WORKING PAPER SERIES

WORKING PAPER NO 6, 2006

Swedish Business School at Örebro University

The Value of a Statistical Life in Sweden

Estimates from Two Studies using the "Certainty Approach" Calibration

by

Mikael Svensson
mikael.svensson@oru.se
Swedish Business School
Örebro University
Sweden

http://www.oru.se/esi/wps

SE-701 82 Örebro
Sweden

ISSN 1403-0586

The Value of a Statistical Life in Sweden
Estimates from Two Studies using the "Certainty Approach" Calibration

Mikael Svensson
mikael.svensson@oru.se

Dept. of Economics, Swedish Business School, Örebro University, Sweden
& Dept. of Health and Environmental Sciences, Karlstad University, Sweden

THIS IS THE WORKING PAPER VERSION OF THE PAPER:
Svensson, M., (2009), "The value of a statistical life in Sweden: Estimates from two studies using the "Certainty Approach" calibration", *Accident Analysis & Prevention*, Vol 41(3): 430-437.

Abstract: Stated preference methods using surveys to elicit willingness to pay have been shown to suffer from hypothetical bias and scope/scale bias. Hypothetical bias usually means that willingness to pay is exaggerated in the hypothetical scenario and scope/scale bias means that there is an insensitivity in willingness to pay with regard to the amount of goods or the size of a good being valued. Experimental results in social psychology and economics have shown that only trusting the most certain respondents can potentially solve the problem with hypothetical bias and scope/scale bias. This paper presents the results of two different surveys in Sweden estimating the willingness to pay to reduce traffic mortality risks by only including the most certain respondents. Using the full sample, estimates of VOSL are $4.2 and $7.3 million. Estimates of VOSL on the subset of the samples only including the most certain respondents are lower and consistent between the two surveys with values of $2.9 and $3.1 million.

Keywords: Value of a Statistical Life; Contingent Valuation; Hypothetical Bias; Calibration; Certainty Approach;

JEL Code: D80; I18; Q51

1. Introduction

Cost-benefit analysis is widely used as a tool by policy-makers to evaluate proposed regulations and public investments. A policy with the purpose of reducing mortality risks requires a monetized value of safety in order to compare the benefits with the economic costs. The monetized benefit of reduced mortality risk is captured in the concept of value of a statistical life (VOSL). VOSL is the willingness to pay (WTP) for a small risk reduction for

each individual in society that overall is expected to prevent one premature death. We may illustrate VOSL using the following example: imagine a population of 1,000 individuals where everyone faces an annual mortality risk of 1/1,000, i.e. during a year there will be one death (everyone has the same risk of dying). If each individual on average is willing to pay $1,000 for an investment that will eliminate the mortality risk, the total willingness to pay for the population is $1 million, which is the population VOSL.[1]

Economists use revealed preference (RP) methods and stated preference (SP) methods to estimate VOSL. An RP method relies on using information from actual behavior to estimate an implicit VOSL. The most common RP approach is to use labor market data to estimate the wage premium demanded for accepting riskier jobs. A review of more than 60 such studies is found in Viscusi and Aldy (2003). Another RP approach is to study consumer behavior. One such example is to estimate the price premium for a safer car, and use the WTP to calculate an implicit VOSL (Andersson, 2005; Atkinson and Halvorsen, 1990).

The SP method instead uses surveys and experiments where respondents are asked to respond to a hypothetical risk reduction program/good. Using different payment mechanisms and designs, the willingness to pay may be measured based on the hypothetical choices made by the respondents. The most common SP approach in the literature on VOSL is contingent valuation (CV) studies, where respondents (usually) are asked a dichotomous choice (yes/no) question; whether or not they would be willing to pay a certain amount for a certain risk reduction. Generally, estimates based on SP studies render higher VOSL estimates (de Blaeij et al., 2003). The RP approach, which is based on actual behavior, is generally regarded to be more reliable (Sugden, 2005), but it may be argued that its use is dependent on the context of the public investment. This could lead policy makers and researchers to turn to SP methods,

[1] VOSL=WTP/Δrisk.

using surveys to directly ask respondents about their willingness to pay for a hypothetical risk reduction.

However, there is criticism of the SP approach, and the major problems are often argued to be (1) *hypothetical bias* and (2) *scope/scale bias*. *Hypothetical bias* implies that respondents will usually overstate their WTP in the hypothetical scenario compared to a real market situation, which implies that estimates of VOSL based on SP studies will be exaggerated; hence public investments will look more beneficial compared to their true benefit/cost ratio. *Scope/scale* bias refers to insensitivity to the amount of goods valued (scope) or the size of the good valued (scale). In a VOSL context scale bias implies that WTP does not increase with a larger risk reduction, and that VOSL will be very sensitive to the risk reduction used in the survey design.

However, recent advancements in the literature have been used to argue that these problems may be overcome by a "certainty calibration" approach, which has shown positive results when dealing with the problems of hypothetical bias and scope/scale bias. The approach involves asking the respondent about his/her level of certainty regarding the hypothetical answer and only trusting positive responses to the WTP question from the most certain respondents as "true" positive responses. In this paper the "certainty calibration" approach is applied to two recent Swedish surveys where conservative estimates of VOSL are based on the most certain respondents.[2] We test the impact on VOSL estimates from the "certainty calibration" approach and also test if it affects the scale bias problem.

The rest of the paper is structured as follows. In the following section methodological issues with the SP approach are discussed in more detail. The third section describes the CV studies and the data used. Section four shows WTP regressions and calculations of VOSL, while the last section concludes with a discussion.

[2] Some results from one of the surveys have previously been reported with a focus on the difference between private and public WTP for risk reductions (Hultkrantz et al., 2006).

2. Methodological Issues

There is a considerable literature on methodological issues in stated preference studies in general, and it is well acknowledged that different types of survey design may have large impacts on survey responses. For example, the payment question and payment mechanism may be important determinants of the answers received. On a theoretical level, Carson and Groves (2007) argue, based on economic theory, that the method that may be incentive compatible is the single dichotomous choice question (yes/no) with coercive payment for a new public good (or choice of which of two new public goods to provide) (Carson and Groves, 2007). Other mechanisms are not consequential, and answers will be less valid. Empirically, it has also been shown that different payment mechanisms in many instances produce different estimates, for instance open-ended questions generally produce lower estimates compared to dichotomous-choice questions (Cameron et al., 2002). Further, regarding estimates of WTP for a risk reduction it has been shown that whether or not the good is described as private or public will have a large impact on the estimate, with a private provision valued higher compared to public provision of the same good (Hultkrantz et al., 2006; Johannesson et al., 1996).

Many recent studies more explicitly focus on the problems and biases (and potential solutions) with estimates based on SP surveys. These problems tend to be common for most of the chosen elicitation procedures (even though the bias may be of different magnitude). Section 2.1 below extends this discussion, while section 2.2 discusses potential remedies for the problems of SP studies that have been discussed in the literature.

2.1 Biased Estimations?

There are many skeptical voices regarding the use of SP methods (Ariely et al., 2003; Carson et al., 2001; Diamond and Hausman, 1994), and in the context of valuing mortality risk

reductions the main problems discussed include hypothetical bias (section 2.1.1.) and scope/scale bias (section 2.1.2).

2.1.1 Hypothetical bias

Hypothetical bias refers to the fact that measures of willingness to pay (WTP) from a hypothetical scenario deviate from measures of WTP in a real market scenario (Hofler and List, 2004; Murphy et al., 2005a). Hypothetical bias is also usually considered to deviate positively from a real market situation, i.e. it is an excess of yes votes in a hypothetical referendum or purchase scenario compared to a real market situation. Harrison (2006) argues that hypothetical bias is the most serious problem to consider with the use of SP data. If researchers estimate VOSL based on SP data, and positive hypothetical bias arises, the use of such an estimate in cost-benefit analysis will exaggerate the consumer surplus from a typical public investment. This may imply that investments that decrease social welfare will be implemented due to the hypothetical bias.

Meta-analyses of experimental tests of hypothetical bias also indicate that the magnitude of bias is often large.[3] One meta-analysis of 39 papers testing for hypothetical bias reported a positive bias in 34 of the papers and a mean (median) hypothetical bias of 300 percent (67 percent) (Harrison and Rutstrom, 2008). Another extended meta-analysis of 59 studies, using the same elicitation mechanism, reported a mean hypothetical bias of 35 percent (Murphy et al., 2005a).

2.1.2 Scope/Scale Bias

Scope/Scale bias refers to the fact that respondents in surveys do not reflect any sensitivity in stated WTP to how many different goods are being valued, here defined as scope bias, or the

[3] To empirically test for hypothetical bias one simple approach is to let one group of respondents answer a hypothetical question regarding the purchase of a particular good, and let another group take part in a real purchase decision in a lab or field setting. WTP from these groups can be compared to detect possible differences between the hypothetical valuation and the real valuation.

amount of a certain good being valued, here defined as scale bias (Hammitt and Graham, 1999). Carson et al. (2001) examined the evidence on scope (and scale) bias and their major conclusion is that poor survey design and administration are the main causes of scope/scale insensitivity found in SP studies. However, they also state one exception to their results, which is the scale insensitivity found in risk reduction studies that rely on individuals stating their WTP for small changes in probabilities. The problem is that individuals have cognitive problems understanding and having preferences for very small changes in probabilities. According to economic theory the WTP for a reduction in the risk of a fatal accident should be close to proportional to the size of the risk reduction (Hammitt and Graham, 1999), and in a recent quasi-experimental study it has been shown that the near-proportionality is predominantly violated by respondents with lower cognitive capacity to understand small probabilities (Andersson and Svensson, 2008). Many SP studies have also included explicit tests for scale bias in VOSL estimations, and the results generally do not show enough sensitivity to the risk reduction (Beattie et al., 1998; Hammitt and Graham, 1999; Persson et al., 2001).

2.2 Certainty calibration

Considering the methodological criticism against SP studies, perhaps strongest when concerning VOSL studies, there is a growing literature on how to overcome the problems with mainly hypothetical bias and scope/scale bias. Regarding the problem of hypothetical bias, different types of calibrations have been proposed. *Cheap-talk calibration* is one approach that can be described as telling respondents in the survey about the hypothetical bias and asking them to carefully consider this in their own answers. The results regarding the success

of cheap-talk calibration are (at best) mixed (Blumenschein et al., 2008; Cummings and Taylor, 1999; Murphy et al., 2005b; Poe et al., 2002).[4]

Another proposed calibration approach is the *certainty calibration* approach, which simply implies including a follow-up question asking respondents how certain they are of their stated answer, e.g. on a scale between 1 and 10. There is a long tradition in social psychology where it has been shown in experiments that the certainty with which an individual holds a certain attitude increases the probability that this attitude will translate to behavior (Fazio and Zanna, 1978; Fishbein, 1963; Fuji and Gärling, 2003; Sample and Warland, 1973; Suchman, 1950).

The certainty calibration approach has been tested in several different SP applications, both in the lab and in the field, and results are promising. Several of the papers have shown that when interpreting an uncertain "yes-answer" as a "no-answer" to the WTP question, mean WTP does not differ between hypothetical groups compared to groups with real transactions of the same good (Blumenschein et al., 2008; Blumenschein et al., 1998; Blumenschein et al., 2001; Champ and Bishop, 2001; Champ et al., 1997; Poe et al., 2002; Vossler et al., 2003). Regarding applications to the road sector, Swärdh (2008) showed that only including the most certain respondents eliminated hypothetical bias in a study where a value of travel time was estimated. The drawback to these studies is that it is unclear what the threshold for defining an uncertain response should be. Some studies have asked the respondents if they are "probably sure" or "definitely sure" of their answer, and interpreted the latter as "true" yes-answers. Other papers have asked respondents to evaluate their certainty on a 0-10 or 1-10 scale, where it has been found that the hypothetical bias is eliminated (compared to a real transaction group) when using a threshold of 7, 8, 9 or 10.

[4] The different results from using cheap-talk calibration could potentially be explained by the fact that the cheap-talk script can take different forms, be of different length and use different wordings to the respondents.

This method has also provided some hope regarding the problem of scope/scale bias. Some evidence that scale bias may be overcome is provided by Alberini et al. (2004), studying VOSL for mortality risk reductions. They found that respondents who stated a high confidence in their answers had a median WTP that increased in proportion to the size of the risk reduction. Similar results were also found in an earlier study, such that respondents with a high confidence in their own answers gave answers almost in line with theoretical predictions for the scale sensitivity (Hammitt and Graham, 1999).

3. Data

3.1 Survey A

The first data for the paper comes from a CV study conducted in the Swedish city of Örebro in 2004. Örebro is located approximately 200 kilometers west of the capital, Stockholm, and has an urban population of 97 000. The starting point for the survey was the Swedish "vision zero" policy drafted by the Swedish Parliament in 1997. It bears a close resemblance to safety goals in e.g. air traffic and safety policies for e.g. nuclear plants. The idea is to construct roads to prevent accidents (fatal and severe injuries) from occurring, but also that when they do occur, due to human nature, they should not result in a severe outcome. Road safety investments aimed at reaching these targets include the use of cable-guard rails, crash cushions, tree removal, speed limits etc. The idea of this "vision" is that the only acceptable goal is zero fatalities and severe injuries due to traffic accidents, i.e. far from an economic cost-benefit perspective.

Urban Örebro is a particularly good location to conduct a CV study using the "vision zero" as a reference point. The "vision zero" policy has been demonstrated along a major walking and biking route in Örebro to show in practice how "vision zero" can be achieved by protecting these road users.

The survey contained the statement that there are four fatal accidents and twelve severe injuries annually in the urban traffic area of Örebro (sixteen persons as a "package"). The accident numbers were thus presented as frequencies, rather than as probabilities, which is easier for the respondents to understand. The public safety good was described as: *"A road traffic safety program that will reduce the number of fatal and severe injuries within the urban area of Örebro by 16 persons during one year. The reduction applies to pedestrians, bicyclists and car users. Outside the urban area the number of road accidents will be unaffected"*.

The good was framed using the "community analogy" concept, where the effect was only to be seen on the local level, and could increase the feeling of participation among the respondents (Kalman and Royston, 1997).[5]

Respondents were consequently told that this would imply a certain cost, and they were asked if they would be willing to pay to a public fund to have this program implemented (a yes/no question). Six different bid levels were used in the survey (SEK 200; 1 000; 2 000; 5 000; 10 000 and 20 000).[6] Since the respondents in reference to the "vision zero" valued a safety package of reductions in both fatal outcomes and severe injuries, measures of VOSL and the value of a severe statistical injury (VSSI) were subsequently estimated with the use of death risk equivalents (DRE), further described in section 4.3. After the WTP question a follow-up question to elicit the preference certainty of each respondent was asked. The question was: *"How certain are you of your answer above, on a scale between 1 and 10, where 1 is very uncertain and 10 is very certain? Please indicate your answer on the line*

[5] The survey was split up into sub samples and *4/5* of the surveys presented the public safety good described in this paper. *1/5* of the surveys presented a private safety good that would eliminate the risks of severe outcomes for the particular user, but for reasons of comparison with Survey B these data will not be used in this paper.
[6] A provision point mechanism was used, which stated that at least 70% of the population had to contribute (qualified majority), otherwise the project would not be implemented (and all money refunded). The payment mechanism/design was identical in Survey B (with the same provision point mechanism etc.).

below". 1,148 mail questionnaires were sent to individuals aged 18 to 75 out of a population of 97 000. After one reminder the response rate was 61 percent.

3.2 Survey B

The second dataset used in the paper comes from a CV survey conducted in the Swedish city of Karlstad in 2006, with a population of 83 000. Karlstad is situated approximately 110 kilometers west of Örebro, in the region Värmland bordering Norway. In this study as well respondents were asked about their WTP for traffic safety improvements that were meant to take place in the local urban environment, hence using the community analogy concept. Respondents were given descriptions of a scenario where 6 traffic deaths take place each year in the urban area of Karlstad (based on mean deaths between 1998 and 2005), as well as a public traffic investment that would reduce fatal accidents by half. Hence, in probabilities the risk reduction corresponded to 3/83 000. The safety investment was described as: *"…To reach this safety improvement a public traffic safety program will be implemented. The safety program applies to all fatal outcomes where at least one vehicle is involved, which implies that the risk reduction applies to car drivers/passengers, motorists, bicyclists, and pedestrians"*. Respondents were then asked whether they would pay for this safety program to be implemented in Karlstad at a given cost (yes/no question). Five different bid levels were used (SEK 200, 500, 1 000, 2 000 and 5 000).

The certainty question, following the dichotomous choice WTP question, was identical to the question posed in Survey A. The complete survey included 1 000 mail questionnaires sent to individuals aged 18 to 75 out of a population of 83 000. After one reminder the overall response rate was 53 percent.

4. Results

4.1 Descriptive Statistics

The dependent variable in the analysis in this paper is *Answer,* which is a binary dummy variable equal to one if the respondent answered yes to the WTP-question. 20 percent of the respondents answered positively to the question in Survey A and 34 percent in Survey B. The latter was theoretically expected to be higher considering that higher bid levels were used in Survey A compared to Survey B.

Explanatory variables used in the analysis are listed in Table 1. It is evident that mean age is almost identical in the two surveys. A large difference can be seen for *Income,* which is 22 240 SEK in Survey A and 36 990 SEK in Survey B. This can be explained by the fact that the latter refers to gross income and the former to net income. Adjusting for the Swedish tax rate, mean income is quite similar in the two surveys.

Table 1
List of Variables and Summary Statistics

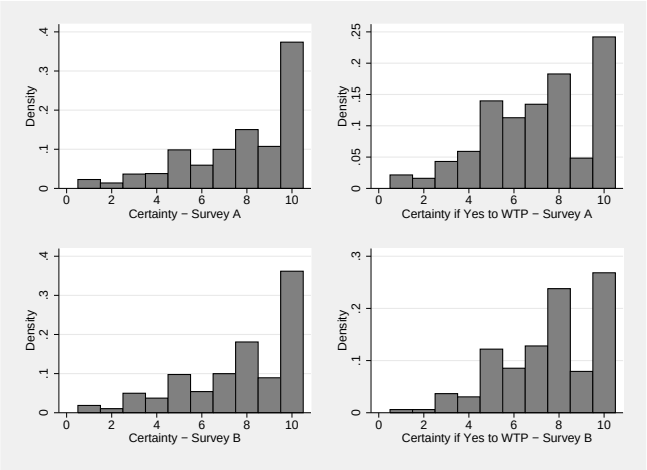
Variable	Description	Mean (std. dev.) Survey A	Mean (std. dev.) Survey B
Dependent variable			
Answer	=1 if bid accepted	0.20 (0.40)	0.34 (0.47)
Independent variables			
Age	Age of the respondent	44.45 (15.96)	44.70 (14.65)
Income	Income in 1,000 SEK	20.39 (11.85)	36.99 (19.42)
Female	=1 if woman	0.52 (0.50)	0.54 (0.50)
Kids	=1 if having under-age children in the household	0.32 (0.47)	0.33 (0.47)
High risk beliefs	=1 if believes own mortality risk higher compared to objective risk	0.07 (0.26)	0.14 (0.35)
Low risk beliefs	=1 if believes own mortality risk lower compared to objective risk	0.25 (0.44)	0.29 (0.46)
Certainty	Stated certainty on scale 1-10	7.85 (2.39)	7.72 (2.41)

Notes: Survey A asked about household disposable income while Survey B asked about gross income. Note that the higher mean of *Answer* is expected given the lower bid levels used in Survey B.

In both surveys it was more common to believe that one's own mortality risk was lower, and not higher, compared to the average mortality risk. This is a common result in the literature. In e.g. Andersson (2007) it was found that Swedish citizens on average under assess their own (traffic) mortality risk.

The preference certainty variable (*Certainty*), evaluated on a 1-10 scale, has a mean of 7.85 in Survey A and 7.72 in Survey B. Since the interest of the paper lies in the effect of stated certainty on VOSL, a more detailed view of stated certainty is shown in Figure 1 below.

Figure 1
Histogram of Certainty



In both surveys the most common response was clearly to state a certainty of 10 (the most certain response). Figure 1 contains histograms of certainty based only on respondents answering yes to the WTP question. To state a certainty of 10 was also the most common response among respondents answering "yes", but certainty levels between 5 and 8 were more common compared to the full sample. In the full sample almost 40 percent stated a 10, and among "yes-respondents" close to 25 percent stated a 10.

Finally in this section, Figure 2 shows predicted probabilities of answering yes to the WTP-question (*Answer*) by level of stated certainty (holding constant other independent variables in Table 1).

Figure 2
Predicted probabilities of Yes/No answer by level of certainty

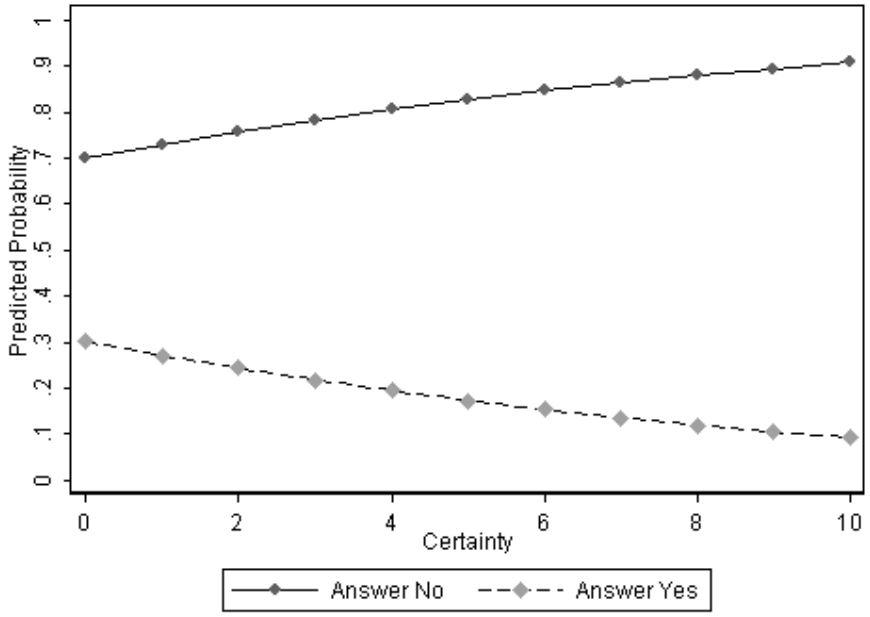


Figure 2 indicates that as the respondent become more certain of the answer to the WTP-question, the probability of answering yes decreased. This could also be interpreted from Figure 1. Hence, we already see indications that more certain respondents were less likely to answer yes, and more likely to answer no.

4.2 Willingness to pay regressions

As outlined in section 4.1 the dependent variable is a binary dummy variable taking the value one if the respondent answered in favor of the safety investment (*Answer*). Table 2 shows marginal effects from logit estimations.

As can be seen the bid level is negative and statistically significant, i.e. higher price is associated with a lower probability of wanting to pay for the investment. Regarding other variables, very few turn out statistically significant, and they show inconsistency between the two surveys. The income variable is positive (statistically significant in Survey A), also according to theoretical expectations. Respondents who have children below the age of 18

living in the household show some (positive) association to wanting to pay for the safety investment.

Table 2
Marginal effects from Logit WTP Regression (standard errors in brackets), dep. var. Answer

	Survey A		Survey B	
	Coeff.	Std. err.	Coeff.	Std. err.
Bid	-2.25E-05***	(0.00)	-8.68E-05***	(2.00E-05)
Age 31-40	0.01***	(0.04)	-0.14**	(0.07)
Age 41-50	-1.97E-03	(0.04)	-0.10	(0.07)
Age 51-60	0.06	(0.05)	0.03	(0.07)
Age 61-75	-0.05	(0.04)	0.08	(0.08)
Income	3.20E-03***	(1.12E-03)	2.22E-04	(1.23E-03)
Female	0.01	(0.03)	-0.01	(0.05)
Kids	0.04	(0.04)	0.17**	(0.06)
High risk beliefs	-0.04	(0.04)	0.06	(0.07)
Low risk beliefs	0.01	(0.03)	-0.02	(0.05)
Certainty	-0.02***	(0.01)	-0.01	(0.01)
N	539		472	
LR-Chi^2	88.93		43.50	
Pseudo-R^2	0.17		0.07	

Notes: *** p<0.01, ** p<0.05, * p<0.10.

The certainty, included as a continuous variable, is negatively related to the dependent variable, i.e. higher certainty implies lower likelihood of answering yes to the WTP-question. As we shall see in the next section, the effect is large for those with the highest certainty. The age categories do not show any consistent pattern between the two surveys. This is not particularly surprising, considering that much of the literature on the VOSL-age pattern largely come to different conclusions, see e.g. Evans and Smith (2006).

4.3 VOSL in survey A & B

To estimate the WTP needed to calculate VOSL in the two surveys, the probability (π) that a respondent answers "yes" to the WTP question is estimated with the logistic model:

$$\pi = \frac{1}{1+e^{-\Delta v}}, \tag{1}$$

where Δv is the change in utility from the public safety good. The mean WTP ($\overline{p}$) is defined as the area under the survivor function for WTP≥0:

$$\overline{p} = \int_0^{\infty} 1/\left(1+e-\Delta v\right) dw = -\frac{1}{\beta}\left[\ln\left(1+e^{\beta x}\right)\right]\cdot \tag{2}$$

Equation (3) outlines how to estimate the mean WTP using a bivariate model:

$$\overline{p} = (-1/\beta_{bid}) \cdot \log\left(1+\exp(\beta_{\text{constant}})\right). \quad (3)$$

I will begin by estimating mean WTP and VOSL including all respondents in the two surveys. *Survey A* asked respondents about their WTP for an elimination in both fatal and severe injuries. Here I am interested in the WTP for the fatal accidents, and to estimate that I make use of death rate equivalents (DREs)[7]:

$$VSL = \frac{WTP}{DRE \cdot \Delta p^{severe} + \Delta p^{fatal}}. \quad (4)$$

DREs are used to express the WTP for different outcomes in a common metric. An estimate of DRE is needed to translate the value of a statistical severe injury (VSSI) reduction into a value of fatality reduction (VOSL). This is done by taking the ratio VSSI/VOSL (Hultkrantz et al., 2006; Viscusi et al., 1991). Swedish official estimates suggest a DRE of approximately 0.15 (Vägverket, 2006), which will be used in this paper.[8]

If the more certain respondents have less positive hypothetical bias, mean WTP should, ceteris paribus, be lower for the most certain respondents. Two different thresholds of stated certainty have been suggested in the literature in attempts to eliminate hypothetical bias. On a 1-10 scale using respondents that state a certainty of 8 and above or only respondents that state a 10 has primarily been suggested. In this section we will explore both thresholds and their effect on VOSL.

Reiterating that the total risk reduction was higher in Survey A, we should expect a higher mean WTP. In Table 3 we see that for the full sample mean WTP is similar in both Survey A and B. This could be interpreted as no sensitivity to the different scale of the proposed goods in Survey A and B. WTP based on respondents with a certainty of 8 and

[7] In equation (4) Δp^{severe} is the risk reduction for a severe injury and Δp^{fatal} is the risk reduction for a fatal outcome due to the safety investment.

[8] The assumption needed here is that WTP(fatal and severe injury reduction) = WTP(fatal reduction) + WTP(severe injury reduction).

above produces significantly lower estimates in both surveys, but more so in Survey A than B. WTP based on only the most certain respondents gives a somewhat different picture. In Survey A WTP is almost the same for the most certain respondents as for respondents with a certainty of 8 and above. In Survey B mean WTP decreases significantly when only including the most certain respondents. The threshold where the lowest estimates are found seems to be at certainty level 10 in Survey B but at certainty level 8 in Survey A.[9]

Table 3
WTP and VOSL estimates in million Swedish kronor (US million in parentheses) in Survey A & B

	All respondents	Certainty ≥ 8	Certainty=10
WTP Survey A	1,774	1,202	1,274
WTP Survey B	1,808	1,430	712
VOSL Survey A	29.43 (4.28)	19.94 (2.90)	21.14 (3.08)
VOSL Survey B	50.00 (7.28)	39.57 (5.76)	19.70 (2.87)
RATIO (VOSL B / VOSL A)	1.70	1.98	0.93

Notes: Values in Swedish Kronor and in 2006 price level. VOSL from Survey A adjusted for price differences between 2004 and 2006. VOSL estimates are in million Swedish kronor. VOSL estimates in million US dollars in parentheses.

Hence, the conclusion is that using only the most certain respondents, the theoretical predictions of (near-proportionality) in WTP for different risk reductions cannot be rejected, i.e. the usual problem with scale bias (scale insensitivity) is eliminated using the most certain respondents. This also implies that VOSL estimates in the two studies are (more or less) equalized.[10] The ratio between the surveys indicates that only including the most certain respondents gives estimates that differ by less than 10%, compared to the full sample where the difference is up to 70%.[11]

[9] In Survey A the certainty threshold 8 is also the lowest threshold that gives valid and consistent estimates; for a more detailed analysis of all the different thresholds (from 1 to 10) only in Survey A see Hultkrantz et al. (2006).
[10] VOSL is calculated as WTP/Δ risk. The problem with scale insensitivity found in most studies on VOSL then implies that given a constant WTP, a lower risk reduction will give higher VOSL. Hence, if WTP does not change (near-proportional) with the Δ risk, VOSL will always depend on the chosen risk reduction in the survey.
[11] Another potential approach is to treat the certainty as a weight (giving a stated certainty of 10, ten times more weight compared to a stated certainty of 1). This approach works less well if there are important "threshold effects" in the certainty scale. Analyzing the data in this paper with this approach gives VSL estimates of 26 milion SEK in Survey A and 43 million SEK in Survey B. Hence, this is lower compared to the main results based on all respondents, but it does not give the same scale sensitive between the two surveys as the certainty approach used in Table 3 and 4.

To indicate the statistical uncertainty of the results Table 4 shows mean estimates (as in Table 3) as well as 95 percent confidence intervals including all respondents and for respondents with a certainty of 10.

Table 4
VOSL in million Swedish kronor with 95% confidence intervals (US million in parentheses)

	All respondents		Certainty=10	
	Estimate	95% CI	Estimate	95% CI
VOSL Survey A	29.43	20.98 – 42.76	21.14	10.52 – 29.1
	(4.28)	(3.05 – 6.22)	(3.08)	(1.53 – 4.24)
VOSL Survey B	50.00	36.68 – 72.46	19.70	12.48 – 30.24
	(7.28)	(5.34 – 10.55)	(2.87)	(1.82 – 4.40)

Notes: Confidence intervals estimated using the bootstrap approach with 1,000 replications of original data.

Comparing estimates only including the most certain respondents, VOSL from Survey A and B are obviously not statistically significantly different from each other. However, considering that the confidence intervals are relatively large, we cannot say that the estimates in the full sample are statistically significantly different from each other at the 95%-level. When performing confidence intervals at the 90%-level, estimates in the full sample are statistically significantly different (not reported here). For Survey B the estimate only including the most certain respondents is a lot smaller and statistically significantly smaller at the 95%-level compared to the estimate from the full sample.

5. Discussion and Conclusions

This paper reports estimates of the VOSL based on two recent surveys in Sweden, using different risk reductions and bid levels.[12] Using the full survey sample the two VOSL estimates are $4.2 million and $7.3 million, a difference of approximately 70 percent. In a next step VOSL estimates are calculated using respondents who stated a certainty of 8 and above as well as those who stated a certainty equal to 10, on a scale between 1-10. This is

[12] The two cities where the surveys were conducted are quite representative for the Swedish population in general. They are clearly larger than the average municipality, which has a population of 15,000. However, regarding income and level of education they are close to the national average (SCB, 2008). In national elections for parliament the municipality of Karlstad is sometimes referred to as the most “typical” or ”average” Swedish municipality.

based on results from social psychological theory and empirical applications in economics that have shown that hypothetical bias can be reduced if trusting only certain respondents. This gives VOSL estimates of $2.9 million and $3.1 million. Hence, it also produces estimates in two completely different surveys, with different risk reductions and bid levels, which are very close to each other. This satisfies the important scale sensitivity criteria, i.e. that the two surveys indicate scale sensitivity according to economic predictions. Further, using the most certain respondents implies a lower VOSL in both surveys, which is in accordance with the expectations of reducing hypothetical bias.

The approach of using the certainty of respondents in pursuit of calibrating for hypothetical bias seems to be able to produce consistent estimates. However, it should also be recognized that nothing necessarily makes these values correct, even if they are very close to each other. Harrison (2006) discusses this line of evidence as: *"…The only claim is that they all might give comparable hypothetical numbers or bounds on the hypothetical WTP."* (Harrison, 2006, p.135). This is a valid point. That two different surveys elicit similar estimates of VOSL is not a necessary and sufficient condition for these estimates to be correct, but it is a necessary condition (Viscusi, 1998). Estimates of VOSL reported in this survey based on all respondents are so different that it becomes hard to defend those using economic arguments.

It should also be noted that the approach here is slightly different compared to recent successful experimental results, where uncertain respondents have all been recoded as answering "no" to the WTP question (Blumenschein et al., 2008). This is understandable in experiments with a very small sample. In this paper I rather focus the estimations on only the certain respondents, hence excluding uncertain respondents. This is more in line with general theoretical arguments for using the certainty calibration, and it is possible here with the larger sample compared to small experiments. It should again be noted that the same qualitative

results are confirmed when recoding uncertain respondents rather than excluding them (estimates in the two surveys converging). The approach in this paper is, however, identical to a recent study that showed an elimination of hypothetical bias compared to a real market test for a value of travel time experiment (Swärdh, 2008).

Finally, further analysis of the determinants of stated certainty indicates that there is a relationship between age and certainty. This may cause some concern, given that using the certainty calibration approach in this case slightly raise the mean age in the sample that VOSL estimates are calculated upon.[13] However, under the assumption that a higher certainty implies less hypothetical bias, the implication is that the elderly have less, or no, hypothetical bias. This result corresponds to previous results showing that to be an efficient utility maximizer induces an effort, which decreases with experience, and the elderly are hence more rational. Weinstein (1968) showed this in a seminal paper where violations of transitivity decrease with age, which was confirmed in later experiments as well (Bradbury and Ross, 1990). To be able to give answers in a hypothetical survey that are not flawed by hypothetical bias, opportunity cost and budget constraint issues need to be carefully considered, something that might come with experience and age. With increased age respondents may also be more knowledgeable and experienced regarding driving, traffic safety etc. A further indication that the calibration worked, and that the more certain (the elderly) have less hypothetical bias is that stating a high certainty is positively associated with having personal experience of traffic accidents. There are indications in the literature that experience of the scenario is important for developing more well-behaved preferences for the good (Bateman et al., 1994). The results of the relationship between age and certainty can potentially have a big impact on the growing literature on the VOSL-age relationship which is one of the most important current research questions within the VOSL-field, especially considering that it is most relevant for

[13] This would create a problem if we believe that this would have an impact on the "true" willingness to pay, e.g. VOSL may be lower among the 50-year olds compared to the 40-year olds. This would underestimate VOSL for the population. However, the "jury is still out" regarding the VOSL-age relationship (Evans and Smith, 2006).

environmental regulations that usually have health benefits for the elderly (Johansson and Kriström, 2006). It has been suggested that age-differentiated measures of VOSL should be used considering that the expected remaining life years for fatalities in road accidents is around 40-45 life years, while it is significantly lower for many other public health investments. To address this concern Alberini et al. (2006b) suggest that: "*The appropriate way to answer the (first) question is to ask people of various ages – including elderly persons – to report their willingness to pay (WTP) for a reduction in their risk of dying"* (Alberini et al., 2006, p.252). If the elderly state a lower WTP, this would then be an argument for using lower estimates of VOSL for e.g. pollution reduction policies and policies aimed at reducing falling accidents. A recent study has found weak evidence that the VOSL for individuals aged over 70 is 20-30 percent less compared to the mean of other individuals (Alberini et al., 2004). However, further research is needed to examine whether the lower VOSL among older respondents is only a result of the elderly having lower hypothetical bias. The conclusion from the indications mentioned here is that reported VOSL-age relationships suggested in the literature so far should be treated with caution in view of this potential bias.[14]

To sum up, the results in this paper constitute one piece of evidence, together with other recent studies as previously discussed, showing that there is a way forward using hypothetical surveys to yield theoretically valid and consistent estimates of willingness to pay. However, more research is needed particularly regarding which type of certainty question works best and how to put the certainty calibration approach into a theoretical (economic) framework.

[14] It should, however, also be noted that it is by no means necessary that the certainty-age relationship holds for other contexts than road safety. And, there is also a very valid discussion of whether age-differentiated VOSL should be based on individual WTP at different ages or on other "fairness" approaches or where individuals are instead asked to respond to questions as "social planners".

Acknowledgements I would like to thank two anonymous referees, Magnus Johannesson, Peter Martinsson, Gunnar Lindberg, Lars Hultkrantz, Henrik Andersson, Björn Sund and seminar participants at VTI, Örebro University and Karlstad University for comments on earlier versions of this paper. Financial support from the Swedish Rescue Services Agency is gratefully acknowledged.

References

Alberini, A., Cropper, M., Krupnick, A., Simon, N., 2004. Does the value of a statistical life vary with age and health status? Evidence from the US and Canada. Journal of Environmental Economics and Management. 48, 769-792.

Alberini, A., Hunt, A., Markandya, A., 2006. Willingness to Pay to Reduce Mortality Risks: Evidence from a Three-Country Contingent Valuation Study. Environmental and Resource Economics. 33, 251-264.

Andersson, H., 2005. The Value of Safety as Revealed in the Swedish Car Market: An Application of the Hedonic Pricing Approach. Journal of Risk and Uncertainty. 30, 211-239.

Andersson, H., 2007. Willingness to pay for road safety and estimates of the risk of death: Evidence from a Swedish contingent valuation study. Accident Analysis & Prevention. 39, 853-865.

Andersson, H., Svensson, M., 2008. Cognitive Ability and Scale Bias in the Contingent Valuation Method. Environmental and Resource Economics. 39, 481-495.

Ariely, D., Loewenstein, G., Prelec, D., 2003. Coherent Arbitrariness: Stable Demand Curves without Stable Preferences. Quarterly Journal of Economics. 118, 73-105.

Atkinson, S.E., Halvorsen, R., 1990. The Valuation of Risks to Life: Evidence from the Market for Automobiles. Review of Economics and Statistics. 72, 133-136.

Bateman, I., Willis, K., Garrod, G., 1994. Consistency between Contingent Valuation Estimates: A Comparison of Two Studies of UK National Parks. Regional Studies. 28, 457-474.

Beattie, J., Covey, J., Dolan, P., Hopkins, L., Jones-Lee, M. et al., 1998. On the Contingent of Safety and the Safety of Contingent Valuation: Part 1 - Caveat Investigator. Journal of Risk and Uncertainty. 17, 5-25.

Blumenschein, K., Blomquist, G.C., Johannesson, M., Horn, N., Freeman, P., 2008. Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment. Economic Journal. 118, 114-137.

Blumenschein, K., Johannesson, M., Blomquist, G.C., Liljas, B., O'Conor, R.M., 1998. Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation. Southern Economic Journal. 65, 169-177.

Blumenschein, K., Johannesson, M., Yokoyama, K.K., Freeman, P., 2001. Hypothetical versus real willingness to pay in the health care sector: results from a field experiment. Journal of Health Economics. 20, 441-457.

Bradbury, H., Ross, K., 1990. The effects of novelty and choice materials on the intransitivity of preferences of children and adults. Annals of Operations Research. 23, 141-159.

Cameron, T.A., Poe, G.L., Ethier, R.G., Schulze, W.D., 2002. Alternative Non-Market Value-Elicitation Methods: Are the Underlying Preferences the Same? Journal of Environmental Economics and Management. 44, 391-425.

Carson, R.T., Flores, N.E., Meade, N.F., 2001. Contingent Valuation: Controversies and Evidence Environmental and Resource Economics. 19, 173-210.

Carson, R.T., Groves, T., 2007. Incentive and informational properties of preference questions. Environmental and Resource Economics. 37, 181-210.

Champ, P.A., Bishop, R.C., 2001. Donation Payment Mechanism and Contingent Valuation: An Empirical Study of Hypothetical Bias. Environmental and Resource Economics. 19, 383-402.

Champ, P.A., Bishop, R.C., Brown, T.C., McCollum, D.W., 1997. Using Donation Mechanisms to Value Nonuse Benefits from Public Goods. Journal of Environmental Economics and Management. 33, 51-62.

Cummings, R.G., Taylor, L.O., 1999. Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method American Economic Review. 89, 649-665.

de Blaeij, A., Florax, R.J.G.M., Rietvald, P., Verhoef, E., 2003. The value of statistical life in road safety: a meta-analysis. Accident Analysis & Prevention. 35, 973-986.

Diamond, P.A., Hausman, J.A., 1994. Contingent Valuation: Is Some Number Better than No Number? Journal of Economic Perspectives. 8, 45-64.

Evans, M.F., Smith, V.K., 2006. Do we really understand the age-VSL relationship? Resource and Energy Economics. 28, 242-261.

Fazio, R.H., Zanna, M.P., 1978. On the predictive validity of attitudes: The roles of direct experience and confidence. Journal of Personality. 46, 228-243.

Fishbein, M., 1963. An Investigation of the Relationships between Beliefs about an Object and the Attitude toward that Object Human Relations. 16, 241-248.

Fuji, S., Gärling, T., 2003. Application of attitude theory for improved predictive accuracy of stated preference methods in travel demand analysis. Transportation Research Part A. 37, 389-402.

Hammitt, J.K., Graham, J.D., 1999. Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability? Journal of Risk and Uncertainty. 18, 33-62.

Harrison, G.W., 2006. Experimental Evidence on Alternative Environmental Valuation Methods. Environmental and Resource Economics. 34, 125-162.

Harrison, G.W., Rutstrom, E.E., 2008. Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods. In: Handbook in Experimental Economics Results. Elsevier Science, New York, pp. 752-767.

Hofler, R.A., List, J.A., 2004. Valuation on the Frontier: Calibrating Actual and Hypothetical Statements of Value. American Journal of Agricultural Economics. 86, 213-221.

Hultkrantz, L., Lindberg, G., Andersson, C., 2006. The value of improved road safety. Journal of Risk and Uncertainty. 32, 151-170.

Johannesson, M., Johansson, P.-O., O'Conor, R.M., 1996. The Value of Private Safety Versus the Value of Public Safety. Journal of Risk and Uncertainty. 13, 263-275.

Johansson, P.-O., Kriström, B., 2006. Introduction. Resource and Energy Economics. 28, 193-198.

Kalman, K.C., Royston, G., 1997. Risk language and dialects. British Medical Journal. 315, 939-942.

Murphy, J., Allen, G.P., Stevens, T.H., Weatherhead, D., 2005a. A Meta-analysis of Hypothetical Bias in Stated Preference Valuation Environmental and Resource Economics. 30, 313-325.

Murphy, J., Stevens, T.H., Weatherhead, D., 2005b. Is Cheap Talk Effective at Eliminating Hypothetical Bias in a Provision Point Mechanism? . Environmental and Resource Economics. 30, 327-343.

Persson, U., Norinder, A., Hjalte, K., Gralén, K., 2001. The Value of a Statistical Life in Transport: Findings from a New Contingent Valuation Study in Sweden Journal of Risk and Uncertainty. 23, 121-134.

Poe, G.L., Clark, J.E., Rondeau, D., Schulze, W.D., 2002. Provision Point Mechanisms and Field Validity Tests of Contingent Valuation Environmental and Resource Economics. 23, 105-131.

Sample, J., Warland, R., 1973. Attitudes and the prediction of behavior. Social Forces. 51, 292-304.

SCB, 2008. Regionalräkenskaper. http://www.scb.se/templates/Product____11078.asp.

Suchman, E., 1950. The intensity component on attitude and opinion research. In: Measurement and Prediction. Princeton University Press, Princeton.

Sugden, R., 2005. Anomalies and Stated Preference Techniques: A Framework for a Discussion of Coping Strategies. Environmental and Resource Economics. 32, 1-12.

Swärdh, J.-E., 2008. Hypothetical bias and certainty calibration in a value of time experiment. Working Paper 2008:7, Swedish National Road & Transport Research Institute, http://swopec.hhs.se/vtiwps/abs/vtiwps2008_2007.htm.

Weinstein, A.A., 1968. Transitivity of Preference: A Comparison among Age Groups. Journal of Political Economy. 76, 307-311.

Viscusi, K., Aldy, J.E., 2003. The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World Journal of Risk and Uncertainty. 27, 5-76.

Viscusi, K., Wesley, A.M., Huber, J., 1991. Pricing Environmental Health Risks: Survey Assessments of Risk-Risk and Risk-Dollar Trade-Offs for Chronic Bronchitis. Journal of Environmental Economics and Management. 21, 32-51.

Viscusi, K.W., 1998, Rational risk policy. Oxford University Press, Oxford.

Vossler, C.A., Ethier, R.G., Poe, G.L., Welsh, M.P., 2003. Payment Certainty in Discrete Choice Contingent Valuation Responses: Results from a Field Validity Test. Southern Economic Journal. 69, 886-902.

Vägverket, 2006. Vägverkets samhällsekonomiska kalkylvärden, Publication 2006:127.