20 QALYs.

How much is society willing to pay for this expected QALY benefit? The answer is: the sum total of population members' WTP for this level of provision of treatment B. Thus if treatment B were provided at this level to a population of 10,000 people, each person's risk of suffering condition X would be reduced by 1 in 1000. If a representative sample expressed a mean WTP of 70€ for such a risk reduction, the implication is that the population would collectively be willing to invest 700,000€ in this programme with an expected benefit of 20 QALYs, giving an average figure of 35,000€ per QALY gained.

For the calculation based on WTP for small but certain health gains, an analogous procedure would apply: First, find the mean QALY value of the small health gain, based on a representative sample of the population. Let us suppose for the sake of example that this turns out to be 0.02 QALYs. To achieve a social benefit of 1 QALY would require a random cross-section of 50 members of the population to experience this health gain: that is, the total value of the gain would be 50 times the average that a representative sample of that population would be willing to pay for such a gain. If that average WTP were, for example, to be 400€, the social value of that 1 QALY gain would be 50 x 400€ = 20,000€

2.3. Hypothesis

In this paper we will test two different sets of hypotheses. One set are those hypotheses that are needed in order to obtain a unique MV-QALY for marginal health gains. They are:

1. MV-QALY has to be constant for any t'.
2. MV-QALY has to be constant for any ΔQ.
3. MV-QALY has to be constant for any p*.

So the above conditions just say that WTP has to be proportional to the size of the health gain for marginal health gains.

The second set of hypotheses test more basic rationality assumptions about WTP questions. They are next:

4. MV-QALY has to be constant irrespective of the order of questions (no order effects).
5. MV-QALY has to be constant irrespective of the length of the payment period.
That is, if WTP is asked in term of monthly outlays, total WTP should not change with the number of months that the respondent has to pay (no temporal embedding).

3. Methods

3.1. Subjects

560 members of the general population were interviewed. They were distributed in 7 subgroups according to a quota sample method (40% between 18 and 40, 30% between 41 and 60, 30% older than 60). They were contacted by telephone through random dialing and those who agreed to be interviewed were visited by an interviewer. Sociodemographic characteristics of the sample can be seen in table 1.

Table 1 Sociodemographic characteristics of the sample

		Group 1A	Group 1B	Group 1C	Group 2A	Group 2B	Group 3A	Group 3B
Number subjects		80	80	80	80	80	80	80
Gender	Women	51,3%	47,5%	50,0%	50,0%	51,3%	50,0%	51,3%
Educational Background	Primary or less	42,5%	52,5%	45,0%	38,8%	43,8%	40,0%	47,5%
	Secondary	35,0%	32,5%	31,3%	37,5%	37,5%	36,3%	30,0%
	University	22,5%	15,0%	23,8%	23,8%	18,8%	23,8%	22,5%
Working status	Active	68,8%	70,0%	65,0%	76,3%	68,8%	71,3%	75,0%
	Retired	17,5%	17,5%	20,0%	18,8%	20,0%	17,5%	18,8%
	Other	13,8%	12,5%	15,0%	5,0%	11,3%	11,3%	6,3%
Income	<600€/month	30,0%	30,4%	31,3%	23,8%	21,3%	26,3%	20,0%
	600€-1200€	41,3%	43,0%	36,3%	55,0%	54,7%	40,0%	40,0%
	1200€-1800€	21,3%	21,5%	27,5%	11,3%	20,0%	22,5%	30,0%
	>1800€	7,5%	5,1%	5,0%	10,0%	4,0%	11,3%	10,0%
Age	18 - 40	40,0%	41,3%	41,3%	42,5%	38,8%	42,5%	42,5%
	41 - 60	35,0%	35,0%	33,8%	32,5%	32,5%	33,8%	35,0%
	> 60	25,0%	23,8%	25,0%	25,0%	28,8%	23,8%	22,5%

3.2 Willingness to pay questions

3.2.1. Health gains

Each option was characterized by some combination of the following attributes:

1. Quality of life.
2. Duration with a certain quality of life.
3. Probability.
4. Cost.

Each option was described as a medical treatment where one treatment implied a lower health outcome than the other but also a lower cost. As the less effective medical treatment was presented as being free at the point of consumption, one can consider this option to be the *status quo*. The framing of the question was such that the option of "no treatment" was not realistic.

The structure of the questions asked in the survey can be summarized as follows:

Table 2. Structure of questions

	Treatment A	Treatment B
Quality of life	Q_A	Q_B
Duration	D_A	D_B
Probability	P_A	P_B
Cost	0	C_B

This structure allowed us to convey the idea of a marginal health gain. This was mainly achieved by using short durations or short probabilities of having the health problem.

We used three Euroqol (EQ-5D) health states in order to illustrate different quality of life situations. EQ-5D is a standardised instrument for use as a measure of health outcome. It has currently 5 dimensions and 3 levels. Level 1 represents having no problems, level 2 having moderate problems and level 3 having severe problems. One its advantages is that it allows to use health states where there are clear relations of dominance that can be used to study the consistency in people's responses. We used health states 21212, 22223 and 11111. The duration of the health problems was short (4 months or 2 months) when we used health problems under certainty and chronic (rest of life) when the risk of the health problem was small (1% or 0.5%). The potential health problems presented to subjects were:

1. (22223, 4 months, 100% probability)
2. (22223, 2 months, 100% probability)
3. (21212, 4 months, 100% probability)
4. (21212, 2 months, 100% probability)
5. (1% risk of 22223, chronic)
6. (0.05% risk of 21212, chronic)

The health gains used in our study are summarized in Table 3. In total, we asked WTP questions for 11 health gains.

Table 3. Health improvements used in the study

Treatment A	Treatment B	Health gain			
		Type	ΔQoL	t'	p
22223, 4 months	(11111, 4 months)	1	22223→11111	4 months	100%
	22223, 2 months + 11111, 2 months	2	22223→11111	2 months	100%
22223, 2 months	11111, 2 months	3	22223→11111	2 months	100%
	21212, 2 months	4	22223→21212	2 months	100%
21212, 4 months	11111, 4 months	5	21212→11111	4 months	100%
	21212, 2 months +11111, 2 months	6	21212→11111	2 months	100%
21212, 2 months	11111, 2 months	7	21212→11111	2 months	100%
1% risk of 22223,chronic	0% risk of 22223,chronic	8	22223→11111	Rest of life	1%
	0.5% risk of 22223,chronic	9	22223→11111	Rest of life	.5%
1% risk of 21212,chronic	0% risk of 21212,chronic	10	21212→11111	Rest of life	1%
	0.5% risk of 21212,chronic	11	21212→11111	Rest of life	.5%

3.2.2. Framing of the WTP questions

For the WTP questions under certainty, subjects were asked to assume that they had been diagnosed with a particular illness that would, if untreated, put them in a specified health state (22223 or 21212) for the rest of their lives. They were told that there was a medicine (treatment A) which, if taken for 1 year, would cure this illness. The cost of this medicine for the patient was zero. They were told that it took some time for the medicine to take effect: 2 months in some cases, 4 months in other cases. They were also informed that there was another medicine (treatment B) that also cured the chronic problem but that was better than A. By "better" we meant one of several things: a) that it worked immediately, so that the lower health state (either 22223 or 21212) would not be experienced at all; b) that it reduced the duration of symptoms from 4 months to 2 months; or c) it reduced the severity of the symptoms from 22223 to 21212 without changing the duration. They were told that none of the medicines had side effects and that both had to be taken for a year. However, treatment B had some monetary cost for them. They were then shown a visual aid of the kind reproduced in appendix 1 in order to decide whether they would pay or not.

9

For the WTP questions under risk, subjects were asked to assume that they had started developing some symptoms that for most people would just disappear in a few weeks. However, there was some chance that the symptoms might be due to an illness that, if not treated, was going to put them in a lower health state for the rest of their lives. The risk of developing this chronic illness was 1%. They were told that they could take a medicine that could reduce the risk, in one case from 1% to 0% and, in another case, from 1% to 0,5%.

3.3. The Questionnaire

The questionnaire had four parts:

1. Introduction

To explain the objective of the survey and to familiarize subjects with the health states, they had to rate health states on a scale from 0 to 10.

2. Standard Gamble

We elicited utilities for health states 21212 and 22223 as chronic (rest of their lives) using the Standard Gamble procedure. We started asking subjects if they would accept a 50% chance of death in order to avoid the chronic health problem. We then used a 'ping-pong' method: if a subject rejected a particular level of risk, that probability was reduced and the question was asked again; if a subject accepted a particular level of risk, that probability was increased and the question was asked again. In this way, we iterated to a range where the indifference point was located. When this range was 5%, we asked the subject to give us the indifference point. For example, if the subject rejected a 5% risk but accepted a 10% risk we asked her to give us the indifferent point between 5% and 10%. In group A we also elicited the utility of 21212 with respect to health state 22223. That is, subjects were faced with the choice between, on the one hand, spending the rest of life in health state 21212, or alternatively taking a medical treatment which, if successful, would put them in full health but which, if it failed, would worsen their health to 22223. So in group A, the utility of health state 21212 was estimated directly, using an SG with death as the worse outcome, and also indirectly, by chaining through health state 22223.

3. WTP questions

The method used to elicit WTP was a card sorting procedure. The cards that were offered to subjects were 6€, 12€, 18€, 30€, 45€, 60€, 90€, 120€, 180€, 240€ and 300€ These were monthly outlays that they had to pay for 1 year, except in case of one group (labeled A-2) who had to pay for 2 years. The interviewer was instructed to shuffle the cards before them and to display all the cards on a table in a random order. Subjects were then asked to state the amounts that they were willing to pay, those they were not willing to pay and those they were not sure about. Those who said they were willing to pay all of the amounts presented to them were asked how much more they would pay.

4. Sociodemographic questions

The interview concluded by collecting personal details of the kind reported in Table 1 above.

3.4. Hypothesis testing and Sub-groups

In order to test our hypothesis we used 7 sub-groups, differentiated in order to allow our various hypotheses to be tested. The different WTP questions asked to each sub-group are shown in Table 4.

Table 4 WTP questions in each sub-group and type of health gain.

	Order 1 – large to small		Order 2 – small to large
Group A-1 Group A-3	[Type 3] (22223,2) → (11111,2) [Type 7] (21212,2) → (11111,2) [Type 4] (22223,2) → (21212,2)	Group A-2	(22223,2) → (21212,2) (21212,2) → (11111,2) (22223,2) → (11111,2)
Group B-1	[Type 1] (22223,4) → (11111,4) [Type 5] (21212,4) → (11111,4) [Type 2] (22223,4) →[(22223,2) + (11111,2)] [Type 6] (21212,4) →[(21212,2) + (11111,2)]	Group B-2	(21212,4) →[(21212,2) + (11111,2)] (22223,4) →[(22223,2) + (11111,2)] (21212,4) → (11111,4) (22223,4) → (11111,4)
Group C-1	[Type 8] (1%, 22223) → (0%, 22223) [Type 10] (1%, 21212) → (0%, 21212) [Type 9] (1%, 22223) → (0.5%, 22223) [Type 11] (1%, 21212) → (0.5%, 21212)	Group C-2	(1%, 21212) → (0.5%, 21212) (1%, 22223) → (0.5%, 22223) (1%, 21212) → (0%, 21212) (1%, 22223) → (0%, 22223)

The 7 sub-groups can be first classified into three main blocks, that we call A, B and C. In each block we used two different orders when asking WTP questions. In sub-groups A-1, A-3, B-1 and C-1 the first question was a WTP for the largest health gain that this subject was going to see and the last question was a WTP question for the smallest health gain. On some occasions there was no objective ranking in the health gain. For example, in block A, it was clear that the largest health gain was health gain Type 3 [(22223, 2 months) → 11111, 2] but there was not a clear ranking between health gains Type 4 and Type 7. In block B, it was clear that health gain Type 1 was the largest and health gain Type 6 was the lowest but there was not a clear ranking between health gains Type 2 and Type 5. The same happens with block C. Nevertheless, as far as possible we had some groups where the order was large to small (A-1, A-3, B-1 and C-1) while other groups were presented with questions in the opposite order.

The only difference between A-3 and A-1 was that in group A-1 subjects were told that they had to pay each month for 1 year and in group A-3 for 2 years.

We can test our hypotheses using these sub-groups as follows:

1. MV-QALY is constant for any t'. This can be done by comparing Types 1 vs 2 and 5 vs 6.

11

2. MV-QALY is constant for any ΔQ. This can be tested comparing all health gains where t' is constant, e.g. Types 1 vs 5, 3 vs 4, 3 vs 7, 4 vs 7, 2 vs 6, 8 vs 10 and 9 vs 11.
3. MV-QALY is constant for any p*. This can be tested comparing Types 8 vs 9 or 10 vs 11.

All these tests are within groups tests.

4. MV-QALY is constant irrespective of the order of questions (no order effects). This can be tested by comparing the same questions for groups were the order has been different, that is, A-1 vs A-2, B-1 vs B-2 and C-1 vs C-2.
5. MV-QALY is constant irrespective of the length of payment period. We will test this by comparing groups A-1 and A-3.

4. RESULTS

4.1. Groups A-1, A-2 and A-3

Table 5 summarizes the results of groups A-1, A-2 and A-3. We can see that:

1. There are important order effects (comparison of groups A-1 and A-2). When we start with the lowest health gain WTP is lower.
2. Monthly WTP is insensitive to the period of payment (comparison of groups A-1 and A-3).

Table 5 Mean WTP per month (€) and mean utilities. In groups A-1 and A-2 the duration of payment was 1 year and in group A-3 it was 2 years.

	A-1	A-2	A-3	A-1 vs A-2	A-1 vs A-3
WTP					
(22223,2) → (11111,2)	112,32	60,13	103,61	*	*
(22223,2) → (21212,2)	59,65	35,48	46,9	*	*
(21212,2) → (11111,2)	67,55	26,38	60,93	*	*
Utilities					
State 22223	0.7026	0.7581	0.7278	ns	ns
State 21212 – direct	0.8538	0.8830	0.8608	ns	ns
State 21212 – chained	0.9353	0.9474	0.9221	ns	ns

* There are statistically significant differences at the 5% level using the t-test;
ns = not significant at the 5% level

Table 6 Median WTP per month (€) and median utilities. In groups A-1 and A-2 the duration of payment was 1 year and in group A-3 it was 2 years.

	A-1	A-2	A-3	A-1 vs A-2	A-1 vs A-3
WTP					
(22223,2) → (11111,2)	60	45	90	*	*
(22223,2) → (21212,2)	30	30	45	ns	*
(21212,2) → (11111,2)	45	18	55	*	*
Utilities					
State 22223	0.750	0.750	0.800	ns	ns
State 21212 – direct	0.900	0.900	0.950	ns	ns
State 21212 – chained	0.964	0.963	0.981	ns	ns

* There are statistical significant differences at the 5% level using the Mann-Whitney test.

As explained earlier, the difference between U(21212)-direct and U(21212)-chained is that in the first case the utility of 21212 is elicited through a gamble where the outcomes are Full Health and Death, whereas in the second case, the utility of 21212 is elicited through a gamble where the outcomes are Full Health and 22223. The fact that the chained procedure produces higher utilities than the direct procedure is a well-known result in the literature (Stalmaier, 2002) and it implies that the "distance" between 22223 and 21212 is perceived as bigger when they are compared directly than when they are compared through common outcomes like Full Health and Death.

These results translate into the MV-QALYs reported in Table 7. As previously discussed, the point estimates have been calculated by combining the means of WTP responses and the mean health utility improvements due to a particular medical treatment. The way the calculation is done can be illustrated with respect to the first entry in Table 5. For Group A-1, the mean monthly WTP to move from 22223 to 11111 for 2 months is 112,33€ Multiply this by 12 to give the total WTP of 1,347.84€ This is the mean WTP for a health improvement worth, on average, a utility gain of 0.2974 (i.e. from 0.7026 to 1.000) for one-sixth of a year: that is, a QALY gain of 0.2974/6 = 0.04956. Using equation [1] gives MV-QALY = 27,192€ The confidence intervals have been obtained by the bias-corrected and accelerated percentile method (Efron 1987). This approach is widely used in the field of health economics to estimate non-parametric bootstrap confidence intervals of incremental cost-effectiveness ratios (Briggs et al. 1997, Tambour and Zethraeus 1998, Lord and Asante 1999, Barber and Thompson 2000). The null hypothesis of equality between MV-QALYs can be tested checking if differences between them are equal to zero. Differences can be directly bootstrapped from the samples, and an approximate one-sided significance level of the differences is obtained by calculating the proportion of negative values in the vector of differences (Efron and Tibshirani 1993). We will use a 10% significance level in order to test if the MV-QALY is constant or not in all cases (for groups A, B and C).

Table 7 Mean-based monetary value of the QALY (1,000€) in groups A-1, A-2 and A-3

Health gain		MV-QALY (1,000€)		
		Group A-1	Group A-2	Group A-3
3	22223, 2 months → 11111, 2 months	27.192 (33.963-21.509)	17.898 (21.963-14.564)	54.803 (63.164-47.830)
7a	21212, 2 months → 11111, 2 months [U(21212)-direct]	33.269 (46.020-23.669)	16.230 (20.714-12.961)	63.032 (82.189-52.103)
7b	21212, 2 months → 11111, 2 months [U(21212)-chained]	75.158 (56.201-98.417)	36.199 (27.083-49460)	112.596 (88.138-158.566)
4a	22223, 2 months → 21212, 2 months [U(21212)-direct]	28.393 (43.342-19.626)	20.454 (26.249-15.772)	50.755 (62.773-41.954)
4b	22223, 2 months → 21212, 2 months [U(21212)-chained]	18.453 (13.959-23.895)	13.484 (11.040-16.296)	34.752 (30.209-40.214)

The largest difference within each group is produced by the procedure used (chained or unchained) to elicit U(21212). Both MV-QALY elicited through the chained procedure are statistically different from the rest. Leaving aside the effect of chaining, the hypothesis that the MV-QALY is constant for any ΔQ is rejected in some but not all cases: the difference between the MV-QALY estimated in 3 vs 7a is statistically significant for groups A-1 and A-3 but not for group A-2. The difference in the MV-QALY in 3 vs 4a is not statistically significant in any group. The difference in the MV-QALY in 4a vs 7a is only statistically significant in group A-2. So there is mixed evidence concerning the effect of ΔQ on the MV-QALY.

There are clearer and more unambiguous differences between groups due to order effects. In all cases, the MV-QALY is significantly higher in group A-1 than in A-2. Also, there is a large and systematic discrepancy between every MV-QALY in A-3 and its counterpart in A-1 and in every case we can reject at the 10% level the hypothesis that the MV-QALY is not influenced by length of payment period: when people have to pay for 2 years instead of 1 year, they hardly adjust their monthly WTP, thereby giving rise to a much higher MV-QALY.

Although we cannot undertake comparable statistical tests, it may be of interest to see the corresponding estimates based on median rather than mean responses. These are shown in Table 7a below.

14

Table 7a Median-based monetary value of the QALY (1,000€) in groups A-1, A-2 and A-3

Health gain		MV-QALY (1,000€)		
		Group A-1	Group A-2	Group A-3
3	22223, 2 months → 11111, 2 months	17.280	12.960	64.800
7a	21212, 2 months → 11111, 2 months [U(21212)-direct]	32.400	12.960	158.400
7b	21212, 2 months → 11111, 2 months [U(21212)-chained]	90.000	35.027	416.842
4a	22223, 2 months → 21212, 2 months [U(21212)-direct]	14.400	14.400	43.200
4b	22223, 2 months → 21212, 2 months [U(21212)-chained]	10.093	10.141	35.801

These figures follow similar patterns (though sometimes to a more extreme extent) to those evident in Table 7. Only in A-2 do the values of MV-QALY look stable across different ΔQ (and then only when direct estimates of utility are used – the figures using the chained estimates appear to be much more volatile). Order effects seem a little more muted for 3, 4a and 4b – but somewhat more extreme for 7a and 7b. And in all cases the impact of payment period: on the basis of means, Group A-3 generally produced figures roughly double those from Group A-1; but on the basis of medians, the A-3 figures are between three and four times bigger than their A-1 counterparts.

4.2. Groups B1, B2

Results for groups B-1 and B-2 are reported in Table 8.

Table 8 Mean and median WTP per month (€). Mean and median utilities

	B1	B2	p	B1	B2	p
WTP	Mean	Mean		Median	Median	
(22223,4) → (11111,4)	156,93	78,3	*	120	47.5	+
(21212,4) → (11111,4)	111,58	51,39	*	60	45	+
(22223,4) →[(22223,2) + (11111,2)]	102,19	44,95	*	60	30	+
(21212,4) →[(21212,2) + (11111,2)]	69,46	27,99	*	37,5	18	+
Utilities						
State 22223	0.7198	0.6826	ns			ns
State 21212	0.8686	0.8269	ns			+

* There are statistically significant differences at the 5% level using the t-test.

+ There are statistically significant differences at the 5% level using the Mann-Whitney test.

These results translate to the values of MV-QALY shown in Table 9. The hypothesis that MV-QALY is constant for any t' is clearly rejected by the comparison between Types 5 vs 6 and 1 vs 2. The hypothesis that MV-QALY is constant for any ΔQ is also rejected by

15

comparing Types 1 vs 5 and 2 vs 6 in both groups. As with groups A-1, A-2 and A-3 the existence of order effects is confirmed in groups B-1 and B-2: the MV-QALY is always bigger in group B-1.

Table 9 Monetary value of the QALY (1,000€) in groups B-1 and B-2.

	Health gain	MV-QALY (1,000€)	
		B-1	B-2
5	21212, 4 → 11111, 4	30.574 (22.716-42.695)	10.685 (8.037-14.548)
6	21212, 4 → 21212, 2 + 11111, 2	38.068 (28.023-55.691)	11.639 (8.823-15.445)
1	22223, 4 → 11111, 4	20.158 (16.684-24.413)	8.881 (7.306-10.770)
2	22223, 4 → 22223, 2 + 11111, 2	26.253 (20.392-33.068)	10.197 (8.508-12.310)

4.3. Groups C1 and C2

We now turn to the case where risk is involved. Results are shown in Table 10.

Table 10 Mean WTP per month (€) and mean utilities. Groups C-1 and C-2

		C-1	C-2	p	C-1	C-2	p
WTP		Mean	Mean		Median	Median	
Risk reduction	Health state						
(1%→0%)	22223, chronic	83.06	54.86	*	75	45	+
(1%→0%)	21212, chronic	74.10	34.95	*	60	30	+
(1%→0.5%)	22223, chronic	61.59	24.6	*	45	18	+
(1%→0.5%)	21212, chronic	39.71	16.05	*	30	12	+
Utilities							
State 22223		0.7084	0.7040	ns	0.7	0.7	ns
State 21212		0.8298	0.7974	ns	0.8	89	ns
Undiscounted QALY life gain[1]							
State 22223		12.23	12.29				
State 21212		7.18	8.40				

(1) Estimated as $\dfrac{\sum_{i=1}^{n}[1-U_i^s]xLE_i}{n}$. That is, we estimated the quality adjusted life expectancy with and without the illness. We took the difference between these two amounts for each individual and estimated the mean. In order to estimate the average health gains we multiplied the undiscounted QALY gain by the risk reduction. E.g. the average health gain in terms of QALYs was 0.1223 for health state 22223 when the reduction was 1% for group C-1.

* There are statistical significant differences at the 5% level using the t-test.

+ There are statistical significant differences at the 5% level using the Mann-Whitney test.

These results translate into the MV-QALYs shown in Table 11. For example, if the average member of the C-1 subset is willing to pay 83.06 per month for 12 months, that comes to a total of 996.72. If we divide that by 0.1223 we get the MV-QALY of 8,151 shown in the appropriate cell in Table 11.

Table 11. Monetary value of the QALY. Groups C-1 and C-2

Health gain		MV-QALY (1,000€)	
		Group C-1	Group C-2
8	22223 1% → 0%	8.151 (6.481-10.525)	5.358 (4.421-6.639)
10	21212 1% → 0%	12.387 (9.522-17.127)	4.992 (4.000-6.317)
9	22223 1% → 0.5%	12.088 (9.280-15.288)	4.805 (3.797-5.970)
11	21212 1% → 0.5%	13.277 (9.725-18.025)	4.585 (3.647-5.794)

The MV-QALY is higher for group C-1 than for group C-2 given the sequence effects observed. The hypothesis that the MV-QALY is constant for any ΔQ can be rejected only in one case [8 vs 10 for sub-group C-1]. The hypothesis that the MV-QALY is independent on the size of risk reduction can be rejected in one case [8 vs 9 for C-1].

4.4. Preliminary conclusions

Our hypotheses can be summarized in table 12 assuming a 10% significance level.

Table 12. Summary of results for the MV-QALY. H_0= MV-QALY is constant.

Hypothesis	Comparison	Large to small	Small to large
MV-QALY has to be constant for any t'	1 vs 2	Reject	Reject
	5 vs 6	Reject	Reject
MV-QALY has to be constant for any ΔQ	1 vs 5	Reject	Reject
	3 vs 4a	Accept	Accept
	4a vs 7a	Accept	Reject
	2 vs 6	Reject	Reject
	3 vs 7	Reject	Accept
	8 vs 10	Reject	Accept
	9 vs 11	Reject	Accept
MV-QALY has to be constant for any p*	8 vs 9	Reject	Accept
	10 vs 11	Accept	Accept

The hypothesis that the MV-QALY is the same for any t' is rejected: shorter durations tend to produce higher MV-QALY. The hypothesis that MV-QALY is the same for any

ΔQ is rejected in 8 of the 14 cases analyzed. When it is rejected, the MV-QALY is always higher when elicited through the milder health state (21212). Finally, it seems that we cannot reject the hypothesis that MV-QALY is constant for different risk reductions.

The survey also shows a clear insensitivity to the period of payment and clear sequence effects. The presence of order effects is well documented in the literature (Stewart, 2002) on Contingent Valuation. One reaction to this problem is to assume that the less biased response is the first one since next questions will be influenced by the previous one(s). So it would be better to test our hypotheses in a between sample study where only one question is asked (or at least, we focus only on the first question). We then decided to conduct another survey aimed at testing some of our previous hypothesis in a between sample design.

5 THE SECOND SURVEY

5.1 Details of the survey

340 members of the general population were interviewed. They were distributed in 4 subgroups according to a quota sample method. They were contacted by telephone and those who agreed to be interviewed were visited by an interviewer. There were not statistically significant differences in sociodemographic characteristics between the four sub-samples. Utilities were elicited again using the SG. The framing of WTP questions was very similar to those asked in the previous survey. The main difference was that the response format was polychotomous since there were five categories of responses (Definitely YES/NO, Probably YES/NO, Not Sure). We will present our results using only the responses to the "Definitely YES" category since this is a conservative estimate and it has been recommended (Arrow et al, 1993) to use conservative estimates in Contingent Valuation studies. Also, it has been shown (Blumenschein, 2008) that hypothetical payments are closer to real payments when we only use the category "Definitely YES". Another difference between both surveys is the number of money cards used: the second survey used 10€, 15€, 20€, 30€, 45€, 60€, 75€, 90€, 120€, 180€, 300€, 450€, 600€ Subjects were asked to allocate every amount to one or other of the five response categories. We elicited the WTP of the following health gains:

- (22223, 4 months) → (11111, 4 months)
- (22223, 2 months) → (11111, 2 months)
- (21212, 2 months) → (11111, 2 months)
- (21212, 4 months) → (11111, 4 months)

In this survey, no questions about risk reductions were included. Instead, we added two more health gains in order to study the MV-QALY for even smaller health gains. They were:

18

- (22223, 2 weeks) → (11111, 2 weeks)
- (21212, 2 weeks) → (11111, 2 weeks)

These two WTP questions were asked always as the second question, so they could have been influenced by the response to the first question. For this reason, the final estimate of the MV-QALY it is less "clean" than the other four questions. However, while acknowledging this potential problem, we thought that it was interesting to have some information on the MV-QALY for health gains that were even smaller than those used in survey 1.

5.2 Results

The results of the second survey are given in Table 13. The results of questions that were asked first were quite similar in both surveys. WTP for (22223, 4 months) was 157€ in the first survey and 163€ in the second survey. WTP for (22223, 2 months) was 112€ in group A-1 and 127€ in this survey. The differences are not statistically significant. For health gains that were not asked in the same order in both surveys, the results were different. In groups D-2 and D-4 (second survey), WTP for health gains (21212, 2 months) and (21212, 4 months) was higher than in groups A and B, and especially higher than in groups A-2 and B-2. This may reflect the fact that in survey 1 the position of these questions was different from their position in survey 2. Once again, there was no sensitivity with respect to duration for health state 21212, while for health state 22223, WTP showed *some* sensitivity to duration but was not proportional.

Table 13 Mean WTP and mean utilities (€). N=83 per group

	Group D-1	Group D-2	Group D-3	Group D-4
WTP				
(22223,4) → (11111,4)	163.54	--------	------	-------
(21212,4) → (11111,4)	------	111.47	------	------
(22223,2) →(11111,2)	------	------	127.37	------
(21212,2) →(11111,2)	-------	------	------	105.38
(22223, 2weeks) →(11111,2weeks)	96.65	------	86.98	-----
(21212, 2weeks) →(11111,2weeks)	------	89.72	-----	58.88
Utilities				
State 22223	.6412	.7010	.5627	.6752
State 21212	.7868	.8292	.7835	.7934

The MV-QALY is given in Table 14. There is a clear pattern here: namely, the lower the duration and the milder the health state, the higher is the estimate of MV-QALY. One particularly worrying result is that the estimates of the MV-QALY are very much larger when the two week time span is used.

19

Table 14. Monetary value of the QALY(€) in second survey

	4 months	2 months	2 weeks
22223	16.407 (12.378-22.081)	21.777 (16.305-26.831)	76.235 (63.140-92.336)
21212	23.487 (16.296-32.049)	37.860 (27.282-54.504)	123.724 (93.678-161.411)

More formally, the hypothesis about duration can be tested by comparing "columns" and the hypothesis about quality of life can be tested comparing "rows". In almost all cases, the hypothesis that the monetary value of the QALY is independent of these factors can be rejected at the 10% significance level. The only case where it cannot be rejected is in the comparison of (22223. 4 months) vs (22223, 2 months). These results reinforce the results of survey 1, suggesting that milder health states and shorter durations both increase the MV-QALY estimates. The inclusion of an even shorter duration (2 weeks) dramatically increases the disparity. These results emerge even more strongly from the between-sample design of survey 2 than in survey 1, where within-subject comparisons may have encouraged *some* greater internal consistency. However, in survey 1, whenever there was a difference in the MV-QALY for 21212 and 22223, it was always the case that it was bigger when elicited through 21212 than through 22223. The results of survey 2 reinforce concerns that the MV-QALY estimate is liable to be systematically influenced by the choice of which features of the scenario are to be varied, and by how much.

6. CONCLUSIONS

The main objective of this paper has been to test the robustness of the various consistency conditions that must hold if we are to obtain a reliable single figure for the MV-QALY from questions involving marginal health gains. Our results are not very encouraging since we find that:

1. There are clear sequence effects, suggesting that responses and estimates are vulnerable to influences that in theory should not make a difference.
2. The MV-QALY varies systematically with the *severity* of the health state used to elicit it, with milder health states resulting in higher estimates.
3. The MV-QALY varies systematically with the *duration* of the health state used to elicit it, with shorter durations resulting in higher estimates.
4. The MV-QALY can be affected dramatically by changing the duration of the period of payment.
5. We find only limited evidence that different risk reductions affect the MV-QALY estimates. This is moderately encouraging – although it should be borne in mind that this result is somewhat out of line with evidence from other areas, such as the value of preventing road fatalities, where estimates often do vary with the size of risk reduction used in the questions asked.

A particularly worrying result of this study is the absolute lack of sensitivity of monthly WTP with respect to the duration of the period of payment. Since this duration is quite arbitrary in many cases, the MV-QALY can be subject to manipulation by just modifying this parameter. This problem has nothing to do with all the assumptions that we need to make in order to elicit a unique MV-QALY. It is a very elementary rationality condition which, if not met, would undermine any serious effort to elicit these monetary values. Unfortunately, there is little existing evidence about this issue in the health economics literature. There is some more evidence on the field on environmental economics (Stevens, 1997; Stumborg, 2001). For example, Stumborg et al (2001) find that total WTP for a public good is higher when the duration of payment is 10 years than when it is 3 years (unless a 40% discount rate is assumed). If the period of payment is arbitrary and the total WTP changes with it, no reliable estimates of WTP can be elicited. However, some others have found greater sensitivity (Johnson et al, 2006).

Our study then suggests further research topics around the relationship between WTP and QALYs. More evidence should be collected to check whether our result that the MV-QALY is higher for milder health states holds, and explore further the relationship between the duration of a health gain and WTP. Finally, we need to understand why WTP is not sensitive to the payment period and whether there is any way of encouraging the appropriate sensitivity. At present, however, we are still a long way from a situation where the MV-QALY can be elicited in a reliable way.

7. REFERENCES

Abellan-Perpinan JM, Pinto-Prades JL, Mendez-Martinez I, Badia-Llach X. Towards a better QALY model. Health Economics. 2006 Jul;15(7):665-76.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation," *Federal Register*, January 15, vol. 58, no. 10, pp. 4601-4614.

Bala MV, Wood LL, Zarkin GA, Norton EC, GafnweA, O'Brien B. Valuing outcomes in health care: a comparison of willingness to pay and quality-adjusted life-years. Journal of Clinical Epidemiology. 1998 Aug;51(8):667-76.

Barber, J. A., and S. G. Thompson (2000). Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. Statistics in Medicine, 19, 3219-3236.

Bleichrodt H, Quiggin J. Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis? Journal of Health Economics. 1999 Dec;18(6):681-708.

Bleichrodt, Han and Jose Luis Pinto. (2005). "The Validity of QALYs under Nonexpected Utility." *The Economic Journal* 115, 533-550.

Blumenschein K, Blomquist G, Johannesson M, Horn N, Freeman P(2008) Eliciting Willingness to Pay Without Bias: Evidence from a Field Experiment, *The Economic Journal* 118 (525) , 114–137.

Briggs, A.H., D. E. Wonderling, and C. Z. Mooney (1997). Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Economics, 6, 327-340.

Byrne MM, O'malley K, Suarez-Almazor ME. Willingness to pay per quality-adjusted life year in a study of knee osteoarthritis. Medical Decision Making. 2005 Nov-Dec;25(6):655-66.

Dolan P, Edlin R. Is it really possible to build a bridge between cost-benefit analysis and cost-effectiveness analysis? Journal of Health Economics 2002 Sep;21(5):827-43.

Donaldson C, Currie G, Mitton C.Cost effectiveness analysis in health care: contraindications. British Medical Journal. 2002 Oct 19;325(7369):891-4.

Effron, B. and Tibshirani, R 1993. An introduction to bootstrap. New York: Chapman & Hall.

Efron, B (1987). Better bootstrap confidence intervals. Journal of the American Statistical Association, 82, 171-200.

Gyrd-Hansen D.   Willingness to pay for a QALY.Health Economics. 2003 Dec;12(12):1049-60.

Gyrd-Hansen D. Willingness to pay for a QALY: theoretical and methodological issues. Pharmacoeconomics. 2005;23(5):423-32.

Johannesson M, Meltzer D. Some reflections on cost-effectiveness analysis. Health Economics. 1998 Feb;7(1):1-7.

Johannesson M, O'Conor RM. Cost-utility analysis from a societal perspective. Health Policy. 1997 Mar;39(3):241-53.

Johannesson M. The relationship between cost-effectiveness analysis and cost-benefit analysis. Social Science & Medicine 1995;41:483-489.

Johnson FR, Fries EE, Banzhaf HS. Valuing morbidity: an integration of the willingness-to-pay and health-status index literatures. Journal of Health Economics. 1997 Dec;16(6):641-65.

Johnson BK, Mondello MJ and Whitehead JC (2006) "Contingent Valuation of Sports: Temporal Embedding and Ordering Effects," *Journal of Sports Economics*, 7 (3), 267-288.

Karnon J, Tsuchiya A, Dolan P. Developing a relativities approach to valuing the prevention of non-fatal work-related accidents and ill health. Health Economics. 2005 Nov;14(11):1103-15.

King JT Jr, Tsevat J, Lave JR, Roberts MS. Willingness to pay for a quality-adjusted life year implications for societal health care resource allocation. Medical Decision Making. 2005 Nov-Dec;25(6):667-77.

Lord, J., and M. A. Asante (1999). Estimating uncertainty ranges for costs by the bootstrap procedure combined with probabilistic sensitivity analysis. Health Economics, 8, 323–333.

Neumann PJ, Sandberg EA, Bell CM, Stone PW, Chapman RH. Are pharmaceuticals cost-effective? A review of the evidence. Health Affairs 2000; 19: 92-109.

Nord E, Pinto JL, Richardson J, Menzel P, Ubel P. Incorporating societal concerns for fairness in numerical valuations of health programmes. Health Economics. 1999 Feb;8(1):25-39.

Stalmeier PF. Discrepancies between chained and classic utilities induced by anchoring with occasional adjustments. *Medical Decision Making*. 2002 Jan-Feb;22(1):53-64.

Stevens, T.H.; DeCoteau, N.E.; Willis, C.E. (1997) Sensitivity of contingent valuation to alternative payment schedules. *Land Economics* 73: 140-148.

Stewart JM, O'Shea E, Donaldson C, Shackley P. Do ordering effects matter in willingness-to-pay studies of health care? Journal of Health Economics. 2002 Jul;21(4):585-99.

Stumborg, Basil E., Baerenklau, Kenneth A. & Bishop, Richard C. (2001) Nonpoint Source Pollution and Present Values: A Contingent Valuation Study of Lake Mendota. *Review of Agricultural Economics* **23** (1), 120-132.

Tambour, M., and N. Zethraeus (1998). Bootstrap confidence intervals for cost-effectiveness ratios: some simulation results. Health Economics, 7, 143–147.

Zethraeus N. Willingness to pay for hormone replacement therapy. Health Economics. 1998 Feb;7(1):31-8.

Appendix 1

OPTION 1		
TAKE MEDICINE X FOR 1 YEAR		
<u>Health State</u> ➢ **I have some problems in walking about** ➢ **I have some problems with self care** ➢ **I have some problems in performing my usual activities** (work, study, housework, family or leisure activies) ➢ **I have moderate pain or discomfort** ➢ **I am very anxious or depressed**	2 MONTHS	**NO** ADDITIONAL EXPENDITURES

OPTION 2		
TAKE MEDICINE A FOR 1 YEAR		
➢ <u>GOOD HEALTH</u>	2 MONTHS	**YES** ADDITIONAL EXPENDITURES