Improving the risk–risk trade-off method for use in safety project appraisal responses

Jytte Seested Nielsen¹ · Susan Chilton¹ · Hugh Metcalf¹

Received: 2 May 2017 / Accepted: 13 June 2018 / Published online: 20 June 2018
© The Author(s) 2018

Abstract

The risk–risk trade-off method is a technique used to elicit the relative trade-off between changes in morbidity and mortality risks in stated preference surveys. The responses can be used to inform the (relative) values or weights that should be given to different accidents in cost–benefit analyses of road-safety projects that reduce the risk of death or injury. While the method has some distinct advantages over eliciting direct monetary measures of value, it is likely to suffer from similar problems that are found in other stated preference surveys, which might mitigate against its more widespread use. This study explores this issue, but shows that the estimates from a risk–risk trade-off study can be improved by employing a pre-survey learning experiment in which respondents make incentivised risky choices and also using a frame that focuses on the total risk or risks that respondents face, thereby broadening the toolkit available to measure preferences over road-safety interventions.

Keywords Risk–risk trade-off · Stated preference

JEL Classification J16 · D6

✉ Jytte Seested Nielsen
jytte.nielsen@ncl.ac.uk

Susan Chilton
susan.chilton@ncl.ac.uk

Hugh Metcalf
hugh.metcalf@ncl.ac.uk

¹ Newcastle University Business School, Newcastle University, 5 Barrack Road,
Newcastle upon Tyne NE1 4SE, England, UK

1 Introduction

1.1 Cost–benefit analysis and stated preference studies

Cost–benefit analyses (CBA) of safety improving initiatives can be used to inform the allocative decision process (e.g., see Treasury 2003; US EPA 2014). In a conventional CBA, the value of a statistical life (VSL) is defined as the aggregated willingness-to-pay (WTP) for small-risk reductions which reduce the expected number of premature fatalities in the group affected by one and equivalent benefit measures can be elicited for non-fatal injuries. Stated preference techniques are often used to directly elicit people's preferences over changes in safety. The majority of such previous stated preference studies have used contingent valuation (CV) techniques (e.g., Johannesson et al. 1996; Andersson and Lindberg 2009); however, discrete choice experiments (DCE) have likewise been applied (e.g., Alberini and Šcasný 2011; Iragüen and de Ortúzar 2004). Both techniques have been subjected to several studies regarding their respective reliability in various valuation contexts (see, e.g., Goldberg and Roosen 2007; Carson et al. 2001; Carlsson et al. 2005), with mixed results. One option available to analysts to deal with unreliable responses may be to discard them. However, discarding responses is undesirable, since it might then compromise the representativeness of the sample, potentially invalidating the resulting CBA recommendation.

In addition to directly estimating the WTP for safety, 'indirect' methods have also been used to estimate the monetary value of preventing fatalities in different contexts. In the UK, the Department for Transport (DfT) commissioned a study to find evidence in favour of applying 'dread' premia to VSLs for accidents other than road accidents (Chilton et al. 2007). Using the risk–risk trade-off approach (Viscusi et al. 1991; Chilton et al. 2006), no strong evidence could be found to support the application of such a premium, resulting in no change to Treasury advice for applying the VSL to accidental death. This finding was also supported in an earlier study, in which a different, but related, relative valuation approach ('matching'), was applied to infer monetary values for the prevention of fatalities in the public sector. Here, the relative monetary valuations for other accidents were 'pegged' to the roads VSL, and the results broadly supported a 1:1 relativity in all cases, implying a single VSL across all sectors (Beattie et al. 2000; Chilton et al. 2002). Finally, the DfT current monetary values for the prevention of different severities of non-fatal road injuries were obtained by applying relative valuations to an absolute monetary 'peg' in the form of the WTP-based monetary value for the prevention of a road fatality (Beattie et al. 1998; Jones-Lee et al. 1995).

Two sources of error might occur when the risk–risk ratios are applied to adjust a VSL, namely, the risk–risk ratios and the VSL itself.¹ Given that policy makers have commissioned relative valuation method studies, particularly in the UK (see above), it is, nevertheless, the case that they have in fact been subject to much less

¹ We thank a referee for pointing out this potential problem.

scrutiny than the direct methods. This paper attempts to close this gap, at least with respect to one indirect method in particular, the risk–risk trade-off method. It has been argued that the risk–risk trade-off method ‘potentially lowers the cognitive burden on respondents by allowing them to compare more similar “commodities”’ (Van Houtven et al. 2008, p. 183), i.e., risks of accidents. This avoids the need for respondents to make a direct money–risk trade-off (as in traditional contingent valuation, for example) given that income and changes in fatal (and non-fatal) risk might be potential incommensurate attributes (Magat et al. 1996). Although a number of empirical studies are reported in the literature (see Table 1, Sect. 2.2), it is striking that only a very few examine potential methodological issues that may affect the reliability and validity of responses elicited using the risk–risk trade-off method.

1.2 Reliability and validity: risk–risk trade-off method

In stated preference surveys, respondents are faced with hypothetical choice situations. This characteristic led to early criticism of the method (e.g., Cummings et al. 1986) and a subsequent body of research exploring the behavioural biases that have been found to influence the reliability and validity of the estimates (summarized in Bateman et al. 2002). This was almost exclusively in the context of WTP. Drawing on this literature, one way to explore the reliability and validity of risk–risk trade-off responses is to focus on the behaviour of the mean relative trade-off under different conditions. For example, the more ‘outliers’ present in the data—in either or both ends of the distribution—the more likely that the central tendency measure will be unstable, with an associated decrease in its reliability and/or validity. In addition, a smaller variance is to be preferred, all other things equal, since it improves the power of subsequent statistical tests. The question then becomes one of whether behavioural biases that affect the reliability and validity of the mean estimates arising from using stated preference mechanisms in general might also be expected to (negatively) affect the risk–risk trade-off method?

A major issue that pervades stated preference elicitation is that of the general lack of sensitivity in survey responses to changes in characteristics that common sense and economic theory would predict should matter to respondents. Examples of this can be found in the related WTP literature and include ‘scope insensitivity’—whereby respondents state identical WTPs for reducing the risk of two very different health outcomes (Gyrd-Hansen et al. 2012; Olsen et al. 2004) or state identical WTPs for two very different sizes of risk reductions for the same outcome (e.g., Carson and Mitchell 1995; Fetherstonhaugh et al. 1997; Goldberg and Roosen 2007) or in the elicitation of a proportionally large number of zero WTP offers (Carson et al. 2001). In the former case, it appears that some respondents anchor completely on their first response, while in the latter, they affix on the current expenditure on the good, ignoring the fact that the quantity or quality has changed. In a similar vein, when eliciting health-related quality of life weights in the standard gamble (SG) surveys, the SG-based values have been found to be biased upwards which could be partly attributed to a ‘certainty effect’, a pervasive psychological phenomenon,

Table 1 Summary of risk–risk trade-off studies

References	Risk outcomes	Indifference elicitation mechanism	Frame
Viscusi et al. (1991)	Fatal car accident and chronic bronchitis	Two areas which varied in the risk of both outcomes. One of the risks in the respondent's preferred area was lowered until indifference was reached	Total
Dolan et al. (1995)	Four non-fatal illnesses	Risk of non-fatal illness was increased marginally. Respondents were asked how much the only non-fatal illness should be reduced to compensate for this increase	Marginal
Magat et al. (1996)	A nerve disease and a fatal car accident compared to cancer of the lymph system and a fatal car accident	Similar to Viscusi et al. (1991)	Total
Clarke et al. (1997)	(Painless) death from a natural disaster and Gaucher Diseases (metabolism disorder)	Two hypothetical cities which varied in the risk of both outcomes. The respondent viewed a sequence of scenarios until indifference was reached	Total
Chilton et al. (2006)	Two different accident types were compared by means of their fatal risks	Risk in the respondent's preferred accident type was increased marginally in a multiple list until indifference was reached	Marginal
van Houtven et al. (2008)	Fatal car accident and fatal stomach illness	Two areas which varied in the two risks. Each respondent received five choices and was asked to indicate their area preference in each choice separately	Total
Cameron et al. (2010)	Injury and death from a landmine accident	Two different communes. The risk of an injury in the respondent's least preferred commune was decreased until indifference was reached	Total
Nielsen et al. (2010)	Different distributions of mortality risks (no context)	Risk in the respondent's preferred distribution was marginally increased until indifference was reached	Marginal
McDonald et al. (2016)	Fatal car accident and fatal cancer	Risk in the respondent's preferred accident type was marginally increased in a multiple list until indifference was reached	Marginal

whereby an individual values positive outcomes more highly when they occur with certainty (Tversky and Kahneman 1981; Law et al. 1998).

Such behaviour appears driven, at least in part, by the decision heuristic of ‘anchoring and adjustment’ (Tversky and Kahneman 1974) and/or to the notion of a reference point (Kahneman and Tversky 1979). Anchoring and adjustments refer to situations in which ‘people make estimates by starting from an initial value, which may be suggested by the formulation of the problem, and then adjust that value to yield the final answer’ (Mitchell and Carson 1989, p. 115). Alternatively, some people may not make any adjustment from their initial ‘anchor’, which in effect serves as their ‘reference point’. Still, others may be susceptible to anchoring and insufficient judgement (Tversky and Kahneman 1974; Quattrone 1982; LeBoef and Shafir 2006).²

While the risk–risk trade-off method is potentially cognitively less difficult than WTP-based methods, it is, nevertheless, likely that it will be affected in its own particular way by this anchoring problem and respondents would appear insensitive to the risk change per se. This might manifest itself in two distinct sets of responses. First, in a reluctance to take any risk increase or decrease at all (from an initial position of equal risk of two different types of injury) or, second, displaying an unwillingness to take any risk increase whatsoever in a less favoured outcome, e.g., a fatal road accident, no matter how large the change in risk of a more preferred outcome (e.g., a non-serious, non-fatal road accident). In the former, the reference point is the initial risk(s) and we will term these ‘non-traders(I)’, while in the latter, it is the more severe outcome and we will term these ‘non-traders(O)’. Either type of non-trader could generate values which may be considered as outliers in data, since they would fall in the extreme tails of the response distribution, thereby potentially significantly impacting the variance and mean of the relative trade-offs for different road accidents if retained in the data set.

Another potential problem that could impact the risk–risk trade-off method relates to framing effects and the resulting violation of the assumption of procedure invariance (Tversky and Kahneman 1981; Tversky and Thaler 1990). This predicts that responses should not vary between two versions of a choice problem that differ in relatively subtle ways, yet evidence shows that strikingly, different choices can be made by respondents (Tversky and Kahneman 1981) which raises concerns about the validity of the survey. Survey respondents have, for example, been found to be very sensitive to whether risk information is presented as absolute or relative risk reductions (Baron 1997; Gyrd-Hansen et al. 2003) and respondents have been found to be sensitive to whether risk reductions or avoided fatalities are presented in a VSL survey (Kjaer et al. 2018). In the risk–risk trade-off literature, two different frames have been used to elicit risk–risk relativities between different outcomes. The risk decrement or increment has either been presented as a marginal change to the current situation (Chilton et al. 2006) or as the (new) total risk that the respondent

² Experimental economists have extensively investigated this phenomenon and have found this heuristic to be pervasive in real, incentivised settings too. Whilst outside the scope of this paper the interested reader is referred to Dhami (2016, pp. 1370–1375) for an excellent and thorough review of this evidence.

would face in the changed situation (Viscusi et al. 1991). How this affects the central tendency measure, and in particular, the anchoring behaviour of non-traders(I) and of non-traders(O) remains an empirical question, one which we also address in this study.

In summary, the ‘risk–risk trade-off’ method could be affected by peoples’ (in) sensitivity to risk changes, framing effects, or both. The aim of our study is to investigate the effect of both in terms of the impact on response reliability, and furthermore, the introduction of a pre-survey experiment in which respondents make (incentivised) risky choices over a familiar good (money), as in a standard economic laboratory experiment the lessons from which can then be taken forward into trade-offs over less familiar goods (here, fatal, and non-fatal accidents). The idea that the lessons learnt in an incentivised setting can be carried over into a non-incentivised survey is similar in spirit to the ‘rationality spill over’ phenomenon introduced in Cherry et al. (2003). It also follows a finding in the literature that behavioural biases are reduced, as people become more familiar with the tasks (Carlsson et al. 2012; Plott 1996). If a learning experiment—based on the same choice framework faced by respondents in the survey—is shown to improve the validity of a risk–risk trade-off survey, the need for discarding responses could be diminished and the representativeness of the samples used to elicit preferences for policy purposes would be improved. The second problem—framing—is addressed in the experimental design by employing two different frames: marginal and total risks. Both frames convey the same information (the risk change), but in two different ways.

The remainder of this paper is organised as follows. Section 2 presents the risk–risk trade-off methodology and a summary of the previous risk–risk trade-off surveys. Section 3 then provides a more detailed description of the experimental design and the four different treatments. The analytical strategy is presented in Sect. 4; results are presented in Sect. 5 and discussed in Sect. 6.

2 Risk–risk trade-off methodology

2.1 Theory

The problem facing the respondent in a risk–risk trade-off scenario can be described in an expected utility framework following Viscusi et al. (1991) and Van Houtven et al. (2008).

We assume that respondents make choices to maximize expected utility:

$$E(U) = r_F U(F, w) + r_S U(S, w) + (1 - r_F - r_S) U(H, w). \quad (1)$$

According to this expression, utility is determined by health outcomes (F, S, or H) and wealth (w). Individuals are assumed to face risks of three mutually exclusive outcomes of a traffic accident within the next period. The first is fatal (F), i.e., dying in a traffic accident with probability (r_F), the second is a serious non-fatal injury (S) with probability (r_S), and the third is ‘normal health’ or all other health outcomes

(H). To simplify the expression, we have assumed that R, S, and F are mutually exclusive events.³

The respondent is asked to consider a choice between two options. In option A, the risk of a fatal road accident is r_F^A , and the risk of a non-fatal serious outcome is r_S^A , while in option B, these probabilities are r_F^B , r_S^B . As such, option A gives the expected utility in equation (Eq. 2), while option B gives the expected utility in equation (Eq. 3).

$$E(U)^A = r_F^A U(F, w) + r_S^A U(S, w) + (1 - r_F^A - r_S^A) U(H, w) \quad (2)$$

$$E(U)^B = r_F^B U(F, w) + r_S^B U(S, w) + (1 - r_F^B - r_S^B) U(H, w). \quad (3)$$

If an expected utility maximiser indicates indifference between options A and B, this suggests that

$$E(U)^A = E(U)^B. \quad (4)$$

Rearranging Eq. (4) gives us the following:

$$U(S, w) = rr_{SF} U(D, w) + (1 - rr_{SF}) U(H, w), \quad (5)$$

where $rr_{SF} = \frac{r_F^B - r_F^A}{r_S^A - r_S^B}$.

The utility of a serious non-fatal injury has thereby been transformed into an equivalent lottery on life with good health and death similar to the model framework setup in Jones-Lee (1976).

Risk–risk trade-off surveys have been carried out previously and Table 1 lists the published studies along with a description of the risk outcomes used, the mechanism used for eliciting indifference point and the framing of the risks, i.e., whether respondents were informed of the new, and total risk they would face or the marginal risk change. In both cases, respondents were aware of the baseline risk they faced. For the purpose of this investigation, we have chosen to include only studies which (1) involve no monetary comparison and (2) elicit an indifference point between two risk reductions (for different outcomes).

2.2 Previous studies

The studies in Table 1 focus largely on reporting the empirical results. An exception to this is the methodological investigation by Clarke et al. (1997) in which the results from the standard gamble method, time trade-off method, and risk–risk trade-off method were compared. Utilising a test–retest methodology, they found

³ In fact, Eq 1 is an approximation of EU which would more accurately be specified as $E(U) = r_F U(F, w) + r_S U(S, w) + (1 - r_F)(1 - r_S) U(H, w)$

$= r_F U(F, w) + r_S U(S, w) + (1 - r_F - r_S - r_S r_F) U(H, w)$

Dropping this interaction term, which results in Eq. 1, greatly simplifies the conceptual framework without significantly altering the empirical findings and is in accordance with the state-of-the art within this literature (Van Houtven et al. 2008).

that the results from the risk–risk trade-off method performed worst with respect to reliability.

Central to the method is the elicitation and/or estimation of indifference points between two risks (as discussed in Sect. 2.1), the measurement and elicitation of which at the individual level could be prone to anchoring-type behaviour and/or reference point effects. Methods to elicit this include variations of the multiple list formats (in which different combinations of risks are presented to the subjects in a list) or the dichotomous choice format (where the different combinations of risk are presented sequentially (randomly or in a decreasing/increasing manner). In this study, we will apply the multiple list format (Chilton et al. 2006; McDonald et al. 2016). It is based on more generic multiple list methods used in a number of domains [for example, in the form of a payment card in a CV survey (Bateman et al. 2002)]. While it might be susceptible to some of the general problems surrounding stated preference mechanisms (Andersen et al. 2007), its primary purpose is to act as a medium by which to help respondents iterate to their indifference point, compared to a question format, e.g., open-ended providing no such help. If successful, it should work to reduce—rather than exacerbate—the anchoring that the risk–risk trade-off method itself is potentially susceptible to, i.e., the non-trading considered in this paper.

With respect to the framing issue, the previous studies have presented the changes in risk as either marginal or total, but to the best of our knowledge, a comparison of the two frames has not been carried out so far in a survey applying the risk–risk trade-off method.

3 Experimental design

Our study is based on an experimental design setup to test whether an incentivised learning section (referred to as the learning experiment) and the frame can have an impact on the spread of responses, by reducing the number of extreme outliers in either or both tails. Since the purpose of this study is to analyse and compare behavioural responses to different stimuli (incentives and framing) and not to elicit preferences to be used for policy purposes, a convenience sample of young adults formed the subject pool.

180 subjects were recruited from the undergraduate student body (59, 33, and 8% on stages 1, 2, and 3, respectively, of their degree programmes) at Newcastle University during the 2013 spring semester. A majority (75%) of the students in the sample studied one of the six economics programmes. A total of 10 sessions were conducted with 7–32 students depending on student and room availability. Students were randomized into experimental treatments and sessions and were instructed to answer on an individual level: no open-ended discussions were introduced during these sessions. A session lasted between 40 min and 1½ h depending on treatment and each session was moderated by one member of the research team, aided by trained assistants. The main data collection was preceded by a piloting phase in which the protocol was refined and amended. A total of 27 students took part in the piloting phase.

Table 2 Experimental design

	Marginal risk frame	Total risk frame
Learning experiment (LE)		
No	T1 (control) <i>n</i> = 50	T2 (control) <i>n</i> = 54
Yes	T3 <i>n</i> = 40	T4 <i>n</i> = 36

Table 3 Injury descriptors

Code letter	Injury severity	Risk per decade
F	Fatal	50 in 100,000
S	Serious ^a	600 in 100,000
NS	Non-serious ^b	5000 in 100,000

^aSerious permanent in the terminology of Jones-Lee et al. (1995)

^bSerious non-permanent in the terminology of Jones-Lee et al. (1995)

Risk–risk trade-offs were elicited in surveys in four different treatments, as illustrated in Table 2. Learning experiments were incorporated into two of the four treatments. The marginal risk frame is tested with and without a learning experiment [T3 and T1 (control), respectively] and the total risk frame likewise, i.e., T2, acts as a control for T4.

3.1 Survey injury severities defined

We employed three potential road accident injuries in the survey—a fatal injury and two non-fatal injuries of differing severities (of which brief descriptions were presented to subjects). The non-fatal injuries descriptions were taken from the previous studies (Jones-Lee et al. 1995; Beattie et al. 1998). Non-serious (NS) below is equivalent to a serious non-permanent injury (with no chronic impacts), whereas serious (S) is equivalent to a serious permanent injury (with chronic impacts). These two types of injuries thereby span a spectrum of non-fatal injuries descriptions previously used in the VSL literature and have been chosen to be in line with the previous studies and to test for any dependence on seriousness of the injury and whether or not it is compared with the risk of a fatal accident. The injury descriptions are reproduced in full in the Appendix, but stylised descriptions, along with associated probabilities of occurrence per decade for the typical car driver or passenger in the UK, are shown in Table 3.

The probabilities were approximated using information from the Department for Transport Statistics to represent broadly what they classify as a serious injury and a slight injury.⁴

⁴ Probabilities for an individual aged 20–29 were estimated for fatal accidents and accidents causing serious and slight injuries.

3.2 Survey context

The most commonly used scenario in the ‘risk–risk trade-off’ literature (see Table 1) is to frame the question as a choice of moving to one of the two different areas which differ only by the mortality and/or morbidity risk. We adopted this scenario in this study and the respondents were given the choice between two employment opportunities (from the same firm) that would be open to them on graduation. Both jobs would be at the same firm, carry the same responsibilities, and pay the same salary.⁵ The only difference is that the two workplaces (and hence, for the purposes of this scenario, place of residence) would be located at one of the two distinct areas of the same city. Housing, alternative employment opportunities, etc., could be considered to be identical; however, the risk of being involved in differing types of road accidents would be different. Each respondent was asked to make three pairwise choices, namely: area 1 versus area 2; area 2 versus area 3; and area 1 versus area 3. Each area was characterized by two risks: the risk of a non-fatal injury (NS or S) and the risk of a fatal injury (F) or the risk of two different non-fatal injuries (NS and S). Area 1 was characterized by an increase in NS, area 2 by an increase in F, and area 3 by an increase in S. Following standard risk trade-off practice, a ‘forced choice’ was imposed on respondents in that staying with the status quo (not moving) was not given as an option. In the survey, we wanted the respondents to rank the injuries according to their own preferences, and hence, we avoided labelling the injuries according to seriousness using the letter R to represent non-fatal serious injury, W non-fatal non-serious injury, and F fatal injury.

3.3 Eliciting risk–risk indifference points

For illustrative purposes, we describe how the risk–risk indifference point between S and F was elicited (SF) in treatments 1 and 3. First, as an introduction, the respondents were informed that currently, where they live their chance of being involved in a traffic accident leading to a non-fatal injury S is 600 in 100,000 per decade, whereas the risk of being involved in a traffic accident leading to a fatal injury F is 50 in 100,000 per decade. The respondents were then given the choice whether they would prefer to move to area 2, where the risk of a non-fatal injury S would increase by 5 in 100,000 or move to area 3, where the risk of fatal injury F would increase by 5 in 100,000. The scenario was presented to them, as illustrated in Fig. 1.

This design was chosen, so that a majority of respondents would chose to move to the area with the risk increase in the less severe outcome (here area 2), thereby maximizing the sample size for subsequent analyses. Respondents were also given the opportunity to be ‘equally happy’ (i.e., indifferent) in which case they were asked to respond to an open-ended question to explain their reasons for this.

⁵ This is standard procedure in such surveys. The survey was subject to the usual piloting and pre-testing and no issues with respect to credibility were raised during those procedures. As such, we do not have any evidence to suggest research participants thought otherwise but must assume that if biases have been introduced due to this they are working in the same direction (or randomly) across all treatments.

RISK (per decade)			
Non-fatal S	600 in 100,000		
Fatal	50 in 100,000

Area 2; An increased risk of Non-Fatal S	
YOUR RISK INCREASE (per decade)	
Non-Fatal S	5 in 100,000
Fatal	0 in 100,000

Area 3; An increased risk of Fatal F	
YOUR RISK INCREASE (per decade)	
Non-Fatal S	0 in 100,000
Fatal	5 in 100,000

Fig. 1 Example of scenario

For those respondents expressing a strict preference for one area, their relative strength of preference between the two areas was elicited by explaining to the subjects that following their initial choice, the firm would write to them to let them know that new information had become available which meant that some of the statistics that they had provided previously required updating and that there would be an increased risk of one of the accident types (specified) in their chosen area, i.e., if the area with an increase of 5 in 100,000 in risk of S was chosen by the respondent, then the updated increase in the risk of S would be greater than 5 in 100,000. Given that increase, the firm would want to know if they would still move to the chosen area or would prefer now to go to the other area. Following Chilton et al. (2006) and McDonald et al. (2016), risk increases were utilised instead of risk decreases because of the relatively low baseline risk for a fatality, leaving little if any scope for decreasing the risk.⁶ To make their choice, each respondent received a table customised to the choice they just made. In the table, their chosen area risk profile was gradually made worse in an exponentially increasing manner.⁷ An example of the table used can be seen in the [Appendix](#). All respondents were asked to go through the table, row by row, and indicate in each row which of the two areas they preferred. Respondents were then asked to record the first row, where they switched

⁶ In addition, risk increases rather than reductions were employed essentially because (in whole number terms) there is a lower bound to the magnitude of feasible risk reductions; see Chilton et al. (2006) for a discussion. In addition, under an expected utility framework, the trade-off for very small-risk increases and decreases should be identical.

⁷ Ideally, the tables would have looked identical for all outcomes. However, since the baseline risks for the three different outcomes were very different we found it necessary to adjust them to take into account that a 5 in 100,000 risk increase would not be perceived as a substantial change in the NS case whereas it would make many subjects switch in the case of F. Hence, if linear increases of for example 5 in 100,000 had been used for all outcomes, the table for the NS case had to be extended extensively to accommodate subjects variations in indifference points. Hence we chose to increase the risks in an exponential manner in accordance with the recommendations in Rowe et al. (1996) for reducing biased answers from payment cards.

Fig. 2 Money wheel

(from area 2 to area 3 in the table included in the [Appendix](#)). We referred to that as the “switching point”. If they failed to switch to indicate a preference for the other area in the table, they were asked to write down their own number, i.e., the risk increases that would induce them to prefer the other area.⁸ The question order was varied in such a way that the non-serious/serious injury trade-off (NS/S) was always presented first followed by either the serious/fatal injury trade-off (S/F) or the non-serious/fatal injury trade-off (NS/F).

The presentation and tasks in treatments 2 and 4 were identical to treatments 1 and 3, respectively. Only the risk framing differed in that risks were presented as total risks, i.e., total risk frame. Respondents were asked to compare areas by making pairwise choices between total risks (including the baseline risk information) in the two areas. As an example, respondents were then given the choice of whether they would prefer to move to area 2, where risk of a non-fatal injury S would increase to 605 in 100,000 or move to area 3, where the risk of a fatal injury F would increase to 55 in 100,000. The tables used for finding indifference points were likewise framed in terms of total risks, but otherwise, identical to the ones used in treatments 1 and 3.

3.4 Learning experiment; incentivised money wheel game (T3 and T4)

The ‘risk–risk trade-off’ survey questions in T3 and T4 (Table 2) were preceded by the incentivised learning experiment. Comparing the results to T1 and T2, respectively, the effect—if any—of the learning experiment could be isolated in either or both the frames. As noted, the learning experiment was based on the same choice

⁸ No tables included multiple switching or obvious errors.

framework faced by respondents in the survey and provided them with experience of trading-off different probabilities of varying (negative) outcomes against each other in an incentivised environment. Respondents' learning in the experiment was facilitated by a 'money wheel'. As illustrated in Fig. 2, the money wheel consisted of 100,000 segments in three different colours and a dial.⁹

All probabilities were expressed in a baseline out of 100,000; the same baseline used in the following survey on traffic risk. The dial was spun and stopped randomly by the moderator pressing a key on the computer which the respondents observed. The colour of the segment that the tip of the dial landed on decided the outcome of that round (if the tip landed on the line, between two colours, the dial was spun again). Landing on a colour had a different effect on what a respondent could finally take home with them. If the dial landed on blue no tokens were lost, green meant three tokens were lost and yellow resulted in the loss of eight tokens.

The experiment consisted of three games: one practice 'money wheel' game and two real, incentivised 'money wheel' games. The wheels varied from game to game, although they were all constructed, such that the largest proportion of the wheel was blue (no loss of tokens) and a smaller proportion green (a small loss). The smallest proportion in the practice game and the first real game was yellow (a substantial loss) and pink in the second real game (a loss of everything). Progressing from the practice game to the first real game, the proportion of yellow decreased and the proportion of pink in the second real game was even smaller. The relative distribution of colours in the wheels matched the relative distributions of the different severities of traffic accidents in the subsequent survey, i.e., a higher chance of landing on blue (survival in the context of traffic), a smaller risk of landing on green (a non-serious injury), and an even smaller risk of landing on yellow or pink (a serious injury or death).

To start the experiment, each respondent received ten experimental tokens, each worth £1. The subjects were informed that any remaining tokens would be redeemable at the end of the session and that the amount received at the end of the session would depend on their choices (however, they were guaranteed their fee for showing up). Each game was made up of two rounds. In the first round of a game, respondents were asked to choose between two different 'money wheels'. The 'money wheels' were presented to them on slides, respondents were asked to choose between them, their chosen wheel was then spun, and the result recorded. As the spin of the wheel could result in a loss of tokens, this experiment allowed respondents to experience the consequence of making choices that were in accord with their preferences (playing the wheel that they preferred) or not (playing the wheel they did not prefer).

In the second round of a game, individual indifference points between the two wheels were elicited by making a respondent's chosen wheel incrementally worse. All respondents were presented with a 6-page booklet. On each page of this booklet, the non-preferred wheel remained fixed, i.e., unchanged and the initially preferred wheel was made worse by removing some blue segments and replacing them with either green or yellow segments. They were asked to indicate on each page

⁹ A sample 'money wheel' is available from the authors.

which wheel they would prefer to play and then to record their switching point (i.e., where the composition of colours in the changed wheel induced them to prefer to play the other, unchanged wheel). After they had made their six choices, the roll of a die determined which set of wheels to play out. These wheels were then spun and the outcome of their chosen wheel was recorded. Again, by incentivising the task, a respondent could potentially experience the consequence of making choices that were in accord with their preferences (playing the wheel that they preferred) or not (playing the wheel they did not prefer if they switched too early or too late).

In total, the wheel was spun six times across the three games (twice in the practice game).¹⁰ At the end of the four money wheel rounds, following procedures common in experimental economics, a random draw decided which of the four wheel spin in the real rounds would determine the subjects' earnings. The different steps in the games are summarized in Table 7 in the [Appendix](#).

4 Analytical strategy and testable hypotheses

Respondents' indifference points between two different risky outcomes will be the focus of the analysis. Each indifference point will be used to calculate the ratio between the risk increases which induced the respondent to switch to the other area. Suppose, for example, that a subject indicates that she prefers a 5 in 100,000 risk increase of a serious injury S to a 5 in 100,000 risk increase of fatal risk F, but prefers a 5 in 100,000 risk increase of a fatal risk F to a risk increase of 41 in 100,000 in a serious injury S (i.e., switched in row number 1 in the example table in the [Appendix](#)). Following (Eq. 5), we will approximate this individual's 'death risk equivalent' between the two injuries, rr_{SF} , by 5/41.

However, for the purpose of the analyses in the next section, we will use the inverse of the 'death risk equivalent'; thus, $RR_{SF} = 1/rr_{SF}$, i.e., in the example above, $RR_{SF} = 41/5 = 8.2$. The interpretation of RR_{SF} is that the disutility of a fatal outcome F is perceived to be 8.2 times higher than the disutility of the serious injury S outcome. The reason for using the inverse of the 'death risk equivalent' is that it is more intuitive: the higher the number (RR) the more the most serious injury outcome is feared relative to the less serious injury outcome.

The geometric mean¹¹ of RR for each of the three different pairwise comparisons will be analysed independently across the different treatments. The

¹⁰ Each game was split into two rounds to ensure that no strategic incentives would appear and hence subjects have the incentive to be truthful in the initial question since that could be played out for real. If the game had only consisted of one round, one potential strategy would be to lie in the answer to the first question (e.g. indicate a preference for B even though the true preference was for A) but then in the subsequent questions on indifference immediately switch back to A. The subject would now indicate a preference for the unchanged A wheel on all pages in the booklet except the first page. Since the roll of a dice will decide which page number to play out; this would be dominant to saying the truth (A) in the initial choice and then making A worse in the subsequent indifference task. Across all the session with the learning experiment, only two students did not lose any tokens at all across the six games.

¹¹ The geometric mean of a data set $\{RR_1, RR_2, \dots, RR_n\}$ is given by: $(\prod_{i=1}^n RR_i)^{1/n} = \sqrt[n]{RR_1 RR_2 \dots RR_n}$ (Sydsaeter and Hammond 1995).

geometric mean is usually considered the most appropriate measure of central tendencies for ratios, as it reduces the influence of extreme outliers (Chilton et al. 2006; Baron 1997). Since the geometric mean is equivalent to the anti-log of the average logged RR, $\ln RR$ will be used as the dependent variable in the regression analyses. In the regression analyses, the standard errors will be clustered at the individual level to take into account that each respondent has answered more than one question, and hence, there could be some interdependence between the question answers. For each of the pairwise comparisons, only those respondents who preferred to increase the risk of the less severe outcome will be included in the analyses, since with two equal sized risk increases, the a priori expectation is that most individuals would prefer to increase the risk of the less severe injury. To choose otherwise would be inconsistent.

First, we will test whether the learning experiment (LE) significantly impacts the mean indifference point between the two risky outcomes. Thus, we set up the following hypothesis:

Hypothesis 1a

$$H_0: \text{mean}(\ln(RR))_{LE} = \text{mean}(\ln(RR))_{no LE} \\ H_A: \text{mean}(\ln(RR))_{LE} \neq \text{mean}(\ln(RR))_{no LE}$$

Following this, we will test whether the learning experiment has a significant effect on the distribution of responses, i.e., it significantly impacts the variance:

Hypothesis 1b

$$H_0: \text{variance}(\ln(RR))_{LE} = \text{variance}(\ln(RR))_{no LE} \\ H_A: \text{variance}(\ln(RR))_{LE} < \text{variance}(\ln(RR))_{no LE}$$

Any difference in variance could potentially be driven by differing anchoring and adjustment behaviour in the two treatments, hence changing of the proportion of outliers. Recall that this type of behaviour can manifest itself in one of the two different ways, either separately or in combination within any one treatment, potentially impacting both ends of the distribution; (1) non-traders(I); individuals who are indifferent between two equal sized risk increases for very different outcomes and unwilling to move away from the initial situation at all (implying a RR of 1) and (2) non-traders(O); people who are very reluctant to switch in the table and hence iterate towards indicating that they would prefer to end up with a less severe injury for certain instead of accepting any risk increase at all in their least preferred outcome (a more severe injury).

To test for the effect of the learning experiment on responses, a probit regression will be carried out to analyse whether the proportion of respondents who display either non-moving or non-trading behaviour, i.e., anchoring is significantly affected by the learning experiment. Following this, non-moving and non-trading behaviour will be analysed separately in two different probit regressions.

Table 4 Geometric means across all treatments

	<i>N</i>	Geometric mean of RR ^a	95% confidence interval	Death risk equivalent (rr)
NS/F	166	1093	(744; 1604)	9.15e–04
NS/S	168	328	(214; 503)	0.003
S/F	163	138	(92; 208)	0.007

^aRR; inverse of death risk equivalent (rr)**Table 5** Test of the effect of the learning experiment (LE)

	<i>N</i>	ln(RR) (std. error)		Mean(ln(RR)) _{LE} = mean (ln(RR)) _{no LE}	Var(ln(RR)) _{LE} < var(ln(RR)) _{no LE}
		LE (T3, T4)	No LE (T1, T2)	<i>t</i> test	Variance-comparison test
NS/F	166 (68 + 98)	7.45 (0.22)	6.68 (0.29)	$p = 0.05^{**}$	$p < 0.01^{***}$
NS/S	168 (71 + 97)	6.59 (0.21)	5.21 (0.33)	$p < 0.01^{***}$	$p < 0.01^{***}$
S/F	163 (68 + 95)	5.20 (0.28)	4.73 (0.30)	$p = 0.27$	$p = 0.022^{**}$

, *Significant at 0.05 and 0.01 levels, respectively

5 Results

Each respondent answered three questions. As expected, in each question, a majority (in the range of 77–83%) preferred to move to the area, where the risk increase of 5 in 100,000 occurred in the best outcome (i.e., in the non-serious injury/fatal injury comparison (NS/F), they preferred the risk increase in the non-serious injury NS. The rest of the analysis will be carried out on this sample of consistent respondents, including those respondents with $RR = 1$ (8–15%). One respondent had problems understanding the tasks (due to very modest English language skills) and has been excluded from the analyses. Table 8 in the [Appendix](#) summarizes the responses to the three different questions.

Table 4 reports the geometric mean of the RRs and 95% confidence intervals pooled across the four treatments and the ‘death risk equivalent’.

The geometric mean of RR_{NSF} is higher than RR_{NSS} followed by RR_{SF} with no overlapping confidence intervals, i.e., on average, the respondents perceived the greatest difference to be between non-serious and fatal injuries (NS/F), followed by non-serious and serious injuries (NS/S) and serious and fatal injuries (S/F). This ranking is in accordance with the results found in Jones-Lee et al. (1995). In addition, by indicating that a ranking of $NS/F > S/F$, i.e., that the difference in utility loss between a non-serious injury and a fatal injury is greater than that between a serious and a fatal injury, subjects on average indicate that a non-serious injury NS is preferred to a serious injury S. Likewise, by indicating $NS/F > NS/S$, subjects

Table 6 Probit regressions. Analysing the impact on anchoring behaviour and number of non-traders across treatments

Dependent variable	ANCHORING (dummy variable) Marg. effects (robust std. error)	NON-TRADER(I) (dummy variable) Marg. effects (robust std. error)	NON-TRADER(O) (dummy variable) Marg. effects (robust std. error)
LE	−0.15 (0.049)***	−0.14 (0.037)***	−4.86e−03 (0.036)
TRF	−0.10 (0.051)**	0.017 (0.038)	−0.12 (0.038)***
NS/S	−5.1e−03 (0.039)	2.8e−03 (0.028)	−1.0e−02 (0.028)
NS/F	0.038 (0.034)	−0.063 (0.021)***	0.098 (0.029)***
Constant	−0.382 (0.148)**	−0.84 (0.17)***	−1.17 (0.20)***
<i>N</i>	497	497	497
Clusters	176	176	176

,*Significant at 0.05 and 0.01 levels, respectively

on average indicate that the difference in utility loss between a non-serious injury NS and a fatal injury F is greater than that between non-serious NS and a serious S injury. Hence, on average, subjects pass the consistency test across trade-offs implied by a ranking of NS to be the best outcome followed by S and finally F.

5.1 Test of research Hypotheses 1a and 1b

Table 5 reports a test of the effect of the learning experiment. Hence, we compare the mean and the variance of the responses in treatments 3 and 4 to the responses in treatments 1 and 2.

Across all three choices, the mean RRs are higher in the treatments with the learning experiments (however, not significant in RR_{SF}) and the variance is significantly lower in these treatments. Hence, on average, the effect of the learning experiment is that it increases mean RR and lowers the variance and Hypotheses 1a and 1b are rejected.¹²

Table 6 reports results of probit regressions with three different dependent variables: (1) NON-TRADER(I), a dummy variable for whether the subject was willing to move away from the initial choice ($RR = 1$); (2) NON-TRADER(O), a dummy

¹² Additionally, linear regressions have been carried out with $\ln(RR)$ as the dependent variable for both the total sample and the sample without subjects displaying anchoring behaviour (not reported). In all regressions both RR_{NS} and RR_{NSF} ratios are found to be significantly higher than the SF. This demonstrates sensitivity to the magnitude of the difference in outcome (i.e. a form of scope sensitivity test). All regressions have also been carried out with the addition of a dummy variable for order. No significant effect from the question order has been found. An interaction term between LE and framing has also been included, but not found to be significant either. Additionally, a dummy variable for stage 1 students, a dummy variable for students on the economics programme and a dummy variable for gender have been included and found insignificant.

variable for whether or not the subject was almost unwilling to take any risk increase in one of the outcomes (subjects who indicated indifference point outside the table range); and (3) ANCHORING, a dummy variable reflecting a type of anchoring behaviour as a combination of (1) and (2).

Here, we further analyse the effect of the learning experiment on anchoring behaviour while at the same time controlling for framing and individual interdependence between question answers. All regressions include dummy variables for the two pairwise comparisons [non-serious/serious injuries (NS/S) and non-serious/fatal injuries (NS/F)], where the serious/fatal injuries' (S/F) dummy variable has been omitted for comparative purposes. In addition, dummy variables for the learning experiment ($LE = 1$ if treatment 3 or 4, else 0) and for the framing effect ($TRF = 1$ if treatment 2 or 4, else 0) has been included. All variables are described in Table 9 in the [Appendix](#).

If we first consider the 'Anchoring' regression, both the LE and TRF variables are negative and significant, implying that both the learning experiment and framing reduce anchoring behaviour. The subsequent regressions, which differentiate between 'non-traders(I)' and 'non-traders(O)', show clearly that the learning experiment has an effect on anchoring by reducing the number of non-traders(I), although it does not significantly influence the number of non-traders(O). This result explains why the mean RRs are higher in the treatment with the LE, as there are significantly fewer non-traders(I) in treatments with the LE and hence fewer respondents who are unwilling to move below the first row in the table (see the [Appendix](#)). On the other hand, framing it in the total risk frame decreases the number of non-traders(O) significantly¹³, whereas there is no significant framing effect on the proportion of non-traders(I).

The results also show that RR_{NSF} has a significantly lower number of non-traders(I) and a significantly higher number of non-traders(O) compared to RR_{SF} . This is in line with expectations, since the non-serious injury NS and fatal injury F outcomes are perceived by respondents to be the most different (see Table 4). However, these results cancel out in the 'Anchoring' regression when the dependent variable is the combined effect of 'non-traders(I+O)'.

To summarise, we rejected Hypothesis 1a of equal mean across the treatments with and without the learning experiment. In addition, we find that the learning experiment lowers the variance of the responses significantly. To further explore this change in variance, we find that it impacted anchoring behaviour by reducing the number of non-traders(I) significantly. In addition, we find that applying the total risk frame impacts the number of non-traders(O) significantly. Non-traders(O) and non-traders(I) can be considered as outliers in data, since they would fall in the extreme tails of the response distribution. In our study, we have thereby shown

¹³ As a sensitivity analysis, an alternative definition of non-traders(O) has been tried as well ($\ln(RR) > 9.2$) with similar results.

that the learning experiment and the application of the total risk frame significantly reduce the number of these outliers.

6 Conclusions and discussion

In this study, we subject the standard ‘risk–risk trade-off’ procedure developed by Viscusi et al. (1991) to a methodologically oriented assessment of traffic accidents of varying severities. We explore the reliability and validity of responses by focusing on the behaviour of the mean relative risk trade-off and its associated variance with and without a learning experiment. By comparing results across experimental treatments, we find that the learning experiment has a significant impact on mean RR and lowers the variance. We further examine the change in mean and variance by analysing the impact of the learning experiment on anchoring behaviour bias and we show that it affects this behaviour by significantly increasing the proportion of respondents who were willing to move away from their initial choice [non-traders(I)]. We have thereby shown that the experience gained trading risks in an incentivised task can spill over to a hypothetical setting, and hence, we are able to present aggregated stated preference survey results which are less influenced by ‘anchoring behaviour biases’ and have a smaller variance and fewer outliers for this particular sample. If this result was replicated in a sample from the general population, then this would increase the validity of the results by decreasing the proportion of outliers in data, potentially reducing the need to discard responses in any regulatory assessments of road-safety projects. A priori, we have no evidence to suggest that the wider population is any less susceptible to anchoring and adjustment behaviour, so it is likely that such a procedure would be beneficial. As such, from a policy perspective, our method could be used if, for example, the DfT wished to update its advice (Treasury 2003) regarding the valuations of different severities of non-fatal road injuries. As the VSL and associated values for these statistical injuries are based on the arithmetic mean—as opposed to the geometric mean used in this paper for technical reasons (see Sect. 4)—it could be argued on grounds of consistency that the arithmetic mean should be used for any new policy values estimated using our method, not least since the impact on the arithmetic mean of these non-traders would also be reduced.

Deploying a learning experiment that improves understanding prior to answering risk–risk trade-offs means that researchers have a choice in that they can decide between creating an incentivised survey environment (as here) in which the understanding gained can spill over to the survey choices (i.e., in the design phase *ex ante*) or to reformulate the model to include ‘behavioural extras’ (e.g., anchoring) in the value function (i.e., the analytical phase *ex post*). While other methods such as cheap talk (Cummings and Taylor (1999) or an oath script (Jacquemet et al. 2013) have also been shown to be helpful in generating more valid stated preference responses,

we feel that a learning experiment of the type deployed here is likely to be more suited to a ‘risk–risk trade-off’ application and the particular problems it poses.

It has been shown that by including such an experiment *ex ante* in the survey, the need to reformulate the model to accommodate such behaviour in any *ex post* analysis is not so necessary. An additional result is that using a frame that focuses on total risk can likewise to some extent ameliorate anchoring; more specifically, using the total risk frame can significantly decrease the proportion of people who affix on the last preferred outcome and hence iterate towards indicating that they would prefer to end up with the non-serious road accident injury for certain instead of accepting any risk increase at all in their least preferred outcome, e.g., a serious injury [non-traders(O)]. It is an interesting question, and one for future research, to further investigate why the learning experiment reduces the number of non-traders(I), although it does not significantly influence the number of non-traders(O), whereas applying risk framing affects decreases the number of non-traders(O) significantly but not non-traders(I). While our results are confined to risk increases, for the reasons outlined earlier, there is no obvious reason to suspect that risk decreases would not be susceptible to similar problems and amenable to similar solutions.

Learning experiments are perhaps best thought of as extensions to the information set in a mortality or morbidity risk valuation survey. For the purpose of this investigation, we have chosen to focus on a setup which (1) involves no monetary comparison and (2) elicits an indifference point between two risk reductions (for different outcomes). However, it is also possible that learning experiments could have a similar impact on the elicitation of WTP for risk changes in iterative choice approaches involving money (see, e.g., Viscusi et al. 2014) or in CV or DCE surveys. In Cherry and Shogren (2007), a significant impact from such a device was shown in that it significantly lowered the mean and median WTP for safer food, but no separate analyses were carried out on its’ impact on any other behaviours, e.g., on the proportion of zero WTP offers. In summary, our findings suggest that utilising a learning experiment is practicable and diminishes the need to discard responses in stated preference surveys and thereby improves the representativeness of the samples used to elicit preferences for policy purposes. It also adds to the valuation ‘toolkit’ available to analysts in regulatory analysis.

Acknowledgements This work was carried out as part of the project IMPROSA (Improving Road Safety) funded by the Danish Research Council (Strategiske Forskningsråd). However, the opinions expressed in the article are solely the responsibility of the authors and do not necessarily reflect the views of the research project’s sponsors. The authors are very grateful to Andrew Burlinson, Justas Dainauskas, and Sam Simpkin for their research assistance in the design phase of this experiment.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

Non-fatal injury descriptions

TYPE NS*

If an accident occurs and you suffer a type W injury you will be:

- In Hospital
 - 2-7 days
 - slight to moderate pain
- After Hospital
 - some pain/discomfort for several weeks
 - some restrictions to work and/or leisure activities for several weeks/months
 - after 3-4 months, return to normal health with no permanent disability

TYPE S*

If an accident occurs and you suffer a type R injury you will be:

- In Hospital
 - several weeks, possibly several months
 - moderate to severe pain
- After Hospital
 - continuing pain/discomfort for the rest of your life, possibly requiring frequent medication
 - substantial and permanent restrictions to your work and leisure activities – possibly some prominent scarring

*Corresponds to injury W in Jones-Lee et al. (1995).

**Corresponds to injury R in Jones-Lee et al. (1995).

Example of table used for eliciting indifference points

I prefer area 2 (2 versus 3)

Row		Risk of non-fatal S 600 in 100,000 (per decade) Risk of fatal F 50 in 100,000 (per decade)			
		Area 2 RISK INCREASES (per decade)		Area 3 RISK INCREASES (per decade)	My choice (area 2 or 3)
0	S	5 in 100,000	OR	0 in 100,000	Area 2
	F	0 in 100,000		5 in 100,000	
1	S	41 in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	
2	S	121 in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	
3	S	258 in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	
4	S	483 in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	
5	S	850 in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	
6	R	1457 in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	
7	S	2493 in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	
8	S	4329 in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	
9	S	7728 in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	
10	S	14,313 in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	
11	S	_____ in 100,000	OR	0 in 100,000	
	F	0 in 100,000		5 in 100,000	

See Tables 7, 8, and 9.

Table 7 Experimental games 1–3

Game 1 (practice)	Round I	Choice between two wheels (A and B) The wheels consist of different proportions of blue, green and yellow segments The chosen wheel is spun and the result noted
	Round II	Individual indifference points between the two wheels were elicited by making the subject's preferred wheel worse (removing blue segments and replacing them with green segments) until indifference is found Six comparative choices were presented to the subjects. Each page in the booklet contained the unchanged not-preferred wheel and a changed version of their preferred wheel The roll of a die decides which wheels [page number (1–6)] to play out for subjects The wheels on the page number decided by the die were spun. The outcome of their chosen wheel was noted down by subjects
Game 2 (real)	Round III	Choice between wheels C and D
	Round IV	The wheels consist of different proportions of blue, green and yellow segments compared to wheel A and B Procedure like game 1
Game 3 (real)	Round V	Choice between wheels E and F
	Round VI	The wheels are made up of different proportions of blue, green and pink segments Procedure like game 1

Table 8 Descriptive statistics

	NSS	NSF	SF
Non-traders(I)	26 (14%)	14 (8%)	24 (13%)
Prefer risk increase in best state	142 (79%)	152 (84%)	139 (77%)
Prefer risk increase in worst state	10 (6%)	9 (5%)	13 (7%)
Missing	0	4 ((2%)	2 (1%)
Language problems	1 (0.6%)	1 (0.6%)	1 (0.6%)
Total	180	180	180
Of which indifference point outside table range	14 (2%)	31 (17%)	15 (8%)

Table 9 Variable description

Variable name	Variable description	Value
LE	(1if T3 or T4, else 0)	0.42
TRF	(1 if T2 or T4, else 0)	0.50
NON-TRADER(I)	(1 if $\ln(\text{ratio})=0$, else 0)	0.12
NON-TRADER-(O)	(1 if indifference point outside the table range, else 0)	0.10
ANCHORING	(1 if NON-TRADER(I)=1, 1 if NON-TRADER(O)=1, else 0)	0.23
NSS	(1 if NSS=1, else 0)	0.34
NSF	(1 if NSF=1, else 0)	0.33

References

- Alberini A, Ščasný M (2011) Context and the VSL: evidence from a stated preference study in Italy and the Czech Republic. *Environ Resour Econ* 49(4):511–538
- Andersen S, Harrison GW, Lau MI, Rustrom E (2007) Valuation using multiple price list formats. *Appl Econ* 39(6):675–682
- Andersson H, Lindberg G (2009) Benevolence and the value of road safety. *Accid Anal Prev* 41(2):286–293
- Baron J (1997) Confusion of relative and absolute risk in valuation. *J Risk Uncertain* 14:301–309
- Bateman JJ, Carson RT, Day B, Hanemann M, Hanley N, Hett T, Jones-Lee MW, Loomes G, Mourato S, Özdemiroglu E, Pearce DW, Sugden R, Swanson J (2002) Economic valuation with stated preference techniques. A Manual. Edward Elgar, Cheltenham
- Beattie J, Covey J, Dolan P, Hopkins L, Jones-Lee MW, Loomes G, Pidgeon N, Robinson A, Spencer A (1998) On the contingent valuation of safety and the safety of contingent valuation: part 1—Caveat Investigator. *J Risk Uncertain* 17:5–25
- Beattie J, Carthy T, Chilton S, Covey J, Dolan P, Hopkins L, Jones-Lee M, Loomes G, Pidgeon N, Robinson A, Spencer A (2000) The valuation of benefits of health and safety control: final report. HSE Contract Research Report CRR 273/2000, London
- Cameron M, Gibson J, Helmers K, Lim S, Tressler J, Vaddanak K (2010) The value of statistical life and cost–benefit evaluation of landmine clearance in Cambodia. *Environ Dev Econ* 15(4):395–416
- Carlsson F, Frykblom P, Lagerkvist CJ (2005) Using cheap talk as a test of validity in choice experiments. *Econ Lett* 89(2):147–152

- Carlsson F, Mørkbak MR, Olsen SB (2012) The first time is the hardest: a test of ordering effects in choice experiments. *J Choice Model* 5(2):19–37
- Carson RT, Mitchell RC (1995) Sequencing and nesting in contingent valuation surveys. *J Environ Econ Manag* 28(2):155–173
- Carson RT, Flores NE, Meade NF (2001) Contingent valuation: controversies and evidence. *Environ Resour Econ* 19:173–210
- Cherry TL, Shogren JF (2007) Rationality crossovers. *J Econ Psychol* 28:261–277
- Cherry TL, Crocker TD, Shogren JF (2003) Rationality spillovers. *J Environ Econ Manag* 45:63–84
- Chilton S, Covey J, Hopkins L, Jones-Lee M, Loomes G, Pidgeon N, Spencer A (2002) Public perceptions of risk and preference-based values of safety. *J Risk Uncertain* 25(3):211–232
- Chilton S, Jones-Lee MW, Kiraly F, Metcalf H, Pang W (2006) Dread risks. *J Risk Uncertain* 33:165–182
- Chilton S, Jones-Lee MW, Metcalf H, Loomes G, Robinson A, Covey J, Spencer A, Spackman M (2007) Valuation of health and safety benefits: dread risks. HSE Books, Norwich
- Clarke AE, Goldstein MK, Michelson D, Garber AM, Lenert LA (1997) The effect of assessment method and respondent population on utilities elicited for Gaucher disease. *Qual Life Res* 6:169–184
- Cummings RG, Taylor LO (1999) Unbiased Value estimates for environmental goods: a cheap talk design for the contingent valuation method. *Am Econ Rev* 89(3):649–665
- Cummings RG, Brookshire DS, Schulze WD (1986) Valuing environmental goods: a state of the arts assessment of the contingent valuation method. Rowman and Allanheld, Totowa
- Dhami S (2016) The foundations of behavioural economic analysis. Oxford University Press, New York
- Dolan P, Jones-Lee M, Loomes G (1995) Risk-risk versus standard gamble procedures for measuring health state utilities. *Appl Econ* 27:1103–1111
- Fetherstonhaugh D, Slovic P, Johnson S, Friedrich J (1997) Insensitivity to the value of human life: a study of psychophysical numbing. *J Risk Uncertain* 14(3):283–300
- 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. *J Risk Uncertain* 34(2):123–144
- Gyrd-Hansen D, Kristiansen IS, Nexoe J, Nielsen JB (2003) How do individuals apply risk information when choosing among health care interventions? *Risk Anal* 23(4):697–704
- Gyrd-Hansen D, Kjær T, Nielsen JS (2012) Scope insensitivity in contingent valuation studies of health care services: should we ask twice? *Health Econ* 21(2):101–112
- Iragüen P, de Ortúzar JD (2004) Willingness-to-pay for reducing fatal accident risk in urban areas: an internet-based web page stated preference survey. *Accid Anal Prev* 36(4):513–524
- Jacquemet N, Joule RV, Luchini S, Shogren JF (2013) Preference elicitation under oath. *J Environ Econ Manag* 65:110–132
- Johannesson M, Johansson PO, O'Connor RM (1996) The value of private safety versus the value of public safety. *J Risk Uncertain* 13(3):263–275
- Jones-Lee M (1976) The Value of life. An economic analysis. The University of Chicago Press, Chicago
- Jones-Lee M, Loomes G, Phillips PR (1995) Valuing the prevention of non-fatal road injuries: contingent valuation vs. standard gambles. *Oxf Econ Pap* 47(4):676–695
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–291
- Kjaer T, Nielsen JS, Hole AR (2018) An investigation into procedural (in)variance in the valuation of mortality risk reductions. *J Environ Econ Manag* 89:278–284
- Law AV, Pathak DS, McCord MR (1998) Health status utility assessment by standard gamble: a comparison of the probability equivalence and the lottery equivalence approaches. *Pharm Res* 15(1):105–109
- LeBoeuf RA, Shafir E (2006) The long and the short of it: physical anchoring effects. *J Behav Decis Mak* 19:393–406
- Magat WA, Viscusi WK, Huber J (1996) A reference lottery metric for valuing health. *Manag Sci* 42(8):1118–1130
- McDonald RL, Chilton SM, Jones-Lee M, Metcalf H (2016) Dread and latency impacts on a VSL for cancer risk reduction. *J Risk Uncertain* 52(2):137–161
- Mitchell RC, Carson RT (1989) Using surveys to value public goods: the contingent valuation method. Resources for the Future, Washington, DC
- Nielsen JS, Chilton S, Jones-Lee M, Metcalf H (2010) How would you like your gain in life expectancy to be provided? An experimental approach. *J Risk Uncertain* 41:195–218

- Olsen JA, Donaldson C, Pereira J (2004) The insensitivity of ‘willingness-to-pay’ to the size of the good: new evidence for health care. *J Econ Psychol* 25:445–460
- Plott, C.R (1996). Rational individual behaviour in markets and social choice processes: the discovered preference hypothesis. In: Arrow KJ, et al (eds) *The rational foundations of economic behaviour: Proceedings of the IEA conference held in Turin, Italy, IEA conference*, vol 114. St. Martin’s Press, New York, Macmillan Press in association with the International Economic Association, London, pp 225–250
- Quattrone GA (1982) Overattribution and unit formation: when behavior engulfs the person. *J Pers Soc Psychol* 42:593–607
- Rowe RD, Schulze WD, Brefle WS (1996) A test for payment card biases. *J Environ Econ Manag* 31:178–185
- Sydsaeter K, Hammond PJ (1995) *Mathematics for economics analyses*. Prentice-Hall, New Jersey
- Treasury HM (2003) *The green book: appraisal and evaluation in central government*. The Stationery Office, London
- Tversky A, Kahneman D (1974) Judgement under uncertainty: heuristic and biases. *Science* 185(4157):1124–1131
- Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. *Science* 211:453–458
- Tversky A, Thaler RH (1990) Anomalies: preference reversals. *J Econ Perspect* 4(2):201–211
- US EPA (2014) *Guidelines for preparing economic analyses*. National Center for Environmental Economics Office of Policy, U.S. Environmental Protection Agency
- Van Houtven G, Sullivan MB, Dockins C (2008) Cancer premiums and latency effects: a risk tradeoff approach for valuing reductions in fatal cancer risks. *J Risk Uncertain* 36:179–199
- Viscusi WK, Magat WA, Huber J (1991) Pricing environmental health risk: survey assessments of risk–risk and risk-dollar trade-offs for chronic bronchitis. *J Environ Econ Manag* 21:35–51
- Viscusi WK, Huber H, Bell J (2014) Assessing whether there is a cancer premium for the value of a statistical life. *Health Econ* 23(4):384–396