TI 2012-046/3
Tinbergen Institute Discussion Paper

Scale, Scope and Cognition: Context Analysis of Multiple Stated Choice Experiments on the Values of Life and Limb

Marija Bočkarjova
*Piet Rietveld**
*Erik Verhoef**

Faculty of Economics and Business Administration, VU University Amsterdam.

** Tinbergen Institute.*

Scale, scope and cognition: context analysis of multiple stated choice experiments on the values of life and limb

M. Bočkarjova[1], P.Rietveld[1,2], E.T.Verhoef[1,2]
[1]Vrije Universiteit Amsterdam,
Faculty of Economics and Business Administration,
Department of Spatial Economics
[2]Tinbergen Institute

Abstract

In this paper we use data from an SP study on flood safety in the Netherlands, and elicit individual preferences for reduction of risk to life and limb. We perform context analysis to test the robustness of fatality risk valuation throughout choice experiments. The main interest of this paper is VOSL sensitivity to the valuation of correlated risks (scope effect). Besides, we explore the role of cognition on the stability of valuation across choice experiments using age and education. We pool data from multiple choice experiments and apply nested and mixed logit models in our analysis.

We confirm statistically significant sensitivity to scope, comparing VOSL estimates for the test group in a choice experiment where correlated risks were present (risks of fatality, injury and evacuation) to an experiment where only fatality risk is valued. We find that the origin of differences in VOSL valuations across the choice experiments lies in differences in age and educational attainment, and may therefore be related to cognitive abilities of respondents. In particular, we conclude that higher VOSL sensitivity to *scope* is most prominently present among respondents of senior age (65 and older) and respondents without college education. This finding has important implications for discrete choice modeling and the use of obtained values in cost-benefit analyses.

Keywords:

stated preferences, value of statistical life, value of statistical injury, value of statistical evacuation, flood risk.

1. Introduction

Value of statistical life (VOSL) is often used in cost-benefit analyses of policies to value potential effects on individual health and safety (Viscusi and Aldy, 2003). Because no explicit market exists for such goods as safety, VOSL can be indirectly derived using revealed preference (RP) methods,[1] or stated preference (SP) methods. In this paper, we shall concentrate on the SP methods, where by means of surveys individuals at risk are asked to indicate their willingness to pay (WTP) for a small safety improvement, or willingness to accept compensation (WTA) for some small increase in risk. A vast literature is devoted to discussion of the *pros* and *cons* of these methods (see for example discussions in Hanley et al., 1998; and Carson and Groves, 2007). One of the major remaining issues is how well the elicitation of preferences can reflect 'true' underlying values.

Among various SP techniques, discrete choice modeling (DCM) in particular has become popular in non-market good valuation in the past couple of decades. It is being increasingly used not only in marketing and transport studies, but also in areas of environmental good valuation and health economics. In other areas, such as hazard analysis, it remains a relatively infrequently used tool. Some of the few examples known to the authors are found in valuation of flood risk (Zhai and Ikeda, 2006; Bočkarjova et al., 2010), risk of avalanches (Leiter and Pruckner, 2006) and air pollution (Vassanadumrongdee and Matsuoka, 2005).

SP methods, as perhaps every inference method, have their weaknesses. An increasing amount of literature is being devoted to the issue of biases. One of these is scale sensitivity of WTP to the changes in the valued risk (in the fields of contingent valuation and conjoint valuation, see Hammit and Graham, 1999; Leiter and Pruckner, 2006). Further, effects of complexity of choice experiment are given attention (DeShazo and Fermo, 2002), as well as effects of the types of good valued – e.g. public vs private (Kahneman and Knetsch, 1992; Johannesson et al., 1996). Other biases include hypothetical bias (Svensson, 2009), payment vehicle (Hackl and Pruckner, 2005), information bias, interviewer bias, non-response bias, protest behaviour bias, strategic bias, symbolic bias, part-whole bias, yeah-saying bias, and reference point bias (see for example de Blaeij, 2003). With this paper, we wish to explore one of the biases pertaining to elicitation of individual preferences with SP methods that should bring us another step closer to answering the more general question of how accurate the obtained valuations are. In particular, we wish to know whether scope sensitivity would be present for the case when multiple risks related to a single event are valued. This is a new subject in hazard analysis; its results will not only be relevant for environmental good valuation or hazard research but also in health studies where DCM's are being increasingly used.

The issue of valuation of correlated risks is not yet extensively discussed in the valuation literature (perhaps except for the cases when it falls under analysis of choice complexity where multiple attributes are shown to have impact on risk valuation). Risks that are related to the same event (like risks of fatal and non-fatal injury due to a calamity) can be seen as "variety of the same commodity that is being valued"[2] and thus can be conceptually interpreted as an instance of scope bias, which is in turn well-known from the environmental valuation literature (Kahneman and Knetsch, 1992). We suspect that when correlated risks are present, VOSL is confounded if valued alone. In particular, in the case

[1] One of the RP methods is hedonic pricing, widely used in risk valuation in labour studies (see Viscusi and Aldy, 2003) but also applied in the valuation of natural hazard (Daniel et al., 2009).
[2] This notion is borrowed from Svensson (2009).

of flood calamity, we suppose that due to embedding of risks, respondents might implicitly value all immaterial risks connected to flooding in general when only fatal incidents are presented in a stated choice experiment, and value 'pure' VOSL when other risks are explicitly distinguished as well. So, a non-confounded (lower) valuation fatality risk (VOSL) would be expected in an experimental setting where other risks to live and limb connected to flooding are presented, such as evacuation and injury.

We analyse risk valuations from respondents in a 'test group' who have completed two sequential choice experiments to explore the presence of scope effects, and compare these results to these from respondents in a 'reference group', who have also completed two choice experiments, one of which was shared with the test group. This allows us to control for other possible effects connected to the experimental setting, such as type of good valued (public vs private), number of choice attribute valued, and the sensitivity to the scale of choice attributes (here: valued differences of the risk and the monetary attributes). Other biases, if present, should be expected to be the same among all our respondents, as the choice experiments in our study share such common elements as time of survey conduction, explanation of actual risk preceding the experiments, form of the choice cards, and the structure of the rest of the questionnaire. Literature offers context analyses for contingent valuation studies (see for example Andersson and Svensson, 2008); however, to our knowledge, this is a first study that provides such systematic analysis within conjoint analysis with a focus on scope (i.e., the valuation of correlated risks).

Another lead that we wish to pursue in this paper is the role of cognition, which has been studied both in the context of choice behavior (like Wierstra et al., 2001; Andersson and Svensson, 2008; Campbell et al., 2008), and for health-related behavior (like Auld and Sidhu, 2005; Cutler and Lleras-Muney, 2010). Notably, both types of studies have shown that cognitive ability is a significant determinant of individual behavior. In this paper, we intend to explore the relation between the stability of risk valuation and cognition of respondents: does higher cognitive ability lead to estimates that are less sensitive to the presence of related risks?

The paper is organized as follows: we shall first review economic valuation literature on the topics of experimental design and scope sensitivity. We shall then proceed with the description of the choice experiments and the methodology to be used. Next follows a description of data, and the analysis of results. Finally, conclusions and implications for research and practice are provided.

2. Literature overview and focus of the study

Measurement of genuine preferences, and eliciting true willingness to pay can be a daunting task due to all sorts of biases. An abundant literature addresses such biases in the field of contingent valuation; also studies on biases in DCM are growing in number as the method is being increasingly used in various contexts. Biases undermine the robustness and in some cases even the validity of valuations, as well as impede comparability of estimates obtained across different studies. Overly simplistically, the bottom line is that eliciting individual preferences with stated preferences methods is a complicated task where more or less 'everything matters' (see Reed Johnson, 2006; echoed by Hensher, 2010).

Two recent reviews of the state of research in DCM are important in the light of our research. Louviere and Lancsar (2009), as well as Hoyos (2010), sketch a general line of development in DCM research in the field of health risk valuation (which has its own valuation record; to name a few: O'Conor and Blomquist, 1997; Olsen and Donaldson, 1998; Lancsar et al., 2007; Bellavance et al., 2009). Essentially, a main message in the two review papers is that the quality of estimates based on DCM is much dependent on the experimental design (in the broad sense) and the cognitive ability of respondents to process the information provided in a questionnaire, and ultimately to translate that into appropriate choices.[3] The roles of the experimental design on the one hand, and cognitive abilities requested from respondents on the other hand, thus need to be studied in more detail in choice experiments.

Hensher (2006b) presents an attempt at a systematic empirical analysis of the impact of different design features on the WTP estimates (considering number of choice sets, number of alternatives in the choice sets, number of attributes per alternative, number of levels of each attribute and the range of attribute levels), which is important for considerations on statistical design behind choice experiments and WTP comparisons across choice experiments. Notably, he concludes that controlling for all design features, no systematic differences in WTP estimates are found due to a specific design dimension; however, when analysed separately, differences in aggregate mean WTP are found to be due to differences in the number of attributes per alternative and the number of alternatives in a choice set.

Two other papers, Foster and Mourato (2003) and Goldberg and Roosen (2007), focus on the theme of scope and scale in the context of environmental good valuation and in health risk reduction, and compare estimates obtained with the contingent valuation method to estimates from choice experiments. Both papers conclude that scope insensitivity problems of WTP persist in DCM, although the method possesses important advantages over the contingent valuation format.

It is important to note, just as other authors did (*inter alios*, Goldberg and Roosen, 2007; Svensson, 2009), that the terms scale, scope, embedding and nesting are often used loosely synonymously in the valuation literature, and therefore need to be clearly defined here. Similar to Norinder et al. (2001), we shall use *scale sensitivity*, perhaps taken most literally, to refer to a sensitivity of VOSL to the extent of change in a choice attribute, such as improvement in a single risk or change in tax to be paid. Scope sensitivity, rather, has to do with the extent of an alternative, and is therefore related to nesting or embedding. So, we shall use *scope sensitivity* as sensitivity of VOSL to the extent of the valued commodity, such as multiple risks related to the same event (in our case, an expected flood event bears risks of fatal or non-fatal injury, and a risk of preventive evacuation). In order to keep the distinction clear, we shall thus refer to the two terms as sensitivity to the scale of choice attributes and sensitivity to the scope of the valued good, respectively. In this paper, we put most emphasis on the analysis of scope. In addition, we shall explore the effect of cognition via the education level and age on the robustness of risk valuation obtained in multiple choice experiments.

The issue of the cognitive burden that is put on respondents when they are asked to fill out choice experiments, and how different respondents go about, is one of the intriguing issues in DCM. On the one hand, the valuation literature provides evidence (de Palma et al., 1994;

[3] Also in health studies cognition is a subject of extended inquiry, see for example Auld and Sidhu (2005) and Cutler and Lleras-Muney (2010).

Hensher, 2006a) that in order to simplify the complexity of considering all choice attributes, some respondents take 'shortcuts' by ignoring one or more of the attributes and so make trade-offs based on only one part of the presented information. To illuminate this, some researchers have been using debriefing questions after the choice experiments, asking respondents about the way choices were made (Campbell et al., 2008). Alternatively, other researchers use modelling tools to account for attribute non-attendance (see Hensher, 2009; Scarpa et al., 2010; Hess and Hensher, 2010). The problem that we have at hand is just the opposite: we hypothesize that some respondents do not ignore some of the choice attributes, but rather implicitly include additional attributes when making trade-offs. In particular, when we offer to value improvements only in fatal risk due to flooding, some respondents might also account for other risks connected to a flood event, such as risks of a non-fatal injury or evacuation. The reason for these confounded valuations might lie in cognition: floods might be perceived by some respondents as events that bring about multiple risks with multiple consequences, from which they fail to separate the risk of fatality for a valuation exercise within a particular choice experiment. We shall use age and education to control for cognitive ability of respondents when testing for the sensitivity of VOSL to the scope of valued risks across the various sub-groups.

We find that in earlier literature, the topic of cognition has already taken an important place in the research agenda, and is coupled both to education and age. The relation between cognitive abilities and the level of education is perhaps the more intuitive one. So, Auld and Sidhu (2005) and Cutler and Lleras-Muney (2010) explore the relation between the amount of schooling, cognition and behaviour, where the latter conclude that education raises cognition, which in turn improves behavioral performance. On the other hand, an extensive medical literature covers the issue of the relation between cognition and ageing. So, a general observation is that "as we age, our brains undergo a series of deleterious changes" (Cabeza et al., 2002, p.1394), that cognitive measures share substantial portions of age-related variance (Verhaeghen and Salthouse, 1997; Allaire and Marsiske, 1999) and that a gradual age-related deterioration is found for a number of measured cognitive tasks (Bäckman et al., 2000). Other evidence exists on the negative age differences in sensory performance that brings about the ageing of complex cognition (Baltes and Lindenberger, 1997). In particular, the processing-speed theory postulates that increased age is associated with a slower speed of performing many activities, which in turn leads to impairments in cognitive functioning (Salthouse, 1996). Similar findings on deficiencies in short-term memory that are found to be most predictive of age differences in higher level cognition are presented by McCabe and Hartman (2008).

We find vast evidence in neuropsychological literature on the relation between age and cognition. This relation is however everything but linear: cognitive ageing cannot be simplified to a general process of progressive mental loss (Reuter-Lorenz, 2002). In-depth research has shown that various tested cognitive domains resemble various degrees (or even none) of age-dependent cognitive decline (Ardila et al., 2000; and Bopp and Verhaeghen, 2007). Also, the elderly population is not homogeneous in cognitive digression; rather, they consist of high- and low-performing older adults. High-performers of senior age are characterized by the presence of multiple mechanisms that counteract brain degradation, and these processes are responsible for their performance in cognitive tasks that is comparable to that of young adults (Cabeza et al., 2002). Finally, there exists a complex relationship between age-related cognitive decline and education, where different patterns may be found depending upon a specified cognitive domain (Ardila et al., 2000).

In the economic valuation literature, age-related VOSL patterns have extensively been studied, see Alberini et al. (2004 and 2006), Krupnick (2007), Kim et al. (2009), Cameron et al. (2010b). There is, however, little evidence on the age-related stability of VOSL valuations. This paper will thus enrich existing economic valuation literature by exploring age and education effects on the robustness of risk valuation across different choice experimental contexts. Basically, we shall take age and education as a shorthand for cognition, with a gross assumption that, on average, cognitive ability decreases with age and increases with the level of education.[4]

3. Description of choice experiments and the sample

The data for the choice experiments was collected by means of an internet-based questionnaire. We have approached only those people who live in flood-prone areas and thus run some risk of dying in flooding. The survey was administered in four flood prone areas in the Netherlands in the fall of 2008. Respondents in our sample were part of TNS NIPO internet panel, and were representative in terms of demographic characteristics for each of the four areas present in the study. A total of 836 respondents completed the questionnaire.

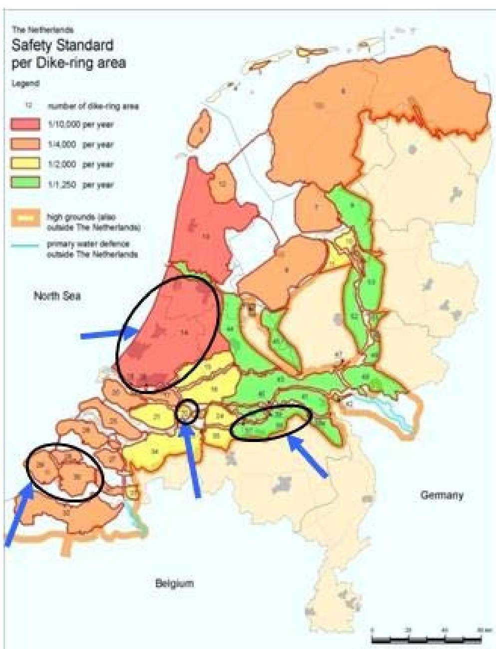

Figure 1. Flood-prone areas (dike-rings) in the Netherlands with study areas circled.

The SP part of the survey consisted of 3 parts. First, respondents were asked about their flood risk perception and knowledge about flood safety in the Netherlands. This was a sort of 'warming up' exercise and was done in order to introduce the respondents to the flood risk topic. During an earlier pilot, most of respondents indicated that they found this part

[4] Other possible interpretations for the effect of higher age are: different preference structure provided less expected life years; stronger perception of vulnerability; income effect; and in our case –prior water calamity experience (a part of elderly respondents have experienced the big disaster of 1953 flood in the Netherlands).

of the questionnaire very helpful in filling out the rest of the survey. In the second part of the questionnaire, we provided respondents with factual flood risk information, specific to the area of residence (the so-called dike-ring). Alongside with numerical information on the average yearly probability of flooding and the probability of flooding in the coming 50 years, as well as the corresponding average probabilities of dying due to flooding (yearly, and in the coming 50 years), we also provided visual aids aimed at better explaining the probabilities. These were later used in the choice experiment, to enable respondents to make well-informed choices. Especially for the case of low probabilities, literature provides repeated evidence for the importance of availability of visual aids (see among others, Hammit and Graham, 1999; Corso et al., 2001; Powe et al., 2005; Vassanadumrongdee and Matsuoka, 2005). This way, information on the levels of risk was communicated to low-numerate respondents for whom non-numeric representations are more appealing (Keller et al., 2009). So, the description of risk was accompanied by probability grids and a so-called "risk ladder" displaying a number of average risks of death in the Netherlands. Finally, the actual average probability of dying due to flooding (specific to the place of residence of each respondent) was compared to the average probability of dying due to a strike of a lightning, the lowest on our risk ladder.

Table 1. Overview of experimental designs: CE_B, CE_R and CE_T.

Choice attributes	**Baseline Choice experiment (CE_B)**	**Reference Choice experiment (CE_R)**	**Test Choice experiment (CE_T)**
Tax [a]	5 levels	5 levels	3 levels
Probability of a fatality	5 levels	5 levels	5 levels
Probability of flooding [a]	5 levels	5 levels	5 levels
Commuting time	-	4 levels	-
Probability of an injury	-	-	5 levels
Probability of an evacuation	-	-	5 levels
Setting	Public good	Private good	Private good
Payment vehicle (yearly tax)	Water board tax	Municipal tax	Water board tax
Alternatives	Generic	Generic	Labelled
N respondents	836	299	537

[a] probability of flooding was shown in all cards and in all choice experiments for reasons of explicability: fatality risk was always 100 times lower than probability of flooding, which facilitated the understanding of (low) risk levels by the respondents.

After the explanation of risk, respondents were split in two groups and each group was to fill out two consecutive choice experiments. Both groups first completed the baseline choice experiment (CE_B); after that, respondents in the reference group (R) have completed the reference experiment (CE_R) and respondents in the test group (T) have completed the test experiment, CE_T. In each of the experiments, respondents were offered 5 choice cards. In Table 1 we provide a brief overview of the three choice experiments and their underlying designs.

All choice experiments included 2 alternatives. 'Status quo' or an 'opt-out' alternatives were not necessary in our case as we were aiming at valuing particular attributes of the choices (like risk of fatality, risk of injury or evacuation) to obtain valuation of changes in risk around the current risk level, and not the value of the entire alternative. Also, the valuation we were performing did not have to do with the provision of a new good (which may or may not be acquired), but rather incremental changes in the existing good (flood protection), which is in fact a public good supplied to all residents of the flood-prone areas. Thus, we were able to avoid the reference point bias connected to the presence of a 'status quo' alternative (Chernev, 2004), and the non-response bias connected to the presence of an 'opt-out' alternative (Kontoleon and Yabe, 2003).

Because the three choice experiments in our study come from the same survey and draw from the same sample of respondents, they are also directly comparable. While they share a number of common elements, the three experiments do differ in terms of experimental design, notably the framing of the choice situation (as a private or public good), the number of attributes per alternative, and the levels and differences in the levels of the valued attributes (see Table 2).

The first choice experiment (CE_B) – a common experiment filled out by both groups of respondents – was the baseline experiment with only two attributes[5], tax and risk of fatality, and with 2 generic alternatives. It was framed as a public good in the context of a choice between two techniques that local Water Boards can use in order to maintain the dikes protecting the area from flooding. These techniques differ in quality (the final level of fatality risk) and price (yearly Water Board tax).

The test choice experiment (CE_T) was described in terms of purchase of a house. Respondents were asked to imagine they would have to move houses for some unspecified reason. Two houses were then described with identical characteristics (such as square meters, number of bedrooms, the garden), but differing in location and flood safety. So, respondents could choose between two polders: the one polder has a 'perfect' evacuation plan, so that all its inhabitants can be timely evacuated (provided full public cooperation following evacuation orders) and thus there is a probability of evacuation in anticipation of flooding, but the inhabitants do not run a risk of dying or getting an injury in flooding; the other polder does not have a possibility of evacuation, so that the residents run some small positive risk of dying, or getting an injury, in case a flood takes place. In both polders the residents have to pay a yearly Water Board tax. This choice experiment CE_T therefore has 2 labelled alternatives, one with a possibility for evacuation and therefore no risk of fatality, and the other one – without a possibility for evacuation but small positive risks of fatal and non-fatal injuries.

The reference experiment (CE_R), just as CE_T, also included a choice situation that was framed in terms of the purchase of a house, conditioned on moving houses for some unspecified reason. Similarly to the test experiment, two houses were described having identical characteristics, but located in two polders, which now differ in terms of flood safety level (the level of fatality risk), yearly payment (yearly municipal tax) and travel time to work (on a weekly basis). The latter was included for two reasons. One is that we

[5] See the Note to Table 1: in all three experiments risk of flooding was shown in all choice cards for the purposes of clarity; it was however always connected to the risk of fatality, and therefore cannot be considered as an independent attribute.

wanted to have a higher number of attributes in CE_R than in CE_B, just as for CE_T (although the numbers are not equal: 3 for CE_R, and 4 for CE_T). The second is that the estimated value of travel time allows a secondary check on the plausibility of our WTP estimates.

Table 2. Overview of demographic characteristics of respondents in the test group and in the reference group.

Demographic variable	Reference Group	Test Group
Males	61.87%	38.55%
Rural inhabitants	9.36%	10.06%
Higher education	31.10%	29.05%
High income [a]	20.74%	16.76%
Income not stated	31.77%	30.73%
Age 18-34	32.44%	13.78%
Age 35-64	56.52%	32.03%
Age 65 and older	11.04%	54.19%
Excellent health condition [b]	68.90%	58.85%
Prior water calamity experience	13.04%	21.04%
N respondents	*299*	*537*

[a] High income measured as respondents with income in 8th-10th deciles of national income distribution. [b] Self-estimated health condition, measured at 8 to 10 on a scale from 0 to 10.

Because weekly traveling time was valued based on 5-day-a-week commuting frequency, only respondents who are actually commuting full-time were asked to fill out the reference choice experiment. In Table 2 we report some demographic characteristics of each group (possible differences in the WTP between the test and the reference groups will be analysed in Section 5). The reference group includes a bigger proportion of males, respondents with higher income and of younger age. This has directly to do with the selection of full-time commuters for the reference experiment (we shall control for these factors in our analyses; however, in most cases they appear to be insignificant). More importantly, however, both groups have a fairly equal proportion of respondents with high education, which facilitates our cross-group comparisons of the effect of cognition.

4. Methodology: Pooling data

In this paper we are mostly interested in the exploration of sensitivity of VOSL to the scope of the valued good, which is in our case valuation of fatality risk alone vs a situation where it is valued together with other risks connected to the same calamity, in particular the risk of non-fatal injury and risk of preventive evacuation. This can be investigated by comparing the VOSL from the baseline choice experiment to the VOSL obtained from the choice experiment where multiple correlated risks are valued (here: CE_B and CE_T). We use a standard discrete choice model, that assumes that respondent *m*'s utility of alternative *i* is defined as:

$$U_{im} = V_{im} + \varepsilon_{im} \quad [1]$$

where V_{im} is observed and thus can be measured by the researcher via the predefined attributes of each alternative in an experiment, and ε_{im} is the unobserved part of respondent *m*'s utility of alternative *i*, which accounts for respondent *m*'s valuation of all other properties of the alternative not observed by the researcher.

The utility function for the baseline choice experiment, CE_B, (shown first) applies to 2 generic alternatives and two attributes – water board tax (x_{Tax}) and risk of fatality (x_{Pf}), and would look like:

$$V_{1,2(CE_B)} = \beta_{Pf} * x_{Pf} + \beta_{Tax} * x_{Tax} \qquad [2]$$

The test experiment, CE_T, concerns two alternatives. One is labeled as a polder with a 'perfect' evacuation plan, described by a probability of evacuation (x_{Pev}) and tax (x_{Tax}). The other alternative is a polder without a possibility of preventive evacuation, so that it is characterized by some small positive risk of dying (x_{Pf}) or getting an injury (x_{Pinj}), and a tax (x_{Tax}). The utility functions can be written as:

$$V_{1(CE_T)} = ASC + \beta_{Pev} * x_{Pev} + \beta_{Tax} * x_{Tax} \qquad [3a]$$

$$V_{2(CE_T)} = \beta_{Pf} * x_{Pf} + \beta_{Pinj} * x_{Pinj} + \beta_{Tax} * x_{Tax} \qquad [3b]$$

We include an alternative specific constant (ASC) in alternative 1 to capture non-linearities, such as discrete preference (either positive or negative) for an option with zero risk of fatality.

From equations [2], [3a,b] the following risk valuation indicators can then be obtained: the value of statistical life (VOSL), the value of statistical injury (VOSI), and the value of a statistical evacuation (VOSE). The VOSL measure, which is a trade-off between the money and the level of risk at the margin, is determined as shown in [4]. VOSE and VOSI can be expressed in a similar way.

$$VOSL = \frac{\partial U / \partial x_{Pf}}{\partial U / \partial x_{Tax}} = \frac{\beta_{Pf}}{\beta_{Tax}}, \qquad [4]$$

A comparison between two choice experiments can be done by pooling the data and running a nested logit model (that is in essence, the 'nested trick' logit model, see Hensher, Rose and Greene, 2007). The need for such a model is dictated by possible differences in the scale of utility across the two experiments, which can immediately be tested in a nested logit setting. When running such a nested logit model, it is necessary to estimate one generic parameter, that is common to all included experiments, to identify the scales of utility. It is usual to assume that risk valuation differs across the choice experiments, and therefore we will assign experiment-specific risk parameters to the respective utility functions, and estimate a generic tax parameter for the two experiments. So, combining equations [2]-[3a,b] and fixing the scale parameter of the test experiment to 1, a pooled nested model can be written as:

$$V_{1,2(CE_B)} = \lambda_{(CE_B)} * (\beta_{Pf(CE_B)} * x_{Pf} + \beta_{Tax} * x_{Tax}) \qquad [5a]$$

$$V_{1(CE_T)} = \quad ASC + \beta_{Pev(CE_T)} * x_{Pev} + \beta_{Tax} * x_{Tax} \qquad [5b]$$

$$V_{2(CE_T)} = \beta_{Pf(CE_T)} * x_{Pf} + \beta_{Pinj(CE_T)} * x_{Pinj} + \beta_{Tax} * x_{Tax} \qquad [5c]$$

Following our hypothesis, if the VOSL obtained from an experiment where a single fatality risk is valued (CE_B) is significantly higher than in the other experiment with multiple risks included (CE_T), then it should serve as a first indication of the presence of sensitivity to scope of the valued commodity. At the same time, however, other factors may play a role when outcomes of two choice experiments are compared, such as differences in experimental setting and statistical design. Experimental setting, as used in this paper, refers to such dimensions of the choice experiment as the payment vehicle, type of good valued (private vs public), complexity (number of alternatives and attributes valued). These effects will be discussed in the next section. Statistical design, and in particular the range of attribute levels that is in turn associated with scale effect, is an essential aspect that can cause substantial differences in risk valuation across choice experiments. The scale effect refers to the sensitivity of the measured marginal willingness to pay to the size of differences in values for the valued attributes in the choice sets (see Hammit, 1999). To avoid misunderstandings, it is important to make a note on the use of the terms 'sensitivity' and 'insensitivity'. Normally, WTP is assumed to increase together with the increase in the valued difference in risk,[6] so WTP should be sensitive to the size of differences in risk. However, the VOSL, which is the ratio between WTP and the change in risk, should be independent of the range of the valued attributes or valued good, and thus be *in*sensitive to the changes in risk or tax attributes. In this paper, we shall rather look at VOSL sensitivity when controlling for scale effects that can be tested for by comparing VOSL's across the choice experiments for the same range of the valued choice attributes, such as the monetary and the risk attributes. Another 'check' is testing for the existence of non-linerarities or 'jumps' in estimated parameters for various ranges of valued attribute differences within a particular choice experiment. This can be done by checking statistical significance of additional interaction terms in the utility function for categories of differences in attribute levels.

Furthermore, results of the analysis of scale sensitivity to the valued attributes and experimental setting for the test group will be compared to a similar analysis performed for the reference group. In the reference choice experiment, CE_R, in addition to a probability of dying in flooding (x_{Pf}) and tax (x_{Tax}), another choice attribute was included, not connected to a flood event, namely commuting time to work (x_{Ttime}). Analogous to [5a-c], a pooled nested model for the reference group would look like:

$$V_{1,2(CE_B)} = \lambda_{(CE_B)} * (\beta_{Pf(CE_B)} * x_{Pf} + \beta_{Tax} * x_{Tax}) \qquad [6a]$$

$$V_{1,2(CE_R)} = \beta_{Pf(CE_R)} * x_{Pf} + \beta_{Ttime(CE_R)} * x_{Ttime} + \beta_{Tax} * x_{Tax} \qquad [6b]$$

For the MNL specification of the utility function for each choice experiment, there exist equivalent mixed logit (MXL) models (this property will be used in Section 6) where the panel structure of the data will be properly accounted for, which is not addressed in a

[6] Proportionality of such increase is, however, contested.

conventional MNL or nested logit framework. To estimate a mixed logit model, some coefficient(s) of choice attribute(s) will be allowed to be random, so that:

$$\beta_{Xi} \sim f\,(\beta_{Xi} |\, \theta_{Xi}) \qquad [7]$$

where θ_{Xi} is a vector of parameters (such as mean and variance) of the respective distribution of a random coefficient β_{Xi} over the population. The assumed mixing distributions are often normal, uniform, triangular or lognormal. In this manner, individual heterogeneity in taste will be captured, which can be reflected by a confidence interval for a specified estimated coefficient.

5. Context analysis

We first explore the contextual differences across the three choice experiments. After running the nested logit models for the test and the reference groups (see model 1 in Tables A1 and A3 in the Appendix) we found that resulting risk valuations between the two pairs of experiments indeed differ from each other. In particular, for the test group we have found VOSL of 9.89 mln € for the baseline experiment CE_B and 6.84 mln € for the test experiment. While the confidence intervals for these two VOSL's overlap (Table 3), the confidence intervals for the respective estimated fatality risk coefficients – $\beta_{Pf(B)}$ and $\beta_{Pf(T)}$ – do not. The same is true for the VOSL's obtained in a pooled nested logit for the baseline and reference experiments (respective values of 7.29 mln € and 11.72 mln €). It is important to note here, that VOSL's for the reference and the test groups obtained from the baseline choice experiment (7.29 mln € and 9.89 mln €, respectively) do not differ in statistical terms based on an MNL model, and thus could have been drawn from the same population. This means that both groups of respondents have provided similar fatality risk valuations in the first experiment, which for both groups is different from their respective valuation in the second experiment (be it reference or test group). These differences in the willingness to pay for personal fatal risk reduction may in general be due to a number of factors, such as demographic differences between the two groups of respondents, as well as differences in context between the various experiments. We shall address these issues.

As we mentioned in section 3, respondents in our two groups appear to have different demographic characteristics which may potentially lead to variations in WTP for safety (in our background analyses we have found that such characteristics as income, age and education level do significantly influence VOSL, see Bočkarjova et al., 2009). So, statistically similar evaluations of VOSL in the baseline experiment by the two groups of respondents testify in favor of relatively high degree of homogeneity between the groups. The modest difference between the two point estimates in the baseline experiment (VOSL=7.29 mln € for the reference group and 9.89 mln € for the test group) could perhaps be clarified by the presence of respondents with an excellent self-estimated health condition, underrepresented in the test group, who tend to have lower fatality risk valuation. This is, however, not the case. A pooled multivariate MNL logit estimation for the baseline experiment (Table A1, model 1 in the Appendix) does not reveal significant health effects, controlling for other demographic variables such as education, income, age, gender, prior calamity experience and place of residence. A factor which is the most likely candidate to explain the differences valuations between the two groups is rather age: a considerably 'older' test group with more than a half respondents in a senior age (65 and older) appear to have significantly higher risk valuation compared to younger respondents (Table A1, model 1). At the same time, the positive effect of income on WTP is not prominent for the differences between the two respondent groups: it should drive the VOSL estimate upwards in the baseline experiment for the reference group, while the opposite is observed. Gender effects were not found to be significant in our sample.

5.1. Experimental setting

It is important to make the necessary reservation about differences in experimental setting that can be responsible for differences in risk valuation between the experiments across the two groups of respondents. Here, it refers to the way in which choice situations were presented. First is the framing of the choice situation: while the choice situation in the baseline experiment CE_B is put as a public good (a choice between two technologies for maintenance of dikes), the situations in CE_R and CE_T are put in terms of a private good (the purchase of a house). This difference in context between the baseline experiment CE_B vs the reference CE_R and the test CE_T, however, is not followed by a clear pattern in mean WTP (see Table 3). That is, point estimate of the VOSL is significantly higher for the reference group in CE_R compared to CE_B as their respective confidence intervals do not overlap; but lower for the test group in CE_T compared to CE_B (yet not significantly lower in statistical terms). The pattern is more in line with differences in the payment vehicle: water board tax (CE_B and CE_T) vs municipal tax (CE_R). The question is, however, whether such significant differences in VOSL between the two pairs of experiments are indeed driven by the differences in the payment vehicle, which in all cases remains a sort of tax.

The next dimension of experimental setting is choice complexity as our experiments differ in the number of valued attributes (see Table 1). Complexity is one of the problematic issues in the environmental valuation literature (Wierstra et al., 2001), where the number of choice attributes in environmental good valuation can be relatively high (4 to 10 attributes would not be unusual). The number of choice attributes in our experiments is thus relatively low – 3 to 5 – and so we believe that the difference between the two pairs of experiments should remain marginal: 1 extra attribute for the reference group as they go from CE_B and CE_R, and 2 extra attributes for the test group respondents as they go from CE_B to CE_T. Another issue connected to choice complexity is cognitive burden that is imposed on respondents. It has repeatedly been shown that the valuation of probabilities – and in particular of low probabilities – requires substantial cognitive effort (Andersson, 2006; Hammit and Graham, 1999). We shall explore this aspect in more detail further in this paper.

Next, the statistical design behind the experiments may influence individual risk valuations. The following issues may play a role: the scale, or the absolute level of the attributes and the valued differences in the attributes. For the test group, the absolute level of the risk attribute included is about the same in both experiments, CE_B and CE_T, and the presented levels of monetary attribute differ slightly. For the reference group, the risk attribute is up to 5 times higher in CE_R, compared to CE_B (which was necessary to balance utility in the presence of the three choice attributes). Also, the design of CE_R includes much higher levels of the monetary attribute compared to CE_B (see Figures A1 and A2 in the Appendix). Literature provides evidence (de Blaeij et al., 2003) that big differences in the attribute levels across experiments may significantly affect elicited values of risk valuation. We shall return to this issue in detail in Section 5.2 where we find that after controlling for scale, the difference in risk valuation between CE_B and CE_T persists.

Table 3.Valuation of VOSL (mln €) by the two groups of respondents - split sample MNL estimates.

	Reference group			TEST group	
	CE_B	**CE_R**		**CE_B**	**CE_T**
Value of statistical life, VOSL (mln euro)	**7.294**	**11.724**		**9.887**	**6.835**
std.VOSL [a]	*0.862*	*1.069*		*1.142*	*1.382*
LB (CI 95%)	5.605	9.630		7.649	4.126
UB (CI 95%)	8.983	13.819		12.125	9.543
Value of time, VOT (euro)	**---**	**5.61**		**---**	**---**
std.VOT [a]	---	*1.42*		---	---
Value of statistical injury, VOSI (euro)	**---**	**---**		**---**	**92,183**
std.VOSI [a]	---	---		---	62,385
Value of statistical evacuation, VOSE (euro)	**---**	**---**		**---**	**2,517**
std.VOSE [a]	---	---		---	*715*
N respondents	*299*	*299*		*537*	*537*
N cards	*1495*	*1495*		*2685*	*2685*

[a] Calculated using the delta method (Goldberger, 1991, p.110-111).

To sum it up, it is not likely that framing of good, payment vehicle or complexity cause the difference between the baseline experiment CE_B on the one hand, and the experiments CE_T and CE_R on the other hand. Scale might play a role, but at the same time, we observe a diverging pattern in VOSL estimates between CE_B and CE_T for the test group, and CE_B and CE_R for the reference group. This makes us suspect that beyond the possible scale effects, also scope effects could be relevant. In the following sections we shall explore in more detail these scale and scope effects.

5.2. Scale and scope effects

Because differences exist in the underlying statistical designs of the three choice experiments we shall test for scale sensitivity of VOSL for both the test and the reference groups.

For the test group, the two experiments CE_B and CE_T are pretty similar in terms of the ranges of the valued risk attribute (Δx_{Pf}). However, as mentioned, there is a slight difference in a higher range of the valued tax attribute in the test experiment CE_T (Δx_{Tax} = 15€ to 70€ per year) compared to the baseline experiment CE_B (Δx_{Tax} = 10€ to 40€ per year) – see Figures A1 and A2 in the Appendix.

So, we apply a pooled nested model for the two experiments to test for non-linearities in the valuation of the monetary parameter in the test experiment. Here, in addition to the basic formulation [5a-c] we include a dummy interaction term in the utility function of the test experiment to distinguish between two ranges of the valued differences in the tax attribute in CE_T: Δx_{Tax}(low) = 10€ to 35€, and Δx_{Tax}(high) = 50€ to 70€. The results of the model (model 1 in Table A2) show a statistically significant coefficient for the dummy interaction term $\beta Tax(CE_T)*D(\Delta x_{Tax}=(50\text{-}70))$, which signals a 'jump' in the valuation of the monetary parameter in the test choice experiment. This means that there is a *scale* effect present: respondents have a decreasing marginal utility of money, and therefore a higher VOSL when higher differences in the monetary attribute are to be valued. Thus, the VOSL is sensitive to the range of the valued differences in the monetary attribute.

Besides, this model reveals one more important finding. We observe that marginal utilities of money differ in the baseline and the test experiments, CE_B and CE_T (the respective coefficients are $\beta Tax(CE_B)$ and $\beta Tax(CE_T)$). Also, the confidence intervals of VOSL's in the baseline and test choice experiments for the same range of valued differences in the tax attribute do not overlap (respective VOSL's are: 9.89 and 3.49 mln €[7]). Because the range of valued differences in the risk attribute, Δx_{Pf}, is the same in both experiments, this means that we also find support for the *scope* hypothesis, claiming that individual risk valuation significantly changes as the valued good changes and multiple correlated risks are valued. We thus find evidence in favour of a suggestion that the scope effect is indeed present in the test group, and that the VOSL may be confounded in the baseline experiment compared to the test experiment, where the risk of non-fatal injury and the risk of preventive evacuation are included alongside with the risk of fatality.

Table 4. Distribution of the risk parameter Δx_{Pf} between choice experiments CE_B and CE_R (the Reference group, N respondents = 537, N cards = 1495).

	Baseline choice experiment (CE_B)	**Reference choice experiment (CE_R)**
Δx_{Pf} (low)	3 to 10 fatalities per 40.000 inhabitants (in 50 years) or 0.00015% to 0.00055% per year (*N=1014 cards*)	---
Δx_{Pf} (mid)	14 to 18 fatalities per 40.000 inhabitants (in 50 years) or 0.0007% to 0.0009% per year (*N=481 cards*)	15 to 25 fatalities per 40.000 inhabitants (in 50 years) or 0.00075% to 0.00125% per year (*N=617 cards*)
Δx_{Pf} (high)	---	50 to 80 fatalities per 40.000 inhabitants (in 50 years) or 0.0015% to 0.0045% per year (*N=878 cards*)

For the reference group, the baseline and the reference experiments differ in terms of valued ranges of the risk and the monetary attributes (see Table 4, as well as Table A6 and Figure A2 in the Appendix). So, we have run two pooled nested models, where we have

[7] Note, that the difference in VOSL becomes bigger between CE_B and CE_T when we restrict attention to the same range or the monetary attribute levels, compared to the difference between the experiments reported in Table 3.

tested VOSL sensitivity to the scale of the risk and the monetary attributes. Results of these models are found in Tables A4 and A5 (model 1 in both cases), again revealing statistically significant dummy interaction terms in both models, testifying for the presence of the scale effect. Furthermore, we have found significantly increasing VOSL for increasing ranges of valued Δx_{Tax} in the reference experiment (Table A4, model 1), as was true for the test group discussed above. Therefore, the scale effect of the monetary attribute on VOSL dominates even despite the presence of the opposite effect of higher Δx_{Pf} on VOSL in CE_R. So, differences between risk valuations in choice experiments CE_B and CE_R are determined by strong sensitivity of VOSL to the much higher valued differences in the monetary attribute in the reference experiment compared to the baseline experiment.

So far we have thus identified non-linearities in choice parameter estimation, subject to control for various aspects of experimental setting, both within and across the experiments, which allows us to conclude that statistically different valuations of personal fatality risk reduction are due to the sensitivity of VOSL to scale of choice attributes, confirmed both for the reference group and for the test group. Besides, after controlling for the scale effect in the test group, a scope effect was detected that is also responsible for differences in VOSL between the two experiments. So, our respondents appear to provide higher (confounded) VOSL estimates when fatal risk reductions are valued alone, and somewhat lower VOSL values when also risk of non-fatal injury and risk of evacuation are explicitly valued as well. In the next section we shall explore further the nature of differences in risk valuation between various sub-groups of respondents.

6. Scope bias: the role of cognition

After having established sensitivity of VOSL to the scale of valued attributes and to the scope of valued commodity, our next question is whether personal characteristics of respondents would clarify these differences in valuation *between* the experiments. We hypothesize that respondents depending on their cognitive abilities will be to a different extent inclined to provide risk valuations that are sensitive to the choice experimental setting, that we here interpret as the presence of correlated risks. We shall use age and education that represent cognitive ability of respondents. Basically, we shall assume that, on average, cognitive ability decreases with age (Cabeza et al., 2002) and increases with the level of education, so we shall particularly focus on the valuations of respondents with lower education and of elderly age as the sub-groups who might presumably provide different valuations of fatality risk under differing conditions. We note here, that age effect is more complex, and beyond the cognition interpretation may include (a combination of) differences in the preference structure related to a shorter life expectancy; perception of vulnerability; income effect; and prior water calamity experience.

6.1. Age effect

In order to identify these effects we have estimated a number of split-sample models. We shall first report the results of split-sample pooled nested models of the type [5a-c] for the test group and [6a,b] for the reference group, for respondents in various age cohorts.[8] We

[8] Recall that due to the need to select full-time commuters for the reference experiment, the reference group consists mainly of respondents aged 18-64 (only 33 out of 299 respondents are 65 or older), and therefore this group will be

can see that VOSL valuations of the two age sub-groups in the reference group (models 4 and 5 in Table A3) seem to be age-dependent within the experiments: young respondents (age 18-34) have a lower VOSL than their older counterparts (age 35 and older) in the baseline experiment (respective VOSLs are 6.07 and 8.06 mln €); and in the reference experiment (respective VOSLs are 10.5 and 12.4 mln €). For the between-experiment valuations, WTP differences are not age-dependent, as both age sub-groups have a higher VOSL in the reference experiment compared to the baseline experiment. The nested models for each sub-group (Tables A4 and A5, models 4 and 5) confirm the presence of scale effect for both age cohorts. We find that for both age sub-groups, respondents are equally sensitive to the higher valued differences in the risk and the monetary attributes between the baseline and the reference choice experiments, which in turn translates into the differences in their risk valuations between the experiments CE_C and CE_R.

We observe more differences in the test group where differences in the valuation of fatality risk across the two experiments seems to depend on the age of respondents. The results of the pooled nested models of the type [5a-c] are reported in Table A1 for each age sub-group (models 4 to 6), and respective VOSL's are depicted in Figure 2. We can see that in the baseline experiment, CE_B, VOSL is low for the young respondents (6.91 mln €), then increases but remains stable for the middle-aged and the elderly respondents (respective VOSLs are 10.71 and 10.29 mln €). This significantly lower valuation of risk by respondents aged 18-34 vs the other two groups of respondents may be interpreted as an income effect, which has the same pattern in the two experiments in the reference group as discussed above.[9]

For the test experiment, CE_T, however, we observe a different age pattern. Here, VOSL is resembling an almost symmetric inverse U-shape: young respondents of 18-34 with VOSL of 5.18 mln €, and elderly respondents of 65 and older with VOSL of 5.72 mln €, peaking for the group of 35-65 year olds (9.86 mln €). This means, that mean point estimates of WTP for fatality risk reduction drops slightly between CE_C and CE_T for the young and middle-aged respondents (from 6.91 to 5.18 mln €; and from 10.71 to 9.86 mln €, respectively). A remarkable drop in VOSL is however observed for the elderly cohort of respondents: 10.29 mln € in the baseline experiment to 5.72 mln € in the test experiment. This almost two-fold drop in VOSL for the respondents of age 65 and older is not in line with the income effect which we saw in the baseline experiment for the test and the reference groups. This fall in VOSL for the elderly respondents in CE_T rather points in the direction of scope sensitivity for this group. This becomes confirmed by the results of model 6 in Table A2, where we see that confidence intervals for the VOSLs in CE_C and CE_T do not overlap for the same valued differences in the risk and the monetary attributes (point estimates are 10.29 and 2.86 mln €, respectively). Besides, this is the only age cohort in the test group for whom scope effect is present. This makes us conclude that respondents of age 65 and older are very sensitive to the scope of risks presented: they systematically provide a substantially higher (confounded) VOSL estimate when fatality risk is valued alone, compared to VOSL that is valued in a context where multiple flood-related risks are simultaneously presented.

divided into two age sub-groups of 18-34 and 35 and older. In the test group, three age sub-groups will be distinguished, 18-34, 35-64 and 65 and older.

[9] The presence of income effect, controlling for other demographic factors, is confirmed, however, only for the reference group: see the significant high income dummy interaction term in the multivariate models for the reference group in Table A1 (models 3 and 5).

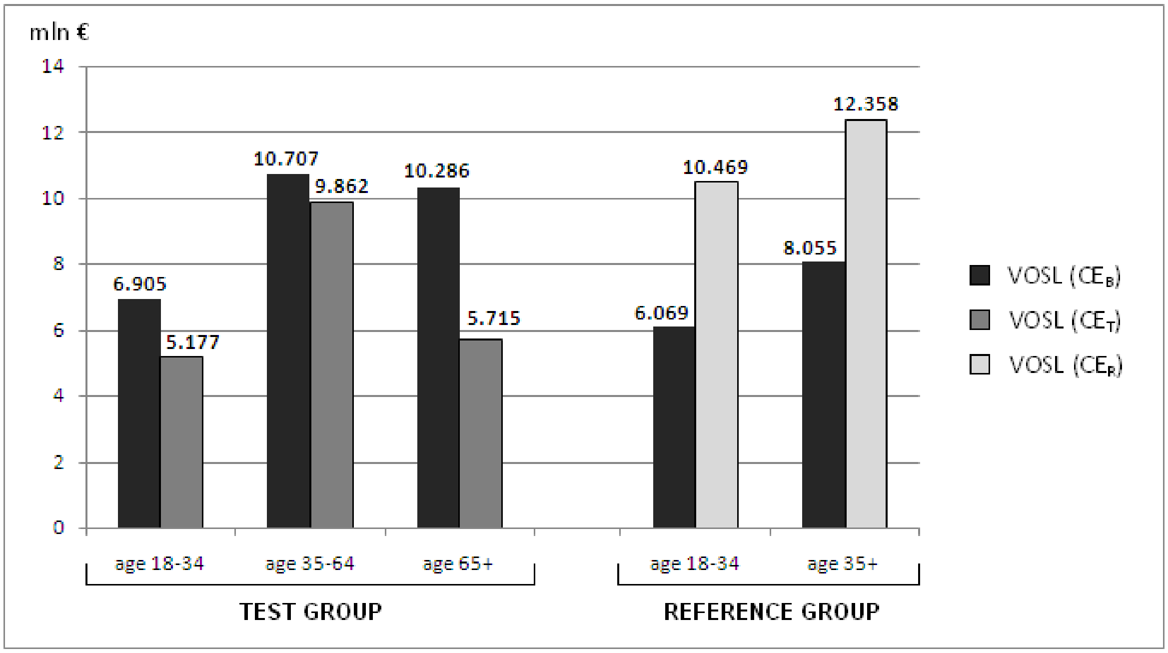

Figure 2. VOSL for the test group (baseline and test choice experiments) and the reference group (baseline and reference choice experiments), split by age (in mln €).

Estimates based on mixed logit models confirm the patterns described above obtained from the nested logit models. In Table A7 we present the results of mixed logit models run for each choice experiment separately, as well as for each age (and education) sub-groups. Here, the panel structure of the data and heterogeneity in taste are taken into account (random parameter in the baseline and the reference experiments is the risk of fatality; in the test experiment, four parameters are random – the constant of the 'evacuation' alternative, risk of fatality, risk of injury and risk of evacuation' all normally distributed).[10] We find that here, all indicators of immaterial damages – VOSL, VOSI and VOSE – from the test experiment are statistically significant for the whole test group. Average values of the indicators are slightly higher than the MNL point estimates, but the patterns are maintained. We also observe significant heterogeneity in the individual valuations of the fatality risk parameter, as well as the evacuation risk parameter. Looking at the three age groups, we see that for the young and middle-aged respondents, only VOSL and VOSE are statistically different from zero. So these two subgroups do not attach a positive value to the reduction in their personal risk of getting an injury in a flood (even respondents of 35-65 years old with a VOSI of 81,700 €); they would rather pay a positive amount for the reduction of the inconvenience connected to evacuation. Markedly, VOSE for these two age sub-groups is almost a double of sample average. For the elderly respondents of 65 and older, on the opposite, it is the risk of injury that they are most willing to avoid with VOSI of 145,900 €, and not the risk of evacuation that on average even takes a negative sign for this age sub-group (and is statistically insignificant). A plausible explanation for this is perhaps higher perceived vulnerability, so that elderly respondents perceive injuries incurred in a flood event as much more severe than their younger counterparts, who would perhaps count on much lighter physical damages in case of a calamity.

[10] All random parameters are assumed to be normally distributed.

Following our hypothesis on embedding of multiple risks when only fatality risk is valued, we may also try to calculate an 'implicit VOSL', based on the estimates of the test experiment to investigate whether our supposition about confounded VOSL in CE_C would be confirmed. According to the design of the test experiment, the presented risk of injury is on average 15 times higher than the risk of fatality; the risk of evacuation is 1500 times higher. This means that by assumption, there are expectedly 15 injured and 1500 evacuated persons per one expected fatal incident. Because we have found the presence of the scope effect for the whole test group, we may assume that VOSL obtained for the test group in the baseline experiment CE_B is confounded, and in addition to the risk of fatality, also implicitly includes other immaterial damages. In turn, VOSL measured in the test experiment, CE_T, should not be confounded as VOSI and VOSE are explicitly measured. An imputed immaterial damage measure[11] would then consist of the values of fatality, injury and evacuation.

$$\text{VOSL(imputed)} = \text{VOSL}_{CE_T} + 15 \times \text{VOSI}_{CE_T} + 1500 \times \text{VOSE}_{CE_T}$$

Thus, we take these indicators from CE_T (for the whole test group, VOSL=7.08 mln €, VOSI=92,000 € and VOSE=1,900 €) and compute the 'imputed' VOSL that equals then 11.32 mln €. We see that, as expected, this number is very close to the VOSL estimate from CE_B (12.28 mln €). For the elderly respondents of age 65 and older (the sub-group where the scope effect was found), this imputed VOSL would then be 7.31 mln € which does not yet close the gap to the VOSL estimate of 12.96 mln € from the baseline experiment, but is already much closer to that number compared to a 'pure' VOSL of 5.12 mln € from the test experiment.

We summarize that our supposition got confirmed that respondents of age 65 and older have indeed proven to be sensitive to the scope of valued risks, stating a systematically lower VOSL in an experiment where multiple risks related to a flood event were valued, CE_T, compared to the baseline measurement of VOSL in CE_B. Notably, this age effect is mirrored by the schooling effect for the lower educated respondents that we discuss in the next sub-section. This suggests that age in this case captures a portion of cognition effect explaining differences across respondents in risk valuation between multiple choice experiments.

6.2. Education effect

We now turn to the effect of education. We have divided respondents into two education sub-groups: with and without college education. We see that VOSL point estimates are much more similar for the respondents with university degree across the two experiments in the reference group (comparing CE_B and CE_R), as well as in the test group (comparing CE_B and CE_T) than for respondents without a college diploma (see results of pooled nested models 2 and 3 in Table A1 and Table A3).

Next, we could confirm the presence of a *scale* effect for both high and low education sub-groups in the reference group (i.e. sensitivity of VOSL estimates to various ranges of the risk and the monetary attributes, within as well as across CE_B and CE_R) – see results of pooled nested models 2 and 3 in Tables A4 and A5. For the test group, however, we could

[11] An interested reader might look in Bočkarjova et al. (2011) for a derivation of a composite valuation of immaterial damage.

not find a scale effect for respondents either with or without college education, which is perhaps due to diminishing statistical power of test in the divided sample (pooled nested models 2 and 3 in Table A2). Notwithstanding that, the *scope* effect was statistically significant for the respondents with a lower education level, which signals sensitivity of their VOSL estimate to the presence of correlated risks in the test choice experiment. Besides, both education sub-groups do not have significantly different VOSL's in the baseline experiment (VOSL is even higher for lower-educated respondents both in the test and in the reference groups, respective point estimates are 10.62 mln € and 7.74 mln €, vs the VOSL of respondents with university diploma in the two experiments, respectively 8.55 mln € and 6.53 mln €), discarding a potential argument in favour of income effect for the higher-educated respondents. This provides an even stronger evidence of high sensitivity of VOSL for the scope of valued good for the lower educated sub-group of respondents.

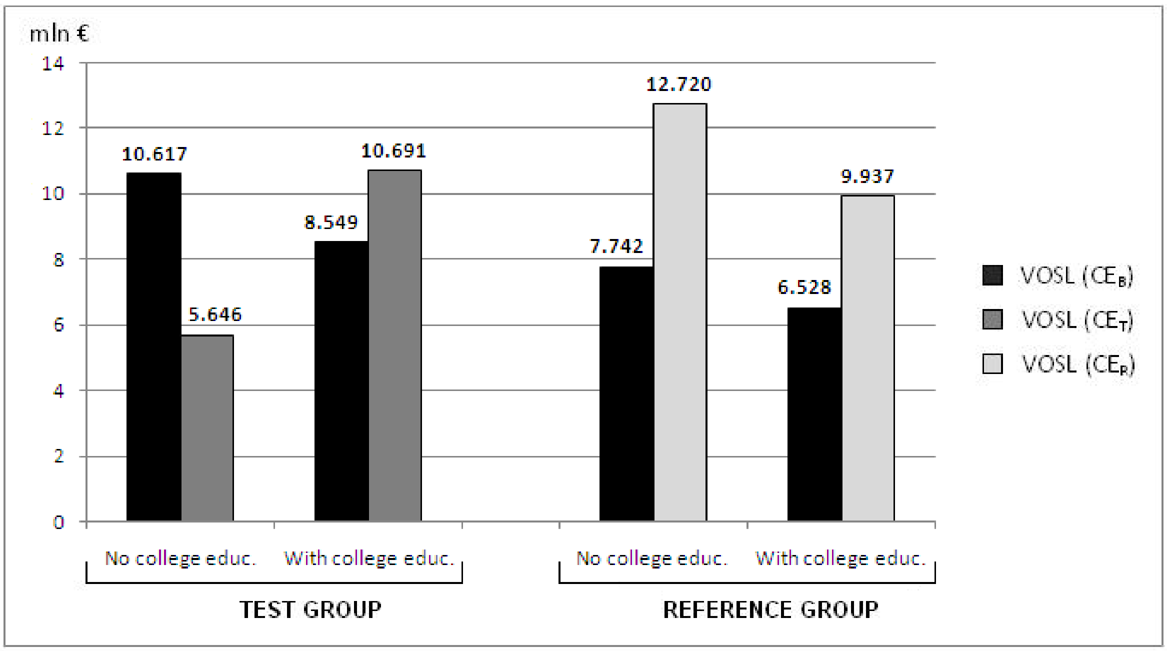

Figure 3. VOSL for the test group (baseline and test choice experiments) and the reference group (baseline and reference choice experiments), split by education (in mln €).

Furthermore, it is remarkable that respondents with university degree have higher point estimates of the values of statistical injury and evacuation, VOSI and VOSE, compared to the respondents without college education in the test group (see models 2 and 3 in Table A1). Mixed logit model estimates for the choice experiments CE_B and CE_T confirm this result (Table A7), and even reveal statistically significant average values of VOSE for both high- and low-educated sub-groups. This is different from the three age sub-groups, where VOSL and VOSE were significant for respondents up to the age of 65, and VOSL and VOSI – for respondents older than 65. So, the stability of VOSL between CE_B and CE_T for the respondents with higher educational attainment in the test group (11.33 mln € and 9.88 mln €, respectively) together with higher average estimates of VOSI (157,400 €) and VOSE (3,240 €) provide an even more pronounced contrast to the respondents without a university degree. These respondents, with lower values of VOSI and VOSE (73,672 € and 1,434 €, respectively), as yet show a significant drop in the average value of VOSL in CE_T (6.18 mln €) compared to CE_B (12.67 mln €). The 'implicit' composite valuation of immaterial damages for the lower-educated respondents, as we have introduced it in Section 6.1, would then be 8.33 mln €, which is closer to their CE_B estimate, as well as is more in line with the VOSL valuations of their higher-educated counterparts. This shows that respondents without a university degree would rather tend to value 'all immaterial damages' when only fatality risk is presented, and provide 'pure' VOSL when multiple risks related to a calamity are valued simultaneously.

We can summarise that indirect evidence on risk valuations by age and education points at the fact that cognitive capacity of respondents indeed plays an important role in the stability of VOSL valuation. Economic valuation literature provides ample evidence that discrete choice experiments impose a substantial cognitive burden on respondents (Viscusi et al., 1987; Wierstra et al., 2001). In line with these findings, we confirm that respondents with supposedly higher cognitive abilities (with higher educational attainment and of younger age) appear to be much better capable of providing stable valuations, independent of changing experimental context.[12] So, all respondents in the test and the reference groups, independent of their age or education level resemble similar sensitivity to the *scale* of valued attributes in the presented choice experiments. However, only respondents of older age (65 and older) and those without a university degree prove to be systematically more sensitive to the *scope* of valued good, and resemble significantly lower VOSLs when correlated risks are valued simultaneously compared to situations where fatality risk alone is valued.

7. Conclusions and implications

In this paper we have addressed the issue of risk valuation stability, performing a context analysis of three stated choice experiments among two groups of respondents (the test group and the reference group). These were conducted in the context of flood safety in the Netherlands. In particular, we have concentrated on the subject of sensitivity of VOSL to the scope of valued good, in particular to the valuation of correlated risks.

Comparing the estimates of VOSL between two pairs of choice experiments, we could not detect a pattern that would be dependent on the type of good measured, be it private or public good. Also, no effect was observed for the complexity of the experiment: VOSL did not systematically change with the number of valued attributes. Rather, the main reason for differences in estimated VOSL's between the CE's for the reference group is the sensitivity of VOSL to the scale of the valued choice attributes in various experiments, both for the risk and the monetary attributes. We found a fairly high sensitivity in the valuation of the risk attribute, so that non-linearities in the risk parameter were already present at very low ranges of valued Δx_{Pf}, which did not exceed 0.005% per year. At the same time, statistically significant differences in the valuation of the monetary parameter were not apparent until Δx_{Tax} reached 450€ per year compared to Δx_{Tax} up to 100€ per year.

[12] Results of mixed logit models (see Table A7, more results are available from the authors) largely corroborate the results of MNL and nested logit models reported so far, but provide a significant improvement in fit. This means that taste heterogeneity captured by the mixed models even within the sub-groups is very substantial. The values of VOSL in all ML estimated models resembled their counterparts from respective MNL/NL models, and remained stable with 5.000 and 10.000 Halton draws both in terms of the estimated sample averages and standard deviations.

Our next finding is that for the test group, controlling for scale effects, the hypothesis about the sensitivity to the scope of the valued good was confirmed. We have found what proved to be embedding after comparing two experiments filled out by the same respondents, in one of which fatality risk alone was valued, and in the other it was valued alongside with the risks of injury and evacuation connected to flooding. We can thus conclude that it is the scope effect that is responsible for the lower VOSL in the test choice experiment CE_T, where multiple risks related to a flood event were valued.

Finally, we conclude that cognition is related to these valuation patterns: all respondents in our sample exhibit (statistically) equal sensitivity of VOSL to the scale of valued attributes, independent of their cognitive abilities (with education and age taken as proxies), but not to the scope of the valued good. Elderly respondents (65 and older), and even more so respondents with lower education level, tend to be more influenced by the presence of multiple correlated risks in choice experiments compared to other respondents. Hence, respondents with an assumedly higher level of cognitive ability systematically prove to have more stable VOSL's across the experiments, and were to a much less extent influenced by the experimental setting in their valuations. Thus, our research confirms earlier findings that completion of choice experiments, even when extended information on the background risks with visual aids is present, requires substantial cognitive effort. In particular, we find that respondents with lower education level as well as elderly respondents seem to be less capable of concentrating on the valuation of a single risk within a specified choice task. Rather, they fail separating risks and value 'all risks' together when correlated risks are present. So, in the case of flooding, multiple risks can be expected to be related to the same event, such as risk of evacuation, risk of injury and risk of fatality. When only fatal risk is valued in such circumstances, we inferred that respondents with lower cognitive ability implicitly value also other risks as well.

This research has important implications for conducting choice experiments. First of all, we establish that fatality risk valuation for the same respondents and conducted under same circumstances may differ across choice experiments. One of the reasons is the sensitivity of VOSL to the scale of valued attributes. Even for very low valued changes in risk, we find non-linearities in the VOSL estimates for various sub-groups in our sample, which suggests that choice experiments should be designed to focus on sizes of risk changes that are of actual interest.

Another theme concerns VOSL sensitivity to the scope of valued good in relation to the presence of correlated risks. In particular, we assumed that this has to do with cognitive abilities of respondents to be able to analytically separate risks and value only those risks presented in a particular choice situation. We thus find that sub-groups of lower-educated and elderly respondents have delivered confounded estimates of VOSL when only fatality risk was valued, and probably less biased estimates when other correlated risks were explicitly valued as well.

Furthermore, this result has an important implication for the use of valuation results for policy purposes. We propose that choice situations are carefully designed, and all relevant risks are included in a choice task when correlated risks are suspected to be present. In this way, the risk of confounded estimates of VOSL can be reduced. Simultaneous valuation of risks related to the same event is also crucial for policy purposes due to possible differences in the composition of total benefits. In Bočkarjova et al. (2011) we show that relative weights of various components in the composite valuation of immaterial damages due to flood can differ substantially across various regions. So, in the areas where preventive massive evacuations are not possible, and therefore higher injury and death tolls might be expected (such as coastal areas in the Netherlands), VOSL would make up the biggest part of immaterial damages, a share ranging between 50% and 80%. However, in areas where flood can be predicted well in advance (like the riverside) and residents can be timely evacuated with low death and injury tolls, the value of evacuation will dominate the composite immaterial damage valuation, and the share of the value of fatalities may vary between 10% and 50%. This implies that more precise valuation of correlated risks is necessary for an accurate application of estimated values in economic assessments to support policy and action.

Acknowledgements

This research was financed by the BSIK project Climate Changes Spatial Planning (Klimaat voor Ruimte, see www.klimaatvoorruimte.nl) and project KBA WV21 from the Dutch Ministry of Transport and Water Management.

References

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

Alberini, A., Hunt, A., Markandya, A., 2006. Willingness to pay to reduce mortality risks: evidence from a three-country contingent valuation study. Environmental and Resource Economics 33, 251–264.

Allaire, J.C., Marsiske, M., 1999. Everyday Cognition: Age and Intellectual Ability Correlates, Psychology and Aging 14(4), 627–644.

Andersson, H., Svensson, M., 2008. Cognitive ability and scale bias in the contingent valuation method. Environmental and Resource Economics 39, 481–495.

Ardila, A., Ostrosky-Solis, F., Rosselli, M., Gómez, C., 2000. Age-Related Cognitive Decline During Normal Aging: The Complex Effect of Education. Archives of Clinical Neuropsychology 15(6), 495–513.

Auld, M.C., Sidhu, N., 2005. Schooling, cognitive ability and health. Health Economics 14(10), 1019-1034.

Bäckman, L., Ginovart, N., Dixon, R.A., Robins Wahlin, T.-B., Wahlin, Å., Halldin, C., Farde, L., 2000. Age-Related Cognitive Deficits Mediated by Changes in the Striatal Dopamine System. American Journal of Psychiatry 157, 635–637.

Baltes, P.B., Lindenberger, U., 1997. Emergence of a Powerful Connection between Sensory and Cognitive Functions Across the Adult Life Span: A New Window to the Study of Cognitive Aging? Psychology and Aging 12(1), 12-21.

Bellavance, F., Dionne, G., Lebeau, M., 2009. The value of statistical life in road safety: a meta-analysis with a mixed effects regression model. Journal of Health Economics 28, 444-464.

Bočkarjova, M., Rietveld, P., Verhoef, E.T., 2009. First results of immaterial damage valuation: value of statistical life (VOSL), value of evacuation (VOSE) and value of injury (VOSI) in flood risk context, a stated preference study (III), IC5 Report within the research programme Climate changes spatial planning, VU Amsterdam, December 2009.

Bočkarjova, M., Rietveld, P., Verhoef, E.T., 2010. Immaterial damage valuation in flooding: value of statistical life, value of statistical evacuation and value of statistical injury, paper presented at the international conference "Delta's in the Time of Climate Change", 29 September – 1 October 2010, Rotterdam, the Netherlands.

Bočkarjova, M., Rietveld, P., Verhoef, E.T., 2011. Composite valuation of immaterial damage in flooding: value of statistical life, value of statistical evacuation and value of statistical injury. TI Working Paper.

Bopp, K.L., Verhaeghen, P., 2007. Age-Related Differences in Control Processes in Verbal and Visuospatial Working Memory: Storage, Transformation, Supervision, and Coordination. Journal of Gerontology: Psychological Sciences 62B(5), 239–246.

Burge, P., Netten, A., Gallo, F., 2010. Estimating the value of social care. Journal of Health Economics 29, 883-894.

Cabeza, R., Anderson, R.D., Locantore, J.K., McIntosh, A.R., 2002. Aging Gracefully: Compensatory Brain Activity in High-Performing Older Adults. NeuroImage 17, 1394–1402.

Cameron, T.A., DeShazo, J.R., Johnson, E.H., 2010a. The effect of children on adult demands for health-risk reductions. Journal of Health Economics 29(3), 364–376.

Cameron, T.A., DeShazo, J.R., Stiffler, P., 2010b. Demand for health risk reductions: A cross-national comparison between the U.S. and Canada. Journal of Risk and Uncertainty 41(3), 245–273.

Campbell, D. Hutchinson, W.G., Scarpa, R., 2008. Incorporating Discontinuous Preferences into the Analysis of Discrete Choice Experiments. Environmental and Resource Economics 41, 401–417.

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

Champ, P.A., Bishop, R.C., 2006. Is Willingness to Pay for a Public Good Sensitive to the Elicitation Format? Land Economics 82, 162-173.

Champ, P.A., Boyle, K.J., Brown, T.C., 2003. A primer on nonmarket valuation. Springer.

Chernev, A., 2004. Goal orientation and consumer preference for the status quo. Journal of Consumer Research 31, 557-565.

Corso, P., Hammit, J.K., Graham, J.D., 2001. Valuing Mortality-Risk Reduction: Using Visual Aids to Improve the Validity of Contingent Valuation. Journal of Risk and Uncertainty 23(2), 165–184.

Cutler, D.M., Lleras-Muney, A., 2010. Understanding differences in health behaviors by education. Journal of Health Economics 29(1), 1–28.

Daniel, V.E., Florax, R.J.G.M., Rietveld, P., 2009. Flooding risk and housing values: An economic assessment of environmental hazard. Ecological Economics 69 (2009), 355-365.

De Blaeij, A., 2003. The Value of a Statistical Life in Road Safety, Ph.D. thesis, Vrije Universiteit Amsterdam.

de Blaeij, A., Florax, R.J.G.M., Rietveld, P., Verhoef, E.T., 2003. The value of statistical life in road safety: a meta-analysis. Accident Analysis and Prevention 35(6), 973-986.

De Brabander, B., 2007. De waardering van dodelijke verkeersslachtoffers in Vlaanderen; Resultaten van een contingente waardering in Vlaanderen, Report Nr.RA-2007-111, Steunpunt Verkeersveiligheid, Diepenbeek, Belgium.

de Palma, A., Myers, G.M., Papageorgiou, Y.Y., 1994. Rational Choice Under an Imperfect Ability To Choose. The American Economic Review 84(3), 419-440.

Foster, V., Mourato, S., 2003. Elicitation Format and Sensitivity to Scope: Do Contingent Valuation and Choice Experiments Give the Same Results? Environmental and Resource Economics 24, 141–160.

Goldberg, I., Roosen, J., 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 34, 123–144.

Goldberger, A.S., 1991. A course in Econometrics. Harvard University Press, Cambridge.

Hackl, F., Pruckner,G.J., 2005. Warm glow, free-riding and vehicle neutrality in a health-relatedcontingent valuation study. Health Economics 14, 293-306.

Hammit, J.K., Graham, J.D., 1999. Willingness to Pay for Health Protection: Inadequate Sensitivity to Probability? Journal of Risk and Uncertainty 8, 33-62.

Hanley, N., Wright, R.E., Adamovics, V., 1998. Using Choice Experiments to Value the Environment. Environmental and Resource Economics 11(3-4), 413-428.

Hensher, D.A., 2006a. How do respondents process stated choice experiments? Attribute consideration under varying information load. Journal of Applied Econometrics 21, 861–878.

Hensher, D.A., 2006b. Revealing Differences in Willingness to Pay due to the Dimensionality of Stated Choice Designs: An Initial Assessment. Environmental & Resource Economics 34, 7–44.

Hensher, D.A., 2009. Attribute processing, heuristics, and preference construction in choice analysis. Working Paper ITLS-WP-09-12.

Hensher, D.A., 2010. Hypothetical bias, choice experiments and willingness to pay. Transportation Research Part B 44, 735-752.

Hess, S., Hensher, D.A., 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. Transportation Research Part B: Methodological 44(6), 781-790.

Hoyos, D., 2010. The state of the art of environmental valuation with discrete choice experiments. Ecological Economics 69(8), 1595-1603.

Jin, J., Indab, A., Nabangchang, O., Thuy, T.D., Harder, D., Subade, R.F., 2010. Valuing marine turtle conservation: A cross-country study in Asian cities. Ecological Economics 69, 2020–2026.

Johannesson, M., Johansson, P.-O., O'Conor, R.M., 1996. The Value of Private Safety versus the Value of Public Safety. Journal of Risk and Uncertainty 13(3), 263-275.

Jones-Lee, M.W., Loomes, G., 1995. Scale and Context Effects in the Valuation of Transport Safety. Journal of Risk and Uncertainty 11(3), 183-203

Kahneman, D., Knetsch, J.L., 1992. Valuing Public Goods: the Purchase of Moral Satisfaction. Journal of Environmental Economics and Management 22, 57-70.

Keller, C., Siegrist, M., Visschers, V., 2009. Effect of Risk Ladder Format on Risk Perception in High- and Low-Numerate Individuals. Risk Analysis 29(9), 1255-1264.

Kim, B.H.S., Kim, E., Lee, J., Son, M., 2009. Elicitation of Health Values from Mortality Risk reduction. The Social Science Journal 46, 776-786.

Kontoleon, A., Yabe, M., 2003 -Assessing the impacts of alternative 'opt-out' formats in choice experiment studies: consumer preferences for genetically modified content and production information in food. Journal of Agricultural Policy and Research 5, 1-43.

Krupnick, A., 2007. Mortality-risk Valuation and Age: Stated Preference Evidence. Review of Environmental Economics and Policy 1(2), 261–282.

Krupnick, A., Alberini, A., Cropper, M., Simon, N., O'Brien, N., Goeree, R., Heintzelman, M., 2002. Age, health, and the willingness to pay for mortality risk reductions: a contingent valuation survey of Ontario residents. Journal of Risk and Uncertainty 24(2), 161–86.

Lancsar, E., Louviere, J., Flynn, T., 2007. Several methods to investigate relative attribute impact in stated preference experiments. Social Science & Medicine 64, 1738–1753.

Layton, D., Hensher, D.A., 2010. Aggregation of common-metric attributes in preference revelation and implications for willingness to pay. Transportation Research Part D: Transport and Environment 15(7), 394-404.

Leiter, A.M., Pruckner, G.J., 2009. Proportionality of Willingness to Pay to Small Changes in Risk: The Impact of Attitudinal Factors in Scope Tests. Environmental and Resource Economics 42, 169–186.

Louviere, J.J., Lancsar, E., 2009. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. Health Economics, Policy and Law 4, 527–546.

McCabe, J., Hartman, M., 2008. An analysis of age differences in perceptual speed. Memory and Cognition 36(8), 1495-1508.

Norinder, A., Hjaltek, K., Persson, U., 2001. Scope and scale insensitivities in a contingent valuation study of risk reductions. Health Policy 57, 141-153.

O'Conor, R.M., Blomquist, G.C., 1997. Measurement of consumer-patient preferences using a hybrid contingent valuation method. Journal of Health Economics 16, 667-683.

Olsen, J.A., Donaldson, C., 1998. Helicopters, hearts and hips: using willingness to pay to set priorities for public sector health care programmes. Social Science and Medicine 46, 1–12.

Reed Johnson, F., 2006. Comment on ''Revealing Differences in Willingness to Pay Due to the Dimensionality of Stated Choice Designs: An Initial Assessment''. Environmental & Resource Economics 34, 45–50.

Reuter-Lorenz, P.A., 2002. New visions of the aging mind and brain. Trends in Cognitive Sciences 6(9), 394-400.

Salthouse, T.A., 1996. The Processing-Speed Theory of Adult Age Differences in Cognition. Psychological Review 103(3), 403-428.

Scarpa, R., Thiene, M., Hensher, D.A., 2010. Monitoring Choice Task Attribute Attendance in Nonmarket Valuation of Multiple Park Management Services: Does It Matter? Land Economics 86(4), 817-839.

Shanmugam, K.R., 2000. Valuations of Life and Injury Risks. Environmental and Resource Economics 16, 379-389.

Svensson, M., 2009. The value of a statistical life in Sweden: Estimates from two studies using the "Certainty Approach" calibration. Accident Analysis and Prevention 41, 430–437.

Vassanadumrongdee, S., Matsuoka, S., 2005. Risk Perceptions and Value of a Statistical Life for Air Pollution and Traffic Accidents: Evidence from Bangkok, Thailand. Journal of Risk and Uncertainty 30(3), 261–287.

Verhaeghen, P., Salthouse, T.A., 1997. Meta-Analyses of Age-Cognition Relations in Adulthood: Estimates of Linear and Non-linear Age Effects and Structural Models. Psychological Bulletin 122(3), 231-249.

Viscusi, W.K., Aldy, J.E., 2003. The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World. Journal of Risk and Uncertainty 27(1), 5–76.

Viscusi, W.K., Magat, W.A., Huber, J., 1987. An investigation of the rationality of consumer valuations of multiple health risks. RAND Journal of Economics 18(4), 465-479.

Viscusi, W.K., Zeckhauser, R.J., 1994. The Fatality and Injury Costs of Expenditures. Journal of Risk and Uncertainty 8, 19-41.

Wierstra, E., Geurts, P.A.T.M., van der Veen, A., 2001 Validity of CVM related to the type of environmental good; an empirical test. Integrated Assessment 2, 1–16.

Wijnen, W., Wesemann P., de Blaeij, A., 2009. Valuation of road safety effects in cost–benefit analysis. Evaluation and Program Planning 32, 326–331.

Zhai, G., Ikeda, S., 2006. Flood Risk Acceptability and Economic Value of Evacuation. Risk Analysis 26(3), 683-694.

Appendix

Table A1. Multivariate MNL models for choice experiments CE_B, CE_T and CE_R([a])

	BASELINE EXPERIMENT (CE_B)			TEST EXPERIMENT (CE_T)	REFERENCE EXPERIMENT (CE_R)
	pooled	Test group	Reference group	Test group	Reference group
	1	2	3	4	5
1 ASC(evacuation)				1.109 *** 0.165	
2 βTax	-23.551 *** 6.484	-15.345 * 8.112	-45.469 *** 9.366	-20.534 *** 5.286	-3.058 *** 0.647
3 βPf	-170.540 *** 26.768	-133.167 *** 33.556	-194.660 *** 33.830	-145.518 *** 35.941	-28.194 *** 5.861
4 βPev				-0.078 *** 0.020	
5 βPinj				-0.364 2.011	
6 βTtime					-20.526 *** 4.776
7 β(Tax * HighINCOME)	2.815 4.338	-5.112 5.636	15.452 ** 6.940	2.998 2.639	1.206 ** 0.551
8 .... * UnknINCOME)	-10.594 *** 3.476	-15.089 *** 4.364	-3.187 5.904	-3.263 2.083	0.017 0.484
9 .... * GoodHEALTH)	1.271 3.177	-1.585 3.941	8.037 5.570	-3.264 * 1.899	-1.089 ** 0.450
10 .... * GENDER)	-0.598 3.189	0.112 4.093	0.851 5.251	-0.780 1.945	0.725 * 0.428
11 .... * CollEDUC)	-15.157 *** 5.060	-14.551 ** 6.298	-18.125 ** 8.665	-2.317 3.207	-1.849 *** 0.564
12 .... * AGE35-64)[b]	2.317 6.004	-2.728 8.720	8.858 8.141	10.542 * 5.629	0.155 0.507
13 .... * AGE65+)	8.022 6.377	2.666 7.997		11.401 ** 5.274	
14 .... * RefGROUP)	-8.453 * 5.111				

	BASELINE EXPERIMENT (CE_B)			TEST EXPERIMENT (CE_T)	REFERENCE EXPERIMENT (CE_R)
	pooled	Test group	Reference group	Test group	Reference group
	1	2	3	4	5
15 β (Pf * CollEDUC)	-70.467 ***	-86.122 ***	-52.057	2.165	-5.032
	22.101	27.929	37.029	25.814	6.035
16 .... * AGE35-64) [b]	-69.390 ***	-137.784 ***	-5.718	15.180	-3.376
	26.099	38.004	35.078	39.618	5.728
17 .... * AGE65+)	-2.926	-45.527		64.264 *	
	27.723	34.486		37.475	
18 .... * RURAL)	-36.856	-45.134	-29.262	-39.402	-21.756 **
	25.009	31.051	42.591	26.302	9.144
19 .... * COAST)	-44.572 ***	-37.689 **	-65.979 ***	-0.646	-12.720 ***
	13.764	17.291	23.213	14.948	4.909
20 .... * EXPERIENCE)	-27.964	4.343	-104.554 ***	49.903 ***	-4.299
	18.579	21.956	37.649	18.598	7.260
21 .... * RefGROUP)	-3.423				
	22.119				
22 β(Pev * AGE35-64)				0.014	
				0.024	
23 .... * AGE65+)				0.071 ***	
				0.022	
24 β(Pinj * AGE35-64)				-0.687	
				2.353	
25 .... * AGE65+)				-1.489	
				2.225	
N cards	4180	2685	1495	2685	1495
N respondents	836	537	299	537	299
LL function	-2385	-1527	-845	-1556	-726
Restr. LL function	-7490	-4811	-2679	-4811	-2679
Degrees of freedom	16	14	12	21	13

*, **, *** - statistical significance at 10%, 5% and 1% level, respectively.
[a] estimated coefficients (standard errors in the parenthesis)
[b] For the Reference group, this age cohort includes respondents of age 35 and older.

Table A2. Pooled nested model (RU1) for choice experiments CE_B and CE_T (Test group)

	All respondents in the TEST group	No college education	With college education	Age 18-34	Age 35-64	Age 65+
	1	2	3	4	5	6
βTax [a]	-14.266 ***	-15.685 ***	-11.025 ***	-26.711 ***	-13.222 ***	-11.922 ***
	(-1.236)	(1.703)	(1.514)	(4.346)	(1.921)	(1.614)
βPf(CE_B) [a]	-141.05 ***	-166.522 ***	-94.253 ***	-184.434 ***	-141.575 ***	-122.633 ***
	(-6.138)	(8.889)	(7.103)	(23.524)	(9.571)	(7.856)
βPf(CE_T) [a]	-97.506 ***	-88.562 ***	-117.876 ***	-138.296 ***	-130.396 ***	-68.138 ***
	(-12.953)	(15.556)	(24.066)	(34.181)	(23.166)	(18.090)
βPinj [a]	-1.315	-1.118	-1.775	0.398	-1.279	-1.947
	(-0.866)	(1.046)	(1.560)	(2.279)	(1.536)	(1.218)
βPev [a]	-0.036 ***	-0.031 ***	-0.046 ***	-0.085 ***	-0.061 ***	-0.006
	(-0.009)	(0.010)	(0.016)	(0.023)	(0.015)	(0.012)
ASC(evacuation) [a]	1.109 ***	1.281 ***	0.717 ***	1.368 ***	1.034 ***	1.078 ***
	(-0.141)	(0.176)	(0.240)	(0.387)	(0.251)	(0.194)
lambda (CE_B) [a]	1.72 ***	1.331 ***	3.216 ***	1.059 ***	2.237 ***	2.237 ***
	(0.745)	(0.963)	(0.399)	(1.211)	(0.573)	(0.573)
wald lambda (CE_B)	0.967	0.344	5.557	0.049	2.159	2.159
N cards	5370	3810	1560	740	1720	2910
N respondents	537	381	156	74	172	291
LL function	-3155	-2240	-904	-459	-956	-1693
Restr. LL function	-7444	-5282	-2163	-1026	-2384	-4034
Chi sqrd	8578	6084	2517	1135	2857	4681
Degrees of freedom	7	7	7	7	7	7

	All respondents in the TEST group	No college education	With college education	Age 18-34	Age 35-64	Age 65+
	1	**2**	**3**	**4**	**5**	**6**
VOSL (CE$_B$), mln € [b]	**9.887** **	**10.617** **	**8.549** **	**6.905** **	**10.707** **	**10.286** **
	(7.649-12.125)	(7.404-13.829)	(5.731-11.366)	(2.968-10.842)	(6.873-14.542)	(6.784-13.788)
VOSL (CE$_T$), mln € [b]	**6.835** **	**5.646** **	**10.691** **	**5.177** **	**9.862** **	**5.715** **
	(4.126-9.543)	(2.956-8.337)	(2.198-19.184)	(1.845-8.510)	(3.287-16.437)	(1.585-9.845)
VOSI (CE$_T$), € [b]	**92,183**	**71,263**	**161,037**	**-14,911**	**96,706**	**163,268**
	(-30,092-214,458)	(-61,643-204,168)	(-138,504-460,577)	(-182,402-152,580)	(-137,816-331,228)	(-53,381-379,954)
VOSE (CE$_T$), € [b]	**2,517** **	**2,001** **	**4,213** **	**3,148** **	**4,626** **	**493**
	(1,116-3,919)	(545-3,457)	(202-8,225)	(1,028-5,341)	(1,147-8,104)	(-1,505-2,491)

*, **, *** - statistical significance at 10%, 5% and 1% level, respectively.
[a] estimated coefficients (standard errors in the parenthesis)
[b] standard deviations for confidence intervals are calculated using the delta method (CI$_{95\%}$ in the parenthesis)

Table A3. Pooled nested logit model (RU1) for choice experiments CE_B and CE_T - scale sensitivity with respect to the valued ranges of the *monetary* parameter, generically estimated risk parameter (Test group)

Variable	All respondents in the TEST group		No college education		With college education		Age 18-34		Age 35-64		Age 65+	
	1		2		3		4		5		6	
βPf [a]	-98.199 (4.375)	***	-89.014 (5.001)	***	-118.912 (8.682)	***	-140.429 (18.129)	***	-131.556 (8.784)	***	-68.255 (4.546)	***
$\beta Tax(CE_B)$ [a]	-9.932 (1.019)	***	-8.384 (1.168)	***	-13.910 (1.996)	***	-20.338 (4.677)	***	-12.286 (1.993)	***	-6.636 (1.046)	***
$\beta Tax(CE_T)$ [a]	-28.148 (7.121)	***	-25.188 (8.761)	***	-31.610 (12.452)	**	-51.503 (18.075)	***	-28.378 (12.377)	**	-23.878 (10.253)	**
$\beta Tax(CE_T) * D(\Delta x_{Tax}=50\text{-}70)$ [a]	9.658 (4.710)	**	6.624 (5.818)		14.273 (8.185)	*	17.229 (11.886)		10.579 (8.211)		8.315 (6.787)	
βPev [a]	-0.034 (0.009)	***	-0.030 (0.010)	***	-0.045 (0.016)	***	-0.083 (0.023)	***	-0.060 (0.015)	***	-0.004 (0.012)	
$\beta Pinj$ [a]	-1.270 (0.867)		-1.081 (1.047)		-1.749 (1.564)		0.155 (2.290)		-1.245 (1.538)		-1.867 (1.221)	
ASCevacuation [a]	1.372 (0.202)	***	1.457 (0.245)	***	1.119 (0.362)	***	1.820 (0.519)	***	1.326 (0.361)	***	1.304 (0.284)	***
lambda (CE_B) [a]	2.471 (0.519)	***	2.490 (0.515)	***	2.549 (0.503)	***	1.391 (0.922)	***	2.408 (0.533)	***	3.195 (0.401)	***
wald lambda (CE_B)	2.835		2.894		3.079		0.424		2.643		5.467	
N cards	5370		3810		1660		740		1720		2910	
N respondents	537		381		166		74		172		291	
LL function	-3153		-2239		-903		-457		-955		-1693	
Restr. LL function	-7444		-5282		-2163		-1025		-2384		-4034	
Chi sqrd	8582		6085		2520		1137		2859		4683	
Degrees of freedom	8		8		8		8		8		8	

	All respondents in the TEST group	No college education	With college education	Age 18-34	Age 35-64	Age 65+
	1	2	3	4	5	6
VOSL (CE_B) [b] *(N cards = 2685)*	**9.887** ** (7.649-12.125)	**10.617** ** (7.404-13.829)	**8.549** ** (5.731-11.366)	**6.905** ** (2.968-10.842)	**10.707** ** (6.873-14.542)	**10.286** ** (6.836-13.736)
VOSL (CE_T - Δx_{Tax}=10-35) [b] *(N cards = 1234)*	**3.489** ** (1.732-5.245)	**3.534** ** (1.094-5.974)	**3.765** ** (0.808-6.716)	**2.727** ** (0.728-4.725)	**4.636** ** (0.627-8.645)	**2.859** ** (0.424-5.293)
VOSL (CE_T - Δx_{Tax}=50-70) [b] *(N cards = 1451)*	**5.311** ** (3.124-7.497)	**4.795** ** (2.312-7.278)	**6.859** ** (1.917-11.800)	**4.097** ** (1.401-6.794)	**7.391** ** (2.399-12.384)	**4.386** ** (1.097-7.674)

*, **, *** - statistical significance at 10%, 5% and 1% level, respectively.

[a] estimated coefficients (standard errors in the parenthesis)

[b] VOSL in mln €, standard deviations for confidence intervals are calculated using the delta method ($CI_{95\%}$ in the parenthesis)

Table A4. Pooled nested logit model (RU1) for choice experiments CE_B and CE_R (Reference group)

	All respondents in the REFERENCE group	No college education	With college education	Age 18-34	Age 35 and older
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**
βTax [a]	-3.416*** (0.217)	-3.019*** 0.250	-4.534*** 0.440	-3.777*** 0.389	-3.260 *** 0.262
βPf(CE_B) [a]	-24.914*** (2.363)	-23.372*** 2.946	-29.597*** 4.199	-22.925*** 3.284	-26.256 *** 3.289
βPf(CE_R) [a]	-40.045*** (2.500)	-38.398*** 2.884	-45.052*** 5.111	-39.545*** 4.427	-40.284 *** 3.028
βTTime [a]	-19.176*** (4.592)	-14.855*** 5.473	-30.592*** 8.572	-19.313** 7.914	-19.336 *** 5.648
lambda (CE_B) [a]	10.478*** (0.1224)	10.432*** 0.123	10.196*** 0.126	11.576*** 0.111	9.897 *** 0.130
wald lambda (CE_B)	77.434	76.720	73.106	95.463	68.662
N cards	2990	2060	930	970	2020
N respondents	299	206	93	97	202
Log likelihood function	-1607	-1142	-458	-517	-1088
Restr. log likelihood function	-4145	-2856	-1289	-1345	-2800
Chi sqrd	5076	3428	1663	1656	3426
Degrees of freedom	5	5	5	5	5
VOSL (CE_B), mln € [b]	**7.294** ** (5.605-8.983)	**7.742** ** (5.453-10.031)	**6.528** ** (4.328-8.728)	**6.069** ** (3.971-8.168)	**8.055** ** (5.706-10.403)
VOSL (CE_R), mln € [b]	**11.724** ** (9.630-13.819)	**12.720** ** (9.931-15.508)	**9.937** ** (7.028-12.846)	**10.469** ** (7.349-13.590)	**12.358** ** (9.695-15.021)
VOT, €/hour [b]	**5.61** ** (2.83-8.40)	**4.92** ** (1.28-8.56)	**6.75** ** (2.83-10.67)	**5.11** ** (0.88-9.35)	**5.93** ** (2.41-9.45)

*, **, *** - statistical significance at 10%, 5% and 1% level, respectively.
[a] estimated coefficients (standard errors in the parenthesis)
[b] standard deviations for confidence intervals are calculated using the delta method ($CI_{95\%}$ in the parenthesis)

Table A5. Pooled nested logit model (RU1) for choice experiments CE_B and CE_R - scale sensitivity with respect to the valued ranges of the *monetary* parameter (respondents in the Reference group)

	All respondents in the REFERENCE group		No college education		With college education		Age 18-34		Age 35 and older	
	Model 1		Model 2		Model 3		Model 4		Model 5	
βPf [a]	-41.541	***	-39.460	***	-48.420	***	-41.093	***	-41.924	***
	(1.962)		(2.296)		(3.866)		(3.567)		(2.369)	
$\beta Tax(CE_B)$ [a]	-5.695	***	-5.097	***	-7.417	***	-6.771	***	-5.205	***
	(0.517)		(0.612)		(1.008)		(0.942)		(0.623)	
$\beta Tax(CE_R)$ [a]	-4.786	***	-4.062	***	-6.640	***	-4.604	***	-4.897	***
	(0.685)		(0.820)		(1.276)		(1.152)		(0.854)	
$\beta Tax(CE_R) * D(\Delta x_{Tax}=300)$ [a]	0.918		0.688		1.257		-0.247		1.482	
	(0.788)		(0.937)		(1.487)		(1.376)		(0.970)	
$\beta Tax(CE_R) * D(\Delta x_{Tax}=450)$ [a]	1.266	*	0.946		1.878		1.080		1.370	
	(0.745)		(0.889)		(1.399)		(1.284)		(0.920)	
$\beta Tax(CE_R) * D(\Delta x_{Tax}=600)$ [a]	2.452	***	1.930	**	3.892	***	1.949		2.705	***
	(0.787)		(0.943)		(1.460)		(1.361)		(0.970)	
$\beta TTime$	-20.819	***	-15.836	**	-34.537	***	-21.765	*	-20.833	***
	(5.945)		(6.957)		(11.014)		(11.206)		(7.078)	
lambda (CE_B) [a]	6.284	***	6.1790	***	6,232	***	6.458***		6.199	***
	(0.204)		(0.208)		(0,206)		(0.199)		(0.207)	
Wald lambda (CE_B)	25.891		24.951		25.425		27.485		25.125	
N cards	2990		2060		930		970		2020	
N respondents	299		206		93		97		202	
Log likelihood function	-1601		-1139		-454		-514		-1083	
Restr. log likelihood function	-4145		-2856		-1289		-1345		-2800	
Chi sqrd	5088		3434		1671		1661		3435	
Degrees of freedom	8		8		8		8		8	

	All respondents in the REFERENCE group		No college education		With college education		Age 18-34		Age 35 and older	
	Model 1		Model 2		Model 3		Model 4		Model 5	
VOSL (CE_B), mln € [b]	**7.294**	**	**7.742**	**	**6.528**	**	**6.069**	**	**8.055**	**
	(5.605-8.983)		(5.718-9.766)		(4.511-8.545)		(4.119-8.020)		(5.964-10.146)	
VOSL (CE_R - Δx_{Tax}=150), mln € [b]	**8.681**	**	**9.715**	**	**7.292**	**	**8.925**	**	**8.562**	**
	(5.990-11.371)		(5.713-13.717)		(4.318-10.266)		(4.292-13.557)		(5.486-11.637)	
VOSL (CE_R - Δx_{Tax}=300), mln € [b]	**10.741**	**	**11.697**	**	**8.995**	**	**8.471**	**	**12.277**	**
	(8.145-13.336)		(8.316-15.078)		(5.547-12.442)		(5.260-11.681)		(8.375-16.178)	
VOSL (CE_R - Δx_{Tax}=450), mln € [b]	**11.801**	**	**12.665**	**	**10.168**	**	**11.660**	**	**11.887**	**
	(9.205-14.398)		(9.422-15.907)		(6.508-13.828)		(6.952-16.369)		(8.768-15.007)	
VOSL (CE_R - Δx_{Tax}=600), mln € [b]	**17.799**	**	**18.507**	**	**17.621**	**	**15.476**	**	**19.126**	**
	(11.531-24.066)		(10.089-26.925)		(7.646-27.596)		(6.410-24.542)		(10.683-27.570)	

*, **, *** - statistical significance at 10%, 5% and 1% level, respectively.
[a] estimated coefficients (standard errors in the parenthesis)
[b] VOSL in mln €, standard deviations for confidence intervals are calculated using the delta method ($CI_{95\%}$ in the parenthesis)

Table A6. Pooled nested logit model (RU1) for choice experiments CE_B and CE_R - scale sensitivity with respect to the valued ranges of the *risk* parameter (respondents in the Reference group)

	All respondents in the REFERENCE group	No college education	With college education	Age 18-34	Age 35 and older
	1	2	3	4	5
βTax [a]	-3.533*** (0.225)	-3.131*** (0.257)	-4.665*** (0.476)	-3.806*** (0.402)	-3.415*** (0.272)
$\beta Pf(CE_B)$ * $D(\Delta x_{Pf}(low))$ [a]	-10.242*** (2.279)	-8.126*** (2.705)	-15.754*** (4.391)	-12.377*** (3.539)	-8.944*** (2.968)
$\beta Pf(CE_B)$ [a]	-21.265*** (2.113)	-20.613*** (2.702)	-23.726*** (3.603)	-18.072*** (2.787)	-23.413*** (3.039)
$\beta Pf(CE_R)$ [a]	-64.254*** (7.604)	-62.902*** (8.771)	-69.500*** (15.779)	-45.140*** (13.175)	-73.899*** (9.381)
$\beta Pf(CE_R)$ * $D(\Delta x_{Pf}(high))$ [a]	26.606*** (7.675)	27.029*** (8.930)	26.549* (15.487)	6.183 (13.424)	36.846*** (9.429)
βTTime	-21.326*** (4.655)	-17.041*** (5.561)	-32.354*** (8.642)	-19.584*** (7.963)	-22.993*** (5.781)
lambda (CE_B) [a]	10.861*** (0.118)	10.636*** (0.121)	11.030*** (0.116)	12.702*** (0.101)	9.990*** (0.128)
Wald lambda (CE_B)	77.434	79.907	86.255	115.895	70.032
N cards	2990	2060	930	970	2020
N respondents	299	206	93	97	202
Log likelihood function	-1589	-1132	-448	-509	-1047
Restr. log likelihood function	-4145	-2856	-1289	-1345	-2800
Chi sqrd	5112	3448	1682	1671	3452
Degrees of freedom	7	7	7	7	7

	All respondents in the REFERENCE group	No college education	With college education	Age 18-34	Age 35 and older
	1	2	3	4	5
VOSL (CE_B - Δx_{Pf}(low)), mln € [b]	**8.917**** (6.763-11.071)	**9.177**** (6.763-11.071)	**8.462**** (5.326-11.598)	**8.001**** (5.008-10.994)	**9.475**** (6.482-12.468)
VOSL (CE_B - Δx_{Pf}(mid)), mln € [b]	**6.018**** (3.289-8.748)	**6.582**** (2.160-11.005)	**5.086**** (2.127-8.044)	**4.749**** (1.912-7.586)	**6.856**** (2.270-11.442)
VOSL (CE_R - Δx_{Pf}(mid)), mln € [b]	**18.185**** (13.397-22.973)	**20.087**** (13.718-26.456)	**14.897**** (7.630-22.164)	**11.861**** (4.645-19.078)	**21.639**** (15.285-27.993)
VOSL (CE_R - Δx_{Pf}(high)), mln € [b]	**10.655**** (8.658-12.652)	**11.456**** (8.658-12.652)	**9.206**** (6.312-12.101)	**10.237**** (6.936-13.538)	**10.850**** (8.348-13.352)

*, **, *** - statistical significance at 10%, 5% and 1% level, respectively.

[a] estimated coefficients (standard errors in the parenthesis)

[b] VOSL in mln €, standard deviations for confidence intervals are calculated using the delta method ($CI_{95\%}$ in the parenthesis)

Table A7. Results of mixed logit models for the respondents in the reference group (CE_B and CE_R, N=299) and in the test group (CE_B and CE_T, N=537), split by age and education level, estimated separately for each choice experiment. (**10.000 Halton draws**, normally distributed coefficients: CE_B and CE_R - fatality risk; CE_T -constant, fatality risk, risk of injury, risk of evacuation).

	All respondents in the TEST group	No college education	With college education	Age 18-34	Age 35-64	Age 65+
Baseline experiment						
VOSL (CE_B) [a]	**12.283** *** (-6.407-30.876) H	**12.673***** (-7.530-32.763)H	**11.326** *** (-3.898-26.465) H	**8.452** *** (-7.688-24.501) H	**12.801** *** (-3.128-28.640) H	**12.962** *** (-7.747-33.554) H
Test experiment						
VOSL (CE_T) [a]	**7.084** *** (1.330-12.850) H	**6.180***** (0.105-12.221)H	**9.878** *** (4.081-15.642)	**6.241** *** (-1.774-14.211) H	**9.908** *** (3.224-16.554)	**5.119** *** (5.023-5.213)
VOSI (CE_T) [a]	**92,014** ** (5,562-177,977)	**73,672** (36,328-110,805)	**157,407** (-235,982-548,578)	**-2,280** (-176,958-171,414)	**81,690** (-14,250-177,088)	**145,884** ** (9,967-281,035)
VOSE (CE_T) [a]	**1,905** *** (-4,786-8,559) H	**1,434**** (-5,249-8,079)H	**3,239** ** (-3,552-9,992)	**3,366** *** (3,255-3,477)	**4,116** ** (-1,071-9,274)	**-382** (-8,282-7,473) H
N resp. / N cards	537/ 2685	381/1905	156/ 780	74/ 370	172/ 860	291/ 1455
	All respondents in the REFERENCE group	**No college education**	**With college education**	**Age 18-34**	**Age 35+**	
Baseline experiment						
VOSL (CE_B) [a]	**8.879** *** (-4.848-22.528) H	**9.461***** (-5.983-24.817)H	**7.907** *** (-3.095-18.848) H	**7.123** *** (-3.346-17.532) H	**10.127** *** (-5.894-26.059) H	
Reference experiment						
VOSL (CE_R) [a]	**13.579** *** (-7.483-34.523) H	**14.367***** (-8.023-36.631)H	**11.966** *** (-5.709-29.541) H	**11.398** *** (-5.744-28.443) H	**14.918** *** (-8.757-38.460) H	
VOT (CE_R) [b]	**6.30** *** (3.59-9.02)	**5.94***** (2.44-9.49)	**6.65** *** (2.51-10.80)	**5.28** *** (1.28-9.28)	**7.04** *** (3.41-10.67)	
N resp. / N cards	299/ 1495	206/1030	93/ 465	97/ 485	202/ 1010	

*, **, *** - statistical significance at 10%, 5% and 1% level, respectively.
H - statistically significant heterogeneity is present in the respective risk coefficient at least at 5%.
[b] VOSL, VOSI and VOSE; std. based on mixed model estimations; confidence intervals are simulated ($CI_{90\%}$ in the parenthesis)
[b] VOT in €/h, standard deviations for confidence intervals are calculated using the delta method ($CI_{95\%}$ in the parenthesis)

Figure A1. Schematic representation of statistical designs of the baseline experiment (left) and the test experiment (right): ranges of valued differences in the risk parameter (upper charts) and the monetary parameter (lower charts).

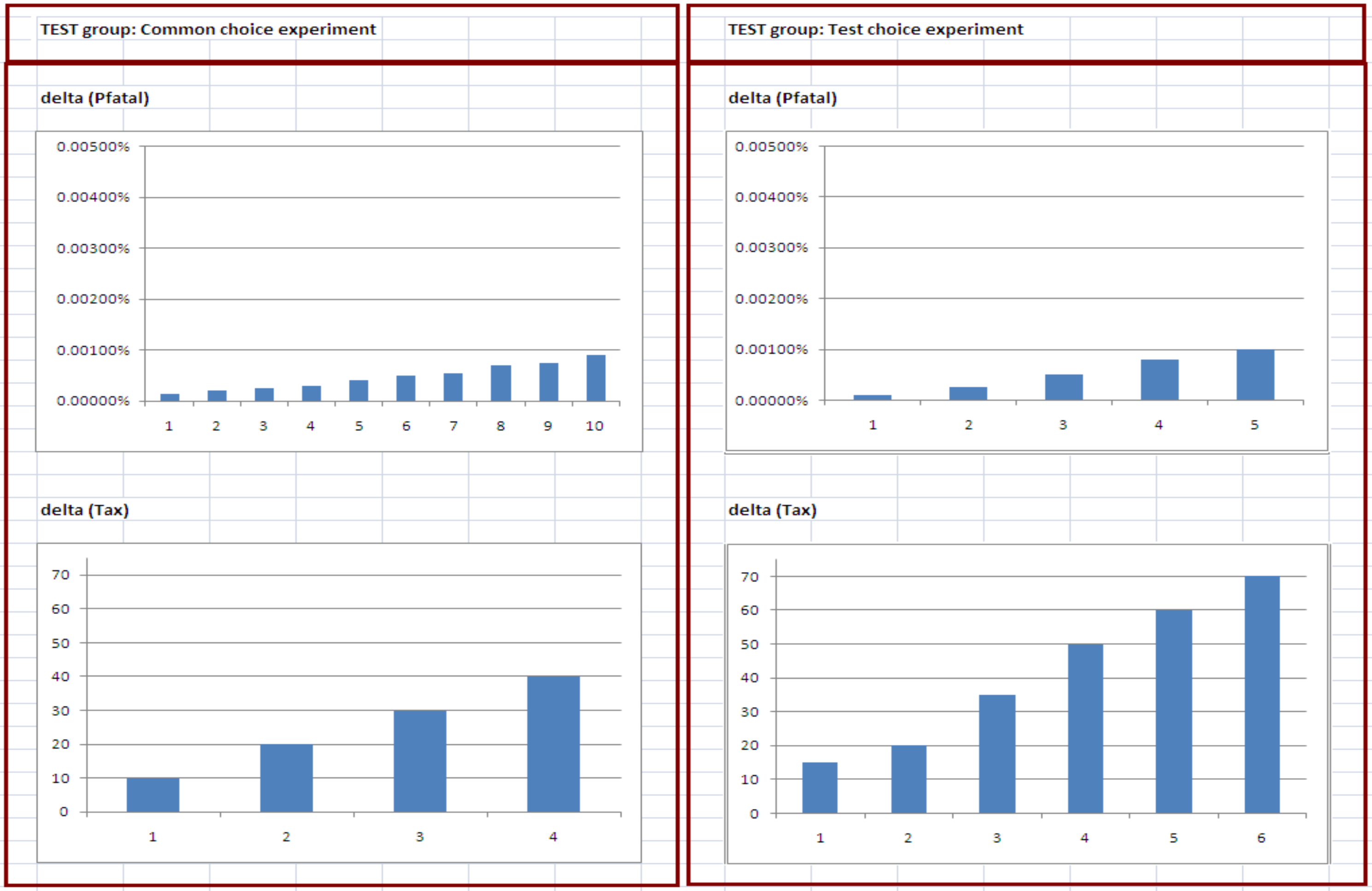

Figure A2. Schematic representation of statistical designs of the baseline experiment (left) and the reference experiment (right): ranges of valued differences in the risk parameter (upper charts) and the monetary parameter (lower charts).