TRYING TO ESTIMATE A MONETARY VALUE FOR THE QALY

January 2009

Jose Luis Pinto Prades*
University Pablo de Olavide
Carretera de Utrera km 1, 41013 Sevilla
Center for Research in Health Economics,
Pompeu Fabra University

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 feasibility of estimating a monetary value for a QALY (MVQ). Using two different surveys of the Spanish population (total n=892), we consider whether willingness to pay (WTP) is (approximately) proportional to the health gains measured in QALYs. We also explore whether subjects' responses are prone to any significant biases. We find that the estimated MVQ varies inversely with the magnitude of health gain. We also find two other (ir)regularities: the existence of ordering effects; and insensitivity of WTP to the duration of the period of payment. Taken together, these effects result in large variations in estimates of the MVQ. If we are ever to obtain consistent and stable estimates, we should try to understand better the sources of variability found in the course of this study.

JEL: I10, I18

Keywords: QALYs, willingness to pay, biases.

*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. Introduction

Much cost-effectiveness and cost-utility analysis has been undertaken to guide health care resource allocation. However, if we wish to go further and conduct cost-benefit analysis – and thereby make health care resource allocation more directly comparable with decision making in other areas of public policy – we need to find some way of attaching monetary values to health benefits. Since much of the health benefit measurement to date has been conducted in terms of Quality Adjusted Life Years (QALYs), one solution – were it to be feasible – would be to estimate the monetary value of a QALY (henceforth, the MVQ).

In the UK, the National Institute for Health and Clinical Excellence (NICE) recognised the desirability of having such an estimate and recently commissioned a study to explore its feasibility. Others, too, have recognized the potential value of such a figure: for example, Johanesson and Meltzer (1998) considered that obtaining this information "should be a research priority" (p. 4). They identified two possible strategies for deriving the Willingness to Pay (WTP) per QALY gained. One would be based on direct elicitation of WTP for some marginal health change(s), while the other would derive this figure from estimates of the value of statistical life (VSL) in the literature. Some work of this latter kind has been attempted[1], but the present study focuses instead on the possibility of estimating the MVQ on the basis of eliciting people's WTP for a range of health benefits.

We acknowledge the possibility that estimating a *unique* MVQ may not be feasible. Several papers, including Bleichrodt and Quiggin (1999) and Dolan and Edlin (2002), have shown that the conditions for a unique MVQ to exist are quite restrictive and are unlikely to hold. However, our objective was to explore just how stable – or variable – such an estimate might be. Our strategy was to use changes that were small relative to those used in other studies in this area[2] in the hope that budget constraints would not cause significant nonlinearities[3].

Abellan-Perpiñan et al. (2006) and Bleichrodt and Pinto (2005) have suggested that a non-linear QALY model may be better than the linear model. If non-linearities are important, this may result in estimates of the MVQ varying considerably according to the basis upon which they are derived. So we wished to check for the possible impact of various non-linearities with respect to severity, duration and the size of risk reduction.

Our empirical work involved two surveys. The first, and larger, of these was designed to investigate several issues fundamental to the robust estimation of a reasonably stable MVQ. The second survey was designed to follow up and clarify certain issues raised by the results of the first.

[1] Hirth et al. (2000); Mason et al. (2008)
[2] For example, Byrne et al. (2005), King et al. (2005), Gyrd-Hansen (2003).
[3] It is always difficult to know how far people feel they are coming up against serious budget constraints, but by using smaller changes than previous studies, we aimed to attenuate this problem.

2. Broad Structure of the Study

Before describing the particular features of each study, we set out the general framework within which we are working and according to which we shall derive the MVQ estimates.

The basic idea of a QALY is to provide a measure that facilitates comparisons across a broad spectrum of health benefits. Ideally, it allows health care decision makers to weigh the total benefit of an intervention that alleviates short-term conditions involving moderate adverse effects against other interventions addressing more severe acute conditions, or chronic conditions of varying degrees of severity, or conditions that threaten to substantially reduce life expectancy. A monetary value of a QALY should therefore be robust across such a spectrum.

So in the first survey, we were concerned to focus on the stability (or otherwise) of the MVQ with respect to:

Variations in the ***severity*** of different conditions;
Variations in the ***duration*** of a given condition;
Variations in ***reductions in the risk*** of chronic conditions involving a substantial total loss of quality of life.

Aware that WTP studies in related areas have been found to be vulnerable to various procedural effects (for examples, see Bateman et al. (2002)), we checked for any possible impact of the order in which questions were asked and the period over which the notional monetary payment was 'collected'.

In the light of questions raised about the relationship between the estimated MVQ and the magnitude of the health gain, we conducted a second smaller (but still substantial) survey to provide further information about that particular key issue.

2.1 Questions and scenarios

The studies revolved around questions which took the following general form. Respondents were asked to assume that they had been diagnosed with a particular illness that would, if untreated, put them in a specified impaired health state 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 impaired health state would not be experienced at all; b) that it reduced the duration of symptoms from 4 months to 2 months; or c) that it reduced the severity of the symptoms but did not change the duration.

In all the above cases, the respondents were asked to consider the prognoses as certain: i.e. the diagnosis was certain and the effects of the treatments were certain. But there was also a form of question involving risk, where respondents were asked to assume that they faced a 1% chance of developing an illness which could put them in an impaired health state for the rest of their lives. They were then 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%.

For all scenarios, they were told that neither of the medicines had side effects and that both had to be taken for a year. However, in all cases the better treatment involved some monetary cost for them. Their willingness to pay for the extra benefits of the better treatment was then elicited.

In order to investigate the relationship between MVQ and the different factors being varied, we considered a total of 11 different scenarios, or 'Types', as set out in Table 1.

We used three Euroqol (EQ-5D) health states to represent different quality of life levels. EQ-5D is a standardised instrument widely used in the computation of QALYs. It has five dimensions, with three levels on each dimension – no problems (1), moderate problems (2) and severe problems (3). So 11111 signifies being in good health with no abnormal problems on any dimension, while 21212 represents moderate problems on some dimensions and 22223 represent moderate problems on most dimensions, with severe problems on the fifth dimension (anxiety/depression). The two impaired health states 21212 and 22223 were chosen because piloting showed that they were not so severe as to create complications with negative scores[4] but were not so mild that many respondents would refuse to accept any risk of death to avoid them.

In all cases where the prognosis and effects of treatment were certain, the duration of the health problems was short (4 months, 2 months, 2 weeks); alternatively, when the prospect was chronic (rest of life), the baseline risk of the health problem was small. Thus, for example, Type 1 was a scenario where, given the 'default' Treatment A (which cost the respondent nothing in monetary terms), she would experience impaired health state 22223 for 4 months before the treatment restored her to full health, whereas Treatment B would work immediately, thereby avoiding the 4 months in that impaired state. Type 2, by contrast, supposed that Treatment B would not work immediately but would reduce the period spent in 22223 from 4 months to 2 months.

[4] Scores below zero signify that the respondent considers a state to be worse than being dead.

Table 1. Treatment Choice Scenarios Used in First Survey

Type	Original state	Treatment A	Treatment B
1	22223, rest of life	4 months in 22223, then move to 11111	Immediately move to 11111
2	22223, rest of life	4 months in 22223, then move to 11111	2 months in 22223, then move to 11111
3	22223, rest of life	2 months in 22223, then move to 11111	Immediately move to 11111
4	22223, rest of life	2 months in 22223, then move to 11111	2 months in 21212, then move to 11111
5	21212, rest of life	4 months in 21212, then move to 11111	Immediately move to 11111
6	21212, rest of life	4 months in 21212, then move to 11111	2 months in 21212, then move to 11111
7	21212, rest of life	2 months in 21212, then move to 11111	Immediately move to 11111
8	22223, rest of life	2 weeks in 22223, then move to 11111	Immediately move to 11111
9	21212, rest of life	2 weeks in 21212, then move to 11111	Immediately move to 11111
10	1% risk of (22223, rest of life)	Stay with original risk	Eliminate risk of (22223, rest of life)
11	1% risk of (22223, rest of life)	Stay with original risk	Reduce to 0.5% the risk of (22223, rest of life)
12	1% risk of (21212, rest of life)	Stay with original risk	Eliminate risk of (21212, rest of life)
13	1% risk of (21212, rest of life)	Stay with original risk	Reduce to 0.5% the risk of (21212, rest of life)

Types 10-13 inclusive are the ‘risk scenarios’, involving either the elimination of the baseline 1% risk (Types 10 and 12) or halving that baseline (Types 11 and 13) for chronic conditions involving either a lifetime of 22223 (10 and 11) or a lifetime of 21212 (12 and 13). The various permutations of benefit in terms of Quality of Life (QoL) improvement, duration and probability can be seen in the right hand columns.

2.2. Computing the Monetary Value of a QALY

With the scenarios described in Table 1 we can investigate the feasibility of estimating a reasonably robust MVQ.

To illustrate, suppose a health care decision maker is considering the benefit to her constituent population of providing (let us say) 6,000 more Treatment B's of Type 1. Each such treatment would reduce the period spent in state 22223 by 4 months, and so would represent a total of 2,000 years of people's time during which the affected people had a higher QoL to the extent that 11111 is better than 22223. If we can measure the QoL improvement on a scale where 0 represents 'as bad as being dead' and 1 represents 'full health', we can 'quality adjust' those 2,000 years accordingly: for example, if the QoL improvement were indexed at 0.6 (i.e. involved moving people from 0.4 to 1), the overall benefit would be judged to be 0.6 x 2,000 = 1,200 QALYs.

If a representative cross-section of the constituent population were willing, on average, to pay Y€ for a single such treatment, this would give a total value of 6,000Y€ for the whole programme of interventions which, collectively, would yield 1,200 QALYs. From this we can derive an estimate for the MVQ: in this example, each QALY would be worth 5Y€. Analogous calculations can be undertaken for Types 2 – 9 to examine the stability of estimates across different scenarios.

For Types 10 – 13, the estimation procedure is similar, but slightly modified. To illustrate, take Type 11. If 6,000 such interventions were delivered, each reducing the risk of the chronic condition from 1% to 0.5%, we should expect the overall effect to be, on average, the prevention of 30 chronic illnesses: that is, 30 people who would otherwise have been in health state 22223 for the rest of their lives would instead be in state 11111[5]. In order to compute the QALY value of this intervention, we need to know not only the index number measuring the QoL difference between 11111 and 22223 but also the remaining life expectancy to which the difference applies: were it to be the case, for example, that the representative recipient had 40 years remaining, the QALY benefit would be 40 x 0.6 = 24 per successful treatment, giving an overall total QALY gain of 720. On the basis of these purely illustrative numbers, the MVQ from questions about Type 11 scenarios would be found by dividing the total WTP of 6,000 representative respondents to this question by the total expected QALY gain of 720.

On this basis, we can set out and test the following 'null hypotheses' as follows:

1. MVQ is constant for any duration. This is tested by comparing Types where the QoL gain is the same but durations differ (e.g. Types 1 vs 2 or 5 vs 6).

[5] In reality, of course, things are not quite so straightforward: ideally what we should compare is the health state profile for the person who suffers the chronic illness with their 'clone' who avoids this illness as a result of the intervention (but who may experience other health impairments during the rest of their life that mean they may not be in state 11111 for the whole of that time – just as the person who succumbs to the chronic condition may also experience other illnesses/impairments on top of that condition).

2. MVQ is constant for any change in QoL. This is tested using pairs where durations are the same but QoL gains differ (e.g. Types 1 vs 5 or 3 vs 4).

3. MVQ is constant for any risk reduction. This is tested directly by comparing Types 10 vs 11 or 12 vs 13; and the estimates derived via these questions can also be compared with those using certainty, based on Types 1 – 9.

All these tests are within-Group tests. All of them rest on the assumption that these health changes are small enough for the utility function to be approximately linear in the neighbourhood of these changes. This assumption will be discussed in the concluding section.

2.3 The Questionnaires

The questionnaires of both surveys had a similar structure, involving four parts:

1. Introduction

This explained the general objective of the survey and familiarised respondents with the health states under consideration by asking them to rate each state on a scale from 0 to 10.

2. Standard Gamble

Utilities for health states 21212 and 22223 were elicited using the Standard Gamble (SG) procedure. Each SG question started by asking subjects if they would accept a 50% chance of death in order to avoid the chronic health problem. A 'ping-pong' procedure followed: 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 that way, the procedure 'homed in' on the risk of death that made the individual indifferent between the two alternatives. Denoting the 'indifference risk of death' for chronic state 22223 by p and the corresponding risk for 21212 by q, the utility scores assigned to those states for that individual are then 1-p for 22223 and 1-q for 21212.

3. WTP questions

The method used to elicit WTP was a card sorting procedure. The cards that were offered to subjects were monthly instalments of 6€, 12€, 18€, 30€, 45€, 60€, 90€, 120€, 180€, 240€ and 300€[6] which they would (in most treatments) have to pay for one year. The interviewer was instructed to shuffle the cards in front of respondents to reassure them that there was no significance to the card that appeared first, or the order in which

[6] In the second survey we also used cards featuring monthly payments of 450€ and 600€.

subsequent cards appeared[7]. Respondents were then asked to identify 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 Appendix 2.

3. First survey

Respondents were 560 members of the general population. They were contacted by telephone through random dialing and those who agreed to be interviewed were visited by an interviewer. They were then assigned at random to one of the seven groups described below (n=80 each group).

3.1. Design

Of the 13 Types listed in Table 1, we used 11 in this first survey: that is, all of them except 8 and 9: the 2 week duration was only addressed in the second survey. The 11 Types were organised into three main blocks which we shall call A, B and C. Within each block we used two different orders when asking WTP questions: as far as possible, the guiding principle was that half of those in each block were presented with questions where they saw the largest health gains first and the smallest health gains last, while for the other half it was the other way round[8]. This generated six groups. A seventh subsample arose from the distinction between Group A-1 and Group A-3. Respondents in these two groups saw exactly the same scenarios in exactly the same order, but Group A-3 were told that they would pay for Treatment B each month for 24 months, whereas those in Group A-1 (and in all other Groups) were told to suppose that their payment would be collected in the form of 12 monthly amounts: the purpose of this manipulation was to see how sensitive overall WTP was to the period of payment.

[7] The intention was to try to control for any possible 'starting point' or 'range' effects – although, as Bateman et al. (2002) have indicated, it may not always be possible to eradicate such effects. Still, by adopting this randomised order, we can at least distribute any bias more or less equally across the groups.
[8] As will become clear, it is not always possible to rank health gains unambiguously: the important point, though, is that in all cases we tested for the independence of answers to the order in which questions were asked.

Table 2: WTP Questions and their Order for Each Subsample Group.

	Order 1 – large to small		Order 2 – small to large
Group A-1 Group A-3	1. [Type 3] 2. [Type 7] 3. [Type 4]	Group A-2	1. [Type 4] 2. [Type 7] 3. [Type 3]
Group B-1	1. [Type 1] 2. [Type 5] 3. [Type 2] 4. [Type 6]	Group B-2	1. [Type 6] 2. [Type 2] 3. [Type 5] 4. [Type 1]
Group C-1	1. [Type 10] 2. [Type 12] 3. [Type 11] 4. [Type 13]	Group C-2	1. [Type 13] 2. [Type 11] 3. [Type 12] 4. [Type 10]

In addition to the hypotheses numbered 1-3 in section 2.2 above, we used this first survey to check for invariance with respect to procedure and to test three further 'null hypotheses' as follows:

4. MVQ estimates from any question are independent of the order of questions. This is tested by comparing A-1 vs A-2, B-1 vs B-2 and C-1 vs C-2.

5. MVQ estimates are independent of the number of payment instalments. This is tested by comparing groups A-1 and A-3.

6. Finally, we also checked for any systematic differences between 'direct' and 'indirect' ways of eliciting QoL indices. In the A Groups, 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. Let the individual's 'indifference risk of 22223' be denoted by r: then in these Groups, the utility of state 21212 was estimated directly as 1-q, and also indirectly, by 'chaining' via health state 22223, as 1-rp. This provides one further check on procedural invariance and the possible impact on estimates of MVQ

3.2 Results

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

These Groups provide *some* of the tests of whether the MVQ is invariant with respect to the size of QoL change used to elicit it. They also allow us to examine all three ways in which procedural invariance might fail: order effects; insensitivity to period of payment; and differences between direct and indirect utility indices for 21212. Table 3 reports the relevant Group means.

Table 3: Mean WTP per month (€) for Treatment B and Mean Health State Utilities from SG Questions. 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 for treatment B					
Type 3	112.32	60.13	103.61	*	ns
Type 7	59.65	35.48	46.90	*	ns
Type 4	67.55	26.38	60.93	*	ns
Utilities					
State 22223	0.703	0.758	0.728	ns	ns
State 21212 – direct	0.854	0.883	0.861	ns	ns
State 21212 – indirect	0.935	0.947	0.922	ns	ns

* There are statistically significant differences at the1% level using the t-test;
ns = not significant at the 5% level

We can see that:

There are important order effects (comparison of Groups A-1 and A-2). When we start with the lowest health gain WTP is lower.

Monthly WTP is insensitive to the period of payment (comparison of Groups A-1 and A-3).

The chained procedure produces higher utilities ($p<0.001$) than the direct procedure. This is a well-known result in the literature (Stalmeier, 2002)

Given the doubts about procedural invariance, we cannot simply pool data from the different A Groups in order to estimate MVQs. We therefore keep them separate and report the various implied mean MVQs in Table 4. In each case, the point estimates have been calculated by combining the means of WTP responses and the mean estimates of the QoL improvements.

The method of calculation can be illustrated with respect to the top-left entry in Table 4. 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 0.297 (i.e. from 0.703 to 1.000) for one-sixth of a year: that is, a QALY gain of 0.04956. This gives MVQ = 1,347.84€/0.04956 = 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 MVQs can be tested by bootstrapping differences from the samples and by obtaining an approximate one-sided significance level of the differences by calculating the proportion of negative values in the vector of differences (Efron and Tibshirani 1993). We will use a 10% significance level in all cases (i.e. for A groups here, and in due

course for B and C Groups). In Table 4, the distinction between figures based on direct estimates of the utility of 21212 and indirect/'chained' estimates is indicated by the qualifiers 'dir' and 'ind' added to the Type label.

Table 4: Mean MVQs and 95% confidence intervals in Groups A-1, A-2 and A-3

Type	MVQ		
	Group A-1	Group A-2	Group A-3
3	27,192 (21,509-33,963)	17,898 (14,564-21,963)	54,803 (47,830-63,164)
7dir	33,269 (23,669-46,020)	16,230 (12,961-20,714)	63,032 (52,103-82,189)
7ind	75,158 (56,201-98,417)	36,199 (27,083-49,460)	112,596 (88,138-158,566)
4dir	28,393 (19,626-43,342)	20,454 (15,772-26,249)	50,755 (41,954-62,773)
4ind	18,453 (13,959-23,895)	13,484 (11,040-16,296)	34,752 (30,209-40,214)

We can see that the violations of procedural invariance – order effects, payment period effects and SG 'chaining`effects – produce a very high volatility in MVQ estimates. These range from 18,000€ to 112,000€. All MVQ comparisons between groups for the same Type were statistically different at the 1% level. If we apply this result to the comparison between A-1 and A-2 this shows evidence of order effects. If we apply this result to the comparison between A-1 and A-3 this shows evidence of insensitivity to the duration of payment. Within each group, MVQ values for the indirect method (7ind, 4ind) were statistically significant at the 1% level from the rest of the MVQ estimates. If we focus on the estimates based on the direct method, the picture is mixed. In group A-1 MVQ for the smallest health gain (Type 7dir) is smaller and statistically different ($p=0.055$) from the MVQ of the largest health gain (Type 3). This suggests that MVQ varies with the change in QoL. However, the differences between Type 3 and 7dir for group 2 is not statistically significant ($p=0.262$), whereas for group A-3 there is a significant difference at the 10% level ($p=0.094$). Overall, there is a tendency for the MVQ to be higher the smaller the size of the health gain.

However, as a test of the sensitivity or otherwise of MVQ estimates to different changes in QoL, we should bear in mind that on the basis of the direct method, there is very little variability across the QoL differences. In direct utility terms, going from 22223 to 21212 is roughly the same as going from 21212 to 11111 – the two mean differences are 0.151 and 0.146 for A-1, 0.125 and 0.117 for A-2 and 0.133 and 0.139 for A-3. As it turned out, the QoL utility differences in these cases are so close that they do not provide a very searching test of the stability of MVQ estimates in the face of rather different QoL improvements.

Groups B-1 and B-2

These two groups provide further evidence about order effects and the impact of different QoL changes on MVQs, but they were also intended to shed light on the stability of MVQs with respect to different *durations* of benefit. Mean WTP and directly elicited health state utilities are reported in **¡Error! No se encuentra el origen de la referencia.**5. These convert into the MVQ estimates reported in the same table.

Table 5: Mean WTP per month (€) for treatment B, mean utilities and MVQ (€).Groups B-1 and B-2

Type	Mean WTP for treatment B		p	MVQ	
	B-1	B-2		B-1	B-2
1	156.93	78.3	*	20,158 (16,684-24,413)	8,881 (7,306-10,770)
2	102.19	44.95	*	26,253 (20,392-33,068)	10,197 (8,508-12,310)
5	111.58	51.39	*	30,574 (22,716-42,695)	10,685 (8,037-14,548)
6	69.46	27.99	*	38,068 (28,023-55,691)	11,639 (8,823-15,445)
	Mean utilities				
State 22223	0.720	0.683	ns		
State 21212	0.869	0.827	ns		

* Statistically significant differences between B-1 and B-2 at the 1% level using the t-test.

First, as we found with the A Groups, the order effects are all statistically significant and all work in the same direction as with the A Groups. On a within-Group basis, however, there appears at first sight to be some sensitivity of monthly WTP to the duration of the benefit: in comparisons between Types 1 and 2 and between Types 5 and 6, halving the duration of benefit is associated with noticeable reductions in both mean and median WTP responses. However, the reductions are generally less than proportional, so that the MVQs in Table 5 derived from the shorter durations are significantly higher ($p<0.01$) in three of the four comparisons (the exception being Types 5 vs 6 in B-2). Moreover, comparing Types 1 vs 5 and 2 vs 6 in both groups leads us to reject the null hypothesis that MVQs are invariant to the size of QoL gain. In the case of Type 1 vs 5, differences are statistically significant at the 1% level. In the case of Type 2 vs 6, differences were statistically significant for group B-1 at the 5% level ($p=0.049$) and at the 10% level for B-2 ($p=0.062$). In conclusion, there may have been some ambiguity on this issue in the A Group data, but here the picture is rather sharper: when MVQs are derived from larger health gains (22223 → 11111) as compared with smaller gains (21212 → 11111), the estimates are significantly smaller. The effect is more muted for B-2, but it is quite clear and strong for B-1.

Groups C-1 and C-2

These two groups were intended to shed light on the stability of MVQs with respect to different *risk reductions*. Mean WTP and directly elicited health state utilities are reported in **¡Error! No se encuentra el origen de la referencia.**6. These convert into the MVQ estimates reported in the same table.

Table 6: Mean WTP per month (€) for treatment B, mean utilities and MVQ (€). Groups C-1 and C-2

Types	Mean WTP for treatment B		p	MVQ	
	C-1	C-2		C-1	C-2
10	83.06	54.86	*	8,151 (6,481-10,525)	5,358 (4,421-6,639)
11	61.59	24.60	*	12,088 (9,280-15,288)	4,805 (3,797-5,970)
12	74.10	34.95	*	12,387 (9,522-17,127)	4,992 (4,000-6,317)
13	39.71	16.05	*	13,277 (9,725-18,025)	4,585 (3,647-5,794)
	Utilities				
State 22223	0.708	0.704	ns		
State 21212	0.830	0.797	ns		
	Undiscounted lifetime QALY gain(1)				
State 22223	12.23	12.29			
State 21212	7.18	8.40			

(1) This is calculated by taking the average number of years of remaining life expectancy, based on the age distribution of the subsample, and weighting this by the index for the QoL gain.
* Statistically significant differences between C-1 and C-2 at the 1% level using the t-test.

The results for the WTP and utilities translate into the MVQs shown in the right hand columns of Table 6. 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€. One hundred individuals would therefore pay 99,672€ to prevent, on average, one case of this chronic illness, thereby achieving an expected benefit of 12.23 QALYs, giving an MVQ of 8,151€, as shown in the appropriate cell in Table 6.

So once again, we observe significant order effects of the kind evident in all other Groups. However, in three out of four comparisons, our respondents appeared to be responsive to the differences in risk reduction. The comparisons between MVQ of Types 10 vs 11 and Types 12 vs 13 mostly showed no statistically significant differences – the only exception being the Type 10 vs 11 comparison in Group C-1, where mean WTP was only about 25% lower for Type 10, resulting in an MVQ which was significantly higher (p=0.000) for Type 11.

4. Second Survey: Design, Implementation and Results

Given the presence of order effects in our first survey, we tried to estimate MVQ for some of these health gains using a between-Group design. This gave rise to a second survey. The second study took the four most straightforward scenarios from the first survey (Types 1, 3, 5 and 7) and added two more (Types 8 and 9) in order to study the derivation of MVQ estimates from even smaller health gains. These two scenarios were always presented as the second question, so they were vulnerable to being influenced by the response to the first question. For this reason, the final estimate of the MVQ from these questions will be less "clean" than from the other four questions. However, while acknowledging this potential problem, we thought that it was interesting to have some information on the MVQ derived from health gains that were even smaller than those used in the first survey. And since these second questions always involved unambiguously smaller benefits, it should have been relatively straightforward to reduce payments proportionally, if that is what respondents had wished to do.

332 members of the general population were interviewed in person after being initially contacted by telephone (see Appendix 2 for characteristics of the sample). They were randomized between 4 subgroups. Utilities were elicited using the direct SG method. The framing of WTP questions was very similar to the first survey, except that the response format was polychotomous, involving five categories of responses (Definitely YES/NO, Probably YES/NO, Not Sure). Respondents were asked to allocate every amount to one or other of the five response categories. We 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".

The key results are summarized in Table 7.

Table 7: Mean WTP (€) for treatment B and Mean Utilities. N=83 per group

Types	WTP				MVQ
	Group D-1	Group D-2	Group D-3	Group D-4	
1	163.54	--------	------	-------	16,407 (12,378-22,081)
5	------	111.47	------	------	23,487 (16,296-32,049)
3	------	------	127.37	------	21,777 (16,305-26,831)
7	-------	------	------	105.38	37,860 (27,282-54,504)
8	96.65	------	86.98	-----	76,235 (63,140-92,336)
9	------	89.72	-----	58.88	123,724 (93,678-161,411)
	Utilities				
State 22223	.641	.701	.563	.675	
State 21212	.787	.829	.784	.793	

The results for the questions that were asked first in both surveys were quite similar. WTP for Type 1 (avoiding 22223 for 4 months) was 156.93€ in Group B-1 in the first survey and 163.54€ in the second survey (group D-1). WTP for Type 3 (avoiding 22223 for 2 months) was 112.32€ in Group A-1 and 127.37€ in this survey (group D-3). These differences were not statistically significant.

Comparing the first responses of Groups D-1 and D-3 relating to Types 1 and 3, there was *some* sensitivity to duration but it was not proportional. In the comparison between D-2 and D-4 relating to Types 5 and 7, there was no significant difference between means, despite one duration being half the other.

The within-Group comparisons between the 4-month or 2-month initial scenarios and the 2-week scenarios in the second questions all showed *some* degree of responsiveness – but again, the reductions in WTP were much less than proportional. Not surprisingly, therefore, the inferred MVQs show a very clear inverse relationship with duration. In the comparison between Types 1 vs 3, the difference is only just significantly different ($p=0.10$); but in all other cases, the differences were clearly significant (Types 1 vs 8, $p=0.00$; Types 3 vs 8, $p=0.00$; Types 5 vs 7, $p=0.02$; Type 5 vs 9, $p=0.00$; Type 7 vs 9, $p=0.00$). In all cases, the estimated MVQ was higher for shorter durations; and even though there was always a within-Group possibility of reducing the WTP for the 2-week duration, the MVQs derived from those questions are especially high.

We also see the consequences of insufficient adjustment of WTP to the smaller QoL gain involved in going from 21212 to 11111 as compared with the improvement from 22223 to 11111: the MVQs inferred from the smaller health gains are all significantly higher

(Types 1 vs 5, p=0.48; Types 3 vs 7, p=0.01 and 8 vs 9, p=0.00). When the insensitivity to duration is compounded with the insensitivity to QoL gain, the result (Type 9) is an estimated MVQ which is even bigger than the highest figure generated in the first survey as the product of payment period effects and SG chaining interacting with a smaller QoL gain of 2 months duration. Thus the results of the second survey reinforce concerns that the MVQ estimate is liable to be systematically influenced by the choice of which features of the scenario are to be varied, and by how much.

5 Conclusions

The main objective of this study was to test the robustness of various conditions that would need to hold if we are to obtain a reliable and reasonably all-purpose MVQ from questions involving different combinations of QoL, duration and risk.

Our results cast quite serious doubt on the possibility of obtaining such value. Besides finding insufficient sensitivity to the duration of health states and the size of QoL improvements, we also found clear and substantial violations of procedural invariance in the form of order effects, payment period effects and 'chaining effects'.

The presence of order effects is well documented in the literature on contingent valuation (Stewart, 2002). One reaction to this problem is to assume that the less biased response is the first one since it cannot be influenced by the previous one(s). Given that several of the MVQ estimated in the first survey could have been "contaminated" by these effects, we conducted a second survey. In this survey, we also observed the same insensitivity even when we compared the MVQ estimates using only the first question. Indeed, we observed even greater problems, since in some groups there was almost no discrimination between two health gains that were clearly different.

The second violation of procedural invariance was 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 MVQ could 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 MVQ. 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 et al., 2001). For example, Stumborg et al. (2001) found that total WTP for a public good was higher when the duration of payment was 10 years as opposed to when it was 3 years (unless a 40% discount rate were 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).

We also found that the particular method used (direct vs chained) to elicit health state utilities had a substantial impact on the MVQ computation. The indirect method raised

the estimate of mean utility of 21212 relative to the direct method, and thus had the effect of reducing the utility difference between 21212 and 11111 while at the same time increasing the difference between 21212 and 22223. Since these differences appear in the denominator of the calculation of MVQs, the effect is to increase (indeed, roughly double) the MVQ based on the Type 7 question while at the same time reducing (to about two-thirds) the MVQ based on the Type 4 question. As a consequence, the indirect method gives both the highest and the lowest MVQ estimates within all three Groups, with the former being in the region of 3 – 4 times as big as the latter.

The question then is which of the two estimates (if either) better reflects preferences. It might be argued that direct methods should be preferred on the grounds that any biases generated in the SG would be compounded by chaining – and certainly, the indirect method is associated with greater volatility of the MVQ estimates. However, it is not obvious that the indirect method is producing 'worse' utility indices for 21212 than the direct method. The indirect method suggests that respondents would be willing to take a mean risk of death of between 5% and 8% to avoid the 21212 chronic condition – figures which some may view as quite high, given the aversion many people express towards many much smaller real-world risks. However, those figures are a good deal less than the means produced by the direct method – between 12% and 15% – the latter being barely better than the odds in Russian Roulette. In order to decide which figure (if either) we should use for policy purposes we need a better descriptive theory that explains how people respond to these questions. Conventional wisdom assumes that people are Expected Utility (EU) maximisers; but the systematic disparities between direct and indirect estimates is incompatible with that assumption – as, indeed, is a large body of other evidence (see Starmer, 2000, for a review).

The various 'effects' listed above all suggest that responses and estimates are seriously vulnerable to procedural influences that in theory should not make a difference. At the same time, there appears to be insufficient sensitivity to features of scenarios (mainly duration and severity) that really ought to matter and that ought to be treated approximately proportionally if a single MVQ is to be widely applicable. We found that estimates of the MVQ vary systematically with the *severity* of the health state used to elicit it, with milder health states resulting in higher estimates. We also found the estimates varying systematically with the *duration* of the health state used to elicit it, with shorter durations resulting in higher estimates. Overall, the patterns resemble the picture to which we have become used in the context of WTP for (road) safety benefits, where, as Jones-Lee and Loomes (2001) have noted, there is on the one hand excessive sensitivity to theoretically irrelevant procedural features while at the same time there appears to be insufficient sensitivity to those variables that really ought to be influential.

Our results are in line with some other studies that have explored the relationship between health gains measured in QALYs and WTP. Van Houtven et al. (2003), expanding the Johnson et al. (1997) meta-analysis, find that WTP increases less than proportionately with changes in duration. Gyrd-Hansen (2003) finds that WTP is not a linear function of QALY values. Hammit and Haninger (2007) find that WTP is only 29% higher for avoiding 7 days vs 1 day of moderate health problems. Donaldson et al.

(2008) found that WTP to avoid 12 months of a health problem was less than 4 times the WTP to avoid 3 months of the same health problem. Overall, it seems that our results reinforce previous findings and they all seem to indicate fairly strong evidence that there may not be a linear relationship between WTP and QALYs even for changes that might be considered 'small'.

However, an alternative interpretation of these findings is that the seeming lack of sensitivity is largely a consequence of hitting budgetary restrictions: that is, it reflects constraints on people's ability to pay. This raises the problem of finding a change small enough to allow a linear approximation. Past research involving one of the authors (Pinto-Prades and Rodríguez, 2001) used even smaller health changes. For example, it elicited WTP to avoid 3 and 15 days in health state 11121 and 22222. Mean WTPs were 12.5€ (3 days, 11121), 34.5€ (15 days, 11121), 65.7€ (22222, 3 days) and 99.3€ (22222, 15 days). These means are in line with the kinds of responses used to estimate the Value of a Statistical Life when people are asked WTP questions for reductions in the risk of death[9] that are considered "marginal". However, in none of the above cases was WTP for the 15 day health gain even approximately 5 times bigger than WTP for 3 days. Of course, it is difficult to be sure that budget constraints have no impact at all: but in many of the studies listed in the previous paragraph, the sums of money rarely account for more than a few percent of people's annual incomes.

Of course, this does not mean that *our results* do not suffer from this problem. To try to explore this possibility within our data, we took our second survey and estimated the ratio between WTP for 2 weeks and 2/4 months for those with income above and below the median. If budget constraints are a major factor, we should find that this ratio is considerably bigger for those with higher incomes, since their ability to express the ratio of benefits is less constrained. In fact, the ratio is 1.94 for those below the median income and 2.20 for those above the median income, a difference that it is not statistically significant. So while we accept that the health gains used in the present study were bigger than some might regard as marginal, we doubt that this is the *main* reason behind our result that MVQ increases inversely with the QALY health gain.

More in line with conventional wisdom, we found only limited evidence that different risk reductions affect the MVQ estimates. This result 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 inversely and significantly with the size of risk reduction used in the questions asked (for a survey of such evidence, see Beattie et al., 1998).

In summary, we have found some effects that produce a wide range of MVQ estimates. These effects can be split into two different groups, namely: violations of the assumption that WTP is proportional to the health gain measured in QALYs; and violations of procedure invariance . The first type of effect might be ascribed to one or more of the

[9] For example, Tsuge et al. (2005) base their estimates of the VoSL in Japan on an average WTP of about 200€ for a risk reduction of 1 in 10,000 in the risk of death.

following: a) limitations of the WTP method (e.g. the impact – to an unkown degree – of budget constraints); b) limitations of the QALY model (e.g. people not evaluating health benefits as a linear function of duration); and/or c) departures from Expected Utility theory (whereas utilities elicited with the SG are usually based on the assumption that subjects are EU maximisers). The logical follow-up to this study would involve trying to disentangle the relative importance of the different possible factors. For example, we might try to estimate the MVQ using different theoretical assumptions. There is some evidence (Bleichrodt et al, 2000; Abellán-Perpiñan et al, 2007; Bleichrodt and Pinto, 2005) suggesting that preferences are better represented if we assume that subjects are not EU maximisers and if we use a non-linear QALY model. Perhaps the volatility of the MVQ may be reduced if we do not use such restrictive assumptions in order to estimate it. However, it would be more problematic if the volatility of the MVQ estimates is due to violations of procedural invariance, since these effects operate outside conventional economic theory. To deal with them we need to understand better the psychological processes that may give rise to such effects.

Overall, then, our study poses a number of quite serious challenges to the idea of finding a single all-purpose monetary value of a QALY. We readily acknowledge that we cannot disentangle all the potential reasons behind our results: indeed, given the multiplicity of 'effects' and (ir)regularites identified by our design, there is every reason to believe that it will require a considerably larger programme of future research to do that. What we *have* been able to do is draw attention to a set of issues that are fundamental to the enterprise of valuing health benefits in a way that makes it possible to compare them more directly with other possible public sector benefits and with the costs of providing them. This is a topic that should be high in the research agenda of health economists. Whether or not it eventually turns out to be possible to establish a means of converting QALYs into money amounts, the exploration of the issues seems likely to give us a better understanding of the ways in which people value health improvements, and will thereby provide a stronger platform for public policy in this area.

7. REFERENCES

Abellan-Perpinan JM, Pinto-Prades JL, Mendez-Martinez I, Badia-Llach X. 2006. Towards a better QALY model. Health Economics. 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.

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.

Bateman, I., Carson, R., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., Pearce, D., Sugden, R. and Swanson, J. 2002. Economic Valuation with Stated Preference Techniques: A Manual, Edward Elgar, Cheltenham.

Beattie, J., Chilton, S., Cookson, R., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Robinson, A., Spencer, A., 1998a. Valuing health and safety controls: A literature review. Health and Safety Executive, HSE books, London, 171.

Bleichrodt H, Quiggin J. 1999, Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis? Journal of Health Economics. 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., Wonderling D. E., and Mooney C. Z. 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. 2005. Willingness to pay per quality-adjusted life year in a study of knee osteoarthritis. Medical Decision Making. 25(6), 655-66.

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

Donaldson C, Baker R, Jones-Lee M, Loomes G, Mason H, Robinson A et a. Weighting and valuing quality adjusted life years: preliminary results from the Social Value of a QALY Project, 2008.

http://www.pcpoh.bham.ac.uk/publichealth/methodology/projects/RM03_JH12_CD.shtml

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. 2003. Willingness to pay for a QALY. Health Economics. 2003 12(12), 1049-60.

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

Hammitt, J.K., K. Haninger. 2007. "Willingness to Pay for Food Safety: Sensitivity to Duration and Severity of Illness." *American Journal of Agricultural Economics* **89**(5): 1170–1175.

Haninger, K., J.K. Hammitt. 2007. "Willingness to Pay for Quality-Adjusted Life Years: Empirical Inconsistency Between Cost-Effectiveness Analysis and Economic Welfare Theory." Working paper draft, Harvard Center for Risk Analysis.

Hirtz RA, Chernew ME, Miller E, Fendrick M, Weissert WG. 2000. Willingness to pay for a Quality Adjusted Life Year: in search for a standard. Medical Decision Making. 20, 332-342.

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

Johnson F.R., E.E. Fries, and H.S. Banzhaf, "Valuing morbidity: an integration of the willingness-to-pay and health-status index literatures," *Journal of Health Economics* 16: 641-665, 1997.

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.

Jones-Lee M and Loomes G, 2001 The Valuation of Health and Safety for Public Sector Decision Making, Chapter 2 in Economic Evaluation: From Theory to Practice, eds J. L. Pinto, G Lopez and Ortún Vl, Springer, Barcelona.

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

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

Mason H, Jones-Lee MW and Donaldson C. Modelling the monetary value of a QALY: a new approach based on UK data. *Health Economics* (forthcoming).

Pinto-Prades JL and Rodríguez R *¿Cuánto vale la pena gastarse para ganar un año de vida ajustado por calidad? Un estudio empírico* [How much is a QALY worth? An empirical study] in Puig J, Dalmau E y Pinto JL, El valor monetario de la salud [The monetary value of health]. Springer-Verlag, Barcelona, 2001, pp.51-81.

Stalmeier PF. 2002, Discrepancies between chained and classic utilities induced by anchoring with occasional adjustments. Medical Decision Making.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. 2002. Do ordering effects matter in willingness-to-pay studies of health care? Journal of Health Economics.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 Zethraeus N. 1998. Bootstrap confidence intervals for cost-effectiveness ratios: some simulation results. Health Economics, 7, 143–147.

Tsuge T, .Kishimoto A and Takeuchi K 2005. A Choice Experiment Approach to the Valuation of Mortality The Journal of Risk and Uncertainty, 31(1), 73–95.

Van Houtven G, Powers J, Jessup A, Yang JC. Valuing avoided morbidity using meta-regression analysis: what can health status measures and QALYs tell us about WTP? Health Econ. 2006 Aug;15(8):775-95.

Appendix 1

OPTION 1		
TAKE MEDICINE **X** FOR **1 YEAR**		
Health State ➢ **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**		
➢ GOOD HEALTH	2 MONTHS	**YES** ADDITIONAL EXPENDITURES

Appendix 2

Sociodemographic characteristics

Survey 1: General population = Catalonia (n=80 per Group)

All numbers are percentages

		Group A-1	Group A-2	Group A-3	Group B-1	Group B-2	Group C-1	Group C-2	General Population
	N	80	80	80	80	80	80	80	
Gender	Women	51.3	47.5	50.0	50.0	51.3	50.0	51.3	51
Educational Background	Primary or less	42.5	52.5	45.0	38.8	43.8	40.0	47.5	61.9
	Secondary	35.0	32.5	31.3	37.5	37.5	36.3	30.0	23.3
	University	22.5	15.0	23.8	23.8	18.8	23.8	22.5	14.7
Working status	Active	68.8	70.0	65.0	76.3	68.8	71.3	75.0	60.7
	Retired	17.5	17.5	20.0	18.8	20.0	17.5	18.8	18.7
	Other	13.8	12.5	15.0	5.0	11.3	11.3	6.3	20.6
Income	<600€/month	30.0	30.4	31.3	23.8	21.3	26.3	20.0	13.6
	600€ -1200€	41.3	43.0	36.3	55.0	54.7	40.0	40.0	26.8
	1200€ -1800€	21.3	21.5	27.5	11.3	20.0	22.5	30.0	17.7
	>1800€	7.5	5.1	5.0	10.0	4.0	11.3	10.0	21.1
Age	18 - 40	40.0	41.3	41.3	42.5	38.8	42.5	42.5	42.0
	41 - 60	35.0	35.0	33.8	32.5	32.5	33.8	35.0	31.5
	> 60	25.0	23.8	25.0	25.0	28.8	23.8	22.5	26.5

Survey 2: General population = Basque country (n=83 per Group)

		Group D-1	Group D-2	Group D-3	Group D-4	General Population
	N	83	83	83	83	
Gender	Women	52.2	47.0	48.2	46.0	51.0
Educational Background	Primary or less	36.1	34.9	36.1	28.9	41.3
	Vocational	20.5	25.3	14.5	25.3	24.3
	Secondary	21.6	18.1	26.5	22.9	11.1
	University	21.7	21.7	21.7	22.9	23.4
Working status	Active	65.1	69.9	68.7	60.2	56.5
	Retired	16.9	19.3	14.5	21.7	19.5
	Other	18.0	10.8	16.8	18.1	24.0
Income	<900€/month	20.5	12.0	19.3	18.1	19.6
	900€ -1500€	22.9	38.6	25.3	34.9	19.2
	1500€ -2100€	18.1	21.7	16.9	21.7	18.2
	>2100€	38.6	27.7	38.6	25.3	43.0
Age	18 - 40	45.8	47.0	39.8	41.0	40.0
	41 - 60	30.1	27.7	41.0	32.5	33.5
	> 60	24.1	25.3	19.3	26.5	26.5