Scale Sensitivity and Numerical Skills in Contingent Valuation Surveys of Risk Reduction Policies*

Justin Quinton† Roberto Martínez-Espiñeira‡
Nikita Lyssenko
Department of Economics, Memorial University of Newfoundland, Canada

Canadian Resource and Environmental Economics
Study Group Annual Conference 2013 Brock University, St. Catharines
Friday-Sunday, September 27-29, 2013

Abstract

This paper considers data from a 2013 study conducted in Newfoundland about the value of reducing mortality risks due to moose-vehicle collisions. Willingness to pay was modeled using a maximum likelihood routine that allowed factors that determine sensitivity to scale to be modeled as functions of covariates. This allows comparison of sensitivity to scale between groups. A unique aspect of the survey design elicited a near continuous distribution of risk reduction levels which allowed for a more in depth analysis of scale sensitivity. There are several key findings from this paper that support previous research findings in assessing scale sensitivity. It was found that previous experience with the risk involved was a significant factor in determining sensitivity to scale. Additionally, we find support for the hypothesis that cognitive ability is also a determinant of scale sensitivity. Individuals with better cognitive skills are more likely to show sensitivity to scale in both the weak and strong form tests. It was found in the weak form test that sensitivity to scale is less likely to be present at lower risk reduction levels. Additionally, own death risk perception was modeled using an OLS regression and the level of math score (on a scale of 0 to 4 correct questions) was modeled using an ordered logit model.

Keywords: moose vehicle collisions; risk reduction; contingent valuation; value of a statistical life; scale sensitivity; cognitive skills

*Honours Thesis

†E-mail: z94jdq@mun.ca

‡Corresponding Author: Department of Economics, Memorial University of Newfoundland, Canada Tel: 1-709-864-3676 Fax: 1-902-867-3610. E-mail: rmartinezesp@mun.ca

Introduction

This paper seeks to further explain scale sensitivity in contingent valuation (CV) surveys. Scale sensitivity is derived from the idea that people should be willing to pay more for more of a good. In this case the good on offer is a reduction in the risk of moose-vehicle collision in Newfoundland. This study focuses exclusively on external tests of scale sensitivity, which involves comparing the responses from independent respondents and testing if the willingness-to-pay (WTP) values elicited are in line with theoretically expected values. Scale sensitivity is considered an essential aspect when considering the validity of a CV study (Arrow et. al 1993) and has been an issue spurring much debate in the literature. Several studies found insensitivity to scale to be a problem (Diamond 1993; Boyle et. al 1994), while Carson (1997) in response found that the problem was with the survey designs rather than any inherent problem with the CV method itself. In a recent survey of the literature, Desvousges et al. (2012) found that a slight majority of the CV papers surveyed pass the weak scale test. However, there are few studies that allow any kind of assessment of the adequacy of scale response, as we attempt here.

The findings in Andersson and Svensson (2008) and Leiter and Pruckner (2009) were what led to this paper's hypothesis tests. Andersson and Svensson (2008) found that individuals with a higher cognitive ability index were more likely to exhibit sensitivity to scale in both the weak and strong form tests. Leiter and Pruckner (2009) found that past experience with the risk reduction on offer is a significant factor when assessing sensitivity to scale. The results from this paper confirmed both of these findings, and suggest that the effect of past experience with the risk may be more important than cognitive ability when assessing scale sensitivity. The first section will provide some background information on the survey method and issues being considered. The second section will describe the survey and its implementation. The third section will discuss the methodology, and the final section will look at results and the suggestions for future research.

1 Background

Contingent Valuation is a survey-based stated preference technique used to elicit the value of a non-market good or resource. The basic idea behind CV is to create a realistic hypothetical scenario that introduces some policy or good to the respondents. If the respondents believe that they will be better off by purchasing this good or (voting for this policy that comes with a tax increase) then they will respond with a “yes” or will state a non-zero willingness-to-pay (WTP). A follow-up question about why the yes or no options were chosen was suggested (Arrow et. al 1993). Follow-up questions allow researchers to gain insight into the reasoning process behind the yes/no response. Any protest responses can be identified here as well. Reminding individuals of the budget constraint serves to reduce the hypothetical nature of the survey, since the individual knows they are not actually required to make any form of payment. Testing for responsiveness to scope is a way of assessing the validity of the CV method in general and for a particular survey as well. If the item being valued is a good, then individuals should place a higher value on more of a good than less (Arrow et al. 1993). Though it has been shown that insensitivity to scope can be consistent with certain economic preferences (Heberlin et. al 2005), this is not always true. For example, it has been shown that WTP for health risk reductions should be strictly increasing and concave (**?**).

The test for responsiveness to the scope of the damage, also known as scale sensitivity, is considered to be an essential aspect in the test for the reliability of a CV survey (Arrow et al. 1993). The reason for the importance of scale sensitivity here being the fact that people should typically place a higher value on more of a good than on less of a good. There are two forms of scale sensitivity identified in the literature (**?**). The weak form is simply stating a higher WTP for a larger risk reduction. Hammitt (2000) shows that for WTP responses to be valid for small mortality reductions they must be "near-proportional", increasing with the size of the risk reduction and strictly concave. This would imply that respondents are expected to place just under double the value on a risk reduction that is twice as large. This near-proportionality is referred to as the strong form of scale sensitivity.

The hypothesis, first posited in Andersson and Svensson (2008), is that respondents scoring highest on cognitive ability questions will be more likely to show sensitivity to scale. It has been noted that difficulty interpreting small probabilities could be a possible reason that scale bias persists in CV studies (Carson 2012), the reasoning here being that people with better computational skills will be less likely to use heuristics. Shah and Oppenheimer (2008) define heuristics as mental shortcuts to reduce the effort associated with a task. We expect that a more calculated approach to valuing small risk reductions will be more sensitive to small changes in risk size. Alternatively, it could be argued that someone with poor numerical and computational skills will be more likely to be placing a value on the solution to the overall issue (MVC) rather than the particular risk reduction offered. This could be interpreted as a framing error in that the respondent is including irrelevant information to the question being asked (**?**). This paper attempts to further explain the factors that influence sensitivity to scale so that future CV surveys can be designed better and elicit more reliable responses.

1.1 Risk Perception

A common risk mis-perception identified in the literature is that individuals tend to overestimate death risk due to low probability events and underestimate death risk due to high-probability events (**?**). Since a MVC is a low probability event we might expect individuals to overestimate their own death risk. However, familiarity with a risk and degree of perceived control (**?**) tend to reduce perception of a risk. Since driving on the highway is both familiar and something individuals tend to feel a lot of control over (driving), this should reduce the risk perception. Andersson (2011) found that females tend to over-assess their own road traffic risk perceptions and males tend to under-assess their own traffic risk. This may explain why females tend to be willing to pay more than males for traffic risk reductions (?). Overall, it is expected that most people will understate their risk, due to the high degree of familiarity and control, and that younger males will overstate risk more than others.

Several studies have found that the WTP for a private risk reduction to be much higher than for a public risk reduction of the same nature and magnitude (Johannesson et. al 1996; Hultkrantz et. al 2006; Svensson and Vredin Johansson 2010). Thus, we might expect the private provision of the good to elicit a higher WTP. However, the nature of the difference in the private and public scenarios may in fact cause the public provision WTP to be higher than the private. The public good is a risk reduction for everyone on the highway, while the private good is a risk reduction applied exclusively to the driver. If we assume that Newfoundlanders value the safety of others and are willing to pay for it, then the publicly provided good should elicit a higher WTP, all else being equal.

1.2 Scale Sensitivity

Insensitivity to scale is a major criticism of the CV method. Carson (2012) suggests the difficulty in understanding and valuing small probabilities (as evidenced in financial planning scenarios) as a possible reason for the lack of scope sensitivity in CV studies. The use of visual aids has been proven to help in abating this problem (Corso et. al 2001) while others have found that the level of education of the respondent influences the extent to which these aids help (?). Another reason could be that the expected utility model is not valid for the way individuals form valuations (?). Yet another explanation for lack of scale sensitivity in WTP surveys is the exclusion of relevant qualities when carrying out sensitivity analysis (Heberlin et. al 2005). For example, when attitudinal factors are included in one study, the study goes from failing the stong scale test for near-proportionality to passing (?). Carson (1994) conducted a review of 27 contingent valuation surveys and found that all but two of them showed significant scope effects on WTP. It was noted that there are several issues with the methods used in these studies as well. The use of open-ended format, problems with the provision of information, and the lack of random sampling are a few examples (?). Also in one of the problematic studies conduceted by DesVousges et. al, once outliers are removed (as is standard practice), a significant scope effect appears. Whether or not these few studies that do not show scope effects are valid, the majority of CV studies do show significant scope effects.

Internal tests of sensitivity involve testing how a single individual responds to risk reductions of different sizes. Internal tests have been criticized as being a weak assessment tool (Arrow et al. 1993), since we would almost certainly expect the same individual to place a different value on two different amounts of a good. External tests of sensitivity involve comparing the mean/median WTP from sub-samples that received different quantities of the good, a risk reduction in this case. These tests are more rigorous in that the varying bid amounts are presented to independent respondents, and the test then considers the values placed on different quantities of a good by different respondents.

Desvousges et al. (2012) have recently criticized the typical application of scope tests in that there is no assessment of the adequacy of response to scope. In their review of 109 CV surveys that conducted scope tests (the majority of which passed the basic scope test), they found only three that permitted an assessment of adequacy of response to scope. The method they implement in these tests is for when goods are incremental to others. For example, suppose good B is made up of good A plus an added benefit and good C includes both A and B. The WTP for A and B combined should be approximately equal to the WTP for C. This is referred to as the "adding up test" which must hold for preferences to be meaningful (**?**). The authors point out that the typical scope test only assesses if the valuation of different levels of the good is statistically different. The test is looking for simply a non-zero scope effect. A test for adequacy of response to scope must include some measure of what the theoretical expectation of the response should be and a test statistic from the sample to compare with this expectation. This will be discussed further in the methodology section.

Hultkrantz et al. (2006) found in a 1998 Swedish data set on WTP for road traffic accident risk reductions insensitivity to scale over the full sample. However, when considering the most confident respondents only, there was weak evidence of scale sensitivity. There was a higher proportion of "yes" responses to the larger risk reduction, and the WTP was near-proportional to the risk reduction in the most confident group. This result implies certainty levels may also be an explanatory factor for scale sensitivity.

Leiter and Pruckner (2009) conducted a CV study that assessed the WTP for a reduction in the risk of death due to avalanches in a the region of Tyrol, Austria. The focus of this study was how attitudinal factors and past experience relate to the sensitivity to scale. Respondents were asked about their past experience with avalanches, including the effect of knowing someone who has been in an avalanche incident. The attitudinal factors included things such as their own perception of an avalanche risk, whether avalanches were natural or anthropogenic, and their degree of risk aversion. Personal factors such as skiing, sports involvement, working a risky job, and smoking habits were also included. Using a weibull distribution assumption, the regression was conducted with and without attitudinal factors. Two versions of each were done, one with "learners" and one without, for a total of four regressions. The "learners" were considered to be those who were proficient with small probabilities. The external test for sensitivity to scale indicated that the ratio of WTP's was non-proportional when not controlling for attitudinal factors in both learners and non-learners groups. The results indicated that the weak scale test of sensitivity was passed (the ratio was statistically different from one) and the strong scale test failed. Interestingly, once attitudinal factors were controlled for, the strong scale test passed for both learners only and non-learners included. This is fairly strong evidence that attitudinal factors and past experience are

Table 1: Variable Definition

Variable	Variable Type	Description
male	Index	= 1 if male; 0 otherwise
publicgood	Index	= 1 if public good; 0 if private good
hitmoose	Index	= 1 if experienced MVC or near-miss; 0 otherwise
SUV	Index	=1 if drives SUV or truck; 0 otherwise
childrenany	Index	=1 if children under the age of 18 are present in household; 0 otherwise
drives30towork	Index	=1 if drives more than 30 Km to work
job12to6am	Index	= 1 if job involves driving between 12 and 6 AM; 0 otherwise
NLander	Index	= 1 if from Newfoundland; 0 otherwise
agegroup	Categorical	= 1 if aged 19 to 34, 2 if aged 34 to 50, 3 if aged 51 to 65, 4 if 65 or older
education	Categorical	= 1 if , 2 if, 3 if, 4 if, 5 if, 6 if, 7 if, 8 if, 9 if missing
mathscore	Categorical	= number of the four math questions correctly answered
income	Categorical	= 1 if , 2 if, 3 if, 4 if, 5 if, 6 if, 7 if, 8 if, 9 if missing
KMyear	Continuous	= self-reported Km driven per year
health	Continuous	= self-reported health status on a scale of 0-100
diffM	Continuous	= Absolute size of mortality risk reduction
WTP	Continuous	= Model estimated willingness to pay

important factors to control for in WTP elicitations.

The most relevant paper to this study is Andersson and Svensson's (2008) study, which looked into how cognitive ability affects the presence of scale bias in contingent valuation surveys. The experiment was carried out on 200 university students in Sweden. They were given 5 minutes to answer 17 questions focused on probabilities, syllogisms and computation, all of which would be relevant when answering a WTP question. They were then given a DBDC-CV survey, followed by questions about demographics. The results showed that respondents scoring higher on the cognitive ability questions showed less scale bias. In particular, questions pertaining to probability skills were significant rather than purely computational methods. They also show that internal tests, typically assumed to be irrelevant since they would obviously pass scale tests, are quite possibly valid as 36.5% showed no scale sensitivity whatsoever. Using factor analysis, the 17 questions were combined into three representative factors: probability, computational, and intuition. In the test for weak scale sensitivity both the computational and probability factors were significant and positive. This means that individuals scoring higher on these factors were more likely to exhibit weak scale sensitivity. Their test for strong scale sensitivity involved the ratio of WTP from the low risk reduction (4 out of 100 000) to the high risk reduction (6 out of 100 000). For the test for strong sensitivity to scale (near-proportionality) only the probability factor was significant. This evidence supports the theory that individuals with a higher level of cognitive ability are expected to show a higher degree of sensitivity to scale.

Table 2: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
age	51.877	13.629	19	85	1383
drives30towork	0.173	0.378	0	1	1568
education	4.847	2.33	1	9	1548
childrenany	0.327	0.469	0	1	1560
hitmoose	0.856	0.351	0	1	1568
income	5.102	2.869	1	9	1568
job12to6am	0.127	0.333	0	1	1435
KMyear	20121.816	23382.136	0	300000	1568
male	0.494	0.5	0	1	1568
mathscore	1.535	0.936	0	4	1568
publicgood	0.495	0.5	0	1	1568

2 The Survey

The survey was completed via telephone using random digit dialing on the island of Newgoundland in the spring of 2013. The survey was initially tested by the researchers and a small group of pre-testers. A field pre-test of 150 respondents was conducted, before a further 1207 respondents were questioned in the main survey. Due to the lower than average proportion of respondents in the 19-30 age range after the pre-test, the survey was modified to ask to speak with the youngest adult in the household. The total number of respondents was 1357 and the response rate was 18.88%.

Respondents were asked specific questions about their driving habits, experiences with moose, and own risk perceptions of a MVC. Typical questions on sociodemographic factors such as age, education, gender, health status, and income were asked. In addition to these typical questions asked in a CV survey, this survey included four questions assessing skills in interpreting decimals and fractions and numerical computation. The intent of these questions is to assess how mathematical and computational skills affect various aspects of the WTP distribution. In this case we were interested in the sensitivity to scale. Each respondent was presented with a random average MVC mortality and injury risk. MVC mortality risks were 4, 6, 8, 10, or 12 out of 100 000. Injury risks were 30 times those amounts to be 120, 180, 240, 300, or 360 out of 100 000. Respondents were then asked to estimate their own MVC mortality and injury risks, given the average risk in Newfoundland. Then their own estimated risk was used as the baseline risk to be reduced by the policy implementation for the private good.

The respondents were presented with a hypothetical public or private method to reduce their risk of death or death and injury due to moose-vehicle collision. A randomly chosen divisor of 2, 3, or 4 was used to reduce the injury and mortality risks. The intent of the emphasis on the device protecting only the driver is to isolate the value placed on the driver's own mortality and injury risk reduction resulting from a MVC making it a pure private good. Version A offered the private good for a reduction in both risk of injury and death in a MVC, and version B offered a private reduction only in mortality risk. It is expected that version A will elicit a higher WTP, since it offers both an injury and a mortality risk reduction, and is also a more plausible scenario. In this study, we focus only on death risk reductions.

The public good was described as fence installed along the highway, implemented by either the federal or provincial government and would last for five years, with the required payment occurring every year. The randomly assigned federal/provincial provision was included to avoid framing effects and payment vehicle bias. Framing would occur if individuals had a tendency to mistrust one level of government more than another. This could result in a biased WTP. Payment vehicle bias occurs when one payment vehicle (higher income taxes) elicits a higher or lower WTP than another (increased license fees). The major difference between the public and private provision is the fact that the risk reductions are now extended to all users of the highway, including passengers in the vehicle. Since this is a public good, the randomly chosen average risk reduction was the value used (rather than the individual's stated risk). This again was produced by using a randomly selected divisor of 2, 3, or 4. Additionally, 50% of the questionnaires in the public good versions received a referendum reminder, explaining that the policy would only be approved if the majority of Newfoundlanders voted for it.

Respondents were then asked "Would you be willing to pay $X for this device/program?" An initial bid was selected at random from $15, $30, $45, $60, $75, $100, $120, or $150. Based on the initial response, a follow-up question was asked with a doubled bid given an initial "yes" or a halved bid given an initial "no". After the WTP questions, respondents were asked to assess their level of certainty about their WTP responses. It has been shown in experimental studies that very certain "yes" responses are much better predictors of actual WTP than less certain responses (Johannesson et. al 1998). This implies considering certainty in responses can produce more reliable WTP distributions. The method used in this paper is to focus on only the respondents who were most certain in their responses, as these are the most reliable predictors of actual WTP. Also, if respondents were unwilling to pay any of the offered amounts they were asked about the reasons why in order to screen protest responses.

The double-bound dichotomous choice method (DBDC) was selected in accordance with the NOAA Panels recommendation. The dichotomous choice aspect mimics the take-it-or-leave-it reality consumers face when making actual market choices or when having to vote in a referendum on a particular piece of legislature (Arrow et al. 1993). The double bound method means a follow-up question is given based on the response to the first. Given an initial "yes" ("no") the follow-up bid is doubled (halved). This significantly improves the efficiency of WTP models as shown by Hanemann (1991). For a sequence no-no or yes-yes responses, the second response gives a narrower interval on which the WTP lies. For the no-yes or yes-no responses, the WTP is now bound from above (the "no" bid level) and below (the "yes" level).

3 Methodology

In this section we analyze the factors affecting respondents sensitivity to scale for the stated WTP responses. The WTP was estimated by writing a maximum likelihood estimation (MLE) routine in Stata 12. The main model was an interval double-bounded probit model. This model assumes that the biases present in DBDC modeling (Alberini et. al 1997; Farmer and Belasco 2011; McFadden 1994) do not affect the main analysis pertaining to scope effects. It should be noted that the single-bounded model was tried but the loss of the efficiency gains resulted in most variables becoming statistically non-significant.

3.1 Data Screening

Given the hypothetical nature of the survey, not all responses are going to convey meaningful information. Any stated mortality risk perception greater than 100 out of 100 000 or equal to zero were omitted. The justification for this is that respondents stating these values (that are orders of magnitude larger than the average risk) are likely unable to fully interpret and place a meaningful value upon small changes in risk. Responses were also screened for protest responses. Protest responses are responses that are seriously biased by the way in which the good is being valued or the proposed program is implemented. Protest responses are typically removed because "it is assumed they are not indicative of respondents' true values" (Jorgensen et. al 1999, p. 131). Some examples of protest responses are if the respondent feels the agency implementing would be wasteful or if the respondent doesn't think the WTP method should be used. Please refer to Appendix III for a complete list of the protest responses screened.

In order to obtain data that is representative of the overall population, sampling weights were used in accordance with standard practice (**?**). The sample weighting method used was to identify the proportion of individuals belonging to a category, based on age and education level. This data was obtained from Statistics Canada. The proportion of respondents in a given category in the survey was then compared with the sample proportion to obtain a sampling weight:

$$sampleweight = \frac{Proportion_{population}}{Proportion_{sample}} \tag{1}$$

Therefore, if a given group is over-represented in the sample, they will receive a weight less than one. If a group is under-represented in the sample, their weight will be greater than one. There was very little change in the model once sampling weights were used, suggesting the sample did a fairly good job of obtaining a sample representative of the population.

3.2 WTP Surveys

Some typical statistically significant variables in risk reduction WTP surveys are income, being female, being married, and being younger (Krupnick 2002; Cameron et. al 2010). The major sociodemographic independent variables included in the regression are age, gender, and previous personal experience of a MVC. As previously mentioned, these variables tend to have an expected effect on WTP elicitation for various reasons. Age, education, the presence of children, and gender were significantly related to the math score in an ordered logit regression. These variables also typically influence risk perception, which is expected to have an effect on WTP. Other variables that were tested in the model but found to have a non-significant effect on WTP were education level, kilometres driven per year, public good, self-reported health status, and the presence of children in the household. The level of risk reduction (*diffM*) and its squared component (*diffM^2*) are the variables of focus in assessing the scope test. Since WTP is expected to be concave in risk reduction, the squared term is a necessary inclusion.

As mentioned earlier, Hammitt (2000) showed that WTP for small reductions in death risk should be near-proportional to the risk reduction and strictly concave. This means that we should expect WTP to increase in near proportion with the size of the risk reduction and at a decreasing rate. There are several expected influences on the WTP function, summarized in the Hammitt (2000). WTP should unambiguously increase with income, as more disposable income implies more money to spend on risk reductions. The effects of increasing baseline risk seem to create a lower WTP for risk reductions, identified as the "dead anyways" effect. If an individual is likely to die in the next few years due to some other reason or illness, we should expect them to be willing to pay less than the average individual, controlling for other factors. A self-reported health and/or smoker index are often used as proxies for this baseline risk. The effects of age on WTP for risk reductions are ambiguous. Some have found a strictly decreasing effect of age on WTP. This is because as one ages, there are fewer years of quality life left. This is similar to the dead anyways effect. Given that the opportunity cost of consumption decreases with age, there could be an inverted U-shaped effect (**?**). Jones-Lee (1974) has shown that WTP for a decrease in risk increases with the initial risk. In other words, those individuals with a higher than average risk of a MVC should be willing to pay more than average.

3.3 Maximum Likelihood Estimation

The MLE code was first built as a regular DBDC interval model. Subsequently, the WTP parameter was broken down to allow for modeling of the coefficients in the WTP parameter as functions of covariates. It is here that this method diverges from other methods. Specifically, the coefficients on *diffM* and *diffM^2* were estimated as functions of covariates such as age group, math score, gender, public good, driving an SUV or large truck and having a near miss or hit a moose with a vehicle before. This is the same as modeling with interaction terms on *diffM* and *diffMsq*, however this method allowed for a clearer exposition of the effects of individual variables.

$$WTP = \beta_0 + \beta_1 \mathit{diffM} + \beta_2 \mathit{diffM}^2 + \epsilon_j \tag{2}$$

$$\beta_0 = \theta_0 + \theta_1 male + \theta_2 publicgood + \theta_3 agegroup + \theta_4 income + \theta_5 SUV + \epsilon \tag{3}$$

$$\beta_1 = \theta_6 + \theta_7 male + \theta_8 publicgood + \theta_9 agegroup + \theta_8 hitmoose + \theta_9 mathscore + \epsilon \tag{4}$$

$$\beta_2 = \theta_{10} + \theta_{11} male + \theta_{12} publicgood + \theta_{13} agegroup + \theta_{14} hitmoose + \theta_{15} mathscore + \epsilon \tag{5}$$

In terms of the scope test, a weak form scope test is to test whether $\frac{\partial WTP}{\partial diffM}$ is statistically different from zero. In this model, the test would be:

$$\frac{VSL = \partial WTP}{\partial diffM} = \beta_1 + 2\beta_2 \mathit{diffM} = 0 \tag{6}$$

The strong form of the scope test involves a more complex calculation. First, the WTP model is estimated using the MLE routine. From these values, we can then calculate the value of a statistical life (VSL) for different groups. The VSL is the price an individual is willing to pay for a given reduction in the risk of death divided by that risk reduction. We can calculate VSL (as estimated by the model) under various specifications and as *diffM* increases. A (more) constant VSL implies (more) proportionality as shown in Equation **??**, which is the equivalent of testing the second derivative of WTP with respect to *diffM* to be zero.

$$\frac{\partial VSL}{\partial diffM} = 2\beta_2 = 2(\theta_{10} + \theta_{11} male + \theta_{12} publicgood + \theta_{13} agegroup + \theta_{14} hitmoose + \theta_{15} mathscore) \tag{7}$$

Proportionality requires that:

$$\frac{\partial VSL}{\partial diffM} = 0 \tag{8}$$

Table 3: Risk Perception OLS Models

Variables	Model 1	Model 2
age	-0.041*	-0.028
hitmoose	1.224	1.368*
drives30towork	1.711**	1.564**
education	-0.165	-0.270**
cartype1	0.936	
NLander	-0.609	
male	-0.344	
smoker	0.457	
childrenany	-0.834	
job12to6am	0.133	
income	-0.107	
Constant	0.397	-0.934
N	938	1020
R^2	0.021	0.016

$*p < 0.10, **p < 0.05, ***p < .01$

which implies that:

$$\beta_2 = \theta_7 + \theta_8 male + \theta_9 hitmoose + \theta_{10} agegroup + \theta_{11} mathscore = 0 \tag{9}$$

4 Results

4.1 Modeling Risk Perception

In order to understand the determinants of respondents' own risk perception (a significant determinant of *diffM*), an Ordinary Least Squares (OLS) model was constructed. The dependent variable was *perception*, which was the difference between the self-reported traffic mortality risk and the average mortality risk presented to that particular respondent.

$$perception = owndeathrisk - averagedeathrisk \tag{10}$$

Therefore, a negative value for perception implies the respondent believes their own risk is below the average and a positive value for perception means the respondent believes their risk is above average. Thus, in the OLS regression identified in Equation **??**, positive coefficients will be factors that tend to increase the perceived risk and negative values will tend to reduce the risk.

Table 4: Hit Moose Logit Models

Variables	Model 1	Model 2
age	-0.021**	-0.012*
drives30towork	2.044***	2.019***
Avalon	-0.959***	-0.951***
KMyear	0.000*	0.000**
huntedmoose	0.995***	1.117***
atemoose	0.887***	0.773***
childrenany	-0.299	
cartype3	-0.462	
knowselse	1.087***	1.073***
mathscore	-0.188*	-0.157
spentless5	-1.178***	-1.250***
perception	0.000	0.000
Constant	2.391***	1.662***
N	1032	1101
Pseudo R^2	0.021	0.016

$*p < 0.10, **p < 0.05, ***p < .01$

$$perception = \beta_0 + \beta_1 agegroup + \beta_2 hitmoose + \beta_3 drives30towork + \beta_4 education + \beta_5 SUV + \beta_6 NLander + \beta_7 male + \beta_8 smoker + \beta_9 health + \beta_{10} childrenany + \beta_{11} job12to6am + \beta_{12} mathscore + \beta_{13} income + \epsilon \quad (11)$$

The significant variables in this regression (shown in Table **??**) were *agegroup*, *hitmoose*, *drives30towork*, and *education*. Surprisingly, *agegroup* was negative and significant at the 5% level. The typical finding is that younger groups tend to understate their risk perception more than older groups. This anomaly could be due to confidence in the control of the situation (driving) that comes with experience (**?**). Having hit a moose before tended to increase the risk perception, which makes intuitive sense. Anyone who has had experience with moose on the highway will see the risk as more credible. Having to drive more than 30 Km to work was a positive variable that was significant at the 1% level. This suggests that individuals who spend more time driving back and forth to work perceive their own risk as being higher. This makes intuitive sense in that individuals who are more exposed to the risk likely do have a higher than average risk. Education was significant at the 5% level and was negative. This means that more educated people tended to perceive their own risk as being lower than average.

The variable *male* was negative as expected and almost significant at the 10% level. In other models, it does become statistically significant at the 10% level. Being from Newfoundland also was negative and almost significant at the 10% level. This suggests that perhaps a higher awareness of or familiarity with the risk at hand means a lower risk perception (**?**). The *mathscore* variable was negative and also almost significant at the 10% level, suggesting that individuals with better numerical skills tend to state a lower risk perception. Interestingly, the *SUV* variable was almost significant at the 10% level and was negative. This suggests that, while controlling for other factors, respondents who drive SUV's or trucks tend to state a higher perceived risk. Initially this may seem counter-intuitive in that individuals who drive a larger vehicle would likely have a lower risk of dying in a MVC. However, the fact that they perceive their risk as being higher could explain the fact that this individual drives a larger vehicle. The other variables that were tested but statistically non-significant at the 10% level were the presence of children in the household, driving frequently between 12 and 6 AM, reported Km driven per year, and income.

In summary, the statistically significant factors that increased the perceived mortality risk perception were having hit a moose and having to drive more than 30 Km to work. Driving an SUV or truck was not significant but was positive. The significant factors that reduced the perceived mortality risk were agegroup and education. The nearly significant and negative factors were male, being from Newfoundland, math score and income.

4.2 Modeling the Index of Cognitive Ability

The *mathscore* variable was based on a series of four questions that were asked at the end of the survey. The questions assessed math skills such as numerical computation and understanding small fractions and decimals, both of which are skills used when assessing various mortality risk scenarios. The *mathscore* term was simply the sum of correctly answered questions by the respondent. For the complete list of questions, please refer to Appendix II.

We first model *mathscore* in an ordered logit regression to assess what covariates affect cognitive ability as measured here by *mathscore*. Both models in Table **??** show high overall significance as the likelihood ratio test (LR) has a p-value of 0.000. The proportionality of odds test is to see if the parameters have a consistent effect on the outcome as *mathscore* ranges from 0 through 4. The rejection of the null in the proportionality of odds test Model 1 suggests that the effect is not consistent across all pairs of categories and an ordered logit model may not be the best choice. Model 2 fails to reject the assumption of proportional odds, suggesting the ordered logit model is an appropriate choice. As shown in Table **??**, the significant covariates are *male*, *age*, *education*, and the presence of children in the household. The coefficient on *male* was significant at the 1% level and positive. This is in line with the typical results that males, on average, have slightly better numerical skills. Age has a negative effect and was statistically significant variable also at the 1%. This was also as expected, since these types of cognitive skills tend to decline with age. Education was significant and positive at the 1%, which makes intuitive sense in that more education would mean better numerical skills. Finally, the presence of children in the household was significant at the 5% and positive. This somewhat surprising result does make sense if we

Table 5: Math Score Ordered Logit Model

male	1.304***
maleeduc	0.026
maleage	-0.015**
age	-0.013**
education	0.255***
childrenany	0.288**
monthofbirth	0.019
income	0.025
cut 1	-0.753*
cut 2	1.051**
cut 3	3.188***
cut 4	5.988***
N	1345
AIC	3379.5
LR Test	251.98***
Proportionality of Odds Test	32.62

$*p < 0.10, **p < 0.05, ***p < .01$

consider the fact that parents who help their children with math homework are routinely exposed to dealing with fractions and decimals, which is exactly what these questions in the survey are assessing. Month of birth, self-reported health status, kilometres driven per year, and the smoker indicator were all statistically non-significant at the 10% level.

$$mathscore = \beta_0 + \beta_1 male + \beta_2 age + \beta_3 education + \beta_4 childrenany + \beta_5 monthofbirth + \beta_6 income + \beta_7 health + \beta_8 KMyear + \beta_9 smoker + \epsilon \quad (12)$$

4.3 Modeling WTP

As shown in Equation **??** WTP is modeled as a function of *diffM* and *diffM*^{2}. The terms β_0, β_1 and β_2 are themselves modeled as functions of covariates, in order to allow us to assess what are the factors that determine the degree of sensitivity to scale.

β_0 can be thought of as the constant component in the WTP equation. The significant factors in the initial model for β_0 are *male*, *agegroup*, and SUV. The *male* coefficient was significant at the 10% level and negative. This result implies that males tend to be WTP about $50 less than females on average. The coefficient on *agegroup* was statistically significant at the 10% level and also negative. This result implies that, all else equal, each step up in the *agegroup* level decreases the WTP by $38.75. The SUV index for if the individual drove an SUV or truck was significant at the 5% level and positive, implying that individuals who drive those types of vehicles were on

Table 6: WTP Models

	Model 1	Model 2
β_0		
male	-16.210	-40.581
drives30towork	38.571*	
cartype1	35.215**	
publicgood	-44.331	
income	5.029	
Constant	27.782	59.360**
β_1		
mathscore	4.517	10.829**
hitmoose	73.127**	87.967***
male	-3.673	14.929
malehit	0.749	
age	-0.379	-0.080
publicgood	10.492	
Constant	-42.458	-78.534***
β_2		
mathscore	-0.005	-1.023
hitmoose	-9.328*	-14.467***
male	12.419	-0.779
malehit	-12.749	
age	-0.002	-0.042
publicgood	-0.679	
Constant	8.740	16.600***
σ	142.691***	162.1519***
N	463	573
aic	1263.093	1403.352
LR test	525.97***	193.90***

$*p < 0.10, **p < 0.05, ***p < .01$

average WTP \$34.90 more than respondents who do not drive those larger vehicle types. These results are consistent with what was expected from the risk perception modeling in that the kinds of respondents who perceived their own risk as being higher (lower) were WTP more (less) than others. The effect of the public good version was positive, as expected, but was not significant at the 10% level, suggesting that the respondents are indifferent between the two provision scenarios. Income was positive as expected, and was also statistically insignificant at the 10% level. When moving to the reduced models, we exclude some of these covariates in order to focus on modeling the covariates in β_1 and β_2.

The β_1 term is the coefficient on the linear component of diffM in the WTP model. When β_1 is modeled as a function of covariates, we get some of the expected results. The coefficient on *mathscore* in β_1 is 12.31. This means that, all else equal, people with higher math scores tend to be more sensitive to scale. Each math score increases WTP by \$12.31 more per increase in diffM.

This result is support for the hypothesis that respondents with more cognitive ability are more sensitive to scale. Their WTP responses tend to increase more for an increase in *diffM* relative to other groups. This is initial support for the hypothesis that individuals with a higher cognitive ability index show more sensitivity to scale in that their WTP responses tend to increase more than other groups implying a closer to proportional increase. The *mathscore* coefficient in β_1 is significant at the 1% level. The attitudinal factor *hitmoose* was included to assess the impact of previous experience on sensitivity to scale, as in Leiter and Pruckner (2009). The positive and significant value of *hitmoose* in β_1 is initial support for the idea that that individuals who have hit a moose before are more sensitive to scale, as their WTP increases by \$47.84 to \$67.59 per increase in diffM, depending on the model selection. The *male*, *agegroup*, and *publicgood* coefficients in $beta_{1)}$ were positive but all non-significant.

The β_2 term in the WTP model is the coefficient on the squared component of *diffM*. The statistically significant and negative term on *mathscore* and *hitmoose* implies that the relationship between WTP and *mathscore* is non-linear and increasing at a decreasing rate. The agegroup coefficient was negative and while statistically non-significant in the full model, it did become statistically significant at the 5% level. This suggests that the relationship between WTP and agegroup is also increasing at a decreasing rate. The value of *male* in β_2 was not significantly different from zero, indicating that *male* has no statistically significant impact on scale sensitivity.

From these results, we can now develop our equations for WTP and VSL, and test for weak and strong scale sensitivity.

4.4 Weak Sensitivity Test

For the weak sensitivity test, we simply want to test if $\frac{\partial WTP}{\partial diffM} > 0$ and statistically significant. A positive and significant value of $\frac{\partial WTP}{\partial diffM} > 0$ indicates that individuals are sensitive to increases in the risk reduction. However, this does not give any indication of adequacy of response as identified by Desvousges et al. (2012). Since β_1 and β_2 are composed of covariates, we can develop tests for various groups and compare them. The following tests are based on Model 3 in Table **??** and are assessed at the 1% statistical significance level.

We will first consider the group that has had previous experience with moose on the highway ($hitmoose = 1$) shown in Table **??**. We allow mathscore to vary from $0 \leq mathscore \leq 4$ and assess the weak sensitivity at each level. Across all *mathscore* values, we see an increasing value of $\frac{\partial WTP}{\partial diffM}$ as diffM increases. This result implies that at lower probabilities sensitivity to scale is more prevalent. This can be explained in part by the fact that difficulty in interpreting small probabilities has been identified as a possible reason that insensitivity to scale exists in studies like this (**?**). Also, as *mathscore* increases, we see the trend of increasing weak scale sensitivity across all ranges of *diffM*, as evidenced by smaller p-values and larger positive VSL's. This can be explained by the hypothesis central to this study, that individuals with stronger numerical skills

Table 7: Weak Scale Test Results

Math Score	hitmoose	p-value	diffM	Weak Scale
0	1	0.3754	2	Fail
0	1	0.0432	5	Pass
0	1	0.0101	10	Pass
1	1	0.01098	2	Fail
1	1	0.0056	5	Pass
1	1	0.0026	10	Pass
2	1	0.0204	2	Pass
2	1	0.0006	5	Pass
2	1	0.0008	10	Pass
3	1	0.0038	2	Pass
3	1	0.0001	5	Pass
3	1	0.0005	10	Pass
4	1	0.0011	2	Pass
4	1	0.0000	5	Pass
4	1	0.0010	10	Pass

$*$ Fail because of a negative value

will exhibit a higher degree of sensitivity to scale.

For the group without previous experience with moose on the highway ($hitmoose = 0$) displayed in Table **??** we see a slightly different result. The values for $\frac{\partial WTP}{\partial diffM}$ are negative when *mathscore* and *diffM* are low, which is a rejection of the null that $\frac{\partial WTP}{\partial diffM} = 0$. There is definite evidence of a trend of $\frac{\partial WTP}{\partial diffM}$ increasing with diffM once again, with the test passing at all levels of *mathscore* when $diffM = 10$. As *mathscore* increases, we observe $\frac{\partial WTP}{\partial diffM}$ increasing across all levels of *diffM* and passing at lower levels of *diffM* when $mathscore = 4$. This too, is evidence that individuals with stronger numerical skills are more likely to show sensitivity to scale in terms of the weak test.

The main difference between the two groups is the fact that when $hitmoose = 1$ the weak scale test passes at all levels of *diffM* when $mathscore \geq 2$ and only at the highest levels of *diffM* when $hitmoose = 0$. This is evidence supporting the result suggested by Leiter and Pruckner (2009) that groups having previous experience with the risk reduction on offer are more likely to show scale sensitivity.

We can conclude from these results that weak scale sensitivity is determined not only by cognitive ability, but by previous experience with the risk being valued and the size of the risk reduction offered.

As shown in Figure **??**, where individuals have had previous experience with a MVC, individuals with a higher math score have a more constant sensitivity to scale, which implies proportionality. The Math Score = 0 group shows an increasing VSL suggesting that proportionality does not hold. In Figure **??**, where no MVC experience is present, there is an almost indistinguishable difference between the slope of the three lines. Also, the slope is clearly positive implying that these groups do

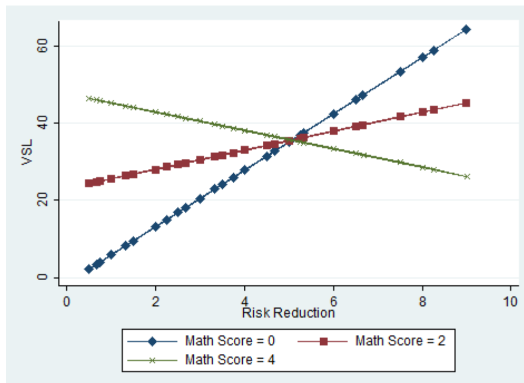

Figure 1: VSL vs. Risk Reduction when *hitmoose* = 1

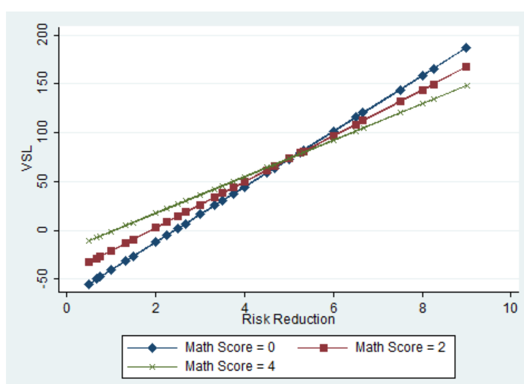

Figure 2: VSL vs. Risk Reduction when *hitmoose* = 0

Table 8: Weak Scale Test Results

Math Score	hitmoose	p-value	diffM	Weak Scale
0	0	0.0000	2	Fail*
0	0	0.6614	5	Fail
0	0	0.0146	10	Pass
1	0	0.0000	2	Fail*
1	0	0.2925	5	Fail
1	0	0.0376	10	Pass
2	0	0.0000	2	Fail*
2	0	0.0991	5	Fail
2	0	0.0000	10	Pass
3	0	0.2838	2	Fail
3	0	0.0324	5	Pass
3	0	0.0000	10	Pass
4	0	0.5876	2	Fail
4	0	0.0130	5	Pass
4	0	0.0000	10	Pass

$*$ Fail because of a negative value

not show scale sensitivity, as supported by the hypothesis testing carried out in Tables **??** and **??**.

4.5 Strong Sensitivity Test

As shown earlier, strong sensitivity to scale implies a constant VSL across values of diffM. Thus, the variable of interest here is β_2. Strong sensitivity to scale implies a value of $\beta_2 = 0$ and a constant VSL. Since failing to reject the null would mean a pass of the strong scale test, these tests will be performed at the 5% statistical significance level.

Table 9: Strong Scale Test Results

Math Score	hitmoose	p-value	Strong Scale
0	0	0.0001	Fail
1	0	0.0001	Fail
2	0	0.0002	Fail
3	0	0.0006	Fail
4	0	0.0020	Fail
0	1	0.418	Pass
1	1	0.625	Pass
2	1	0.966	Pass
3	1	0.640	Pass
4	1	0.362	Pass

Again, we will first consider the group with previous experience with moose on the highway ($hitmoose = 1$). As shown in Table **??**, we reject the null that $\beta_2 = 0$ for the groups with $mathscore \leq 2$ at the 5% level. For the groups with $mathscore = 3, 4$ we fail to reject the null hypothesis that $\beta_2 = 0$. This implies that the groups with the highest math scores show near-proportionality which is in line with our hypothesis.

For the group without experience with moose on the highway ($hitmoose = 0$) we reject the null hypothesis for $0 \leq mathscore \leq 4$. As with the tests on $\frac{\partial WTP}{\partial diffM}$, we reject the null that $\beta_2 = 0$ for all $0 \leq mathscore \leq 4$. The results are in line with expectations in that the higher math score groups show a β_2 closer to zero than groups with lower math scores. Additionally, this is further evidence that past experience with the risk reduction on offer is a significant factor when assessing scale sensitivity.

4.6 Conclusion

To summarize the major findings, we can conclude that cognitive ability, the size of the risk reduction offered, and past experience with the risk involved are all important factors when assessing sensitivity to scale. These results suggest that sensitivity to scale is not homogeneous across all groups and that this heterogeneity should be accounted for. Additionally, assessing respondents' past experience with the risk on offer and some measure of cognitive ability would be of importance in future risk reduction CV surveys. Future research in search of better communication aids for small probabilities would be valuable. Future studies could look at what specific aspects of cognitive ability are important in determining sensitivity to scale, and if past experience with similar risks would be significant. For example, in this case it might be relevant to ask if the respondent has been in a serious motor-vehicle accident. Also, researchers may be interested in replicating the survey design in terms of the range of risk reductions offered as it allows researchers to assess the adequacy of response in terms of scale sensitivity through the strong form test.

Appendix 1: Summary Statistics

Table 9: Distribution of response patterns by initial bid (CAD $) in %

	No-No	No-Yes	Yes-No	Yes-Yes	Total
$15	34.78	5.59	12.42	47.20	100.00
$30	31.37	9.80	19.61	39.22	100.00
$37.5	0.00	0.00	0.00	100.00	100.00
$45	44.14	5.52	22.07	28.28	100.00
$50	0.00	0.00	0.00	100.00	100.00
$60	41.58	10.53	23.16	24.74	100.00
$75	44.85	6.06	18.79	30.30	100.00
$100	44.77	8.14	16.86	30.23	100.00
$120	45.77	9.15	19.72	25.35	100.00
$150	50.75	8.21	20.15	20.90	100.00
$300	100.00	0.00	0.00	0.00	100.00
Total	42.06	7.91	19.05	30.99	100.00

Table 10: **Math Score**

Math Score	Number	Per cent
0	200	17
1	385	32
2	449	37
3	163	14
4	10	1
Total	1,207	100

Table 11: **Q12** ***How high do you think is your own risk of dying in a car accident involving a moose?*** **(out of 100 000)**

Own Risk	Number	Per cent
1	240	29
2	99	12
3	53	6
4	91	11
5	52	6
6	68	8
7	5	1
8	45	5
9	2	0
10	91	11
11	1	0
12	42	5
13	1	0
15	6	1
16	1	0
20	11	1
25	6	1
30	1	0
50	16	2
60	1	0
80	1	0
90	1	0
99	1	0
Total	835	100

Appendix II: Math Score Questions

27) What do you think is more likely to happen:

1. something that happens 3 times in 10,000 or
2. something that happens 6 times in 100,000?
3. don't know/no answer

28) Which number in the following group of numbers represents the smallest amount?

1. 3/4
2. 0.8
3. 31
4. 0.33
5. Don't know/no answer

29) A 20% reduction in a 40% risk level results in a new risk level of:

1. 32%
2. 20%
3. 35%
4. 80%
5. Don't know/no answer

30) A baseball bat and a ball together cost \$11. The baseball bat costs \$10 more than the ball. How much is the ball?

1. \$1
2. \$0.5
3. \$0
4. any other number
5. don't know/no answer

Appendix III: List of Reasons for Protest Responses

- I dont believe the money would be spent on that.
- I would not trust the government to do the job properly.
- It should not be financed through taxes/not everyone should have to pay their share to protect drivers.
- I should not have to pay individually: the province/government should pay for that without raising taxes.
- I do not believe that the program would be effective.
- The drivers should pay for that themselves.
- The drivers' insurance should pay for that.
- The government should fund the program with existing revenues, and not ask for additional taxes.
- Brush should be trimmed from roadsides to enable visibility.
- Moose population should be decreased/culled.
- MVC prevention should focus on driver safety/awareness.
- Need more information/proof/evidence of effectiveness.
- The problem exists because moose are not native to the area.
- Should not have to pay for the poor habits of other drivers.

Acknowledgement

I would first like to thank Dr. Roberto Martínez-Espiñeira for being a most patient supervisor and mentor in the writing of this paper. Always ready and willing to answer a question or help solve a problem. I would like to thank Dr. Martínez-Espiñeira and his colleague Dr. Nikita Lyssenko for allowing me the use of their novel data set. A big thanks to the entire Department of Economics Faculty and Staff at Memorial University for continually spurring my interest in the field of Economics and helping me realize my academic potential. Thanks to my colleague Aja Oki for his feedback and editing help. Finally, a big thank you to my partner, friends, and family for their endless support of my academic endeavours.

References

Alberini, A., B. Kanninen, and R. T. Carson (1997). Modeling response incentive effects in dichotomous choice contingent valuation data. *Land Economics 73*(3), 309–324.

Andersson, H. (2011). Perception of own death risk: An assessment of road-traffic mortality risk. *Risk Analysis 31*(7), 1069–1082.

Andersson, H. and P. Lundborg (2007). Perception of own death risk : An analysis of road-traffic and overall mortality risks. *Journal of Risk and Uncertainty 34*(1), 67–84.

Andersson, H. and M. Svensson (2008). Cognitive ability and scale bias in the contingent valuation method: An analysis of willingness to pay to reduce mortality risk. *Environmental and Resource Economics 39*(4), 481–495.

Armantier, O. (2006). Estimates of own lethal risks and anchoring effects. *Journal of Risk and Uncertainty 32*(1), 37–56.

Arrow, K. J., R. Solow, P. Portney, E. Leamer, R. Radner, and H. Schuman (1993). Report of the NOAA panel on contingent valuation. Technical Report 58: 4601-4614, Federal Register.

Boyle, K. J., W. H. Desvousges, F. R. Johnson, R. W. Dunford, and S. P. Hudson (1994). An investigation of part-whole biases in contingent-valuation studies. *Journal of Environmental Economics and Management 27*(1), 64–83.

Cameron, T., J. DeShazo, and P. Stiffler (2010). Demand for health risk reductions: A cross-national comparison between the u.s. and canada. *Journal of Risk and Uncertainty 41*(3), 245–273.

Carson, R., L. Wilks, and D. Imber (1994). Valuing the preservation of australia's kakadu conservation zone. *Oxford Economic Papers Vol. 46, Special Issue on Environmental Economics (Oct., 1994),*, 727–749.

Carson, R. T. (2012, September). Contingent valuation: A practical alternative when prices aren't available. *Journal of Economic Perspectives 26*(4), 27–42.

Corso, P. S., J. K. Hammitt, and J. D. Graham (2001). Valuing mortality-risk reduction: Using visual aids to improve the validity of contingent valuation. *Journal of Risk and Uncertainty 23*(2), 165–184.

Desvousges, W., K. Mathews, and K. Train (2012). Adequate responsiveness to scope in contingent valuation. *Ecological Economics 84*(C), 121–128.

Diamond, P. (1993). Testing the internal consistency of contingent valuation surveys. Working papers 93-1, Massachusetts Institute of Technology (MIT), Department of Economics.

Diamond, P. (1996). Testing the internal consistency of contingent valuation surveys. *Journal of Environmental Economics and Management 30*, 337–347.

Farmer, M. and E. Belasco (2011). A finite mixture model of heterogeneous anchoring with distinct anchoring patterns. *Applied Economics Letters 18*(2), 137–141.

Hammitt, J. (2000). Valuing mortality risk: Theory and practice. *Environmental Science and Technology 34*(8), 1396–1400.

Hanemann, M., J. Loomis, and K. B. (1991). Statistical efficiency of double-bounded dichotomous choice contingent valuation. *Agricultural and Applied Economics Association 73*, 1255–1263.

Hanemann, W. M. (1994). Valuing the environment through contingent valuation. *The Journal of Economic Perspectives 8*(4), 19–43. Symposia: Contingent Valuation.

Heberlein, T. A., M. A. Wilson, R. C. Bishop, and N. Schaeffer, Cate (2005). Rethinking the scope test as a criterion for validity in contingent valuation. *Journal of Environmental Economics and Management 50*(1), 1–22.

Hultkrantz, L., G. Lindberg, and C. Andersson (2006). The value of improved road safety. *Journal of Risk and Uncertainty 32*(2), 151–170.

Johannesson, M., P. Johansson, and R. O'Conor (1996). The value of private safety versus the value of public safety. *Journal of Risk and Uncertainty 13*(3), 263–275.

Johannesson, M., B. Liljas, and P. Johansson (1998). An experimental comparison of dichotomous choice contingent valuation questions and real purchase decisions. *Applied Economics 30*(5), 643–647.

Jones-Lee, M. (1974). The value of changes in the probability of death or injury. *Journal of Political Economy 82*, 835–849.

Jorgensen, B. S., G. J. Syme, B. J. Bishop, and B. E. Nancarrow (1999). Protest responses in contingent valuation. *Environmental and Resource Economics 14*(1), 131–150.

Krupnick, A. (2002). The value of reducing risk of death: A policy perspective. *Journal of Policy Analysis and Management 21*(2), 275–282.

Leiter, A. M. and G. J. Pruckner (2009). Proportionality of willingness to pay to small changes in risk: The impact of attitudinal factors in scope tests. *Environmental and Resource Economics 42*(2), 169–186.

McFadden, D. (1994). Contingent valuation and social choice. *American Journal of Agricultural Economics 76*, 689–708.

Peters, E. and I. Levin (2008). Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. *Judgement and Decision Making 3*, 435–448.

Shah, A. and O. Oppenheimer (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin 134*, 207–222.

Shepard, D. S. and R. J. Zeckhauser (1982, August). The choice of health policies with heterogeneous populations. In *Economic Aspects of Health*, NBER Chapters, pp. 253–298. National Bureau of Economic Research, Inc.

Sjoberg, L. (2002). Factors in risk perception. *Risk Analysis 20*, 1–11.

Sund, B. (2009). Sensitivity to scope in contingent valuation: Testing two aids to communicate mortality risk reductions. Technical report, Swedish Business School, Orebro University, Sweden.

Svensson, M. and M. Vredin Johansson (2010). Willingness to pay for private and public road safety in stated preference studies: Why the difference? *Accident Analysis and Prevention 42*(4), 1205–1212.

Yansaneh, I. (2003). Construction and use of sample weights. Technical report, United Nations Secretariat: Statistics Division.