



# The value of statistical life in the context of road safety: new evidence on the contingent valuation/standard gamble chained approach

Fernando-Ignacio Sánchez-Martínez<sup>1</sup> · Jorge-Eduardo Martínez-Pérez<sup>1</sup> · José-María Abellán-Perpiñán<sup>1</sup> · José-Luis Pinto-Prades<sup>2</sup>

Accepted: 18 August 2021  
© The Author(s) 2021

## Abstract

This study estimates the value of statistical life (VSL) on a road traffic accident using the Contingent Valuation/Standard Gamble chained approach. A large representative sample ( $n = 2020$ ) is used to calculate a VSL for use in the evaluation of road safety programmes in Spain. The paper also makes some methodological contributions, by providing new evidence about the consistency of the chained method. Our main results are: (1) A range from 1.3 million euro to 1.7 million euro is obtained for the VSL in Spain in the context of road accidents. This range is in line with the values used in the same context in other European countries, although it is lower than those obtained in different contexts and with other methods. (2) The method performs much better in terms of scope sensitivity than the traditional contingent valuation method, which asks subjects about their willingness to pay for very small reductions in the risk of death. (3) We introduce a new ‘indirect’ chaining approach which reduces (but does not remove) the disparity between direct and indirect chaining approaches. More extreme VSL estimates are still obtained with this indirect method than with the direct one. (4) VSL estimates depend on the injury used. More specifically, we obtained a lower VSL when a more severe injury is used. (5) Framing the risk of death in the modified standard gamble question as “*10n* in 10,000” instead of “*n* in 1000” influences the value of VSL. We attribute this effect to the Ratio Bias.

**Keywords** Value of statistical life · Contingent valuation · Standard gamble · Ratio bias effect

**JEL classification** J17 · D61 · D81 · R41

---

✉ Fernando-Ignacio Sánchez-Martínez  
fernando@um.es

<sup>1</sup> Applied Economics Department. Faculty of Economics and Business, University of Murcia, Murcia 30100, Spain

<sup>2</sup> Department of Economics. School of Economics and Business, University of Navarra, Pamplona 31009, Spain

## 1 Introduction

The value of a statistical life (VSL) is the monetary valuation of a reduction in the risk of death that would prevent one ‘statistical’ death. Two approaches are available to estimate the VSL. On the one hand, revealed preferences methods seek to identify the trade-off between wealth and mortality risk from the decisions that people make in actual markets. Within this approach, most research has focused on the labour market (Viscusi and Aldy 2003; Viscusi 2018b). On the other hand, the approach based on stated preferences aims to elicit the trade-offs between money and the risk of death from individuals’ responses to surveys where hypothetical markets are recreated. This second approach has been used mostly in the context of health risks (about 50% of all the studies that followed these methodologies), as well as in road safety (30%) and environmental risks (20%) contexts (Viscusi and Masterman 2017). In this paper, we focus on the estimation of the VSL to be used in the context of the design and evaluation of road safety policies. We use the CV/SG chained method proposed by Carthy et al. (1999), which combines both health risks and road safety contexts. In addition, the weaknesses and strengths of this method will be examined by a series of consistency tests.

Contingent valuation (CV) has been the traditional approach to estimate the VSL. Using surveys, CV elicits subjects’ willingness to pay (WTP) to reduce the risk of death or their willingness to accept compensation (WTA) for an increase in the risk of death in a road accident. According to the theory (Jones-Lee 1974), the Marginal Rate of Substitution (MRS) between wealth and the risk of death ( $m_d$ ) can be estimated from WTP and WTA valuations. One of the main problems of this method is that it asks subjects to value very small risk reductions since this is the real risk of death that people face of dying in a car accident. Nevertheless, there is plenty of evidence that CV methods entail problems of scope insensitivity (Beattie et al. 1998; Hammitt and Graham 1999). Specifically, while the theory (Jones-Lee 1974) predicts that individuals’ WTPs should be proportional to the size of risk reduction (if risk is marginal or very small), the usual finding is that WTP is much less than proportional to small reductions in the risk of death (Andersson et al. 2016; Sjøgaard et al. 2012). Thus, VSL estimates change with the size of risk reduction (Dubourg et al. 1997; Jones-Lee et al. 1995).

To overcome those problems Carthy et al. (1999) developed a new method that they called the contingent valuation/standard gamble (CV/SG) chained approach. It is “chained” because the VSL is estimated by combining (i.e., chaining) the responses to two types of questions. Very briefly (the method is explained more in depth in Sect. 2), this approach requires the interviewee to respond to the following questions: (1) WTP for the certainty of a quick and complete cure for a given non-fatal road injury  $i$ ; (2) WTA for the certainty of sustaining the same injury; and (3) a “modified” Standard Gamble (MSG) question where two lotteries are compared: one with Normal Health and Death as outcomes and the other one with injury  $i$  and Death as outcomes. As a consistency check, they incorporated another set of WTP/WTA and MSG questions with another injury  $j$  less

severe than  $i$ . The MSG involved a choice between one lottery with injury  $i$  and Normal Health as potential outcomes, and a second lottery with injury  $i$  and injury  $j$  as outcomes.

From questions 1 and 2, the authors estimated the MRS between wealth and the risk of injury  $i$  ( $m_i$ ). The important point here is that those questions did not involve the use of small probabilities, as was the case of the traditional CV approach. From question 3, they estimated the so-called “relative utility loss” (RUL) which was the ratio  $m_d/m_i$ . So VSL can be obtained as the product of  $m_i$ , calculated from WTP and WTA questions, and the RUL ( $m_d = m_i \times m_d/m_i$ ). This is the “direct chaining”. From questions involving the milder injury  $j$ , they estimated  $m_j$  and the RUL between  $i$  and  $j$  ( $m_i/m_j$ ). So  $m_d$  could be estimated as  $m_j \times m_i/m_j \times m_d/m_i$ . This is the “indirect chaining”, and consistency would require that both methods lead to the same  $m_d$ .

Their results were encouraging in various aspects. First, they did not have problems of subjects not willing to pay anything or subjects willing to pay the same for the reduction in the risk of injuries of different severity, that is, WTP and WTA responses did not suffer from insensitivity to scope. Second, responses to MSG also showed face validity since subjects were willing to accept higher risks to recover from more severe injuries. For these reasons, some researchers consider this approach to be the best method available to estimate VSL (Nellthorp et al. 2001; Spackman et al. 2011) and it has been used to estimate the official VSL for the UK (DETR 1998; Jones-Lee and Spackman 2013).

However, the method is not without its problems. The main limitation of the CV/SG approach is that the direct and indirect chaining procedures result in different VSL estimates. The indirect method gives extremely large and implausible values for a small (around 10%) group of subjects, producing means that were around 15–30 times larger than those obtained with the direct method. Even excluding those subjects, VSL estimates were clearly higher using indirect chaining. Based on this, Thomas and Vaughan (2015) questioned the internal consistency of the CV/SG, raising a debate on its validity (Chilton et al. 2015; Jones-Lee and Loomes 2015; Olofsson et al. 2019; Balmford et al. 2019).

In this paper we apply the CV/SG method in a large survey ( $n=2020$ ), conducted with a random stratified sample representative of the Spanish adult population in 2009, to estimate the official VSL in Spain. The study was commissioned by the Spanish Department of Transport (Dirección General de Tráfico). Besides obtaining a VSL based on the CV/SG method, a basic objective of this paper is to provide more evidence about the consistency of the method. Several consistency tests were conducted by splitting our sample into several groups. In two groups ( $n_3=243$ ,  $n_7=233$ ) we followed, as closely as possible, the methods used by Carthy et al. (1999) ( $n=167$ ). We tried to replicate their results, including their problems of internal consistency. In the other groups: (a) we introduced some changes in an attempt to overcome some of their problems by applying an alternative indirect chaining procedure; and (b) we included new consistency checks, for which we use two different health states (i.e. non-fatal injuries) to elicit WTP and WTA estimates, and we also use two different ways of framing risks ( $n$  in 1000 vs.  $10n$  in 10,000).

A range from €1.3 million to €1.7 million for the VSL in the context of road safety in Spain is obtained based on the ‘direct’ CV/SG approach. Regarding consistency checks, we conclude that the disparity between the direct and indirect methods can be reduced by using an alternative indirect chaining procedure, although differences between results from the two approaches persist, thus questioning the consistency of the CV/SG method. We also find that VSL estimates depend on the non-fatal injury used in the direct method, in the sense that more severe health states yield lower VSL estimates. Likewise, we observe that risk framing affects VSL estimates, which could be explained, at least in part, by the so-called ‘ratio-bias’ effect.

In the following section we explain the CV/SG method and describe what we consider to be its main problems according to the literature. In Sect. 3, we detail how we tried to avoid these problems, and the new consistency checks. The empirical study is described in Sect. 4. The main results are reported in Sect. 5. The Discussion closes the paper.

## 2 The contingent valuation/standard gamble (CV/SG) chained approach and its limitations

As has been said, the CV/SG chained method splits the valuation of risk reductions into three steps:

1. The MRS between wealth and the risk of a non-fatal injury  $i$  (denoted by  $m_i$ ) is inferred from WTP and WTA estimates by assuming a specific functional form of the utility function.
2. A modified standard gamble (MSG) question is used to estimate what Carthy et al. (1999) call the ‘Relative Utility Loss’ (RUL) or  $m_d/m_i$ , where  $m_d$  is the MRS between wealth and the risk of death. This MSG is a lottery equivalent (McCord & de Neufville 1986) question, which is called ‘modified’ because, contrary to the traditional standard gamble (SG) method, it does not involve comparing between one lottery and one certain option, but between two non-degenerated lotteries. The SG has been shown to be influenced by the ‘certainty effect’ (Kahneman et al. 1979). For example, Jones-Lee et al. (1995) observed this effect, since a large percentage of subjects (more than 50% in some injuries) were not willing to accept any risk of death using the traditional SG. The MSG is intended to avoid this effect (Abellán Perpiñán et al. 2012). More specifically, Carthy et al. (1999) asked subjects to establish indifference between two binary lotteries X and Y. Assume that  $X = (q, \text{Death}; \text{Injury } i)$  and  $Y = (p, \text{Death}; \text{Normal health})$ , where  $q$  and  $p$  are the probabilities of death in X and Y respectively. Assume that  $q$  is fixed at a predetermined value (1 in 1000 in Carthy et al.). Subjects have to state the value of  $p$  that leaves them indifferent between X and Y. The important contribution of Carthy et al. (1999) was to show that if subjects are Expected Utility maximizers,

$$\frac{m_d}{m_i} = \frac{1 - q}{p - q} \tag{1}$$

3. Finally, since we know  $m_i$  from WTP/WTA questions and  $\frac{1-q}{p-q}$  from MSG, the theoretical result in (1) makes it possible to infer  $m_d$ :

$$m_d = \frac{1 - q_i}{p_i - q_i} \times m_i \tag{2}$$

Note that the relevant advantage of this method is that we can estimate  $m_d$  without asking subjects WTP questions about very small risk reductions (e.g., 0.00001). This is, we believe, a huge advantage of the CV/SG method.

The main internal problem that was observed in the study by Carthy et al. (1999) was the large discrepancy between VSL values obtained through what they called the ‘direct’ and the ‘indirect’ chaining procedures. Direct chaining is the procedure explained above. The indirect chaining method adds another step, namely, it estimates two RULs, one between two non-fatal road injuries ( $i$  and  $j$ ) and one between  $i$  and death. The authors also estimate two MRS between wealth and the risk of the two ( $i$  and  $j$ ) non-fatal injuries. They implemented the indirect method as follows:

1. They chose a health state ( $j$ ) that was milder than the health state they used for the direct format and estimated  $m_j$  by combining WTP and WTA responses. As  $j$  is milder than  $i$ , we would expect  $m_j < m_i$ . This is a first consistency check.
2. They asked a second MSG question, where injury  $i$  becomes the worst outcome in both lotteries:  $(\theta_j, \text{Injury } i; \text{Injury } j)$  vs  $(\Pi_j, \text{Injury } i; \text{Normal health})$ . From this, we have that

$$m_i = \frac{1 - \theta_j}{\Pi_j - \theta_j} \times m_j \tag{3}$$

3. Consistency requires that

$$m_d^{ind} = \frac{1 - q_i}{p_i - q_i} \times \frac{1 - \theta_j}{\Pi_j - \theta_j} \times m_j \tag{4}$$

where  $m_d^{ind}$  is the MRS between wealth and the risk of death calculated using the indirect chaining approach. The theory predicts that  $m_d = m_d^{ind}$ , but they found that  $m_d^{ind} > m_d$  by a large margin. They explained this difference in terms of “compounding of errors” in the indirect chaining.

The second problem that Carthy et al. (1999) observed was that about 10% of subjects did not want to accept any risk to cure injury  $i$ . It seems that MSG could not totally avoid the certainty effect, although it was significantly reduced in relation to Jones-Lee et al. (1995) where they used SG. These were excluded from the analysis along with two outliers who gave extremely large values for VSL. In the

case of indirect chaining, they had to exclude about 20% of subjects. These new exclusions were mainly justified in terms of the implausibly large values of VSL that the indirect method yielded. This is worrying since the survey was administered by members of the research team, who were able to explain it carefully and to answer queries.

### 3 Objectives and hypotheses

We explain here the objectives of our study and the hypotheses to be tested.

The first objective is simply to replicate the original study, which will allow us to derive a VSL for Spain that may be recommended for use in the evaluation of road safety policies. This may have some interest in itself, since one of the main problems in studies using experimental and survey data is that sometimes results are not replicated (Anderson and Kichkha 2017; Berry et al. 2017; Maniadis et al. 2017; Mueller-Langer et al. 2019). In our case, the surveys were conducted by professional interviewers and not by members of the research team. Interviews were face-to-face using computers and physical visual aids. We hypothesized that we would reproduce the main results of (Carthy et al. 1999):

- MSG and  $m_i$  would show scope effects.
- The indirect chaining method would produce larger VSL than the direct method.

These hypotheses will be tested between-subjects and not within-subjects because our survey did not include only the CV/SG questions. We also included questions on the standard method and asked subjects their WTP to reduce marginally the risk of death in a road traffic accident. Due to the larger number of questions, we only asked each subject one WTP and one WTA question.

The second objective was to try to overcome one of the limitations observed in Carthy et al. (1999), namely, the disparity between direct and indirect methods. We explain how we tried to do this. Carthy et al. (1999) found that  $m_d^{ind} > m_d$ , which implies:

$$\frac{m_i}{m_j} < \frac{1 - \theta_j}{\Pi_j - \theta_j} \quad (5)$$

There are two possible reasons for this result: (a) a lack of sensitivity in the CV part of the CV/SG method ( $\frac{m_i}{m_j}$  is too low); (b) RUL between  $i$  and  $j$  derived from MSG response is too large.

Carthy et al. (1999) do not provide a very detailed analysis of the reasons behind this disparity. Our hypothesis is that the large VSL estimates resulting from double chaining were related to RULs (i.e., with MSG answers) rather than to the CV component. This was based on their result that 10% of their sample did not want to accept any risk and that the median risk in the lottery where death was the worst

outcome was very small (about 1%). Since RUL is a ratio, if the denominators  $(p_i - q_i)$  or  $(\Pi_j - \theta_j)$  are very small, the product  $\frac{1-q_i}{p_i-q_i} \times \frac{1-\theta_j}{\Pi_j-\theta_j}$  can be extremely large. For this reason, one of our groups was designed to: (a) minimize the disparity between direct and indirect methods, (b) isolate the effect of MSG in the comparison between direct and indirect chaining if the disparity persisted.

Our indirect chaining method works as follows:

- a. We first ask subjects to establish indifference between lotteries  $(p_k, \text{Death}; \text{Normal health})$  and  $(q_k, \text{Death}; \text{Injury } k)$ . From this we get

$$m_d = \frac{1 - q_k}{p_k - q_k} \times m_k \tag{6}$$

where  $k$  is a health state worse than the health state we used in direct chaining ( $k < i$ ).

- b. We ask subjects to establish indifference between lotteries  $(\Phi_i, \text{Injury } k; \text{Injury } i)$  and  $(\Pi_i, \text{Injury } k; \text{Normal health})$ . From this we get

$$m_k = \frac{1 - \Phi_i}{\Pi_i - \Phi_i} \times m_i \tag{7}$$

and from (6) and (7)

$$m_d^{ind} = \frac{1 - q_k}{p_k - q_k} \times m_k = \frac{1 - q_k}{p_k - q_k} \times \frac{1 - \Phi_i}{\Pi_i - \Phi_i} \times m_i \tag{8}$$

There are two features of our indirect chaining method that are important for this paper. One is that, unlike Carthy et al. (1999), the health state ( $k$ ) used in indirect chaining is more severe than the health state  $i$  used in direct chaining. We hypothesized that we would have fewer subjects accepting very small risks of death or no risks at all using a more severe health state. This might reduce the number of subjects producing a very small denominator in (6). Smaller RULs would make  $m_d^{ind}$  and  $m_d$  more similar. The second feature is that, if we compare (2) and (8), we can see that any potential discrepancy between  $m_d^{ind}$  and  $m_d$  can only be due to the MSG part of the CV/SG method. In summary, our hypothesis is that the use of this alternative double chaining will reduce the discrepancy between the direct and indirect approaches.

The third objective of our study was to conduct two additional consistency tests. One was the internal consistency of the direct method and the second was the susceptibility of the method to framing effects.

- 1. Consistency of the direct method. The question we address in our study is to what extent the VSL remains constant when we use the direct method with two different health states. The null hypothesis is that there are no differences between the two VSL calculated with the direct method.

2. Framing effects. Carthy et al. (1999) used a value of  $q=0.001$  in the status quo lottery, which was framed as 1 in 1000. Subjects had to provide values for  $p$  larger than 0.001, framed as a risk of “ $n$  in 1000”. In our case, we also use  $q=0.001$  but, in some groups, we frame it as “ $10n$  in 10,000” and subjects have to provide values of  $p$  framed as “ $n$  in 10,000”. The main reason that led us to study this framing effects is that we have some experience with the phenomenon known as ‘ratio bias’ (Pinto-Prades et al. 2006). According to this bias, when probabilities are presented as a ratio, respondents tend to focus more on the numerator than on the denominator of the ratio. If subjects are influenced by this framing effect, they tend to regard a good outcome with probability specified as  $10n$  in 10,000 as preferable to the same good outcome with probability specified as  $n$  in 1000 and regard a bad outcome with probability specified as  $10n$  in 10,000 as worse than the same bad outcome with probability specified as  $n$  in 1000. As a result, the risk of treatment failure that respondents will be willing to accept in a standard gamble will tend to be smaller if risks are framed as  $10n$  in 10,000 rather than  $n$  in 1000, which will result in increasing RULs and VSLs. We wanted to see to what extent this effect could influence the VSL calculated using the CV/SG method. The null hypothesis is that there will be no framing effects.

## 4 The valuation study

### 4.1 Participants

A random multi-phase sampling stratified by autonomous communities (Spanish regions), population size, age groups and gender was carried out, with the aim of obtaining a sample that was representative of the Spanish adult population. The participants ( $n=2020$ ) were randomly assigned to one of eight subsamples. The survey was administered face to face through computer assisted personal interviews (CAPI) by GFK Emer, S.A. in 2009. Average time per interview was about 40 min.

### 4.2 The questionnaire

The questionnaire consisted of 42 questions grouped in five sections. Sections 1 and 5 were identical for all the subsamples, while Sects. 2, 3 and 4 were different. The structure of the questionnaire for each of the groups is summarized in Table 1.

Although the sample is split into eight groups, we will combine groups 1 and 2 (group 1&2 from now on) and groups 4 and 5 (group 4&5 from now on) since they share the same tasks in Sect. 3, which is the section devoted to the CV/SG chained method. They are different in some of the tasks conducted in Sect. 2 and will be analyzed elsewhere.

Section 1 informs subjects about the risk of road accidents in Spain using a risk ladder in logarithmic scale (Corso et al. 2001). We also collected data about car use, attitudes toward road safety and subjective risk perceptions.

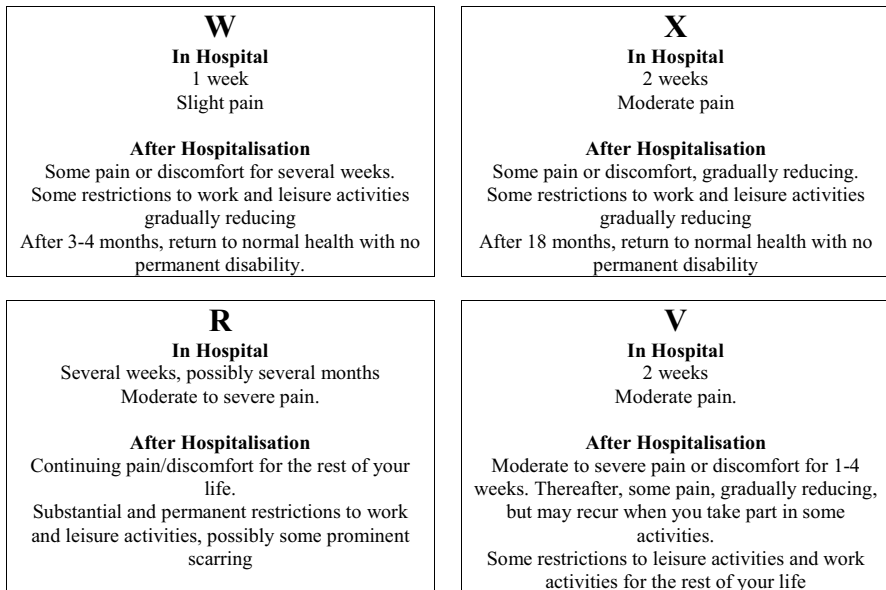


**Table 1** Structure of the questionnaire and experimental design

	Groups/subsamples							
	1	2	3	4	5	6	7	8
Section 1. Introduction								
Section 2. Visual Analogue Scale (VAS) and willingness to pay (WTP) for risk reductions								
(a) VAS scoring								
Health states/injuries	X, W, R, V and Death							
(b) WTP for a risk reduction <sup>a</sup>	Death and injury X							
Type of risk	1 in 10,000	1 in 10,000	1 in 10,000	Death and injury V	1 in 10,000	1 in 10,000	Death	Death
Reduction of risk	1 in 10,000	1 in 10,000	1 in 10,000 (Death) 10 in 10,000 (X)	1 in 10,000	1 in 10,000	1 in 10,000 (Death) 10 in 10,000 (V)	1 in 100,000	1 in 10,000
Section 3. WTP and willingness to accept (WTA) under certainty								
(a) WTP to avoid a health state								
Health state/injury	X			V			W	X
(b) WTA for suffering a health state								
Health state/injury	X			V			W	X
Section 4. Modified standard gamble								
Treatment A outcomes	X/Death			V/Death			X/Death	R/Death
Treatment B outcomes <sup>b</sup>	NH/Death							
Base of probability	10,000		1000	10,000		1000	1000	
Treatment A outcomes	—						1000	
Treatment B outcomes <sup>b</sup>							W/X	X/R
Base of probability							NH/X	NH/R
Section 5. Sociodemographic questions							100	

<sup>a</sup>Results from Sect. 2b not reported in this paper

<sup>b</sup>NH stands for Normal Health



**Fig. 1** Description of the Health States (non-fatal injuries) used throughout the questionnaire

In Sect. 2, the survey showed four cards (Fig. 1) describing four hypothetical road accident injuries or ‘health states’, anonymously labeled as X, W, R and V, at random, to the respondents. The health states were almost literal translations of the cards used by Carthy et al. (1999) and Jones-Lee et al. (1995). Next, participants had to score the states on a 0–100 Visual Analogue Scale (VAS): first they rank the health states according to their preferences and then place them on the scale such that the intervals between states reflect the differences in preference (Drummond et al. 2015). This was followed by WTP questions for small risk reductions for fatal and non-fatal road injuries (depicted by states X or V, depending on the group). These questions aimed to estimate the VSL through the “standard” CV method.

Sections 3 and 4 asked the questions related to the CV/SG chained approach. In Sect. 3 we asked the maximum WTP to avoid non-fatal road injury (X, V or W, see Table 1) under certainty. Participants were also asked for their minimum monetary compensation (WTA) that would produce a utility gain equivalent to the utility loss of the same non-fatal road injury as before under certainty. A set of payment cards was used to ask WTP and WTA questions. Each card represented a number of euro from the following quantities: 10, 30, 50, 100, 150, 300, 600, 1000, 3000, 6000, 10,000, 30,000, 100,000 and 300,000. The two questions (WTP and WTA) were displayed in a random order. The framing of WTP questions, was as follows: A payment card showing a certain amount of money randomly appeared in the top left corner of the screen, and respondents had to place the card into one of the following categories: (a) “I would pay this amount for sure”; (b) “I would not pay this amount for sure”; and (c) “I am not sure whether I would pay or not”. This method produced an interval defined by the highest amount that they would pay for sure and

the lowest amount that they would not pay for sure. An open-ended question asked about the amount they would pay within the range. This open response is the WTP that we use in the analysis. A similar procedure was used to derive WTA values.

The values obtained from WTP and WTA questions were used to infer the MRS between wealth and the risk of non-fatal road accident  $i$  ( $m_i$ ) on the assumption of a homogeneous underlying utility of wealth function (see Sect. 5.2.2. for a discussion on the choice of the utility function). We obtained three  $m_i$  for three different health states: X (groups 1&2, 3 and 8), V (groups 4&5 and 6) and W (group 7). Estimates of  $m_X$  from groups 1&2 and 3, and  $m_V$  from groups 4&5 and 6, were used to obtain VSL estimates with the direct CV/SG chained approach. Estimates for  $m_X$  and  $m_W$  from groups 7 and 8, respectively, were used to derive VSL estimates with the indirect chained procedure.

Section 4 includes the modified standard gamble (MSG) questions. Respondents were asked to imagine that they had suffered a road accident that would have fatal consequences if they were not treated. In this situation, two treatments were available. Both might be successful, with a certain probability, or they might fail. Outcomes and probabilities in case of success and failure of the were as follows:

- (a) *Groups 1&2 and 3*—Subjects had to choose between treatment A and B. If treatment A was successful, it would result in non-fatal health outcome X with probability 0.999. It could fail with probability of 0.001 and this would result in the death of the subject. Treatment B, if successful, would lead a complete recovery in 3–4 days with probability  $(1 - q_X)$  but it could also fail, and this would result in the death of the subject with probability  $q_X$ . The elicitation procedure asked the subject about the risk of death in Treatment B ( $p_X > 0.001$ ) that would make them indifferent between the two treatments.

A choice-based matching (CBM) procedure (Pinto et al. 2018) was used to find this probability. This procedure estimates  $p_X$  using a convergent sequence of choices between Treatments A and B using the ‘bisection’ method at first, followed by a ‘ping-pong’ procedure until an interval containing  $p_X$  is produced. Afterwards, respondents were asked to establish the exact value of  $p_X$  within the interval in an open question. The same CBM procedure was used in the remaining groups of the sample.

In group 1&2, probabilities  $(p_i, q_i)$  were framed as “10n in 10,000”, while in group 3 they were framed as “n in 1000”. VSL was estimated according to Eq. (1) as follows:

$$m_d = \frac{1 - 0.001}{p_X - 0.001} \times m_X \quad (10)$$

- (b) *Groups 4&5 and 6*—Everything was the same (risk framing of “10n in 10,000” for group 4&5 and of “n in 1000” for group 6) as in group 1&2 except that the non-fatal health state was V instead of X. The VSL was estimated according to Eq. (1) as follows:

$$m_d = \frac{1 - 0.001}{p_V - 0.001} \times m_V \quad (10)$$

- (c) In group 7 we reproduced the double chaining procedure used by Carthy et al. (1999). The first MSG question was exactly the same as in group 3, namely, to estimate  $p_X$  such that subjects were indifferent between treatment A (with  $q_X = 0.001$ ) and treatment B ( $p_X > 0.001$ ). Probabilities were framed as “ $n$  in 1000”. One additional MSG question was asked between treatments C and D. Treatment C, if successful, led to a non-fatal outcome W but failure led to condition X, which the vast majority of subjects evaluated as more severe than W (96,5% in the ranking task, 94% in VAS scoring and 100% in WTP questions). Treatment D, if successful led to a rapid recovery, but failure led to X. The probability of failure for Treatment C ( $\theta_W$ ) was set at 1 in 100 and subjects had to provide the probability of failure for treatment D ( $\Pi_W$ ) that made them indifferent between C and D. Given that in Group 7 we also obtained  $m_W$ , we estimated VSL using Eq. (3) as follows:

$$m_d^{ind} = \frac{1 - 0.001}{p_X - 0.001} \times \frac{1 - 0.01}{\Pi_W - 0.01} \times m_W \quad (11)$$

- (d) In group 8 we used a different double chaining. The first MSG question was between treatments A and B. If treatment A was successful, it would result in non-fatal health outcome R with probability 0.999. It could fail with probability of 0.001 and this would result in the death of the subject. Treatment B could result in death with probability  $p_R$  and a quick recovery with probability  $(1 - p_R)$ . The value of  $p_R$  was changed until indifference was reached between treatment A and treatment B. Probabilities were framed as “ $n$  in 1000”. The second MSG question was between treatments C and D. Treatment C was a lottery that resulted in injury R with probability 1 in 100 and in a milder injury X with probability 0.99. Treatment D led to Death with probability  $\Pi_X$  or to a quick and full recovery with probability  $(1 - \Pi_X)$ . Probabilities were framed as “ $n$  in 100”. Since respondents in group 8 were also asked about their WTP and WTP for injury X, we estimated VSL using Eq. (7) as follows:

$$m_d^{ind} = \frac{1 - 0.001}{p_R - 0.001} \times \frac{1 - 0.01}{\Pi_X - 0.01} \times m_X \quad (12)$$

The final part of the questionnaire (Sect. 5) collects socio-demographic information: age, sex, educational attainment, income level, employment status, etc.

### 4.3 Objectives and hypothesis testing

Hypotheses were tested between-subjects. This is different from Carthy et al. (1999), since they observed the disparity between the direct and the indirect methods within the same sample. We could not include more questions in the survey to make the

comparison within subjects since ours also included questions of the standard CV method, namely, asking subjects their WTP to reduce marginally the risk of death in a road traffic accident. Due to the larger number of questions, we only asked each subject one WTP and one WTA question using the CV/SG approach. With the results of the above survey, we test our hypotheses as follows. All between-subject comparisons are performed by using parametric (ANOVA and t-test) and non-parametric tests (Kruskal Wallis and Wilcoxon–Mann–Whitney—WMW-tests) over individual data:

1. *Objective 1: replicating* (Carthy et al. 1999).
  - a. Sensitivity to scope using MSG. This will be tested by comparing  $p_R$ ,  $p_X$  and  $p_V$ . We should find that  $p_R$  (group 8)  $>$   $p_V$  (group 6)  $>$   $p_X$  (group 3).
  - b. Sensitivity to scope in CV. This will be tested by comparing  $WTP_X$  (groups 1–3 and 8),  $WTP_V$  (groups 4–6) and  $WTP_W$  (group 7), as well as  $WTA_X$ ,  $WTA_V$  and  $WTA_W$  (same group distribution).
  - c. Discrepancy between direct and indirect chaining. This will be tested by comparing VSL of groups 3 and 7. We expect VSL in indirect chaining to be larger (group 7) than VSL using direct chaining (group 3).

2. *Objective 2: reconciling the direct and indirect method.*

We will test if our new indirect chaining method avoids the discrepancy between direct and indirect chaining by comparing the RUL and VSL of groups 3 and 8.

3. *Objective 3: testing the consistency of the direct method.*

We will compare VSL estimates obtained by using direct chaining through X (groups 1&2 and 3) or V (groups 4&5 and 6).

4. *Objective 4: testing the influence of framing effects*

To test whether RUL depends on risk framing, we will compare  $p_X$  framed as “10n in 10,000” (group 1&2) to  $p_X$  framed as “n in 1000” (groups 3). We will also compare  $p_V$  framed as “10n in 10,000” (groups 4&5) to  $p_V$  framed as “n in 1000” (group 6).

## 5 Results

### 5.1 Sample characteristics and homogeneity of subsamples

Socio-demographic and attitudinal characteristics of our sample can be seen in Table 2, which also shows the distribution of adult population with respect to age and gender, according to the Spanish census, and with respect to education, marital status and employment status, according to the labour force survey (LFS). In

**Table 2** Percentage distribution of sample characteristics (n = 2020)

Variables	Sample (%)	Spanish population <sup>a</sup>
Gender		
Male	50.4	49.3
Female	49.7	50.6
Age		
18–29	21.6	18.0
30–39	20.5	22.0
40–49	19.1	20.5
50–65	22.9	24.4
≥ 66	15.9	15.2
Education		
No education, Prim. or Lower Sec	40.3	54.8
Upper secondary	32.7	20.3
Tertiary	26.9	24.8
Employment status		
Inactive	38.5	40.1
Employed	49.1	47.1
Unemployed	12.4	12.7
Household income (€)		
Up to 2.000	70.5	69.5
€2001–3000	25.2	19.5
More than 3000	4.3	11.0
Marital status		
Single	30.1	31.4
Married	56.9	55.4
Divorced	6.5	5.7
Widowed	6.5	7.5

<sup>a</sup>Spanish 2011 census and Spanish Labor Force Survey (1st quarter of 2011)

general, our sample resembles the characteristics of the general population. More information was collected about other characteristics and can be seen in Table 2.

Since most of the subsequent analysis will be performed on a between-subject basis, homogeneity of subsamples must be assessed. According to parametric (ANOVA) and non-parametric (Kruskal Wallis) tests, we could only reject the null hypothesis for the variable age, with individuals in groups 4 and 8 being slightly older than the other groups. There are no differences in the rest of the characteristics among groups. Homogeneity in health preferences has been tested using VAS scores. Both ANOVA and Kruskal Wallis tests ( $p > 0.05$ ) suggest that no statistically significant differences exist between groups regarding valuations of health states through VAS scoring, which are presented in Table 3.

**Table 3** Visual Analogue Scale scores for non-fatal injuries and death, by group and on aggregate

Injuries/ health states <sup>a</sup>		Groups								Whole sample
		1	2	3	4	5	6	7	8	
R	Mean	17,28	17,60	17,51	17,66	16,96	15,84	17,11	16,20	17,04
	St. Dev	10,58	12,00	11,79	13,04	12,28	12,14	13,92	9,88	12,02
V	Mean	33,25	32,59	32,37	33,25	32,20	31,12	33,10	31,50	32,43
	St. Dev	13,24	13,13	12,14	13,93	14,00	12,84	13,72	12,69	13,23
X	Mean	56,47	54,86	55,73	55,73	52,84	55,31	56,13	54,99	55,25
	St. Dev	13,45	13,75	14,17	14,72	15,37	14,10	14,79	15,81	14,52
W	Mean	71,83	70,74	72,87	71,44	69,56	70,71	72,19	70,31	71,19
	St. Dev	15,18	16,12	15,02	16,68	17,97	15,06	15,60	17,96	16,23
Death	Mean	2,05	2,41	2,55	2,93	2,17	2,68	2,11	1,54	2,32
	St. Dev	5,01	5,57	9,71	9,17	5,00	5,40	5,35	4,04	6,49
	n	251	283	243	269	254	259	233	228	2020

<sup>a</sup>Health states or non-fatal injuries R, V, X and W are described in Fig. 1

**Table 4** Willingness to pay and willingness to accept in euro to avoid/suffer a health state, and marginal rates of substitution between wealth and health ( $m_i$ )

	Group and injuries/ health states <sup>a</sup>					
	1&2	3	4&5	6	7	8
	X	X	V	V	W	X
<b>WTP</b>						
Mean	3226	2838	7497	5980	1993	2572
Median	500	500	800	725	250	450
St Dev	12,562	8740	22,423	16,773	6013	6907
n	531	239	509	258	233	227
<b>WTA</b>						
Mean	106,492	115,879	227,655	228,426	100,689	112,176
Median	11,000	10,000	30,000	32,500	8,000	10,000
St Dev	312,960	361,479	682,040	553,688	361,986	548,946
n	528	241	517	254	233	225
$m_i$						
Mean	3817	3586	10,650	8837	2076	3553
Median	787	909	1198	1200	397	659
St Dev	12,095	11,033	32,842	24,154	6,000	10,479
n	525	237	504	253	233	224

Marginal Rates of Substitution between wealth and health ( $m_i$ ) derived from individuals' WTP and WTA responses (estimation based on the homogeneous utility function):

$$m_i = \frac{2 \times WTP \times WTA}{WTP + WTA}$$

<sup>a</sup>Health states or non-fatal injuries X, V and W are described in Fig. 1.

## 5.2 Results of contingent valuation (CV) questions

### 5.2.1 WTP and WTA estimates

Table 4 shows WTP and WTA statistics based on usable data. There are no valid WTP responses for 23 (1.1%) subjects who were willing to pay any amount of money, no matter its magnitude, to avoid the health state ( $WTP \rightarrow \infty$ ). Likewise, in 22 (1.1%) cases, respondents stated that no amount of money could compensate them for suffering the injury ( $WTA \rightarrow \infty$ ). Those 55 subjects were excluded from the analysis. On the other hand, 96 respondents refused to pay any amount of money for any treatment ( $WTP=0$ ). Those subjects were not excluded from the analysis.

We can see that WTP and WTA values are sensitive to scope. Non-parametric tests (Wilcoxon–Mann–Whitney test) confirm that there are statistically significant ( $p < 0.05$ ) differences in WTP and WTA values between the three health states (i.e. groups 1&2, 3 and 8 for state X; groups 4&5 and 6 for state V; and group 7 for state W). The results are also confirmed by the t-test, except for the comparison between X and W ( $p=0.17$  for WTP and  $p=0.74$  for WTA). The direction of the effect is consistent with the severity of injuries observed in the VAS scores.

In addition, theoretical validity of WTP estimates have been tested by regression analysis (see Appendix in the Electronic Supplementary Material). The results show a positive and significant relationship between individuals' income and their WTP. The analysis also confirms the sensitivity to scope since a positive relationship between the size of the health gain and the WTP is observed.

### 5.2.2 Marginal rates of substitution (MRS)

Individual MRS ( $m_i$ ) are inferred from WTP and WTA values by applying a homogeneous utility function, one of the four functional forms used by Carthy et al. (1999). Results do not essentially change when other functional forms (i.e., negative exponential, logarithmic and nth root) are assumed. We show the results in detail for only one functional form, for the sake of simplicity, and we choose the homogeneous one because it minimizes the number of observations that we have to discard due to missing values. The MRS based on all the utility functions (and, hence, the corresponding VSL estimates) will be considered later when a recommended VSL for the evaluation of road safety policies in Spain is proposed in Sect. 5.5. Basic statistics for MRS are shown in the bottom part of Table 4. All the binary comparisons between MRS for states X, V and W are significant ( $p < 0.05$ , t-test and WMW) and correspond to the ranking of severity of those injuries.

## 5.3 Results of the MSG part of the CV/SG chained approach

### 5.3.1 Risk of death accepted to avoid a non-fatal injury

Indifference probabilities elicited in the different groups, as well as the RULs constructed from them are shown in Table 5. The main results are:



**Table 5** Modified standard gamble results: probabilities ( $p_i, \Pi_j$ ) and relative utility losses (RULs)

$p_i(i=X,V,R); \Pi_w;$ $\Pi_X^a$	Groups							
	1&2	3	4&5	6	7	8		
	$p_X$		$p_V$		$p_X$	$\Pi_W$	$p_R$	$\Pi_X$
Mean	0.114	0.097	0.235	0.315	0.126	0.464	0.365	0.257
Median	0.004	0.006	0.025	0.125	0.006	0.470	0.356	0.110
St Dev	0.237	0.203	0.314	0.348	0.261	0.354	0.347	0.281
Take minimum risk (%)	26.0	31.7	13.8	9.7	33.9	16.3	8.3	21.9
Take $\leq 0.002$ (%)	43.5	31.7	23.9	9.7	33.9	–	8.3	–
Injuries/health states <sup>b</sup>	X		V		W		X	
Relative utility loss (RUL) <sup>c</sup>	$\frac{1-0.001}{p_X-0.001}$		$\frac{1-0.001}{p_V-0.001}$		$\frac{1-0.001}{p_X-0.001} \times \frac{1-0.01}{\Pi_w-0.01}$		$\frac{1-0.001}{p_R-0.001} \times \frac{1-0.01}{\Pi_X-0.01}$	
Mean	3095	407	1,680	145	11,888		5,800	
Mean (max multiplier 999)	511	–	304	–	–		–	
Median	309	200	42	8	749		93	
St Dev	4,223	425	3429	296	28,926		19,822	
N	534	243	523	259	233		228	

<sup>a</sup>  $p_i$ : indifference value for probability of death in the first modified standard gamble question (with X, V and R).  $\Pi_w, \Pi_X$ : indifference value of the probability of suffering X or R, respectively in the second modified standard gamble (with W and X, respectively)

<sup>b</sup> Health states or non-fatal injuries X, V, W and R are described in Fig. 1

<sup>c</sup> In groups 7 and 8, they are compound RULs

1. MSG responses are sensitive to scope. Mean and median values of the risk of death that participants were willing to take to avoid a non-fatal injury (i.e., probabilities  $p_i, i = X, V, R$ ) change according to the severity of the condition ( $p_R > p_V > p_X$ ) at 0.05 level (t-test and WMW).
2. The way that probabilities were framed (“n in 1000” vs “10n in 10,000”) had an effect on the risk assumed ( $p_i$ ). There are significant differences ( $p < 0.05$ , WMW) both for injury X (i.e., group 1&2 vs. groups 3 and 7) and for injury V (i.e., group 4&5 vs. group 6). This conclusion holds when t-test comparisons are performed, but only for state V. Effects are in line with predictions of the ratio bias effect, namely, individuals accept less risk when probabilities are framed as 10 in 10,000 than when they are framed as 1 in 1000.
3. There is a good number of people whose indifference point coincides with the minimum risk they saw on the screen (see row “Take minimum risk” in Table 5). Note that in groups 3 and 6 the minimum risk that participants may state is 2 in 1000 (the lowest value above  $q=0.001$  when this risk is framed as 1 in 1000). In contrast, in groups 1&2 and 4&5, the minimum level is 11 in 10,000 (remember that in the lottery used as stimulus the risk of death was framed as 10 in 10,000).

The percentage of subjects who take the minimum risk is about the same (20%) in groups 1&2 and 4&5 as in groups 3 and 6.

4. We show in row “Take  $\leq 0.002$ ” the percentage of subjects not willing to take a risk larger than 0.002. Note that in groups 1&2 and 4&5 this percentage is much larger than in groups 3 and 6. We add this information because it confirms the influence of the ratio bias. The reason is that, in practice, the minimum risk subjects could take in groups 3 and 6 was 2 in 1000. This corresponds to 20 in 10,000 in groups 1&2 and 4&5. The change in framing largely increases the percentage of subjects who do not want to accept a risk above 2 in 1000 when it is framed as 20 in 10,000.

#### 5.4 Relative utility losses (RULs)

The main results regarding RULs are:

1. Groups where risk was framed as “10n in 10,000” (groups 1&2 and 4&5) produced much bigger RULs than groups 3 and 6. All differences are statistically significant ( $p < 0.001$ ; t-test and Wilcoxon–Mann–Whitney test). These results are expected since RULs are a sort of monotonic transformation of the probabilities presented above.
2. Means and medians are very different within each group, especially in groups where risks were framed as “10n in 10,000”. This reflects the fact that some subjects accepted very small increases in the risk of death, producing very large values of RULs.
3. In order to better understand the differences between groups 1&2 and 4&5, on the one hand, and groups 3 and 6, on the other, the RULs of groups 1&2 and 4&5 are capped at 999 (see row “max RUL 999”, Table 5), which is the maximum RUL that a subject could produce in groups 3 and 6 since it corresponds to a risk of 2 in 1000. We can see that, even with capping, the mean RULs in groups 1&2 and 4&5 are still larger than in groups 3 and 6 ( $p < 0.005$ ).
4. There is a very large difference between means and medians within groups 7 and 8. This shows the effect of the multiplication of two ratios on these ‘compound’ RULs when some people are willing to take only very small risks in the two MSG: the means are huge. For this reason, we can expect that the mean VSL in these two groups will be much larger than in groups with direct chaining. Take into account that the RUL can reach a value of 99,999 if a subject takes the minimum risk in the two questions. In group 7 there are 22 subjects (9.4%) who take the minimum risk in the two lotteries while in group 8 there are only nine individuals (3.9%). It is not surprising that group 7 is more affected than group 8 by this sort of behavior, since state X is milder than R. The influence of those subjects on the mean is very large.
5. RULs of groups 3 and 8 are different. However, the mean and the median go in the opposite direction. The mean RUL of group 8 is much larger than the

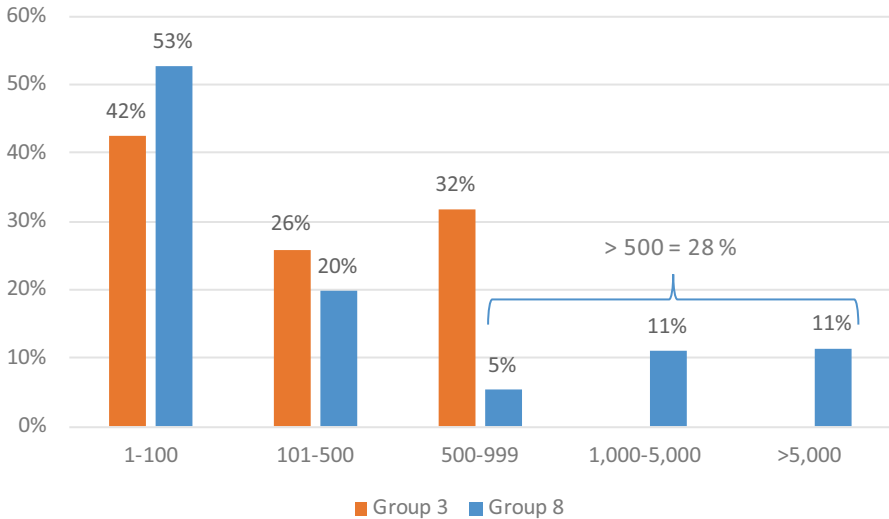


Fig. 2 Distribution of relative utility losses (RULs) in groups 3 and 8

mean of group 3, but the median is smaller. Statistical tests are inconclusive. The t-test rejects the hypothesis of equal means ( $p < 0.05$ ), but the non-parametric test (WMW) cannot reject the null hypothesis ( $p = 0.188$ ). The reason for this disparity is seen in Fig. 2. The double chaining of group 8 generates very large values in a group of subjects. About 22% of subjects have RUL > 999 in group 8, which is the maximum RUL in group 3. Those subjects generate large RULs, producing a large mean. It is also observed that our indirect chaining produced lower RULs than the direct method in a large proportion of subjects. In spite of that, the multiplication of two ratios generates very large RULs in a (significant) minority of subjects. This result suggests that the mean values for VSL estimated using the CV/SG method will depend on the number of steps (MSG chainings) that we use.

### 5.5 Value of statistical life (VSL) estimations based on the SG/CV chained approach

The VSL for groups 1–6 is obtained by combining individuals’ MRS ( $m_x$  or  $m_v$ ) with the ratio  $m_d/m_i$  derived from MSG responses. The VSL estimates for groups 7 and 8 have been calculated by combining  $m_w$  and  $m_x$ , respectively, with the corresponding ‘compound’ RUL, at the individual level. The existence of some extreme values led us to discard what we defined as outliers. Carthy et al. (1999) trimmed out values above an ad-hoc threshold of £15 million. We excluded individuals whose VSL values were larger than three standard deviations of the mean. As a result of

**Table 6** Value of statistical life (VSL) estimates in euro based on the CV/SG chained “direct” and “indirect” approaches<sup>a</sup> by groups

Groups	Direct X-based			Direct V-based			Indirect	
	1&2	3	1 to 3	4&5	6	4 to 6	7	8
Mean	8,675,951	1,130,138	5,837,279	6,366,670	311,410	4,536,131	11,600,000	4,159,875
Median	286,963	113,094	226,189	69,860	13,258	47,557	238,289	40,834
St. Dev	27,200,000	2,763,374	22,500,000	23,400,000	1,201,420	19,700,000	40,900,000	15,000,000
n	522	236	723	499	250	718	232	222
Restricting max. multiplier to 999 in Groups 1&2 and 4&5								
Mean	1,529,958	1,130,138	1,377,085	1,320,773	311,410	1,059,461	11,600,000	4,159,875
Median	191,857	113,094	179,739	62,570	13,258	43,206	238,289	40,834
St. Dev	3,674,120	2,763,374	3,335,320	3,847,300	1,201,420	3,609,492	40,900,000	15,000,000
n	521	236	723	497	250	716	232	222

<sup>a</sup>“Direct X-based” refers to the CV/SG chained approach that uses injury X both in the CV part of the method (to derive the MRS) and in the MSG part (to obtain the RULs). “Direct V-based” is the result of applying the CV/SG chained approach by using injury V in the valuation procedures (CV and MSG). The “Indirect” approach uses injuries W (group 7) and X (group 8) in the CV part and then different MSG procedures involving injuries W and X in group 7 and states X and R in group 8

this trimming procedure, 15 values were discarded, which represents less than 1% of the sample. Table 6 shows trimmed VSL estimates. The main results are:

1. VSL estimated through X (groups 1&2 and 3) is larger than VSL estimated through V (groups 4&5 and 6). The t-test rejects the null hypothesis at the 10% level ( $p=0.07$ ). The nonparametric test (WMW) clearly rejects the null hypothesis ( $p<0.001$ ), so challenging the consistency of the direct version of the CV/SG method. The results do not change when comparisons are restricted to groups with the same framing (i.e., 1&2 vs. 4&5 and 3 vs. 6).
2. VSL estimated using the indirect chaining method used by (Carthy et al. 1999) (group 7) is much larger than VSL obtained with the direct approach (group 3) ( $p<0.05$ ; t-test and Wilcoxon–Mann–Whitney test).
3. In group 8 the results for VSL are similar to those obtained regarding RULs, namely, the t-test rejects the hypothesis of equality of means ( $p<0.05$ ), but the non-parametric test (WMW) cannot reject the null hypothesis ( $p=0.53$ ). This result implies that using a more severe health state (R) to perform the double chaining does not fully prevent the discrepancy between the indirect and the direct CV/SG estimates of the VSL. However, the discrepancy between the mean VSL estimates obtained in groups 8 and 3 is almost 73% lower than between groups 7 (the indirect chaining used by (Carthy et al. 1999)) and 3 (direct chaining).

To obtain the recommended VSL to be used in the economic evaluation of road safety programmes in Spain we consider the results of groups 1–6 jointly, that is, only the VSL estimates obtained on the basis of direct chaining are used for that purpose. Carthy et al. (1999) concluded that the direct approach is less vulnerable to

**Table 7** Value of Statistical Life (VSL) estimates in euro based on the CV/SG chained “direct” approach<sup>a</sup> using alternative wealth utility functions for deriving MRS

	Homogeneous	Logarithmic	Negative Exponential	Nth root
Mean	1,195,543	1,957,797	1,026,982	2,624,536
Median	78,687	166,083	75,479	216,186
StdDev	3,341,072	5,343,087	2,813,742	7,322,648
n	1504	1444	1444	1444

<sup>a</sup>VSL estimates resulting from the aggregation of individual’s values of groups 1–6, after restricting multipliers to 999 in groups 1&2 and 4&5 and after trimming values to exclude outliers

compounding of errors due to its fewer steps, compared to the indirect chaining. Our results confirm this conclusion, since more extreme values are obtained when the indirect or double chaining approach is applied. There being no normative criterion for choosing a specific functional form, we decided to obtain the summary value as the average of the four estimates. Table 7 shows VSL estimates for groups 1–6 considered altogether, based on each of the utility functions. The figures in the table have been obtained after capping RULs from groups 1&2 and 4&5, as explained in Sect. 5.4, and then trimmed by discarding VSL values larger than three standard deviations of the mean.

According to welfare economics, it is the mean values which should be regarded as representative of aggregate preferences. The simple average of mean estimates in Table 7 gives a VSL of 1.7 million euro. Notwithstanding, as Carthy et al. (1999) suggested, at least some weight could be given to the estimates based on median values. Hence, given the large dispersion of the estimates and the notable disparities between mean and median values, a simple arithmetic average of the minimum of the median values (the one derived from the negative exponential function) and the maximum of the mean values (that obtained when the nth root function is used) is calculated, resulting in a VSL of 1.35 million euro. On the other hand, some authors have raised concerns regarding the meaningfulness of WTA values in this context (Viscusi 2018a). Considering this, individual VSL estimates have also been obtained by using only WTP responses to derive the MRS. In that case, no assumption must be made regarding the functional form. A unique VSL estimate is derived for each subject and the mean VSL estimate is 1.6 million euro (median value 0.5 million euro). In view of these results, the recommended VLS for the evaluation of road safety programmes should be in a range from about €1.3 million to €1.7 million.

Finally, theoretical validity of the VSL estimates has been tested by regression analysis (see Appendix in the Electronic Supplementary Material), yielding figures for income elasticity from 0.64 to 0.70, depending on the utility function used to derive MRS from WTP and WTA values. These values are similar to those in a recent meta-analysis of elasticities of studies estimating VSL through stated preferences. Specifically, for middle and high-income countries, a range between 0.55 and 0.85 is reported (Viscusi and Masterman 2017). In any case, the value of the income elasticity is below 1, as is the case in most studies so far (Miller 2000; Viscusi and Aldy 2003).

## 6 Discussion

A first outcome of the study is the estimation of a VSL for Spain in the context of road accidents. The estimated range from €1.3 to €1.7 million is in line with VSL estimates used in other European countries to evaluate road safety programmes. A recent study including official costs per road traffic fatality in 29 European countries indicates that this range is between 0.7 and 3 million euro (adjusted for PPP and referring to the year 2015) (Wijnen et al. 2019). As it is known, stated preference VSL estimates are usually lower than those implied by revealed preference studies. Our estimate is not an exception to that rule and, in fact, is lower than the market estimates of VSL provided by (Martínez and Méndez 2009) for Spain, which range from €2.8 to €8.3 million. Likewise, our VSL estimate is also lower than estimates based on automobile purchase decisions in the U.S. (O'Brien 2018), from which an inverted-U shape to the age-VSL function that ranges from \$1.5 to \$19.2 million is obtained. Moreover, since Viscusi and Gentry (2015) find that the current VSL estimate of \$9.2 million used by the U.S. Department of Transportation's, based on labor market studies, is appropriate for valuing transport-related fatalities, it can also be stated that our VSL estimated in the specific context of traffic fatalities is much lower than that value. Moreover, the values in the range estimated in this study are lower than those obtained in Spain through revealed preferences in the context of the labour market, which are between €2.8 and €8.3 million (Martinez and Mendez 2009). They are also clearly below the value recently suggested to assess the cost of the current COVID pandemic based on the estimated VSL for the US, using an income elasticity of 1, which is \$6.6 million (Viscusi 2020).

Besides this outcome, the main objectives of our study were to understand better the problems observed in Carthy et al. (1999), to overcome some of those problems, and to conduct new consistency checks. Regarding our objectives and hypotheses, we find that:

1. We have reproduced the main results of Carthy et al. (1999). Our survey confirms that CV shows sensitivity to scope. WTP/WTA are larger for more severe conditions. The same happens with MSG, namely, subjects accept higher risks of treatment failure (resulting in death) to avoid more severe injuries.
2. We observe consistency in the relative values of health states using CV and MSG at the aggregate level. Mean WTP (WTA) values for state V (groups 4&5 and 6) are between 2.1 (1.97) and 2.3 (2.1) times the mean values for state X (groups 1&2, and 3). When we assess the relative value V/X implicit in the responses to MSG questions, the results are remarkably similar: ratio of means of RULs for V and X are between 1.8 and 2.8. In summary, the CV/SG method performs much better in terms of scope sensitivity than the standard method of asking WTP questions for very small risk reductions.
3. Regarding objective 2, the attempt to reconcile the direct and the indirect methods, we have managed to reduce the disparity between them. The differences between the means of the direct and indirect methods were smaller using our method (group 8). Statistical tests are not so conclusive about the difference between the

direct and indirect approaches. However, we also observe more extreme values in our indirect chaining method than in the direct method, leading to very large means. It seems that the more steps we use in MSG, the more extreme values we will observe. The implication is that the more intermediate steps (chained MSG lotteries) we use to estimate RUL, the larger the mean values of VSL we will find. The problem seems to lie in the skewness of MSG. There is a significant number of subjects who only accept very small increases in risk. This generates very large values of RULs. When these values are combined with the MRS, they may result in very large VSLs. The distribution of MRS is also highly skewed. The combination, at the individual level, of extreme values for  $m_i$  and RUL, generates extremely large values of VSL in a minority of subjects. It looks as if some arbitrary “trimming” will always be necessary.

4. Regarding objective 3, to test the consistency of the direct method when different non-fatal injuries are used, we found that VSL estimates depend on the health state (i.e., injury) used. It looks as if more severe health states produce lower VSLs. This result was also observed by Olofsson et al. (2019). However, their study does not follow as closely as ours does the methods used by Carthy et al. (1999). Also, their VSL figures were based on mean estimates whereas our results are based on individual chaining.
5. Regarding objective 4, the influence of framing effects, we observe an important influence of risk framing in the RUL and, by extension, in the VSL estimates. This impact is particularly strong for the subsamples who value state V. The mean RUL of group 4&5 is about 12 times higher than the corresponding value in group 6. This discrepancy is also large between medians (five times greater) and it cannot be attributed to the influence of a few outliers. One explanation for this finding is the ratio bias effect.

To conclude, responses to WTP, WTA and MSG questions, which are required to apply the CV/SG chained method, are sensitive to scope. Average means and medians are consistent ordinally and cardinally. In this respect, this method is better than the standard CV survey that asks subjects their WTP to reduce very small risks. However, we have observed that the method has two problems. One is that RULs are extremely sensitive to the responses of subjects who accept small risks. For this reason, a simple manipulation such as changing the base of the probabilities generates extremely different responses. This effect is amplified when we use double chaining, since we need to compound two ratios to calculate RUL. The second problem is that the two parameters we need to calculate VSL ( $m_i$  and RUL) tend to have very skewed distributions. When we work with individual data, this may produce very large VSLs. This is one of the reasons why very consistent measures of central tendency for  $m_i$  and RUL can generate problematic VSL values. We conclude that the CV/SG method produces more consistent results than the standard CV method. However, it is very susceptible to changes in the way that it is operationalized.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11166-021-09360-x>.

**Acknowledgements** The authors thank the Spanish Road Traffic Directorate General (Dirección General de Tráfico), for an unrestricted grant, the Spanish Ministry of Economy, Industry and Competitiveness, Grant/Avard Numbers: ECO2016-75439-P, PID2019-104907GB-I00 and Seneca Foundation (Science and Technology Agency of the Region of Murcia), Grant/Award 20825/PI/18. The authors thank the Editor-in-Chief, Kip Viscusi and anonymous reviewer for their helpful comments on earlier versions of the paper.

**Author contributions** All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by all authors. The first draft of the manuscript was written by F-I S-M and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** Support for this research comes from the Spanish Road Traffic Directorate General (unrestricted grant), the Spanish Ministry of Economy, Industry and Competitiveness (grants ECO2016-75430-P and PID2019-104907GB-I00), and the Seneca Foundation (Grant 20825/PI/18).

**Data availability** The dataset generated during this study is available from authors upon request.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abellán Perpiñán, J. M., Sánchez Martínez, F. I., Martínez Pérez, J. E., & Méndez, I. (2012). Lowering the “floor” of the SF-6D scoring algorithm using a lottery equivalent method. *Health Economics (United Kingdom)*. <https://doi.org/10.1002/hec.1792>
- Anderson, R. G., & Kichkha, A. (2017). Replication, meta-analysis, and research synthesis in economics. *American Economic Review*, *107*(5), 56–59. <https://doi.org/10.1257/aer.p20171033>
- Andersson, H., Hole, A. R., & Svensson, M. (2016). Valuation of small and multiple health risks: A critical analysis of SP data applied to food and water safety. *Journal of Environmental Economics and Management*, *75*, 41–53. <https://doi.org/10.1016/j.jeem.2015.11.001>
- Balmford, B., Bateman, I. J., Bolt, K., Day, B., & Ferrini, S. (2019). The value of statistical life for adults and children: Comparisons of the contingent valuation and chained approaches. *Resource and Energy Economics*, *57*, 68–84. <https://doi.org/10.1016/j.reseneeco.2019.04.005>
- Beattie, J., Covey, J., Dolan, P., Hopkins, L., Jones-Lee, M., 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. *Journal of Risk and Uncertainty*, *17*(1), 5–26. <https://doi.org/10.1023/A:1007711416843>
- Berry, J., Coffman, L. C., Hanley, D., Gihleb, R., & Wilson, A. J. (2017). Assessing the rate of replication in economics. *American Economic Review*, *107*(5), 27–31. <https://doi.org/10.1257/aer.p20171119>
- Carthy, T., Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., & Spencer, A. (1999). On the contingent valuation of safety and the safety of contingent valuation: Part 2—The CV/



- SG “Chained” approach. *Journal of Risk and Uncertainty*, 17(3), 187–214. <https://doi.org/10.1023/A:1007782800868>
- Chilton, S., Covey, J., Jones-Lee, M., Loomes, G., Pidgeon, N., & Spencer, A. (2015). Response to “Testing the validity of the ‘value of a prevented fatality’ (VPF) used to assess UK safety measures.” *Process Safety and Environmental Protection*, 93, 293–298. <https://doi.org/10.1016/j.psep.2014.11.002>
- Corso, P. S., Hammitt, 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. <https://doi.org/10.1023/A:1011184119153>
- DETR. (1998). Valuation of the benefits of prevention of road accidents and casualties. Highway Economics Note No 1.
- Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press.
- Dubourg, J.-L., & Loomes, G. (1997). Imprecise preferences and survey design in contingent valuation. *Economica*, 64(256), 681–702. <https://doi.org/10.1111/1468-0335.00106>
- Hammitt, J. K., & Graham, J. D. (1999). Willingness to pay for health protection: Inadequate sensitivity to probability? *Journal of Risk and Uncertainty*, 18(1), 33–62. <https://doi.org/10.1023/A:1007760327375>
- Jones-Lee, M. (1974). The value of changes in the probability of death or injury. *Journal of Political Economy*, 82(4), 835–849. <https://doi.org/10.1086/260238>
- Jones-Lee, M., & Loomes, G. (2015). Final response to Thomas and Vaughan. *Process Safety and Environmental Protection*, 94(C), 542–544. <https://doi.org/10.1016/j.psep.2015.01.006>
- Jones-Lee, M., Loomes, G., & Philip, P. (1995). Valuing the prevention of non-fatal road injuries: Contingent valuation vs. standard gambles. *Oxford Economic Papers*, 47(4), 676–695. <https://doi.org/10.1093/oxfordjournals.oep.a042193>
- Jones-Lee, M., & Spackman, M. (2013). The development of road and rail transport safety valuation in the United Kingdom. *Research in Transportation Economics*, 43(1), 23–40. <https://doi.org/10.1016/j.retrec.2012.12.010>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291.
- Maniadiis, Z., Tufano, F., & List, J. A. (2017). To replicate or not to replicate? Exploring reproducibility in economics through the lens of a model and a pilot study. *The Economic Journal*, 127(605), F209–F235. <https://doi.org/10.1111/econj.12527>
- Martinez Perez, J. E., & Mendez Martinez, I. (2009). ¿Qué podemos saber sobre el Valor Estadístico de la Vida en España utilizando datos laborales? *Hacienda Pública Española*, 191(4), 73–93.
- McCord, M., & de Neufville, R. (1986). “Lottery equivalents”: Reduction of the certainty effect problem in utility assessment. *Management Science*, 32(1), 56–60. <https://doi.org/10.1287/mnsc.32.1.56>
- Miller, T. (2000). Variations between countries in values of statistical life. *Journal of Transport Economics and Policy*, 34(2), 169–188.
- Mueller-Langer, F., Fecher, B., Harhoff, D., & Wagner, G. G. (2019). Replication studies in economics—How many and which papers are chosen for replication, and why? *Research Policy*, 48(1), 62–83. <https://doi.org/10.1016/j.respol.2018.07.019>
- Nellthorp, J., Sansom, T., Bickel, P., Doll, C., & Lindberg, G. (2001). Competitive and sustainable growth (growth) programme unification of accounts and marginal costs for Transport Efficiency UNITE Valuation Conventions for UNITE. [www.its.leeds.ac.uk/unite](http://www.its.leeds.ac.uk/unite).
- O’Brien, J. (2018). Age, autos, and the value of a statistical life. *Journal of Risk and Uncertainty*, 57(1), 51–79. <https://doi.org/10.1007/s11166-018-9285-3>
- Olofsson, S., Gerdttham, U. G., Hultkrantz, L., & Persson, U. (2019). Value of a QALY and VSI estimated with the chained approach. *European Journal of Health Economics*, 20(7), 1063–1077. <https://doi.org/10.1007/s10198-019-01077-8>
- Pinto, J., Sanchez, F., Abellan, J., & Martinez, J. (2018). Reducing preference reversals: The role of preference imprecision and nontransparent methods. *Health Economics (United Kingdom)*, 27(8), 1230–1246. <https://doi.org/10.1002/hec.3772>
- Pinto-Prades, J.-L., Martinez Perez, J. E., & Abellan Perpiñan, J. (2006). The influence of the ratio bias phenomenon on the elicitation of health states utilities. *Judgment and Decision Making*, 1(2), 118–133.
- Søgaard, R., Lindholt, J., & Gyrd-Hansen, D. (2012). Insensitivity to scope in contingent valuation studies. *Applied Health Economics and Health Policy*, 10(6), 397–405. <https://doi.org/10.1007/bf03261874>
- Spackman, M., Evans, A., Jones-Lee, M., Loomes, G., Holder, S., Webb, H., & Sugden, R. (2011). *Updating the VPF and VPIs: Phase 1: Final Report Department for Transport*. NERA.

- Thomas, P. J., & Vaughan, G. J. (2015). Testing the validity of the 'value of a prevented fatality' (VPF) used to assess UK safety measures: Reply to the comments of Chilton, Covey, Jones-Lee, Loomes, Pidgeon and Spencer. *Process Safety and Environmental Protection*, 93, 299–306. <https://doi.org/10.1016/j.psep.2014.11.003>
- Viscusi, W. K. (2018a). *Pricing Lives: Guideposts for a Safer Society*. Princeton University Press.
- Viscusi, W. K. (2018b). Best estimate selection bias in the value of a statistical life. *Journal of Benefit-Cost Analysis*, 9(2), 285–304. <https://doi.org/10.1017/bca.2017.21>
- Viscusi, W. K. (2020). Pricing the global health risks of the COVID-19 pandemic. *Journal of Risk and Uncertainty*, 61(2), 101–128. <https://doi.org/10.1007/s11166-020-09337-2>
- 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. <https://doi.org/10.1023/A:1025598106257>
- Viscusi, W. K., & Gentry, E. P. (2015). The value of a statistical life for transportation regulations: A test of the benefits transfer methodology. *Journal of Risk and Uncertainty*, 51(1), 53–77. <https://doi.org/10.1007/s11166-015-9219-2>
- Viscusi, W. K., & Masterman, C. J. (2017). Income elasticities and global values of a statistical life. *Journal of Benefit-Cost Analysis*, 8(2), 226–250. <https://doi.org/10.1017/bca.2017.12>
- Wijnen, W., Weijermars, W., Schoeters, A., van den Berghe, W., Bauer, R., Carnis, L., Elvik, R., & Martensen, H. (2019). An analysis of official road crash cost estimates in European countries. *Safety Science*, 113, 318–327. <https://doi.org/10.1016/j.ssci.2018.12.004>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.