WORKING PAPER SERIES

WORKING PAPER NO 1, 2009

Swedish Business School at Örebro University

Certainty calibration in contingent valuation

- exploring the within-difference between dichotomous choice and open-ended answers as a certainty measure

Björn Sund

Swedish Business School, Örebro University, Sweden

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

SE-701 82 Örebro
SWEDEN

ISSN 1403-0586

Certainty calibration in contingent valuation

- exploring the within-difference between dichotomous choice and open-ended answers as a certainty measure

Björn Sund[1]
Swedish Business School, Örebro University, Sweden

Abstract: Hypothetical bias is a serious problem of stated preference techniques. The certainty approach calibrates answers by assessing different weights to remedy respondents' valuation. However, very little research has been done to find a link between economic theory and empirical treatment of uncertainty through certainty calibration. We use a combination of dichotomous choice (DC) followed by an open-ended (OE) question to examine the relation between the degree of confidence and the distance between the DC bid and the OE answer. The results show that the OE bid difference is significantly correlated to the certainty level in one of our two contingent valuation (CV) surveys, with the probability of stating the highest confidence value increasing between 5-19 percent per SEK 1000 (~$170/€106) that the answer to the OE question and the bid differ. The second CV survey shows a significant relation for the no-responders.

Keywords: Contingent valuation; Hypothetical bias; Calibration; Certainty approach; Value of a statistical life; Traffic safety; Cardiac arrest

JEL Code: H43, I18, R41

Acknowledgements

I would like to thank Daniela Andrén, Lars Hultkrantz, Thomas Laitila, Lena Nerhagen, Mikael Svensson, Tore Söderqvist and seminar participants at Örebro University for helpful comments and Sandra Wallberg and Debbie Axlid for research assistance. Financial support from the Swedish Rescue Services Agency is gratefully acknowledged.

[1] Corresponding address:
bjorn.sund@oru.se
Örebro University, Swedish Business School, SE – 702 82 Örebro, Sweden
Tel: +46 (0) 31-786 52 49

1. Introduction

As a technique to value non-market goods or services the contingent valuation (CV) method has been widely used, but not unilaterally accepted. The technique is exposed to considerable criticism regarding the ability to measure individual preferences that are consistent with economic theory. A summary of the most important evidence against and in favour of the method is presented in Carson et al. (2001).

Answering hypothetical questions in a CV study may be difficult and there is a risk that the respondent answers in a way that reflects attitude more than real commitment. This hypothetical bias is found to be a serious problem (e.g. Harrison & Rutström, 2005; Murphy et al., 2005). Harrison (2006) argues that ‘assessment of the extent of hypothetical bias is, without doubt, the most important area of application in the field of environmental valuation.’ Several meta-analyses confirm that CV often overstates real economic values by as much as 135 to 300 percent (List & Gallet, 2001; Little & Berrens, 2004; Harrison & Rutström, 2005; Murphy et al., 2005). Instead of simply dismissing the method, researchers are now searching for a way to eliminate or adjust for this bias. So, is there a way?

One of the most promising alternatives is the certainty approach, where it is assumed that the more confident the respondents are about their attitudes, the more their answers are a good predictor of actual behaviour. In this paper we study data from two CV surveys using a discrete-continuous CV format, where both dichotomous choice (DC) and open-ended (OE) questions are asked to the same sample of respondents. The combination is used to examine the relation between the degree of confidence and the distance between the DC bid and the answer to the OE question (from here on referred to as the ‘gap variable’). Our hypothesis is that the larger the gap variable, the higher the certainty level. The purpose is to shed some light on the determinants of the certainty level and not just accept that some empirical adjustments of respondent WTP seem to give results closer to ‘real’ WTP than others.

It is increasingly found that incorporating respondent uncertainty has a potential to improve the predictive power of CV data (e.g. Champ et al., 1997; Blumenschein et al., 1998, 2001, 2008; Johannesson et al., 1999; Champ & Bishop, 2001; Poe et al., 2002; Vossler et al. 2003), but the causes of respondents’ uncertainty and its implications for valuation are yet not well known (Murphy & Stevens, 2004). Different empirical treatments to account for uncertainty have been suggested (see e.g. Shaikh et al., 2007). One example is to use a follow-up certainty scale of 1 (very uncertain) to 10 (very certain) and recoding all yes-responses to no-responses if the respondents were not completely certain (i.e.<10). The no-responses remain unchanged. This empirical approach is quite common, but can we be sure that respondents stating a certainty level of 10 really are more certain than someone stating a 9? The certainty scale can be perceived differently between individuals and also, is it correct to recode someone who states a 9 on a certainty scale (almost as certain as one can be) from a yes to a no, which is to totally reverse a respondents' answer?

As long as the causes of uncertainty are not understood it may be hard to agree on an overall empirical solution. It may well be the case that there is no single solution but instead a number of different courses of action. That individuals know their preferences with certainty is a common economic assumption and in this paper we try to explain the link between stated WTP, real WTP and a follow-up certainty question in a CV survey. Without knowing the real economic value of a good or service it is important to know how to account for uncertainty in deriving welfare measures and perhaps to develop better means for eliciting preference uncertainty.

The setting in our two CV surveys is valuing mortality risk reductions for road traffic safety and sudden out-of-hospital cardiac arrest (OHCA). Valuation of statistical lives (VSL) in road traffic is a common objective in valuation surveys (e.g. de Blaeij et al., 2003), and due to the media attention, large public expenditure on accident prevention and about 400-500 deaths per year in Sweden (link: Swedish Road Administration), it is a quite interesting field. OHCA is a condition with a low probability of survival, often below 5 percent, and is one of the most frequent causes of mortality in the Western world (Hollenberg, 2008). VSL in this setting has not yet, to our knowledge, been explored and it is also interesting to study because of the large number of incidents (about 5000-10 000 persons suffer an OHCA annually in Sweden), the age pattern (median age is about 70 years) and the possibilities to improve survival by increasing the density of defibrillators (ibid.).

The results from our first CV survey (road traffic safety) clearly show that the gap variable is correlated to the certainty level, with the probability of stating a 10 (highest confidence value) increasing by 5-19 percent per SEK 1000 increase in the gap variable.[2] Our second CV survey (out-of-hospital cardiac arrest) shows a

[2] Exchange rates: $1 = SEK 5.89, €1 = SEK 9.40 (2008-04-16).

significant relation for the no-responders: the probability of stating a 10 increases by 4 percent per SEK 1000, while there is no clear relation for yes-responders. In both surveys, the predicted probabilities of stating a certainty level lower than 10 decreases the higher the gap variable gets. By assuming that the OE answer is closer to ‘real’ WTP than the DC bid, this result mainly strengthens the theoretical argument of the certainty approach, i.e. the higher the confidence of the responders the more we can trust that stated WTP is correlated to actual WTP. However, we do question the recoding of yes-responses of less than totally certainty (i.e. <10) to no-answers.

We also find interesting differences in the value of a statistical life (VSL) regarding road traffic safety and cardiac arrest. VSL in the former case is SEK 10 million, while it is SEK 41 million for cardiac arrest. Restricting the sample to only include the most confident responders (=10) gives VSL estimates of SEK 9 million and SEK 53 million respectively. The estimates are very sensitive to different distributional assumptions, but the fact that VSL for cardiac arrest is higher than that for road traffic seems to be robust. This is surprising because statistical lives on average are both longer and ‘healthier’ for road traffic.

The next section describes the literature on correcting for hypothetical bias in stated preference methods with a focus on calibration of responses. It also includes a sub-section on differences between dichotomous choice and open-ended questions. Section 3 presents our research hypothesis, CV surveys and estimation methods, and is followed by the results in Section 4. A discussion in Section 5 concludes the paper.

2. Previous literature

Preference uncertainty can be introduced in several ways into the CV model. Let’s assume that individuals have a true value (y_i) for the risk reduction, but do not know this with perfect certainty. Then the respondent will arrive at some value $\bar{y}_i = y_i + \varepsilon_i$, where ε_i is a stochastic disturbance term arising from uncertainty (Li & Mattsson, 1995). If $\bar{y}_i \geq t_i$, where t_i is the bid level, then the individual would say yes to the offered risk reduction at this cost. This model opens up for the possibility that an individual answers yes (no), even though the true value is below (above) the bid depending on the sign and size of ε_i. Certainty calibration of responses may help improve the estimation and accuracy of CV analysis (Li & Mattsson, 1995; Ready et al., 1995).

2.1. Certainty calibration of responses

Harrison (2006) separates the procedures to correct for hypothetical bias into instrumental and statistical calibration. We will focus only on instrumental calibration and one of the most important innovations within this field: the certainty approach.[3] The certainty approach can be said to be applied ex post (Hofler & List, 2004; Blumenschein et al., 2008) as a follow-up on a respondent's answer in an attempt to find out whether he/she really would pay the stated amount. Generally, two versions of the certainty approach have been used (Blumenschein et al., 2008): The first assesses respondents' hypothetical WTP certainty based on a follow-up question with two degrees of certainty, i.e. 'probably sure' and 'definitely sure'. In the second version a numerical scale is used i.e. a 1-10 scale from 'very uncertain' to 'very certain'.

In a series of laboratory and field experiments, Blumenschein et al. (1998, 2001, 2008) divided the WTP responses into two degrees of certainty ('probably sure' and 'definitively sure'). Only the 'definitely sure' yes-responses were treated as yes-responses, while the 'probably sure' yes-responses were treated as no-responses. No treatment was carried out with the no-responses. All three studies show a close correspondence between 'definitely sure' yes-responses and real yes-responses, indicating that this can be an effective method to eliminate hypothetical bias. Blumenschein et al. (2008) also tested the effect of a cheap talk script to remove hypothetical bias, but did not find it to be effective.

The numerical version of the certainty approach, a 1-10 or a 0-10 scale, has shown similar results as the 'definitely/probably sure' version (Champ et al., 1997; Johannesson et al., 1999; Champ & Bishop, 2001; Poe et al., 2002; Vossler et al., 2003). When only treating very sure yes-responses as real yes-responses, no significant difference from real WTP values was detected. The question one has to consider in this version of the certainty approach is how to treat the numerical assessment of uncertainty. If we choose to use a cut-off level of certainty, then where

[3] Another important instrumental calibration approach is 'cheap talk' introduced by Cummings & Taylor (1999). 'Cheap talk' is applied ex ante and aims at removing bias through better study design and implementation by including an explicit discussion about hypothetical bias.

is it large enough, i.e. at 5, at 8 or at 10? Johannesson et al. (1999) estimated a calibration function that takes into account both the degree of certainty and the price level. After calibration, there was no statistical difference between hypothetical and real WTP responses.

Shaikh et al. (2007) compared six different empirical treatments of incorporating uncertainty in CV into the traditional random utility model (RUM) with assumed certainty. The treatments were: (1) a weighted likelihood function model (Li & Mattsson, 1995), (2) an asymmetric uncertainty model (Champ et al., 1997), (3) a symmetric uncertainty model (Loomis & Ekstrand, 1998), (4) a random valuation model (Wang, 1997), (5) multiple-bounded discrete choice (Welsh & Poe, 1998; Alberini et al., 2003) and (6) a fuzzy model (van Kooten et al., 2001). They concluded that empirical treatments can potentially increase goodness of fit, but also could introduce additional variance.

In a field test, Samnaliev et al. (2006) compared the 10-point certainty scale to an alternative where respondents had the opportunity to choose a 'not sure' option (recoded as no-answers). Generally, the two approaches were found to produce different results where the certainty scale reduced WTP by half and the 'not sure' option did not reduce WTP at all. In the latter case 'not sure' seems to represent no-responses at high bid levels, while at low bid levels 'not sure' represents both yes- and no-responses.

Which of the certainty approach versions performs best? According to Blumenschein et al. (2001) there was no clear statistical difference between the 'probably/definitely sure' calibration and the calibration function from Johannesson et al. (1999), but the former performed better in terms of magnitude. However, the calibration function seems to work better than earlier 'probably/definitely sure' approaches by Johannesson et al. (1998) and Blumenschein et al. (1998).

Svensson (2009) estimated the value of a statistical life (VSL) of two Swedish CV surveys and found that when using only the most certain responders the road traffic safety VSL values were very close.[4] It was also found that age is a significant determinant of certainty, with older respondents expressing higher confidence in their answers. It may therefore be the case that lower VSL values

[4] Including all respondents, the VSL estimates were SEK 29 million and SEK 50 million. Using only the most certain responders (certainty=10), the VSL estimates were SEK 21 million and SEK 20 million respectively.

among older respondents are not due to age per se but to less hypothetical bias. The only other study that we know of, that investigates the determinants of the certainty levels is Wang (1997), who assumed that individuals' preferences are uncertain and concluded that uncertainty is expected to be large for bids close to real WTP and small for bids decidedly smaller or larger than real WTP.

2.2 Differences between dichotomous choice and open-ended questions

The reason why we choose between these two question formats is that we have a situation where we expect that respondents truly do not know their valuation of a good or service. Without uncertainty, we would use an open-ended (OE) question format to elicit the value. In general, answers to OE questions include more information than dichotomous choice (DC) answers without revealing the cost level (avoids yea/nay-saying and anchoring). On the other hand, DC questions show more resemblance to real market transactions and are incentive compatible. The National Oceanic and Atmospheric Administration (NOAA) panel (Arrow et al., 1993) concluded that DC questions were preferred in contingent valuation surveys. In most studies DC yields higher WTP estimates than OE.

The explanation for this observation may fall into two major groups: (1) differences in how respondents perceive the various valuation formats and (2) variations in the statistical efficiency and robustness of the WTP estimates (Halvorsen & Sælensminde, 1998). According to their study, the main source of difference between WTP estimates from DC and OE questions is the latter, i.e. violation of the assumptions made about the random utility function. They used two valuation studies that both applied a discrete-continuous CVM format. The DC estimates turned out to be very unstable and the assumption about a homoscedastic distribution was crucial. Any correlation between the error terms and the cost lead to severe biases for the DC estimations. They argued that the conclusion of the NOAA panel, that DC is preferred to OE, depends only on how respondents perceive the elicitation formats and that NOAA's conclusion therefore can be modified.

Kealy & Turner (1993) used a discrete-continuous CVM format for both public and private goods and developed a test to investigate whether the different mechanisms lead to significantly different results. For the public good the DC/OE ratio was between 1.4 and 2.5, depending on the specification, leading them to conclude that individuals indeed respond differently to DC and OE. For the private good they did not find any differences in WTP estimates. Balistreri et al. (2001) compared laboratory experimental data from more than 800 individuals on DC and OE contingent values with actual auction values. The valued good was an insurance policy (private good) and they found that DC leads to overestimated values. While OE (if not trimmed for outliers) did too, it approximated the auction values better than DC.

Kriström (1993) compared the results from DC and OE in a non-parametric way and also examined the anchoring effect. He split his sample into two parts: sub-sample A, which received both a DC and an OE question (discrete-continuous CV format), and sub-sample B, which received only the OE question. By using a chi-square test, Kriström rejected the hypothesis that DC and OE answers are generated from the same distribution and therefore suggested that the elicitation formats are perceived differently. On the other hand, the DC bids did not seem to have any effect on the distribution of the OE answers in the two sub-samples. Therefore, the hypothesis of anchoring was not supported.

3. Materials and method

3.1 Research hypothesis

The general WTP elicitation setups of both CV surveys are similar. The initial dichotomous choice bid is followed by a question asking the respondents to indicate their degree of certainty about their WTP response on a 1-10 scale rated from 'very uncertain' to 'very certain'. Respondents were then asked an open-ended WTP question, also followed by a certainty question (answered on a 1-10 scale). The second certainty question is mainly included to signal that there are two separate WTP questions and not one WTP question followed by two indicators that measure the strength of the attitude to the first one. In summary the elicitation format looks like this:

1. A scenario to elicit dichotomous choice WTP for reduced risk in the case of traffic safety or cardiac arrest
2. A certainty approach question (1-10 scale)
3. A follow-up open-ended question to elicit maximum WTP
4. A certainty approach question (1-10 scale)

The point of this set-up is to see whether the differences between questions 1 and 3 correspond with the confidence expressed in question 2. Our research hypothesis is that*: Individuals with large differences between the DC bid and the OE response (our gap variable) also show high confidence in their answers.* For example: Suppose that a respondent faces a dichotomous choice of SEK 200 in question 1. If he/she answers yes to this bid level it gives no further information of the confidence in the answer or how far from this bid level his/her 'real' WTP may be. The only thing we are sure of is that the respondent's *hypothetical WTP is at least* SEK 200, i.e. $y_i + \varepsilon_i \geq 200$. This is why the three follow-up questions are used. If the respondent answers SEK 200 or 600 to the open-ended question (see Table 1), would it not be more likely that the yes-answer to the dichotomous choice question was more confident if the answer were SEK 600 than if it were SEK 200? By comparing this difference with the stated certainty level in question 2, the relation between them can be examined.

[Insert Table 1 here]

Given that individuals have a true value (y_i) for the risk reduction and estimate this value at $\bar{y}_i = y_i + \varepsilon_i$, we would like to minimise the probability that respondents might answer yes (no) even though the true value is actually less (greater) than the bid. We assume that the OE answer is closer to actual WTP than the DC bid and that the error term is homoscedastic. If we find a correlation between the gap variable and the certainty level, then we can trust that stated WTP is a better predictor for actual WTP the higher the certainty level (m). The conditional probabilities for yes and no responses are, respectively:

$$\Pr(\bar{y}_i \geq t_i \mid y_i \geq t_i \,\&\, m = k) > \Pr(\bar{y}_i \geq t_i \mid y_i \geq t_i \,\&\, m = k-1)$$
$$\Pr(\bar{y}_i < t_i \mid y_i < t_i \,\&\, m = k) > \Pr(\bar{y}_i < t_i \mid y_i < t_i \,\&\, m = k-1)$$

k = 2,…, 10

where t_i is the bid level for individual i.

3.2 Survey administration and structure

We use data from two CV surveys conducted in 2006 and 2007. Survey A is about willingness to pay to reduce the risk of fatal traffic accidents and is representative for a Swedish middle-sized city, while survey B deals with willingness to pay to reduce the risk of dying from out-of-hospital cardiac arrest and is representative for Sweden as a whole. Both surveys begin with an introduction explaining the aims of the study, some facts about traffic safety/cardiac arrest in general as well as local circumstances, and the random sampling procedure (explains how the respondent were chosen). Then some socio-economic characteristics questions are asked, including a question eliciting the individual baseline risk compared to the average inhabitant. Following this section the valuation scenario and WTP questions begins (see Appendix Table A4-A5).

3.2.1 Survey A (traffic safety)

Our first survey was randomly distributed among residents aged 18-75 in the municipality of Karlstad. Karlstad is located in the west of Sweden, with a population of approximately 82 000 (Statistics Sweden). The first questionnaires were sent out in November 2006, and a reminder was mailed three weeks later. The total number of questionnaires was 1000, split into a main sample and a scope test sample (see Table 2).[5] In total 517 surveys were returned, so the overall response rate was 53 %.

[Insert Table 2 here]

We used a public traffic safety programme to present the hypothetical risk reduction scenario. The programme was not explicitly specified in terms of what kinds of measures that may be taken, so as to avoid eliciting WTP for attitudes towards certain measures. However, the programme was said not to influence quality of travelling, average speed in traffic, environment or possibility to choose means of conveyance. A provision condition was used in order to minimise strategic bias, and reminders of the respondents' budget restrictions were also included.

Communicating small changes in low-level risks is very difficult and that is why we used a 'community analogy' (Calman & Royston, 1997). The key

[5] 16 questionnaires were returned because the address was wrong. No compensation was made for this, so the actual total number was 984.

phrase in the valuation scenario was: "*The road traffic safety programme will reduce your own and others' risk in Karlstad by half, meaning that the number of fatal traffic injuries will be reduced from 6 to 3 on average per year.*" Directly after this phrase the WTP elicitation questions begun. A test of scope sensitivity was included in the sample, where the number of fatal traffic injuries was reduced from 6 to 1 instead of from 6 to 3. Only one bid level (SEK 1000) was used in this case.[6]

3.2.2 Survey B (cardiac arrest)

The second CV survey was sent out in June 2007 and a reminder was mailed in September of the same year. We sent 1400 questionnaires to residents aged 18-75 in Sweden and the sample was split into a main sample and a scope test sample.[7] Also, two different communication aids to present the risk reduction were used and this is pooled in the data. Table 3 presents the sub-samples. The bid levels were determined by a pilot survey from a sample of 100 individuals in May 2007. The overall response rate in survey B was 43 %.[8]

[Insert Table 3 here]

A public programme to increase the survival rate after out-of-hospital cardiac arrests was valued. The programme contained an increased density of defibrillators in the municipality. Defibrillation was explained to be initiated by firemen, policemen, security guards or nurses, and public access defibrillators may be located in hotels, shopping malls, sports centres or theatres. The willingness to pay for an increased survival rate was elicited and the key phrase was: '*The programme will reduce your own and others' risk* [of dying from cardiac arrest] *and the survival rate will be increased from 5 to 10 percent on average*'. A scope sensitivity test assumed

---

[6] There exists some terminological confusion in this field, where scope/scale bias, embedding and part-whole bias are often used synonymously. We adopt the general distinction of Goldberg & Roosen (2007), following Carson & Mitchell (1995), that *scope insensitivity* 'is present when respondents do not sensitively react to the extent of improvements in a single risk to consumer safety but value the risk reduction in general'. In our view, this can be interchangeably used with 'scale bias'. *Embedding* 'refers to the phenomenon that consumers do not respond adequately to health risk reductions for different diseases or symptoms.'

[7] The population in Sweden was 9 166 604 in September 2007 (Statistics Sweden).

[8] 590 questionnaires were returned. 21 addresses were wrong, so the total sample was actually 1379. The response rates of the two surveys are close to the rates in similar surveys recently made in Sweden, suggesting that a local mail survey coming from the local university attains a 10 percentage unit higher response rate than a national survey.

an increase to 15 instead of 10 percent, and two bid levels were used in that case. Also, a provision condition was provided in the scenario.

3.3 Estimation method

We use the following parametric estimation methods to calculate our results: (1) mean/median WTP – exponential WTP constant-only bid function with normally distributed error terms (lognormal model), (2) variations in estimated WTP – fully parameterised lognormal models, and (3) variations in certainty level – ordered probit model.[9] The reason why the lognormal model was chosen was because it restricts WTP to be positive (>0) and it results in the highest value on the likelihood function ('best fit'). The model restricts WTP to be non-negative by using an exponential WTP function:

$$WTP_k = \exp(\beta x_k + \varepsilon_k) \qquad \varepsilon \sim \mathrm{N}(0, \sigma^2),$$

where x_k is a vector of covariates, β is the corresponding parameter vector and ε_k is an error term. The probability of accepting a certain bid (t_k) is then:

$$P[Yes] = 1 - \Phi\left(\frac{\ln t_k - \beta x_k}{\sigma}\right) = 1 - \Phi(\lambda \ln t_k - \beta^* x_k), \qquad \lambda = 1/\sigma,\ \beta^* = \beta/\sigma .$$

$\Phi(x)$ is the standard normal cumulative distribution for ε and we calculate median and mean WTP in the following way:[10]

$$Median\ WTP = \exp(\beta x_k) = \exp\left(\frac{\beta^* x_k}{\lambda}\right),$$

$$Mean\ WTP = \exp(\beta x_k)\exp\left(\frac{1}{2}\sigma^2\right) = \exp\left(\frac{\beta^* x_k}{\lambda}\right)\exp\left(\frac{1}{2\lambda^2}\right).$$

In the Appendix, the explanatory variables are described in Table A1 and their means and standard deviations are presented in Table A2. Finally, when analysing variations in certainty level we use the ordinal probit model, where the

[9] All data analyses are made in Stata/SE 9.1.

[10] For a constant-only bid function, median WTP is equal to $\exp(-\beta_{constant}/\beta_{logbid})$ and mean WTP is equal to $\exp(0.5\times(1/\beta_{logbid})^2-\beta_{constant}/\beta_{logbid})$. Mean and median WTP for continuous data (OE) are calculated by taking the logs of WTP, performing the calculations of mean/median and then transforming the results back to the original scale.

structural function is the same as for probit and the error term is assumed to follow a standard normal distribution. We have a threshold function with several cut points ($\mu_1, \mu_2, \ldots, \mu_9$) to be estimated:

$$y_i = \begin{cases} 1 \; if \; -\infty \le y_i^* \le \mu_1 \\ 2 \; if \; \mu_1 < y_i^* \le \mu_2 \\ \ldots \\ 10 \; if \; \mu_9 < y_i^* \le \infty \end{cases},$$

and the predicted probability of observing $y_i = m$ (a specific certainty level, m=1, 2,…, 10) for given values of x's is:

$$\Pr(y_i = m \mid x_i) = \Phi(\mu_m - \beta' x_i) - \Phi(\mu_{m-1} - \beta' x_i).$$

4. Results

This section starts with results from the discrete-continuous WTP distributions by treating them as separate samples, calculating mean/median WTP and comparing the effects of explanatory variables. Then, a deeper analysis will be made about the within-difference between the DC and OE answers (the gap variable). The intention is to measure the relative difference between DC and OE answers to see whether it is a relevant factor for explaining the level of confidence.

4.1 DC/OE answers and variations in WTP

Table 4 shows the survival functions of the bid levels. As expected, the acceptance level falls as the bid levels increase. The point estimates of OE responses follow the same pattern, although generally below the corresponding DC level. The scope-test samples, representing a reduction in the annual traffic mortalities of 5 instead of 3 (survey A) and an increase in the survival rate to 15 instead of 10 percent (survey B), indicate opposite results.[11,12] For survey A, an increase in the frequency of yes-responses from 22.1 to 27.8 percent (DC) is in line with the expectations of an increasing frequency. However, in a statistical test the scope parameter is not significant, neither for DC or OE data (Table A3, Appendix). For survey B, we see an

[11] We only test for 'weak' scope sensitivity, i.e. the WTP for a risk reduction increases in the amount of risk reduction (Goldberg & Roosen, 2007).

[12] A limitation in our data is that the scope sensitivity is only tested at one (survey A) and two (survey B) bid levels.

unexpected pattern in the DC frequency levels in the sense that the scope levels actually are *lower* than for the main sample, but the parameter estimate is not significant in this case either. Sund (2009) further analyses the sensitivity to scope in survey B.

[Insert Table 4 here]

If we want to treat DC and OE data as separate samples, can we statistically prove that they are not generated from the same distribution? Kriström (1993) used a simple chi-square test in which the difference between expected (E_t) and observed (O_t) number of no-answerers in each group are evaluated. OE answers are regarded as the expected number and DC answers as the observed number.[13] Since we have five bid levels, the relevant statistic is:

$$\sum_{t=1}^{5}\left[(O_t - E_t)^2 / E_t\right].$$

The values of the statistic are 48.15 (survey A) and 91.42 (survey B). Comparing these values with a chi-square distribution with four degrees of freedom at the 95% level (critical value = 9.49) rejects the hypothesis that OE and DC answers are generated from the same distribution for both surveys. As Kriström (1993) also notes, there may exist an in-sample bias where respondents answer the DC and OE questions 'consistently'. In this case the rejection of the hypothesis is even stronger. The conclusion is that DC and OE results are to be treated as two separate samples in the further analysis.

This difference can also be seen when calculating the mean and median WTP for both distributions. Table 5 shows mean and median WTPs in groups with different levels of confidence. For all levels, the mean WTP values are higher for DC than for OE.[14] The mean WTP is larger than the median WTP for DC, indicating a positively skewed distribution. The median is often chosen instead of the mean of the distribution, since the latter is very sensitive to outliers in the data and to distributional

[13] The scope samples are excluded in this test.
[14] This result, that VSL(DC)>VSL(OE), is found in most comparative studies (Halvorsen & Sælensminde, 1998).

assumptions. In survey A, the estimated WTP for DC generally decreases the more confident the respondents are, but the same pattern cannot be seen in survey B.[15]

[Insert Table 5 here]

The results of the lognormal models with covariates estimated from DC and OE answers are presented in the Appendix (Table A3). We find that the bid level influences WTP with a decreasing probability of stating yes the larger the bid is (for DC). For OE, the parameter estimate is significantly positive. For survey B we find that estimated WTP is negatively affected when: (1) own perceived risk of OHCA is lower than average, (2) the number of inhabitants in the respondents' municipality is large (DC), and (3) the education level is low (OE). Interestingly, age and age^2 are also significantly negative/positive for survey B, which implies a U-shaped relation between age and WTP.

4.2 The certainty levels and DC-OE differences (gap variable)

In this section, different specifications of the gap variable are tested to explain variation in the certainty levels. We use five different specifications as explanatory variable: (1) absolute differences, (2) only yes-responses, (3) only no-responses, (4) relative gap, only yes-responses, and (5) relative gap, only no-responses.[16] Figure 1 shows the fractions of the gap variable and a gap variable-certainty box plot for the first three specifications of both surveys. The hypothesis is that the larger the gap variable, the higher the confidence. By studying the box plots, the median, quartiles, adjacent values and outside values for the certainty levels can be followed.[17] We display median and quartiles instead of mean and standard deviation because there are very few observations for some certainty levels, especially the lower ones. Median and quartiles are more robust.

[Insert Figure 1 here]

[15] The confidence intervals are constructed using bootstrapping with 1000 replications.
[16] Relative gap=(WTP_OE-bid)/bid is included to give more weight to relatively large differences at low bid levels.
[17] Upper adjacent value is the largest data value that is less than or equal to the third quartile plus 1.5×IQR, where IQR (interquartile range) is the difference between the first and third quartiles. Lower adjacent value is the smallest data value that is greater than or equal to the first quartile minus 1.5×IQR.

We use multiple regressions to further analyse the relation between certainty levels and other covariates (in particular the gap variable). Since the certainty variable is naturally ordered from 1 to 10, the ordered probit model is used to take account of this information. Table 6 presents the result from the regressions on survey A. Age has a significant positive effect in model 1, model 3 and model 4. We tested to include age^2 as well, but it came out insignificant in all five models. The bid level is not included in the analysis in this section and the reason is that we expect it to be correlated to our gap variable. Interestingly, Table 6 shows that the gap variable is significant with positive (model 1, 2, 4) and negative (model 3, 5) coefficients respectively. The sign on the gap variable coefficient is negative because absolute terms are not used in the latter models. Overall, the result implies that the larger the gap variable, the more confident the respondents. However, the full models for yes-responses (models 2 and 4) break down, which could be due to few observations (n=102) and many empty cells.[18] Restricting the independent certainty variable to include only certainty levels of 6-10 results in a significant ordered probit model.[19]

[Insert Table 6 here]

Continuing to survey B, the corresponding results are shown in Table 7. Again, age has a significant and positive influence on certainty for models 1, 2 and 4. Age^2 was also tested separately and resulted in significant negative parameter estimates for models 2 and 4 ($p<0.05$), implying an inverted U-shaped relation. The dummy for low education is significant and negative in models 1, 3 and 5, implying that low educated respondents are less confident. If the respondent has suffered from heart disease it increases the confidence for the yes-responses models. Our specifications of the gap variable do not result in significant parameter estimates for any model except the two no-responses models.

[Insert Table 7 here]

---

[18] The Wald Chi2 test indicates that the joint test that none of the coefficients are different from zero (p=0.707 and p=0.787).

[19] The p-values are now p=0.076 and p=0.027, respectively. We also tried to reduce the number of certainty levels by merging them together into three levels. It turned out that the result was very sensitive to the way the grouping was made and the Wald Chi2 test often turned out with insignificant p-values.

From the coefficients in Tables 6 and 7 it is not easy to say what the marginal effects are. Larger values correspond to a higher probability of a high certainty level, but we do not know by how much. Marginal changes in probability can be computed for the ten levels of certainty separately, and in Tables 8-9 marginal effects for certainty level 10 is presented. The reason for only presenting level 10 is that the marginal changes in predicted possibilities are largest for the gap variable at this level. For survey A (Table 8) we can see that the probability of stating a 10 increases by 5 percent per SEK 1000 increase in the gap variable (model 1). For the yes-responses (model 2) and the no-responses (model 3) the increasing probabilities are 19 and 6 percent respectively.

[Insert Table 8 here]

We calculate the marginal changes in predicted probabilities for survey B in the same way (Table 9) and see that the probability of stating a 10 increases by 4 percent per SEK 1000 that the gap variable for no-responses increases (model 3). Also, in the same model, we see that a low educated respondent reduces the probability of stating a 10 by 24 percent. In the yes-response model the dummy variable heart increases the probability by 29 percent. Age is significant and positive for both surveys (models 1-2) and increases the probability by 5-6 percent per 10 year increase in respondent age.[20]

[Insert Table 9 here]

The share of the respondents in the certainty level 10 group can also be illustrated by graphing the predicted probabilities for each certainty level and the gap variable (Figure 2). For all certainty levels below 10 a decreasing trend is shown in all models. A totally different picture is shown when looking at certainty level 10, except for the absolute gap in survey B, and instead we see that the predicted probability approaches large shares in all models. The pattern is less distinct for survey B, but the increasing certainty with a larger gap is obviously driven by the certainty level 10 group.

[20] Including age^2 in model 2 results in parameter estimates for age=0.031 (p=0.001) and for age^2=-0.00027 (p=0.011).

[Insert Figure 2 here]

5. Discussion

From the results of survey A it is obvious that the gap variable is correlated to the certainty level, with the probability of stating a 10 increasing by 5-19 percent per SEK 1000 that the answer to the OE question and the bid differ. If we assume that the OE answer is closer to real WTP than the DC bid and that the error term is homoscedastic, this result strengthens the theoretical argument of the certainty approach, i.e. the higher the confidence of the respondents the more we can trust that stated WTP is correlated to real WTP. This implies that the hypothesis from Wang (1997), that uncertainty is expected to be large for bids close to real WTP and small for bids decidedly smaller or larger than real WTP, is supported. The results of survey B are not as clear, but we do find a significant and negative effect of the gap variable for the no-responses that is close in magnitude to survey A.

One crucial assumption we make is that OE answers are closer to real WTPs than DC bids. Since we do not measure real WTP, we cannot be sure about the accuracy of this statement. The empirical evidence is mixed in the respect that while the NOAA panel (Arrow et al., 1993) recommends DC over OE, the main source of difference between WTP estimates from DC and OE questions is argued to be variations in the statistical efficiency and robustness and not differences in how respondents perceive the formats (Halvorsen & Sælensminde, 1998). Intuitively, it is not an unfamiliar argument to suggest that on the individual level, the respondents who receive a bid that is far below or above the real WTP would also have a larger gap. At least on average, our assumption seems to be plausible for the discrete-continuous CV format.

The link between theoretical and empirical treatment of respondent uncertainty is not easy to find in the literature concerning the certainty approach. One of the most applied empirical treatments is to recode all yes-responses less than 10 as no-responses ('the asymmetric uncertainty model, or ASUM'), i.e. a no-answer is always interpreted as a no, but a yes may be converted to a no. The interesting sub-group here is then the yes-responders. Can we assume that a respondent with a lower than 10 certainty actually means no to that particular WTP question? If we study the predicted probabilities of certainty levels (Figure 2) with respect to the gap variable, we can see that at a gap of SEK 1000, the probability of a certainty level of 10 is approximately 40-50 percent. This also implies that, at this relatively large gap, 50-60 percent of the respondents are still predicted to state a certainty level lower than 10. From this result we would not *theoretically* recommend the approach of recoding yes-responses of certainty levels lower than 10 in such a way that we assume that they all have a real WTP that is less than the bid.

Another empirical treatment is to only use the sub-group of the most confident respondents (certainty=10). This significantly reduces the sample size, but the results in this study indicate that the certainty 10 group could be a watershed between a certain yes and a certain no. Using this approach also opens up for the possibility that estimated WTP can either increase or decrease (as we see examples of in our two surveys), instead of using the ASUM where estimated WTP per definition decreases. The results also signal that different goods may have different needs for calibration, even though the dichotomous choice approach in CV generally has been found to overestimate WTP.

Our intuitive interpretation of the empirical treatments of certainty calibration is not in accordance with the previously found close correspondence between ASUM calibration and real WTP (e.g. Blumenschein et al., 1998, 2001, 2008). We do not have a measure of real WTP, but we can study the effect of different certainty calibration treatments on mean and median WTP values. Table 10 presents information on the total value of a statistical life (VSL) regarding road traffic safety and out-of-hospital cardiac arrest.[21] We can see that VSL estimates are very sensitive to assumptions about the distribution, choice of DC/OE answers, mean/median values and calibration treatments.[22] In general, our VSL estimates are in the standard intervals found in previous studies on road safety (e.g. de Blaeij et al., 2003).

The non-parametric model of Kaplan-Meier-Turnbull (K-M-T) usually estimates a lower-bound for VSL, but by using ASUM calibration we arrive at even lower estimates (except for the mean WTP in survey B). The calibration treatments have very different effects on our two data sets and as long as we cannot explain why we would recommend a cautious approach to this form of calibration. K-M-T shows more resemblance to the VSL estimates from OE. Maybe an interesting idea for the

[21] We use mean/median WTP and divide by the marginal risk reduction (Δp) to calculate VSL. Survey A: $\Delta p=3/82\ 000$, Survey B: $\Delta p=3.35/100\ 000$.

[22] The lognormal model was used for both calibration treatments.

future would be to test whether real WTP corresponds to OE estimates in a discrete-continuous CV format or to explore the possibility to calibrate the OE answers?

[Insert Table 10 here]

The fact that the VSL for cardiac arrest is higher than the VSL for road traffic safety seems to be robust. This is surprising because statistical lives are both longer and 'healthier' for road traffic. Statistical lives for cardiac arrest last only for approximately six additional years and have a health utility of 0.7 (Rauner & Bajmoczy, 2003). We can only make speculations about why this unexpected and relatively large difference exists. Differences between questionnaire designs and contexts are possible, even though the designs are very similar and the VSL values for road traffic safety are consistent with previous surveys (Svensson, 2009).[23] A second possibility is that we really measure some kind of preference for 'individual freedom', where further road traffic safety measures are perceived as limiting freedom of action (e.g. speed cameras, seat belts, helmets) whereas an increased density of defibrillators does not affect individuals in this way. Thirdly, we may capture solidarity with older and helpless individuals while road-users are perceived to have more controllable risks to manage.

In a follow-up of the responses to the discrete-continuous CV format we early revealed some inconsistencies. The most apparent inconsistency is the case where a respondent answered yes (no) to a DC bid and then gave an OE answer that was lower (higher) than the bid. Obviously, this is not in accordance with expected economic behaviour, and these responses should hence be excluded. However, if we look at the certainty levels of the DC and OE answers for these respondents a striking cognitive pattern is revealed. Of the 24 respondents in survey A and 54 respondents in survey B who were inconsistent, 16 and 26 respondents (i.e. 67 and 48 percent) respectively answered in a way that cognitively is explained by their answers regarding certainty levels. That is, they did not maximised their WTPs in the OE question, as was asked for, but instead tried to further increase the certainty level. A totally inconsistent answer is explained by a cognitive mistake, which makes the

[23] Svensson (2009) compared the data from survey A with data from another road traffic safety CV study and found consistent VSL values among the most confident respondents.

response only ‘weakly’ inconsistent. A better design might help guide these respondents.[24]

There is one alternative interpretation of this inconsistency and that is the fact that payment in the DC question is uncertain and conditional on what the responder believes other respondents will answer (we used a provision condition). On the other hand, payment in the OE question is certain and not conditional on other respondents answers. According to Krüger & Svensson (2009) a high degree of objective or perceived subjective uncertainty in the outcome results in exaggerated WTP values. If this is the case in our surveys, WTP from DC may not be directly comparable to WTP from OE in the way we assume it to be. The inconsistency we found can theoretically be explained by exaggeration of the DC values. We may also question if the gap between DC and OE is underestimated, which makes our results from survey A even stronger, and if the degree of uncertainty may explain some of the difference in VSL values between survey A and B.[25] However, although we recognise the difference in certainty we do not find it likely that the respondents perceived the difference and behaved according to this hypothesis.

Finally, our results are somewhat inconclusive since the findings from survey A and survey B do not coincide on all matters. We call for further development of the theory regarding certainty calibration and the determinants of the certainty levels. Different goods and services obviously show different patterns regarding certainty calibration. Nevertheless, the reasons for the differences in VSL values regarding road traffic safety and out-of-hospital cardiac arrest are interesting from a policy perspective, suggesting that individuals may value factors other than number of years saved and the quality of these years (e.g. ‘individual freedom’ and ‘controllable’ risks).

---

[24] In the data analysis, all inconsistent responses are excluded.

[25] Provision condition in survey A was ≥70 %, while it was ≥50 % in survey B. This implies that the respondents might have perceived more uncertainty in provision in survey A.

References

Alberini, Anna, Boyle, Kevin & Welsh, Michael, (2003). Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. Journal of Environmental Economics and Management, 45, 40-62.

Arrow, K., Solow, R., Portney, P.R., Learner, E.E., Radner, R. & Schuman, H., (1993). Report from the NOAA panel on contingent valuation. Federal Register 58 (10): 4601-14.

Balistreri, Edward, McClelland, Gary, Poe, Gregory & Schultze, William, (2001). Can hypothetical questions reveal true values? A laboratory comparison of dichotomous choice and open-ended contingent values with auction values. Environmental and Resource Economics 18: 275-292, 2001.

de Blaeij, Arianne, Florax, Raymond J.G.M., Rietveld, Piet & Verhoef, Erik, (2003). The value of statistical life in road safety: a meta-analysis. Accident Analysis and Prevention 35 (2003), 973-986.

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., Yokohama, K.K. & Freeman, P.R., (2001). Hypothetical versus real willingness to pay in the health care sector: results from a field experiment. Journal of Health Economics 20, 441-457.

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.

Calman, Sir Kenneth C. & Royston, Geoff, (1997). Risk language and dialects. British Medical Journal 1997, 315, 939-942.

Carson, R.T., Flores, N.E. & Meade, N.F., (2001). Contingent valuation: Controversies and evidence. Environmental and Resource Economics 19: 173-210, 2001.

Carson, Richard T. & Mitchell, Robert Cameron, (1995). Sequencing and nesting in contingent valuation surveys. Journal of Environmental Economics and Management, 28(2), 155-173.

Champ, P.A., Bishop, R.C., Brown, T.C. & McCollum, D.W., (1997). Using donation mechanism to value nonuse benefits from public goods. Journal of Environmental Economics and Management 33, 151-162.

Champ, P.A. & Bishop, R.C., (2001). Donation payment mechanisms and contingent valuation: an empirical study of hypothetical bias. Environmental and Resource Economics 19, 383-402.

Cummings, R.G. & Taylor, L.O., (1999). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. Economic Review 89(3), 649-665.

Goldberg, Isabell & Roosen, Jutta, (2007). Scope insensitivity in health risk reduction studies: A comparison of choice experiments and the contingent valuation method for valuing safer food. Journal of Risk and Uncertainty (2007) 34: 123-144.

Halvorsen, Bente & Sælensminde, Kjartan, (1998). Differences between willingness-to-pay estimates from open-ended and discrete-choice contingent valuation methods: The effect of heteroscedasticity. Land Economics, May 1998, 74 (2): 262-82.

Hanemann, W.M., (1984). Welfare evaluations in contingent valuation experiments with discrete responders. American Journal of Agricultural Economics, 66, 332-341.

Harrison, Glenn W., (2006). Experimental evidence on alternative environmental valuation methods. Environmental & Resource Economics 34:125-162.

Harrison, Glenn W. & Rutstrom, Elisabeth E., (2005). Experimental evidence on the existence of hypothetical bias in value elicitation methods. In: C. Plott and V. Smith (eds.), Handbook in Experimental Economic Results. Elsevier Science, New York.

Johannesson, M., Liljas, B. & Johansson, P-O., (1998). An experimental comparison of dichotomous choice contingent valuation questions and real purchase decisions. Applied Economics 30, 643-647.

Johannesson, M., Blomquist, G.C., Blumenschein, K., Johansson, P-O., Liljas, B. & O'Conor, R.M., (1999). Calibrating hypothetical willingness to pay responses. Journal of Risk and Uncertainty 18, 21-32.

Kealy, Mary Jo & Turner, Robert W., (1993). A test of the equality of closed-ended and open-ended contingent valuations. American Journal of Agricultural Economics, 75 (May 1993), 321-331.

Kriström, Bengt, (1993). Comparing continuous and discrete contingent valuation questions. Environmental and Resource Economics 3:63-71.

Krüger, Niclas & Svensson, Mikael, (2009). The impact of real options on willingness to pay for mortality risk reductions. Journal of Health Economics, forthcoming.

Li, Chuan-Zhong & Mattsson, Leif, (1995). Discrete choice under preference uncertainty: An improved structural model for contingent valuation. Journal of Environmental Economics and Management, 28, 256-269.

List, John A. & Gallet, Craig A., (2001). What experimental protocol influence disparities between actual and hypothetical values? Evidence from meta-analysis. Environmental and Resource Economics 20: 241-254.

Little, Joseph & Berrens, Robert, (2004). Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis. Economics Bulletin , vol.3, no.6, pp.1-13.

Loomis, John & Ekstrand, Earl, (1998). Alternative approaches for incorporating respondent uncertainty when estimating willingness to pay: the case of the Mexican spotted owl. Ecological Economics, 27 (1998), 29-41.

Murphy, James J. & Stevens, Geoffrey, (2004). Contingent valuation, hypothetical bias, and experimental economics. Agricultural and Resource Economics Review, October 2004, v. 33, iss. 2, pp. 182-92.

Murphy, James J., Allen, P. Geoffrey, Stevens, Thomas H. & Weatherhead, Darryl, (2005). A meta-analysis of hypothetical bias in stated preference valuation. Environmental and Resource Economics 30: 313-325.

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.

Rauner, Marion S. & Bajmoczy, Nikolaus, (2003). How many AEDs in which region? An economic decision model for the Austrian Red Cross. European Journal of Operational Research, 150(1), 3-18.

Ready, Richard C., Whitehead, John C. & Blomquist, Glenn C., (1995). Contingent valuation when respondents are ambivalent. Journal of Environmental Economics and Management, 29, 181-196.

Samnaliev, Mihail, Stevens, Thomas H. & More, Thomas, (2006). A comparison of alternative certainty calibration techniques in contingent valuation. Ecological Economics, 57 (2006), 507-519.

Shaik, Sabina L., Sun, Lili & van Kooten, Cornelis, (2007). Treating respondent uncertainty in contingent valuation: A comparison of empirical treatments. Ecological Economics 62 (2007), 115-125.

Sund, Björn, (2009). Sensitivity to scope in contingent valuation – introducing a flexible community analogy to communicate mortality risk reductions. Working Papers No 2009:2, Swedish Business School, Örebro University.

Svensson, Mikael, (2009). The value of a statistical life in Sweden: Estimates from two studies using the 'certainty approach' calibration. Accident Analysis and Prevention, forthcoming.

Swedish Road Administration (8 September 2008), http://www.vv.se/templates/page3wide____1827.aspx

van Kooten, G. Cornelis, Mrcmar, Emina & Bulte, Erwin H., (2001). Preference uncertainty in non-market valuation: a fuzzy approach. American Journal of Agricultural Economics, 83, 487-500.

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.

Wang, Hua, (1997). Treatment of "don't-know" responses in contingent valuation surveys: a random valuation model. Journal of Environmental Economics and Management, 32, 219-232.

Welsh, Michael P. & Poe, Gregory L., (1998). Elicitation effects in contingent valuation: comparisons to a multiple bounded discrete choice approach. Journal of Environmental Economics and Management, 36, 170-185.

Tables

Table 1. Hypothesis of the relation between the differences in the discrete-continuous CV format and confidence

	Answer to question 1: DC=SEK 200	*Answer to question 3: OE*	*Meaning in terms of confidence (question 2)*
Individual 1	Yes	200	Certainty<10
Individual 2	Yes	600	Certainty=10

Table 2. Sub-samples of survey A

		Magnitude of risk reduction	*Bid levels*	*Number of questionnaires*
1	Main sample	From 6 to 3 deaths per year	All levels	850 (170 per bid*)
2	Scope test sample	From 6 to 1 death per year	SEK 1000	150

** Bid levels are SEK 200, 500, 1000, 2000, 5000.*

Table 3. Sub-samples of survey B*

		Magnitude of risk reduction	*Bid levels*	*Number of questionnaires*
1	Main sample	From 5 to 10 percent survivors per year	All levels	1,000 (200 per bid**)
2	Scope test sample	From 5 to 15 percent survivors per year	SEK 500 and 1000	400

Notes: ** Half of the questionnaires used an array of 10 000 dots as a visualisation of the risk, while the other half used a 'flexible community analogy'. ** The bid levels are SEK 200, 500, 1000, 2000, 5000.*

Table 4. Proportions of yes-responses and WTP (in percent) at different bid levels

Bid level (SEK)	*Main sample (DC)*	*Scope sample (DC)*	*Main sample (OE)*	*Scope sample (OE)*
Survey A				
200	68.9 (n=61)		59.0	52.9
500	43.3 (n=83)		41.8	38.2
1000	22.1 (n=77)	27.8 (n=72)	18.0	27.9
2000	13.4 (n=82)		7.2	5.9
5000	11.6 (n=78)		2.2	1.5
			n =361	n=68
Survey B				
200	85.3 (n=75)		82.9	81.4
500	75.7 (n=70)	66.2 (n=74)	64.7	70.2
1000	64.8 (n=71)	51.7 (n=58)	42.5	33.9
2000	42.3 (n=52)		14.4	4.8
5000	16.1 (n=62)		5.1	0
			n=292	n=124

Notes: *n=number of respondents.*

Table 5. Mean/median WTP (SEK) in groups with different levels of confidence

Certainty level	*Survey A Dichotomous choice*	*Survey A Open-ended*	*Survey B Dichotomous choice*	*Survey B Open-ended*
Main sample	1722/368	476/500	4821/1390	642/648
95 % CI, mean	[1017, 4403]	[68, 3331]	[2764, 10980]	[73, 5646]
95 % CI, median	[243, 487]	[71, 3502]	[1094, 1809]	[74, 5698]
	n=381	n=237	n=330	n=260
Level ≥ 5	2026/399	484/500	6431/1737*	641/648
	n=334	n=209	n=279	n=240
Level ≥ 6	1738/411	490/500	6568/1794*	636/600
	n=299	n=187	n=246	n=209
Level ≥ 7	1953/390	496/500	6079/1781*	640/600
	n=283	n=178	n=226	n=197
Level ≥ 8	1449/388	489/500	7774/1901**	621/600
	n=246	n=151	n=195	n=169
Level ≥ 9	724***/361	438/500	5126/1716*	606/600
	n=175	n=95	n=139	n=118
Level = 10	630***/346	438/500	5586/1772*	616/600
	n=140	n=69	n=108	n=90

Notes: *Based on main sample. Significantly different (α=0.01***, α=0.05**, α=0.10*) from the mean/median WTP for the main sample, n=number of observations.*

Table 6. Ordered probit regression on certainty level for survey A

Variable	*Model 1 (absolute values)*	*Model 2 (yes-responses, certainty levels 6-10)*	*Model 3 (no-responses)*	*Model 4 (relative gap, yes-responses, certainty levels 6-10)*	*Model 5 (relative gap, no-responses)*
Gender	0.0034	0.093	-0.071	0.19	0.039
	(0.975)	(0.670)	(0.592)	(0.394)	(0.772)
Age (10 years)	0.072*	0.014	0.083*	0.15*	0.025
	(0.059)	(0.108)	(0.061)	(0.092)	(0.602)
High education	0.031	0.34	-0.0022	0.31	-0.092
	(0.778)	(0.149)	(0.987)	(0.184)	(0.497)
Low education	0.081	-0.14	0.029	-0.075	0.011
	(0.697)	(0.776)	(0.906)	(0.876)	(0.967)
High risk	0.050	-0.21	0.15	-0.18	0.21
	(0.746)	(0.479)	(0.414)	(0.565)	(0.262)
Low risk	0.12	0.27	0.18	0.31	0.19
	(0.321)	(0.311)	(0.196)	(0.236)	(0.191)
Income (SEK 10 000)[26]	-0.040	-0.0065	-0.088	0.019	-0.092
	(0.514)	(0.952)	(0.233)	(0.860)	(0.270)
Scope	0.16	-0.073	0.25	0.063	0.062
	(0.290)	(0.842)	(0.148)	(0.859)	(0.726)
Gap variable	0.00014***	0.00054**	-0.00015***	0.46***	-2.18***
	(0.000)	(0.021)	(0.001)	(0.002)	(0.000)
Log-likelihood	-772.10	-146.37	-536.16	-144.04	-522.25
N	415	102	294	102	294

Notes: *Significant for α = 0.01***, α = 0.05**, α = 0.10*, based on robust standard errors, p-values in parentheses.*

[26] We are aware that it is theoretically problematic to include income as an independent variable in the WTP regression for DC questions, since utility is assumed to be linear in income (Hanemann, 1984). However, we do not interpret income as 'income per se' but instead as a proxy for household characteristics and focus on the empirical relationship.

Table 7. Ordered probit regression on certainty level for survey B

Variable	*Model 1 (absolute values)*	*Model 2 (yes-responses)*	*Model 3 (no-responses)*	*Model 4 (relative gap, yes-responses)*	*Model 5 (relative gap, no-responses)*
Gender	0.050	0.018	0.0097	0.0092	0.14
	(0.645)	(0.906)	(0.954)	(0.952)	(0.396)
Age (10 years)	0.14***	0.16***	0.087	0.16***	-0.0026
	(0.000)	(0.001)	(0.223)	(0.001)	(0.972)
High education	-0.029	-0.070	-0.036	-0.074	-0.11
	(0.808)	(0.673)	(0.845)	(0.657)	(0.543)
Low education	-0.32*	0.19	-1.01***	0.16	-0.85***
	(0.077)	(0.398)	(0.000)	(0.465)	(0.007)
High risk	-0.13	-0.048	-0.044	-0.042	-0.18
	(0.423)	(0.821)	(0.881)	(0.842)	(0.540)
Low risk	-0.089	0.14	-0.16	0.13	-0.14
	(0.459)	(0.405)	(0.385)	(0.443)	(0.470)
Income (SEK 10 000)	0.069	0.019	0.047	0.021	0.074
	(0.221)	(0.808)	(0.536)	(0.795)	(0.335)
Population (in 100 000)	-0.023	-0.033	0.0046	-0.031	0.00079
	(0.359)	(0.270)	(0.899)	(0.301)	(0.983)
Heart	0.18	0.73***	-0.35	0.72***	0.048
	(0.445)	(0.005)	(0.335)	(0.007)	(0.893)
Aid	0.061	-0.048	0.070	-0.049	0.11
	(0.582)	(0.751)	(0.710)	(0.744)	(0.549)
Scope	-0.020	0.023	0.088	0.021	0.11
	(0.868)	(0.876)	(0.698)	(0.890)	(0.600)
Gap variable	-0.000034	0.00016	-0.0011*	0.031	-2.45***
	(0.439)	(0.176)	(0.051)	(0.620)	(0.000)
Log-likelihood	-733.76	-377.38	-314.09	-378.30	-303.95
N	386	233	153	233	153

Notes: *Significant for α = 0.01***, α = 0.05**, α = 0.10*, based on robust standard errors, p-values in parentheses.*

Table 8. Coefficients of marginal changes in predicted probabilities for survey A

Variable	*Model 1 (absolute values) Certainty 10*	*Model 2 (yes-responses, Certainty levels 6-10) Certainty 10*	*Model 3 (no-responses) Certainty 10*
Gender	0.0013	0.033	-0.027
Age (10 years)	0.027*	0.050	0.032*
High education	0.012	0.12	-0.00083
Low education	0.031	-0.049	0.011
High risk	0.019	-0.071	0.060
Low risk	0.045	0.098	0.072
Income (SEK 10 000)	-0.015	0.0023	-0.034
Scope	0.059	-0.025	0.098
Gap variable (SEK 1000)	0.053***	0.189**	-0.058***
Predicted (cert.=10)	0.36	0.31	0.40

Notes: *Significant for α = 0.01***, α = 0.05**, α = 0.10*.*

Table 9. Coefficients of marginal changes in predicted probabilities for survey B

Variable	*Model 1 (absolute values) Certainty 10*	*Model 2 (yes-responses) Certainty 10*	*Model 3 (no-responses) Certainty 10*
Gender	0.018	0.0067	0.0031
Age (10 years)	0.049***	0.061***	0.028
High education	-0.011	-0.026	-0.012
Low education	-0.11*	0.073	-0.24***
High risk	-0.047	-0.018	-0.014
Low risk	-0.032	0.052	-0.053
Income (SEK 10 000)	0.025	0.0072	0.015
Population (in 100 000)	-0.0082	-0.012	0.0015
Heart	0.068	0.29***	-0.10
Aid	0.022	-0.018	0.023
Scope	-0.0071	0.0087	0.029
Gap variable (SEK 1000)	-0.012	0.060	-0.037*
Predicted (cert.=10)	0.33	0.36	0.26

Notes: *Significant for α = 0.01***, α = 0.05**, α = 0.10*.*

Table 10. Value of a statistical life (VSL) using various assumptions (in million SEK)

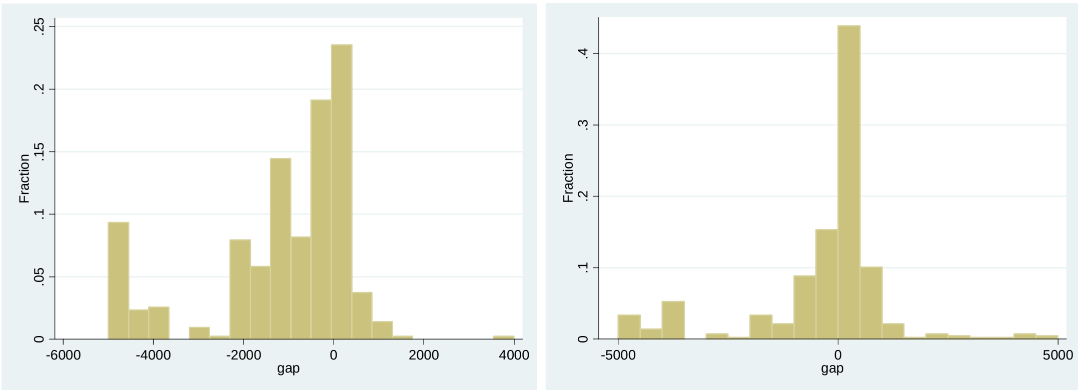
	Survey A Mean	*Survey A Median*	*Survey B Mean*	*Survey B Median*
Dichotomous choice				
Lognormal model	47	10	144	41
Kaplan-Meier-Turnbull (non-parametric model)	24	5	49	30
Open-ended				
Original data	14	5	29	15
Original data, WTP>0	21	14	33	19
log OE	13	14	19	19
Calibration treatments				
ASUM calibration	7	2	95	3
Only certainty=10	17	9	167	53

Figures

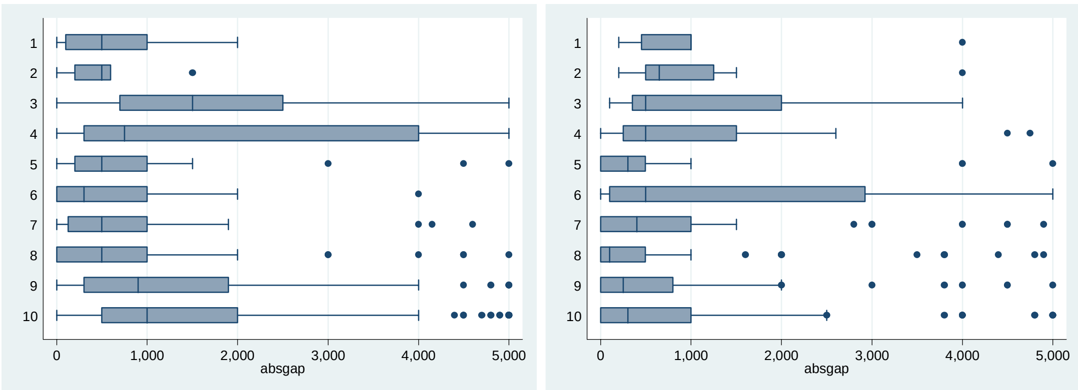
Figure 1. Fraction of the gap variable and a gap variable-certainty box plot

Survey A *Survey B*

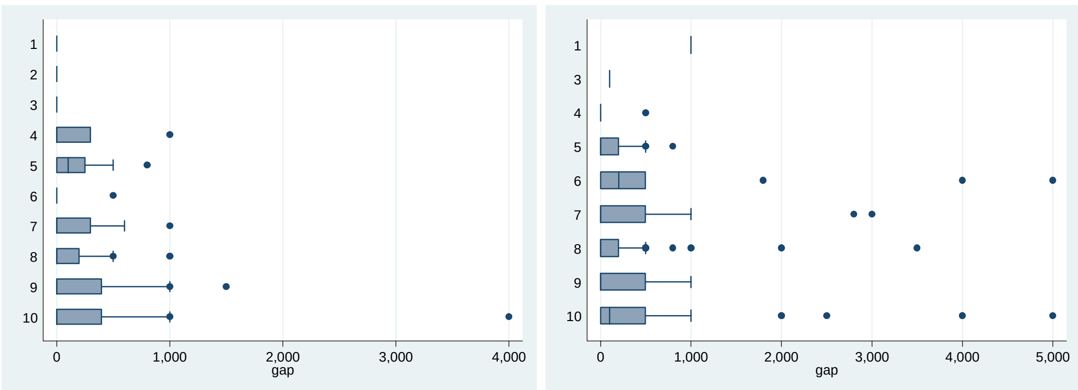
Fraction



Absolute gap



Yes answers



No answers

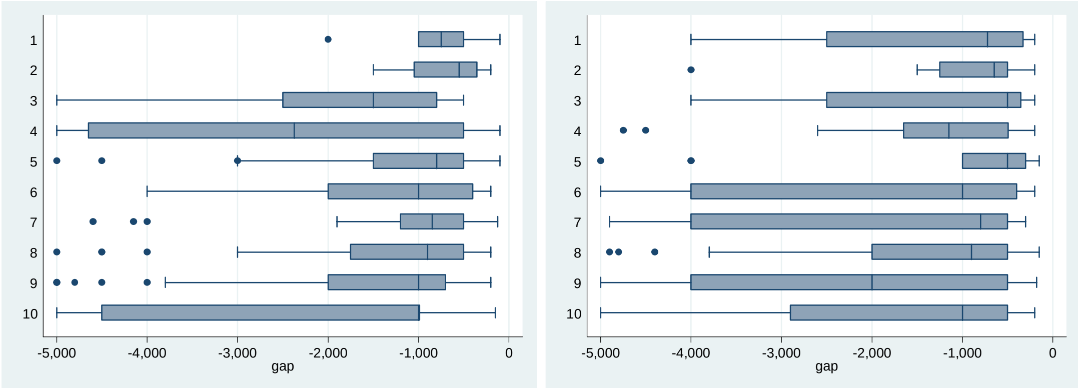
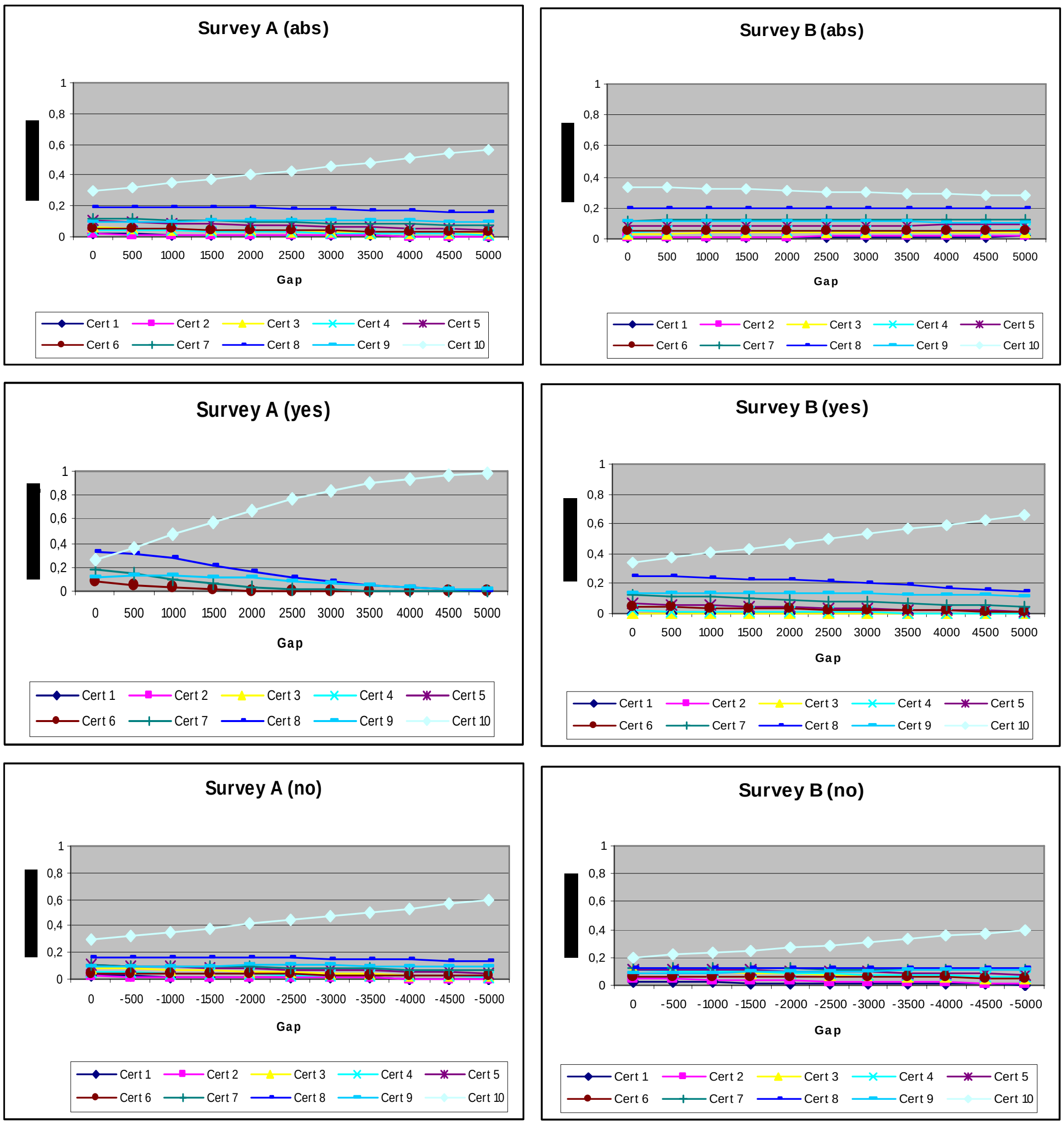


Figure 2. Predicted probabilities of certainty levels by the gap variable



Appendix

Table A1. Characteristics of the variables

Variable	*Characteristics*
Gender	Unit dummy variable for gender of the respondent; one if female
Age	Age of respondent, between 18-75
High education	Unit dummy variable if education level is at least one term at a university; zero otherwise
Low education	Unit dummy variable if education level is at most nine-year compulsory school; zero otherwise
High risk	Unit dummy variable if the own perceived traffic risk/risk of cardiac arrest is higher than average; zero otherwise
Low risk	Unit dummy variable if the own perceived traffic risk/risk of cardiac arrest is lower than average; zero otherwise
Income	The income per consumption unit given by the total household income* divided by the number of household members weighted as follows: adult person # 1 = 1.16, adult person # 2 = 0.76, children 0-3 years old = 0.56, children 4-10 years old = 0.66, children 11-17 years old = 0.76
Population	Number of inhabitants (self assessed by respondents) in the municipality
Heart	Unit dummy variable if the respondent has suffered from heart disease; zero otherwise
Aid	Unit dummy variable if visual aid is an array of dots; zero if 'flexible community analogy'
Bid	The predetermined bid level: SEK 200, 500, 1000, 2000 or 5000
Scope	Unit dummy variable for a larger risk reduction
Certainty (DC)	The respondent's own certainty when replying to the dichotomous choice (DC) valuation question on a scale from one (very uncertain') to ten ('very certain')
Certainty (OE)	The respondent's own certainty when replying to the open-ended (OE) valuation question on a scale from one ('very uncertain') to ten ('very certain')

** The respondents were asked to mark an interval with a range of SEK 4999.The income was then approximated by using the mid value of the interval.*

Table A2. Mean and standard deviation of the variables

Variable	*Survey A Main sample*	*Survey A Scope sample*	*Survey B Main sample*	*Survey B Scope sample*
Number of returned questionnaires*	382	72	333	135
Gender (1=female)	0.54 (0.50)	0.61 (0.49)	0.50 (0.50)	0.50 (0.50)
Age (18-75)	44.8 (14.0)	47.2 (15.1)	48.3 (15.3)	47.5 (16.1)
High education	0.54 (0.50)	0.38 (0.49)	0.44 (0.50)	0.41 (0.49)
Low education	0.09 (0.28)	0.15 (0.36)	0.18 (0.39)	0.14 (0.35)
High risk	0.15 (0.35)	0.17 (0.38)	0.16 (0.36)	0.13 (0.34)
Low risk	0.31 (0.46)	0.35 (0.48)	0.41 (0.49)	0.52 (0.50)
Income	19 216 (10 224)	18 225 (10 914)	19 223 (10 992)	19 584 (11 753)
Population	-	-	147 676 (227 607)	150 425 (242 404)
Heart	-	-	0.11 (0.31)	0.10 (0.30)
Aid	-	-	0.47 (0.50)	0.50 (0.50)
Certainty (DC)**	8 (6, 10)	8.5 (7, 10)	8 (6, 10)	8 (6, 10)
Certainty (OE)**	8 (7, 10)	9 (7, 10)	8 (7, 10)	9 (7, 10)

*Totally blank questionnaires, WTP>0.05×Income and inconsistent respondents are not included. The number of respondents in these three groups is 33+6+24=63 (survey A) and 45+12+54=111 (survey B). **Ordinal data, median (Q_1, Q_3) reported.

Table A3. Variations in WTP. Estimated with exponential probit models. Marginal effects.

Variable	*Survey A DC*	*Survey A OE*	*Survey B DC*	*Survey B OE*
Constant		4.96***		7.13***
		(0.000)		(0.000)
Gender	-0.040	-0.039	0.080	0.082
	(0.379)	(0.737)	(0.122)	(0.417)
Age (10 years)	0.0083	0.20	-0.18*	-0.58***
	(0.941)	(0.483)	(0.076)	(0.003)
Age2 (10 years)	0.00029	-0.00058	0.0021*	0.0071***
	(0.817)	(0.864)	(0.054)	(0.001)
High education	0.029	0.086	-0.055	0.064
	(0.565)	(0.485)	(0.341)	(0.586)
Low education	-0.071	-0.22	-0.056	-0.30**
	(0.349)	(0.400)	(0.487)	(0.039)
High risk	0.083	0.035	-0.066	0.065
	(0.226)	(0.825)	(0.447)	(0.651)
Low risk	-0.0089	-0.018	-0.13**	-0.42***
	(0.860)	(0.892)	(0.024)	(0.001)
Income (SEK 10 000)	0.0059	0.0065	0.027	0.066
	(0.791)	(0.902)	(0.318)	(0.234)
Log(bid)	-0.19***		-0.27***	
	(0.000)		(0.000)	
Bid (SEK 1000)		0.26***		0.20***
		(0.000)		(0.000)
Scope	-0.012	0.17	-0.066	0.13
	(0.845)	(0.340)	(0.245)	(0.236)
Population (in 100 000)			-0.025**	0.010
			(0.014)	(0.633)
Heart			0.012	0.024
			(0.894)	(0.853)
Aid			-0.014	-0.020
			(0.786)	(0.848)
Log-likelihood	-230.05	-359.67	-235.75	-465.48
n	438	273	426	348

Notes: *Significant for α = 0.01***, α = 0.05**, α = 0.10*, based on robust standard errors, p-values in parentheses.*

Table A4. The valuation scenario and WTP questions for survey A (translated from Swedish)

In the municipality of Karlstad, an average of 6 individuals each year dies in road accidents. Imagine that there is a possibility to **halve the risk** of road fatalities in Karlstad. We will ask you about your willingness to pay for such measures. Remember that the money you are willing to pay for security improvements reduces your possibilities for other consumption.

To achieve this safety improvement a public road safety program is considered. The safety improvement is for all accidents where **at least one vehicle** may be involved, which means that the risk is reduced for car users, motorcyclists, moped riders, cyclists and for pedestrians.

A prerequisite for the road safety program to be implemented is that **at least 70 %** of the individuals in the municipality of Karlstad pay a fee to an especially dedicated road fund used for the program and managed by the municipality. **If the individuals do not contribute enough with the fee, the road safety program can not be imposed, and your fee will be refunded in full.**

The road safety program will not affect the possibility to choose transportation, the quality of travelling, average speed or e.g. the urban environment in Karlstad.

The implementation of the road safety program means that **your own risk as well as the risk of all other individuals** in the municipality of Karlstad being halved and the number of traffic deaths on average will be **reduced from 6 to 3** in Karlstad for a year.

Question 7. Would you be willing to pay SEK 200 **per year** in fees to a special road fund for this road safety program to be implemented in Karlstad?

□ Yes □ No

Question 8. How confident are you in your answer to the above question, where 1 is very uncertain and 10 is very certain? Circle your answer.

1 2 3 4 5 6 7 8 9 10

very uncertain ... *very certain*

Question 9. Provided that the programme is carried out, how much would you most be willing to pay **annually** for the implementation of the program, which halves your own risk as well as the risk of all other individuals in your municipality of death in traffic?

Answer: ……………SEK per year

Question 10. Same as question 11.

Table A5. The valuation scenario and WTP questions for survey B (translated from Swedish)[27]

A number of individuals suffer from cardiac arrests each year in your municipality. Imagine that there exists a possibility to **reduce mortality risks** for cardiac arrests. We will ask you about your willingness to pay for such measures. Remember that the money you are willing to pay for security improvements reduces your possibilities for other consumption.

To reduce the mortality risk a public programme to increase the density of defibrillators is considered. One possibility is to equip and educate employees within certain professions in the municipality which may respond faster than the ambulance. These professions might be firemen, policemen, security guards or nurses. Public access defibrillators may also be located in hotels, shopping malls, sports centres or theatres.

A prerequisite for the programme to be implemented is that at least 50 % of the individuals in your municipality are positive to the introduction of the programme. The cost is paid as an annual fee. If the individuals will not contribute enough with the fee, the programme will not be imposed.

What is the effect of the programme?

The programme will result in **your own risk as well as the risk of all other individuals** in your municipality being reduced, and **the survival rate will increase from 5 % to 10 % on average**. In the table the effect of the programme for various municipality sizes are presented.

Observe that the table represents effects over 10 years!

Inhabitants	Number of out-of-hospital cardiac arrests over 10 years	Number of survivors over 10 years (before), 5 %	Number of survivors over 10 years (after), 10 %	Difference
10 000	70	3	7	+4
20 000	130	6	13	+7
30 000	200	10	20	+10
50 000	330	16	33	+17
75 000	500	25	50	+25
100 000	670	33	67	+34
150 000	1000	50	100	+50
250 000	1670	83	167	+84
500 000	3350	167	335	+168
750 000	5020	251	502	+251

[27] Survey B is divided into two sub-samples that use two different aids to communicate the risk reduction. We present the valuation scenario of the 'flexible community analogy' (FCA) and for a presentation of the array of dots we refer to Sund (2009).

Example from the table: In a municipality of 10 000 individuals, 70 persons will suffer from out-of-hospital cardiac arrest during a 10 year period on average. Now 3 persons will survive and after the programme 7 persons will survive, which implies an increase of 4 persons over 10 years.

Question 10. How would you vote if your personal fee was SEK 200 **per year** (i.e total SEK 2000 for 10 years), for this programme to be implemented in your municipality?

I would vote: □ Yes □ No

Question 11. How confident are you in your answer to the above question, where 1 is very uncertain and 10 is very certain? Circle your answer.

1 2 3 4 5 6 7 8 9 10

very uncertain … *very certain*

Question 12. Provided that the programme is carried out, how much would you most be willing to pay **annually** for the implementation of the programme, that reduces your own risk as well as the risk of all other individuals in your municipality for a cardiac arrest mortality?

Answer: ……………SEK per year

Question 13. Same as question 11.