Economic valuation of preventing fatal and serious road injuries. Results of a Willingness-To-Pay study in four European countries

Schoeters, Annelies; Large, Maxime; Koning, Martin; Carnis, Laurent; Daniels, Stijn; Mignot, Dominique; Urmeew, Raschid; Wijnen, Wim; Bijleveld, Frits; van der Horst, Martijn

published in
Accident Analysis and Prevention
2022

DOI (link to publisher)
10.1016/j.aap.2022.106705

document version
Publisher's PDF, also known as Version of record

Accident Analysis and Prevention 173 (2022) 106705

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Economic valuation of preventing fatal and serious road injuries. Results of a Willingness-To-Pay study in four European countries

Annelies Schoeters[a,*], Maxime Large[b], Martin Koning[b], Laurent Carnis[c], Stijn Daniels[a], Dominique Mignot[b], Raschid Urmeew[d], Wim Wijnen[e], Frits Bijleveld[f], Martijn van der Horst[g]

[a] Vias Institute Chaussée de Haecht, Haachtsesteenweg 1405, 1130 Brussels, Belgium
[b] Université Gustave Eiffel, Campus de Lyon Cité des Mobilités 25, Avenue François Mitterrand, Case24, F-69675 Bron Cedex, France
[c] Université Gustave Eiffel, Campus de Marne-la-Vallée 5 Boulevard Descartes, Champs-sur-Marne, F-77454 Marne-la-Vallée Cedex 2, France
[d] Federal Highway Research Institute (BASt), Brüderstrasse 53, 51427 Bergisch Gladbach, Germany
[e] W2Economics Verlengde Hoogravenseweg 274, 3523 KJ, Utrecht, The Netherlands
[f] SWOV Institute for Road Safety Research Bezuidenhoutseweg 62 2594 AW Den Haag, The Netherlands
[g] KiM Netherlands Institute for Transport Policy Analysis, Bezuidenhoutseweg 20, 2596 AV The Hague, The Netherlands

ARTICLE INFO

Keywords:
Value
Cost
Time
Risk
Road injury
WTP

ABSTRACT

This paper presents the results of a stated choice study for estimating the Willingness-To-Pay of respondents in four European countries (Belgium, France, Germany and the Netherlands) to reduce the risk of fatal and serious injuries in road crashes. Respondents were confronted with hypothetical route choices that differ in respect of travel costs, travel time and crash risk. The survey was completed by 8,002 respondents, equally spread over the four participating countries and representative for each country with regards to gender, age and region. Possible biases caused by problematic choice behaviour such as inconsistent, irrational or lexicographic answers were addressed.

The resulting values were estimated by means of a mixed logit model allowing to account for the panel nature of the data. The Value of a Statistical Life (VSL) was estimated at 6.2 Mill EUR, the Value of a Statistical Serious Injury (VSSI) at 950,000 EUR, and the Value of Time (VoT) at 16.1 EUR/h. Consequently, the relative value of avoiding a fatal injury is estimated to be around 7 times higher than the value of an avoided serious injury. The study revealed differences between countries with France showing values that are significantly lower than the average and Germany showing values that are significantly higher. The estimated VSL values are considerably higher than the values currently used in the four countries, but they are within the range of values found in similar stated choice studies. The results can be used as an input in a broad range of socioeconomic studies including cost-benefit analysis and assessments of socioeconomic costs of road crashes.

1. Introduction

Estimates of socioeconomic road crash costs are frequently used in road policy-making. These costs represent the negative consequences of road crashes[1] on society and can be used as an input for budget allocation, to justify road safety investments or for cost-benefit analyses (CBA) of projects with road safety impacts. Socioeconomic road crash costs consist of economic costs (such as medical costs, production loss, property damage and administrative costs) and intangible costs or "human costs". The latter have no explicit market value since they represent the pain, grief and the loss of (quality of) life that is caused by (fatal) injuries due to road crashes (Bahamonde-Birke et al., 2015). Different methods exist to estimate the monetary value of human costs of which Willingness-To-Pay (WTP) is the most commonly used method when the values are used as an input for CBAs (Freeman et al., 2014). A WTP study estimates the amount a potential victim is willing to pay for a reduction of personal risk (Bahamonde-Birke et al., 2015). Most WTP studies are dedicated to calculating the Value of a Statistical Life (VSL)

* Corresponding author.
E-mail addresses: annelies.schoeters@vias.be (A. Schoeters), laurent.carnis@univ-eiffel.fr (L. Carnis), frits.bijleveld@swov.nl (F. Bijleveld), martijn.vander.horst@minienw.nl (M. van der Horst).

[1] In accordance with the recommendations by Stewart and Lord (2002) we use "crashes" instead of "accidents" because the term crash includes a larger variety of potential causes than accidents.

https://doi.org/10.1016/j.aap.2022.106705
Received 6 January 2022; Received in revised form 3 May 2022; Accepted 5 May 2022
Available online 22 May 2022

and thus the estimation of the costs of fatalities. Despite the increased importance of addressing serious injuries in road safety policy, only few WTP studies (a.o. O'Reilly et al., 1994; De Brabander, 2006; Persson, 2004) have addressed the monetization of the immaterial costs caused by injuries (Schoeters et al., 2020).

Cost estimates that are used in official guidelines are often outdated, not estimated using recommended methods or based on standardized figures such as those developed in Bickel et al. (2006). It has been recommended (Wijnen et al., 2019) to improve the quality and comparability of crash cost estimates. Since human costs of fatalities and serious injuries constitute a major share in the total crash costs, this study focusses on the estimation of human costs of fatalities and serious injuries. In accordance with the existing literature, it makes use of the WTP method.

Two methodologies are available to conduct a WTP study: in Revealed Preference (RP) methods WTP values are derived from people's actual behavior (e.g. the price paid for safety options in cars) and in Stated Preference (SP) methods surveys are used to ask people how much they are willing to pay for a hypothetical risk reduction (Boardman et al., 2017). While SP methods suffer from different types of bias due to the hypothetical and intentional nature of the survey, they have several advantages over RP methods. The main advantage is the broader applicability due to the flexibility of using questionnaires. Different types of road safety issues can be assessed simultaneously and the method is not dependent on the availability of data on the amount of money people actually pay for safety. Moreover, SP methods provide the opportunity to explain small risk reductions and test the respondents' risk understanding, while RP methods assume that individuals correctly understand the changes in (very small) risks associated with their choices (Boardman et al., 2017; Freeman et al., 2014; de Blaeij, 2003). In general there are two types of SP methods: while Contingent Valuation (CV) surveys ask directly how much respondents are willing to pay for a risk reduction compared to a reference situation, Stated Choice (SC) surveys use a more indirect way to elicit respondents' WTP by asking them to make choices in hypothetical situations. Comparisons of both methods in the literature have shown that SC is regarded as superior to CV, mainly because the indirect way of asking people's preferences reduces several types of bias (de Blaeij, 2003; Bahamonde-Birke et al., 2015; Rizzi & Ortúzar, 2006). Most SC surveys have been dedicated to estimating the Value of Time (VoT) and the application to estimate the VSL is relatively new. Bahamonde-Birke et al. (2015) have identified 13 SC studies that estimated the VSL in the context of road safety. These studies mention different challenges in the practical application of the SC method, such as the hypothetical character of the survey, difficulties in risk understanding and problematic choice behaviour (such as irrational, inconsistent or lexicographic answers), and propose approaches how to deal with these issues.

The main objectives to set up the current study were (a) to provide an update of the VSL based on current preferences in four European countries, (b) to include serious injuries in the estimation of human costs, (c) to use the best method available according to current knowledge, (d) to take stock of the knowledge developed in previous SC studies to create a robust design, and (e) to produce results that can be used for applications outside the setting of this study by applying a value transfer method.

This paper presents the results of a SC study conducted in four European countries (Belgium, France, Germany and the Netherlands) with the objective to estimate the WTP of respondents for reducing the risk of fatal and serious injuries in road crashes. This study provides updated estimates of the VSL (Value of a Statistical Life), the VSSI (Value of a Statistical Serious Injury) and the VoT (Value of Time) for each of the participating countries and common values for all countries combined. The rest of this paper proceeds as follows. The next section describes the empirical model that is used to analyse the data resulting from the SC survey. The third section describes the survey design, including the valuation context and the selection of choice sets. The fourth section explains how we dealt with problematic choice behaviour such as irrational and lexicographic answers. The fifth section describes the data collection process and presents descriptive statistics of the sample. The sixth section provides the resulting model estimates that are discussed in the seventh section. The final section presents our conclusions.

2. Empirical model

In a SC study respondents don't state the amount they are willing to pay directly, they have to indicate their preference by making choices in different hypothetical choice sets. Each choice set consists of two or more alternatives, that each consist of different attributes with varying attribute levels (Louviere et al., 2000). In our survey respondents were presented with 8 choice sets in a hypothetical route choice scenario[2]. Each choice set consisted of two alternatives that had four attributes: travel cost, two risk attributes representing the risk of having a fatal injury and the risk of having a serious injury, and travel time. The underlying empirical model is a Random Utility Model (RUM) which assumes that individuals maximize their utility when making choices (Thurstone, 1927; McFadden, 1974). By successively choosing the alternative that gives them the highest utility, respondents reveal their preferences and implicitly their WTP (Bahamonde-Birke et al., 2015).

2.1. Random utility model

In a RUM, utility is modelled as a function of preference weights and the levels of the attributes. The choices of a sample of respondents are aggregated to generate statistically reliable parameter estimates, which are used to determine the independent influence of the attributes on utility. These parameter estimates reflect the marginal disutility of a higher crash risk, a higher travel time and a higher travel cost. WTP values can consequently be estimated by dividing the parameter estimate for risk or time by the parameter estimate for travel cost, which is the marginal rate of substitution between income and risk or income and time (Louviere et al., 2000).

A RUM can be specified as (1) where V_{nj} represents the deterministic utility of individual n, which is a function of X_j, the attributes of route j and of Z_n, the individual characteristics of individual n. The error term ε_{nj} is unknown and randomly distributed.

$$U_{nj} = V_{nj}(X_j, Z_n) + \varepsilon_{nj} \tag{1}$$

In our study the RUM is specified as (2). The vector X_j includes the risk of being fatally injured (f_j), the risk of being seriously injured (I_j), the travel cost (c_j) and the travel time (t_j). Given that a linear specification is assumed, the different preference parameters (β, γ, θ and ρ) describe the marginal disutilities of the different travel attributes.

$$V_{nj} = \alpha + \beta f_j + \gamma I_j + \theta c_j + \rho t_j + \varphi Z_n \tag{2}$$

In the case where individual n would prefer route 1 over the alternative route 2, the probability (P_{1n}) of observing this choice is given by (3), where individual characteristics Z_n disappear because they remain constant across alternatives.

$$P_{1n} = \mathrm{P}(\beta(f_1 - f_2) + \gamma(I_1 - I_2) + \theta(c_1 - c_2) + \rho(t_1 - t_2) > \varepsilon_{n1} - \varepsilon_{n2}) \tag{3}$$

Assuming that the error terms are IID (independent and identically distributed) and Type 1 extreme value distributed, a simple Multinomial Logit (MNL) model can be determined by (4) where σ_n is a scale parameter normalized to 1 (Train, 2009).

$$P_{1n} = \frac{\exp(\sigma_n V_{n1})}{\exp(\sigma_n V_{n1}) + \exp(\sigma_n V_{n2})} = \frac{\exp(V_{n1})}{\exp(V_{n1}) + \exp(V_{n2})} \tag{4}$$

[2] Out of the 8 choice sets that were presented to respondents, only 7 could effectively be used for analysis. One choice set included a dominant alternative in order to test for irrational behavior.

Imagine that you have to make a trip of 50 km by **car** for a **leisure activity**. You make this trip **alone**, there are no passengers in the car. You can choose between two routes, both routes go over a **motorway** where there is **usually a lot of traffic**, but rarely traffic jams. On both routes there are **20 million cars per year**, this equals 55,000 cars per day.

All characteristics of these routes are similar, for example the driving comfort and the scenery. These routes only differ according to:

- The **costs** to make the trip (operating costs, fuel, toll, etc.).
  You have to pay yourself for these costs.
- The **time** to make the trip
- The **risk** you personally face to be hit by another driver:
  - The number of **fatal** car drivers per year.
  - The number of **seriously injured** car drivers per year.

    Someone who is seriously injured must be hospitalised for treatment. The injuries have short and/or long term consequences for daily functioning and are sometimes even life-threatening. For example concussion with loss of consciousness, (partial) amputations, skull fracture, open fractures, spinal cord contusion or severe organ injuries.

Fig. 1. Description of the choice context.

WTP values are then calculated as the ratio between the non-monetary coefficients (β, γ and ρ) and the cost coefficient (θ) (5). In the choice experiment the risk values are presented in absolute values instead of probabilities. Therefore it is necessary to multiply the average WTP values of avoiding one fatality or one serious injury by the number of trips that is considered in the choice experiment (20 million) in order to define the VSL and VSSI values.

$$\begin{aligned} VSL &= 20Million \times \frac{\beta}{\theta} \\ VSSI &= 20Million \times \frac{\gamma}{\theta} \\ VoT &= 60 \times \frac{\rho}{\theta} \end{aligned} \tag{5}$$

2.2. Considering preference heterogeneity

The simple MNL model assumes that all observations are independent and that preferences are homogenous between individuals, however this is not a valid assumption for SC data where one respondent is responsible for multiple observations (Iraguën & Ortúzar, 2004). An alternative is the Mixed Logit (ML) model in which heterogeneity in preferences is considered. ML models allow for taste variations between respondents by adding an error component, specific for each respondent, to the coefficient of the preference parameters. With regards to the marginal disutility of the risk of being fatally injured, the parameter can be described as (6) where $\overline{\beta}$ is the constant part of the marginal disutility of being fatally injured and ε_n^{β} is the random component that differs among the N individuals.

$$\beta_n = \overline{\beta} + \varepsilon_n^{\beta} \tag{6}$$

By using random parameters, for which a distribution (e.g., normal, log-normal, zero-bounded triangular) needs to be specified, the ML model allows for correlation among the choices by the same respondent. In this way the panel structure of the data is taken into account and dependencies in the error terms of the RUM are allowed (ε_{nj} in (1)) (Train, 2009; Sillano & Ortúzar, 2005). The probability function of the ML model is presented in (7) and shows that the probability that individual n would prefer route 1 in choice t is conditional on the random parameter (β).

$$P_{1nt}(\beta) = \frac{\exp(V_{n1t})}{\exp(V_{n1t}) + \exp(V_{n2t})} \tag{7}$$

The unconditional choice probability P_{1nt} is calculated in (8) by integrating over the distribution $f(\beta)$.

$$P_{1nt} = \int P_{1nt}(\beta) f(\beta) d\beta \tag{8}$$

3. Survey design

The design of the survey was based on a literature review of SC studies that were carried out in the field of transport safety, and was further improved during discussions between experts from the participating countries. Feedback was received from researchers from TU Dresden who implemented a similar pilot study in 2018 (Obermeyer & Hirte, 2020). Early 2020 focus groups were organized in Belgium, France and Germany to test the wording and understanding of the questions[3]. >30 people with different sociodemographic backgrounds participated. Subsequently a pilot survey was conducted with a representative sample of 400 respondents, 100 per country. Both the feedback from the focus groups and the analysis of the pilot survey data showed potential weaknesses which lead to a revision of the survey. The questionnaire was initially developed in English and in a final stage translated to German (DE), French (FR), French (BE), Dutch (BE) and Dutch (NL). A thorough comparison of the different language versions was done to ensure that all questions would be interpreted in the same way.

Apart from the choice experiment, the questionnaire included questions regarding sociodemographic variables, travel behaviour, experience with road crashes, attitudes and opinions towards risk-taking, spending money and dealing with time pressure, the credibility of the SC exercise, the perception of road safety on motorways and the potential influence of the COVID-19 pandemic.

3.1. Valuation context

A crucial part of a SC study is the design of the valuation scenario in which the respondents have to make hypothetical choices. According to Pearce & Özdemiroglu (2002) the validity of SC studies depends highly on the credibility and realism of the valuation scenarios. A description of the choice context that was presented to respondents in our survey is illustrated in Fig. 1. As in most SC studies that are conducted in the field

[3] At that time the Dutch partner did not join the project yet, so there were no focus groups organized in the Netherlands. The pilot survey was conducted in all four countries.

Table 1
Trip motives presented in the choice scenarios.

Trip motive	% of respondents
Going to work	17%
Going to school	5%
Leisure activities	17%
Dropping someone off/picking someone up	10%
Running errands/services	17%
Visiting someone	17%
Vacation	16%

of road safety (a.o. Rizzi & Ortúzar, 2003; Henscher et al., 2009; Veisten et al., 2013; González et al., 2018), a route choice scenario was used.

While the purpose of the study is to determine general WTP values, it appeared to be difficult to design a valuation context that was applicable to all road users. Previous SC studies that used a route choice scenario were mostly designed from the perspective of car drivers. A Norwegian study (Flügel et al., 2019) compared the WTP values between bus passengers and car drivers and found no significant differences. For our study a context of a car driver on a motorway was chosen. The advantage of a car driving scenario is that it's generally familiar to most road users in the participating countries and that real world payment vehicles such as toll and operating costs exist. To decrease the hypothetical character of the choice experiment, only respondents that had driven on a motorway at least once during the past 12 months were selected to participate in the survey.

Respondents were presented with 8 choice sets in a hypothetical route choice scenario. Each choice set consisted of two alternatives that both covered a 50 km trip on a motorway and only differed with respect to four attributes: the travel cost, two risk attributes and travel time. To increase the realism of the choice scenario, a trip motive was added based on a previous question in which respondents had to indicate their two most frequent trip motives when driving a car on a motorway (Table 1). The trip motive that was presented in the choice scenarios only differed over respondents, but remained the same for both alternatives and for all 8 choice sets. The motives"other" and "professional trips" were not used since "other" was meaningless to program in the text, and for "professional trips" some respondents may not pay themselves, which is an important prerequisite for a WTP study. In these cases the trip motive was replaced with another one from the list of available options. If there was no suitable alternative, the motive "leisure activity" was used.

3.1.1. *Presentation of risk attributes*

A challenge for WTP studies in general is to deal with the fact that people have problems understanding small risks (de Blaeij et al., 2003). When risks are explicitly expressed as probabilities, it is likely that respondents cannot interpret such risks correctly. Therefore, most SC surveys use absolute values instead of probabilities (Rizzi & Ortúzar, 2003). However to be able to calculate VSL and VSSI, it's necessary to have the probability of a fatal (or serious injury), i.e. the absolute numbers related to an exposure variable. Previous studies have calculated the actual risk (probability of dying) afterwards by making assumptions about the traffic volume on the roads that were presented in the scenario (Henscher et al., 2009) but do not include this in the scenario that is presented to the respondent. Obermeyer & Hirte (2020, p.7) argued that "people should at least be informed about the objective level of risk, even if the concept is difficult for some people to understand" and included therefore both the absolute number of victims per year and the probability (one victim per number of trips) in their SC survey.

In our study the risk is presented in absolute values as the number of fatally (or seriously) injured car drivers per year. Information about the volume of the total traffic flow (20 million vehicles per year) was presented as well, so that respondents were correctly informed about the objective risk level (number of fatalities or serious injuries per year divided by the annual traffic flow) of the routes. The number of trips is an estimated average of the real traffic flow on motorways in the four participating countries. This estimation is based on the length of motorways (Eurostat, 2020a) and the number of kilometres driven by vehicles on motorways per year (OECD, 2020). The traffic situation was described as "usually a lot of traffic, but rarely traffic jams" and the traffic flow per day was mentioned. In that way respondents have intuitively a more or less correct idea about the number of trips.

Some studies show that the preference for safety increases when there is a passenger (Rizzi & Ortúzar, 2003). Therefore, in our survey it is clearly stated that a respondent has to make the trip alone, to make sure that they will only take the WTP for their personal risk reduction into account. Furthermore the attribute description emphasizes that it concerns a risk "you personally face to be hit by another driver" to avoid that respondents think they can control the risk by driving more carefully.

3.1.2. *Definition of seriously injured victims*

Serious injuries can be defined in several ways, for example by hospital admission or by injury severity levels. The use of different definitions in European countries was identified as one of the potential causes for the large variation in crash cost estimates (Schoeters, et al., 2020). In our survey seriously injured victims are defined as MAIS3 + victims, according to the definition that the European Commission established in 2013. This definition includes (hospitalized) traffic victims with injuries that have a score of 3 or more on the Abbreviated Injury Scale (AIS) (European Commission, 2013)[4].

To our knowledge, there exists no official definition that gives specific information or examples about the impact and the consequences of these injuries for the MAIS3 + victims. However, a correct understanding by respondents of what the risk attributes entail is an important feature of WTP studies. Therefore we developed an operational definition of serious injuries that was presented to respondents. An examination of the population of MAIS3 + victims (Nuyttens & Van Belleghem, 2014) showed that this population is very diverse regarding their injuries and the consequences for their daily functioning. Since most of the MAIS3 + victims are MAIS3, the probability of death is not extremely high and should not be over-emphasized.

The operational definition of serious injuries therefore included:

- Hospitalisation;
- Impact: the injuries have short and/or long-term consequences for daily functioning;
- Probability of death: the injuries are sometimes life-threatening.

Some examples of injuries are added to the definition (see Fig. 1), which are derived from the AAAM/EC conversion table for ICD10 to AIS (AIS3 +) (Loftis et al., 2016).

3.1.3. *Other attributes*

In order to calculate WTP values, a SC scenario should minimally include a payment vehicle. According to Pearce & Özdemiroglu (2002) appropriate payment vehicles are credible, relevant, acceptable and coercive. Different route choice surveys have used a road toll, however, except in France, most respondents in the participating countries are not familiar with tolls and this could provoke some aversion which could

[4] The AIS scores are developed by the Association for the Advancement of Automotive Medicine (AAAM) and the scale is "an anatomically based, consensus derived, global severity scoring system that classifies an individual injury by body region according to its relative severity on a 6-point scale (1=minor and 6=maximal). The MAIS is the highest (i.e. most severe) AIS code in a patient with multiple injuries." (AAAM, 2020). The scores are determined by a group of experts and are mainly based on the probability of death, but also take consequences of the injuries into account such as permanent impairment, treatment period and energy dissipation (MacKenzie et al., 1988).

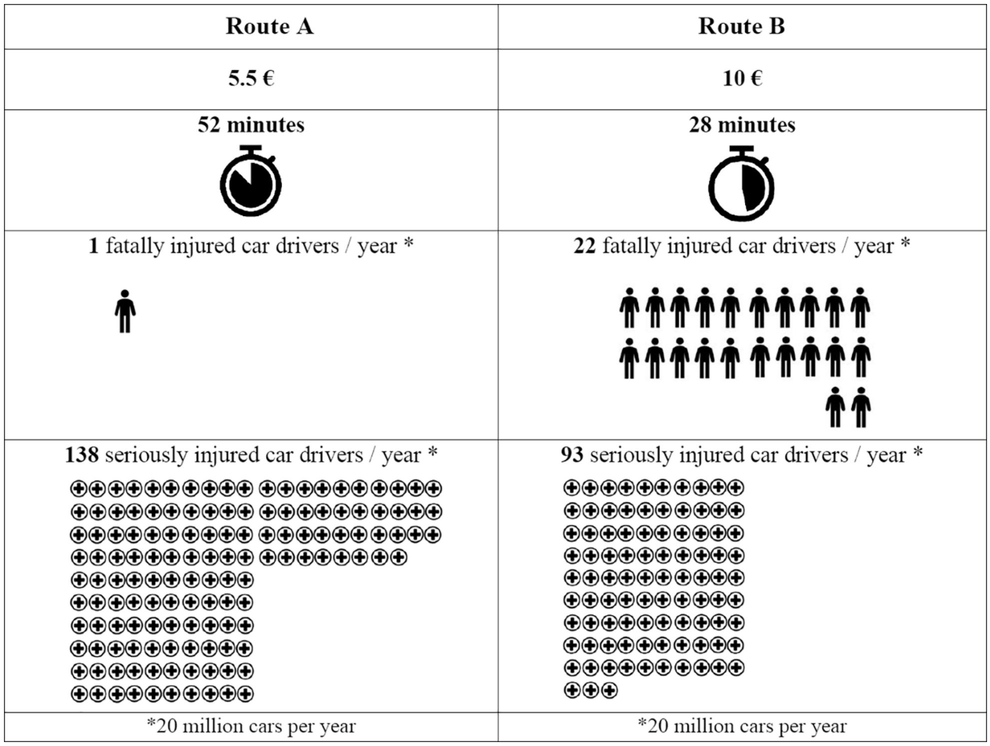

Fig. 2. Example of a choice scenario.

Table 2
Attribute levels, prior parameter estimates and corresponding expected variation of utility per level difference.

	Level 1	Level 2	Level 3	Level 4	Prior estimate	Expected variation in utility
Travel cost	5.50	7	8.50	10	/	1.50 EUR
Travel time	28	36	44	52	11.50 EUR/ hour	1.53 EUR
Number of fatally injured car drivers	1	8	15	22	4.586 Mill EUR	1.60 EUR
Number of seriously injured car drivers	3	48	93	138	0.688 Mill EUR	1.55 EUR

potentially lead to strategic behavior (Hess & Rose, 2009). Ultimately the payment vehicle is defined in a more general way as "the costs to make the trip (operation costs, fuel, toll, etc.)" and it is specified that respondents have to pay themselves for the travel costs. Furthermore travel time was included as a fourth attribute since it's mostly an important variable when choosing a route and would thus increase the realism of the choice experiment. Moreover, including travel time enables us to calculate the Value of Time (VoT) as well.

3.1.4. Visualisation

Studies have shown that visualisation can help people understand small changes in the attribute levels which results in more reliable estimates (Lindhjem et al., 2010; Wiktor et al., 1998). During the focus groups we experienced that respondents had difficulties to correctly grasp the differences in the attribute levels, especially for the time and risk attributes. Differences in travel cost were easier to understand since people in real life are more familiar with comparing costs than with comparing the levels of the other attributes. For that reason different symbols were added in the choice scenarios to visualize the levels of travel time, the risk of a fatal injury and the risk of a serious injury, as shown in Fig. 2.

3.2. Experimental design

The choice sets in our survey are composed of two alternative routes that have four attributes, each attribute can have four different levels. In total there are 256 possible routes and 32,768 combinations of two different routes[5]. A selection of these combinations was done by applying an experimental design. In our survey a Bayesian D-efficient design was created using the "idefix" package (Traets et al., 2020) in R 4.0.1. The experimental design consisted of 14 choice sets, divided over two groups of 7 sets (see Appendix). The groups of choice sets were randomly assigned to respondents. In D-efficient designs choice sets are selected that have the lowest D-error which is a measure for the inefficiency of the design. The D-error is based on the determinant of the variance–covariance matrix, and requires to have prior information on the parameter estimates. The prior estimates are based on a literature review and presented in Table 2. Based on these parameter estimates, choice sets were selected that force respondents to make trade-offs and that maximize the information that is gained from every choice (Traets et al., 2020).

3.2.1. Prior parameter estimates

The prior estimate for the risk of a fatal injury was based on recommendations made by OECD (2012) about the use of VSL estimates in CBAs. Based on a *meta*-analysis of SP studies, OECD recommends to use 3,614,506 USD (2005) as a base value to calculate national VSL estimates in EU-27 countries. Applying the formula for value transfer that is proposed by OECD using data on GDP per capita and Purchasing Power Parities (Eurostat, 2020b) and an income elasticity of 0.8, the VSL was

[5] 256 possible routes A multiplied by 256 possible routes B divided by 2 equals 32,768.

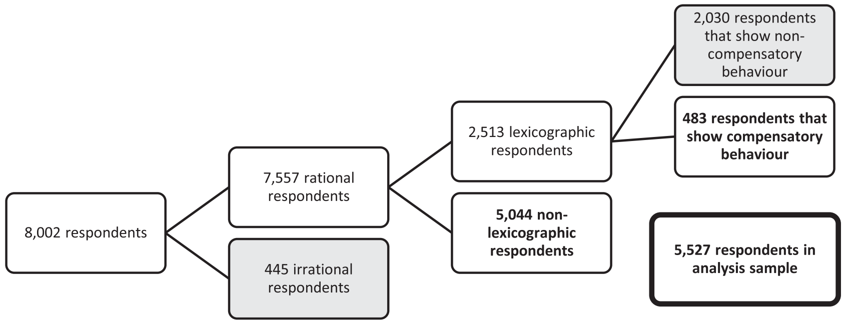

Fig. 3. Selection of the sample for the empirical analysis.

estimated for each of the participating countries. This lead to an average VSL estimate for the four countries of 4.586 Million EUR (2019). Subsequently the prior estimate for the risk of a serious injury was defined by looking at the ratio of VSL and VSSI in other studies. Values presented in the European HEATCO project (Bickel et al., 2006), the European 'Handbook of external costs of transport' (Korzhenevych et al., 2014) and a study in the United States (Blincoe et al., 2015) varied around 15%. Applying this ratio to the prior VSL estimate results in a VSSI estimate of 688,000 EUR (2019).

The prior parameter estimate for travel time was based on values from the HEATCO project (Bickel et al., 2006). In this project a *meta*-analysis was conducted to determine a standard VoT for European countries, which were also used in the subsequent 'Handbook on estimation of external costs in the transport sector' (Maibach et al., 2007). The unweighted mean of the VoTs of long and short distance trips, as well as commuting and leisure trips per country is used. Calculating a weighted mean was not possible due to the lack of data on distance travelled or number of trips by trip length and travel motive. Subsequently the standard values are updated to income and price level 2020 using GDP per capita (Eurostat, 2020b) and an income elasticity of 0.7 as recommended by HEATCO. The resulting average VoT for the four countries is 11.5 EUR per hour or 0.19 EUR per minute (2020).

3.2.2. *Attribute levels*

In our survey each attribute can have four different levels which are presented in Table 2. While some SC studies (Veisten et al., 2013; Niroomand & Jenkins, 2016; González et al., 2018) have pivoted their choices on actual travel behavior by asking about a recent trip, this approach is not used in our survey since there is a risk that these designs induce respondents to exhibit inertia or non-trading behaviour which causes model estimation problems (Hess & Rose, 2009), as was the case for the SC study by González et al. (2018). According to Pearce & Özdemiroglu (2002) the attribute levels in a SC survey should include both the current situation and realistic levels above and below the current levels. However, in the case of road safety, actual variation of risk levels is too small to include them with enough variation to encourage respondents to make a trade-off. Since respondents mostly don't know the actual risk figures, presenting a higher number of fatalities and serious injuries would not influence the realism of the experiment. Therefore only the lowest risk attribute levels are based on realistic numbers of fatalities and serious injuries and the higher attribute levels are chosen in order to promote trade-offs. This means that the lowest level of fatalities (which is 1) and the lowest level of serious injuries (which is 3) are based on an average of real crash figures of car drivers per 50 km of motorways in the participating countries (European Commission, 2020).

Respondents do however have a realistic idea about the actual time and – to a lesser extent – cost of a certain trip. For these attributes an estimation of the actual values is used. We looked into gas prices in the four countries (ANWB, 2020) and the average gas consumption for a 50 km car trip on a motorway (Milieu Centraal, 2020) and estimated that fuel costs equal to 5.25 EUR. We added 0.25 EUR for operating costs. To estimate a realistic value for travel time we made the assumption that the average speed of a car on a trip on a motorway without congestion is 100 km/h. A trip of 50 km would then result in a travel time of 30 min. As both travel costs and time can easily increase during a trip, our estimates for the travel cost and time are used as the lowest attribute levels.

The other attribute levels were determined using the prior parameter estimates that were calculated for the experimental design (see Section 3.2.1). To increase the likelihood that respondents will make a trade-off by considering all four attributes, the differences between the attribute levels reflect a variation in the expected utility that is similar for all attributes. This means that an increase of 8 min travel time generates a similar decrease in utility as an increase of 1.5 EUR travel costs or an increase of 7 fatally injured car drivers or 45 seriously injured car drivers (for a total traffic flow of 20 million trips).

4. Dealing with problematic choice behaviour

Models that are used to analyse choice data assume utility maximizing respondents who are fully informed and are able to assign utility levels to alternatives and choose the alternative with the highest utility (Sælensminde, 2006). Choice behaviour that is not in line with these assumptions has been a source of criticism for SC studies (Elvik, 2016). The occurrence of problematic choice behaviour is however a common problem in SC studies. When individuals face complex choice situations, in which the exercise of utility maximization is difficult, they tend to resort to "heuristics". These are mental strategies aimed to reduce the complexity of a decision process (Kahneman & Tversky, 1972). In our study we have analysed the impact of three types of heuristics, i.e. inconsistent behaviour, irrational behaviour and lexicographic behaviour. Irrational respondents and lexicographic respondents that show non-compensatory behaviour are identified and removed from the sample (Fig. 3). The impact of inconsistent behaviour will be discussed in Section 6.2.

4.1. *Irrational behaviour*

Respondents that made irrational choices were identified by means of an extra choice set in which one of the route alternatives was clearly superior over the other one. This approach was recommended by Pearce & Özdemiroglu (2002) and conducted by Burge & Rohr (2004). In the sample, 445 out of a total of 8,002 respondents (5.6%) chose the inferior route and were considered as showing irrational behaviour. Since random utility models assume rational decision making, these respondents could bias the results (Hess et al., 2010) and were excluded from the sample. The dominated choice situation was removed from the estimation data for the other respondents.

Table 3
Descriptive statistics of the base sample and the analysis sample.

Variables	Base sample (8,002 respondents)	Analysis sample (5,527 respondents)	Difference
Country			
Belgium	25%	24.8%	−0.2%
France	25%	25.1%	+0.1%
Germany	25%	26%	+1%
Netherlands	25%	24.1%	−0.9%
Gender			
Female	49.1%	48.5%	−0.6%
Age			
18–19	1.7%	1.7%	0%
20–29	15.2%	15.1%	−0.1%
30–39	17.1%	16.6%	−0.5%
40–49	15.3%	15.3%	0%
50–59	17.8%	17.5%	−0.3%
60–69	20.8%	21%	+0.2%
> 69	12.1%	12.8%	+0.7%
Average income per capita	1760.3 EUR	1762.3 EUR	+2.0 EUR (+0.1%)
Professional occupation			
White collar or office employee	29.5%	29.7%	+0.2%
Blue collar or manual worker	15.1%	14.8%	−0.3%
Executive	10.7%	10.4%	−0.3%
Self-employed/ independent professional	6.3%	6.3%	0%
Currently no professional occupation	38.3%	38.8%	+0.5%
Number of km travelled by car over the last year			
< 10,000 km	43.1%	43.9%	+0.8%
Between 10,000 km and 20,000 km	40.5%	40.3%	−0.2%
Between 20,001 km and 30,0000 km	11.2%	10.8%	−0.4%
> 30,000 km	5.2%	5%	−0.2%

4.2. Lexicographic behaviour

Subsequently 2,513 out of 7,557 respondents (that did not show irrational behaviour) have been identified as showing lexicographic behaviour. These are respondents that systematically evaluate the alternatives on the basis of the same attribute. Their share (33.3%) is comparable to the percentage found in other SC studies (Veisten et al., 2013; Hojman et al., 2005; Iraguën & Ortúzar, 2004). The number of lexicographic respondents is not equally distributed over the attributes: 991 respondents (13.1%) always chose the alternative with the lowest travel time, 548 respondents (7.3%) always chose the cheapest alternative, 705 respondents (9.3%) always chose the alternative with the lowest number of fatalities and 269 respondents (3.6%) always chose the alternative with the lowest number of serious injuries. The total share of lexicographic respondents was similar in all four countries.

For that reason most SC studies exclude all respondents that show lexicographic behaviour from the analysis. However, some authors state that these respondents are not always behaving in a non-compensatory way (Sælensminde, 2006; Elvik, 2016; Hess et al., 2010). While lexicographic choice patterns can be the result of simplification, to reduce the mental effort of the choice task, other causes have been identified in the literature. Lexicographic choices can result from strategic behaviour, which is especially prominent when politically sensitive attributes such as a road toll are presented. These respondents hope to influence political decisions by their responses. Like respondents who use simplification, they are also not making a trade-off between the attributes. Besides lexicographic behaviour can also be caused by the experimental design. Respondents with an extreme preference for a certain attribute show apparent lexicographic behaviour simply because their individual-specific thresholds are not presented in the choice sets. For these respondents the attribute level differences are not adapted to show the upper or lower boundary of their valuations of the attributes. As opposed to respondents who use simplification or strategic behaviour, respondents with extreme preferences are behaving in a compensatory way (Hess et al., 2010).

Very few SC studies have investigated the motivation of lexicographic respondents. Elvik (2016) advises however to look into these motivations, because simply excluding all lexicographic respondents could bias the results. In our study we applied two approaches to distinguish the lexicographic respondents that are behaving in a compensatory way and those that are not. Only the latter group is excluded from the sample.

Firstly lexicographic respondents were asked an additional question about the reason for their lexicographic choices. Out of 2,513 lexicographic respondents, 370 respondents (15%) indicated that they used a simplification strategy. Additionally 1,203 respondents (48%) chose the option "I would always take the [safest] route, even if the [cost or travel time] would be (very) high" which indicates that they are unwilling to make a trade-off. Both groups are not showing utility-maximizing behaviour and are excluded from the sample.

Secondly the motivation of the group of lexicographic respondents that indicate to have an extreme preference for one of the attributes (768) is tested by means of a comparison with a CV question. This test that is based on Sælensminde (2006) and recommended by Elvik (2016), entails a comparison of respondents' individual WTP values based on a SC survey and the values based on CV questions. Two closed CV

Table 4
Results of the MNL and ML models, applied to all respondents and to the country-specific samples.

	MNL - All	ML - All	ML - Belgium	ML - France	ML - Germany	ML - The Netherlands
Sample size	5,527	5,527	1,368	1,385	1,441	1,333
Cost	−0.148***	−0.240***	−0.241***	−0.263***	−0.223***	−0.232***
Time	−0.042***	−0.065***	−0.069***	−0.057***	−0.071***	−0.064***
Fatality	−0.047***	−0.074***	−0.072***	−0.071***	−0.082***	−0.073***
Serious injury	−0.008***	−0.011***	−0.011***	−0.011***	−0.012***	−0.011***
VSL (M€)	6.29	6.19	5.94	5.35	7.35	6.29
[Min, Median, Max]		*[3.21, 6.20, 8.98]*	*[3.12, 5.92, 8.44]*	*[3.07, 5.38, 7.69]*	*[3.78, 7.34, 10.44]*	*[3.38, 6.35, 8.95]*
St. deviation		*0.86*	*0.79*	*0.76*	*1.04*	*0.87*
VSSI (M€)	1.02	0.95	0.94	0.83	1.10	0.98
[Min, Median, Max]		*[0.43, 0.95, 1.37]*	*[0.48, 0.94, 1.33]*	*[0.42, 0.83, 1.16]*	*[0.55, 1.09, 1.56]*	*[0.48, 0.97, 1.37]*
St. deviation		*0.19*	*0.19*	*0.19*	*0.23*	*0.19*
VoT (€/h)	17.0	16.1	17.2	12.9	19.0	16.4
[Min, Median, Max]		*[-6.8, 14.7, 36.6]*	*[-7.1, 16.3, 37.3]*	*[-4.4, 12.2, 29.5]*	*[-8.6, 17.2, 41.9]*	*[-5.8, 14.6, 35.9]*
St. deviation		*12.9*	*10.9*	*8.1*	*12.4*	*10.0*
Adjusted AIC	51,619	48,734	12,029	12,433	12,396	11,788
Adjusted Log-likelihood	−25,805	−24,362	−6,010	−6,212	−6,193	−5,889

Significance levels: *** < 0.001, ** < 0.01, * < 0.05.

Table 5
Robustness tests of the ML model.

	Benchmark ML model	ML model including lexicographic respondents	ML model excluding potential hypothetical bias	ML model excluding inconsistent respondents	ML model excluding potential free-riders
Sample size	5,527	7,557	3,619	4,390	4,816
Cost	−0.240***	−0.291***	−0.220***	−0.415***	−0.253***
Time	−0.065***	−0.085***	−0.060***	−0.111***	−0.066***
Risk of death	−0.074***	−0.092***	−0.070***	−0.123***	−0.076***
Risk of injury	−0.011***	−0.012***	−0.010***	−0.019***	−0.012***
VSL (M€)	6.19	6.35	6.40	5.91	6.06
[Min, Median, Max]	*[3.21, 6.20, 8.98]*	*[2.44, 6.39, 9.45]*	*[3.34, 6.40, 9.08]*	*[2.51, 5.90, 8.96]*	*[3.49, 6.05, 8.57]*
St. deviation	*0.86*	*1.06*	*0.88*	*1.08*	0.87
VSSI (M€)	0.95	0.81	0.97	0.92	0.94
[Min, Median, Max]	*[0.43, 0.95, 1.37]*	*[0.38, 0.80, 1.22]*	*[0.49, 0.97, 1.36]*	*[0.34, 0.91, 1.37]*	*[0.45, 0.93, 1.33]*
St. deviation	*0.19*	*0.17*	*0.18*	*0.23*	0.19
VoT (€/h)	16.1	16.9	16.3	15.9	15.7
[Min, Median, Max]	*[-6.8, 14.7, 36.6]*	*[-12.5, 14.1, 42.0]*	*[-8.3, 14.6, 38.4]*	*[-4.7, 15.0, 32.1]*	*[-6.4, 14.0, 35.2]*
St. deviation	*12.9*	*12.9*	*11.2*	*9.1*	*10.0*
Adjusted AIC	48,734	46,880	48,924	47,350	48,629
Adjusted Log-likelihood	−24,362	−23,436	−24,455	−23,668	−24,309

Significance levels: *** < 0.001, ** < 0.01, * < 0.05.

Table 6
Responses to additional questions regarding the impact of the COVID-19 pandemic.

Do you have the feeling that road safety in [country] has decreased, increased or has remained the same in comparison with the period before the outbreak of COVID-19?	
Decreased	16%
Increased	23%
Remained the same	53%
I don't know	8%
Has road safety become more, less or equally important to you personally since the outbreak of COVID-19?	
More important	11%
Less important	4%
Equally important	80%
I don't know	5%

Table 7
VSL, VSSI and VoT estimates.

	VSL (Mill EUR)	VSSI (Mill EUR)	VoT (EUR/h)
Four countries	6.19	0.95	16.1
Belgium	5.94	0.94	17.2
France	5.35	0.83	12.9
Germany	7.35	1.10	19.0
The Netherlands	6.29	0.98	16.4

questions were added to the survey from which WTP values are calculated for the remaining 768 lexicographic respondents. If their choice pattern is the result of an extreme preference, we expect their individual WTP values to be above the average WTP of the group that did not answer lexicographically towards that specific attribute. Out of 343 respondents that systematically chose the fastest route, 244 were excluded from the sample because their VoT based on the CV question was lower than the average of the respondents that did not answer lexicographically towards time. The same was done for 59 out of 178 respondents that always chose the alternative with the lowest number of fatalities.

By means of these questions, we were able to identify 483 respondents, out of 2,513 respondents that were initially identified as showing lexicographic behaviour, that appeared to be motivated by extreme preferences and were behaving in a compensatory way. These respondents were kept in the sample. Ultimately the sample that is used for the analysis consisted of 5,527 respondents that show rational, utility-maximizing behaviour[6].

5. Data collection and descriptive statistics

The sample for our study was drawn from an internet panel by an external panel provider, Profacts. For each country the panel consisted of 100,000 or more possible respondents, aged 18 years or older. The sample was drawn based on a simple random probability sampling method, taking into account different criteria for representativeness. By assigning quotas prior to the further selection the raw sample was representative with regards to age, gender and region[7]. The survey was conducted between 22nd of October and 13th of November 2020 simultaneously in the four participating countries. The average duration to fill in the questionnaire was 14 min and 8 s (median 10 min and 12 s).

The final sample consists of 8,002 respondents including 2,005 Belgian, 2,000 French, 2,000 German and 1,997 Dutch respondents. By means of a selection question at the start of the survey only respondents that had driven a car on a motorway at least once in the past 12 months were included in the final sample. Even though the quotas were assigned to the sample prior to the further selection based on the selection question, the final sample was still representative in the respective countries with regards to gender and age, and with regards to respondents' geographic origin.

[6] The ultimate sample size per country varied between 1,333 (Netherlands) and 1,441 (Germany). According to Orme (2010), a sample size between 200 and 300 respondents is sufficient to conduct robust discrete choice analysis. More formally, according to Johnson & Orme (2003) the minimum sample size can be calculated as 500*c/(t*a) where c is the largest number of attribute levels (which is 4 in our study), t is the number of choice tasks per respondent (7 in our study) and a is the number of alternatives per choice task (2 in our study). Applying this formula, we can conclude that a sample of around 150 respondents would have been sufficient for our design.

[7] These quotas include 12 categories in which gender (male, female) and six age categories (18–24, 25–34, 35–44, 45–54, 55–64, 65+) are crossed. Additionally soft quota were assigned to the regions in the different countries. The quotas are based on the most recent statistics provided by the United Nations (2018) or a national source (Statbel, 2019; INSEE, 2019; CBS Statline, 2019; Genesis Census, 2011).

Based on the identification of problematic choice behaviour discussed in the previous section, 445 respondents were removed from the sample for the analysis because they showed irrational behaviour and 2,030 were removed because they showed non-compensatory lexicographic behaviour. Consequently the sample that was used in the analysis consisted of 5,527 respondents. A comparison of the sociodemographic composition of the base sample and the analysis sample is presented in Table 3 shows that both samples only differ very slightly with regards to country, age, gender, income and occupation. Also respondents' car travel behaviour, indicated by the reported number of kilometer driven in a person car during the last year, was similar in both samples.

6. Results

6.1. Main results

Different econometric models have been estimated using the "mlogit" package (Croissant, 2020) in R 4.0.1. The results of two models, MNL and ML are presented in Table 4. In the simple MNL model all coefficients are statistically significant at the 1% level and, as they represent marginal utilities associated with additional travel time, travel cost and injury risks, they all have a negative sign. WTP values are calculated as the ratio between the non-monetary coefficients and the cost coefficient. For the time attribute this gives a corresponding mean VoT of 0.28 EUR/minute or 17.0 EUR/hour. Since the risk values are presented in absolute values instead of probabilities, it is necessary to multiply the average WTP values of avoiding one fatality or one serious injury by the number of trips that is considered in the choice experiment in order to define the VSL and VSSI values. For each country the same number of trips was used (20 million) which was based on the average of the real traffic flow on motorways in the four participating countries. According to the MNL model the mean VSL is 6.3 Million EUR and the mean VSSI is about 1 Million EUR.

Secondly preference heterogeneity has been taken into account by estimating a ML model[8] in which an error term is introduced in the coefficients. The distribution of the error terms in the risk coefficients is defined by a triangular distribution bounded by zero[9] and the error term in the time coefficient is defined by a normal distribution[10]. The cost coefficient is assumed to be non-random[11], in order to avoid having positive marginal utilities and to simplify the calculation of the WTP values. The results show that all coefficients remain negative and significant at the 1% level. Both the AIC criterion and the log-likelihood improve significantly compared to the MNL model which indicates a better fit of the ML model. The WTP values decrease slightly compared to the MNL model. The average VSL is estimated at 6.19 Million EUR, with a confidence interval between 3.21 and 8,98 Million EUR, the average VSSI is estimated at 0.95 Million EUR, with a confidence interval between 0.43 and 1.37 Million EUR and the average VoT is estimated at 16.1 EUR/hour, with a confidence interval between −6.8 and 36.6 EUR/hour. The negative values in the VoT can be explained by the fact that a normal distribution was applied to the error term of the time attribute.

The same ML model was applied to the specific samples of each participating country in order to elicit national VSL, VSSI and VoT values. Based on a Wilcoxon-Mann-Whitney test we can conclude that the distributions of the WTP values differ significantly between the four countries. German respondents have significantly higher WTP values than respondents in the other three countries: the VSL is 19% higher than the average, the VSSI is 16% higher and the VoT is 18% higher. French respondents on the other hand have significantly lower WTP values than respondents in the other three countries. Their VSL is 14% lower than the average, the VSSI is 13% lower and the VoT is 20% lower.

6.2. Robustness tests

Several robustness tests were performed on the ML model and the results are presented in Table 5. The application of the ML model to the sample of 7,557 respondents *including* lexicographic respondents yields results that are very similar to those of the sample omitting lexicographic respondents. A Wilcoxon-Mann-Whitney test showed that the distribution of individual values of both groups is not significantly different. While some differences can be observed, they remain limited. A second test investigated the potential impact of hypothetical bias on the results. After they assessed the different hypothetical choices in the survey, respondents were asked if they thought parts of the presented choice situations were not realistic. 1,908 out of 5,527 respondents (34.5%) of the analysis sample considered (parts of) the scenarios as unrealistic. Excluding these respondents from the sample appears to have a minor impact on the results, with a very moderate increase in WTP values.

The impact of inconsistent behaviour was also analysed. Inconsistent behaviour is defined as respondents that make choices that are inconsistent with previous or subsequent choices and can be problematic since it constitutes a violation of the transitivity axiom (Sælensminde, 2002). In the benchmark sample 1,137 out of a total of 5,527 respondents (20.6%) were identified as showing inconsistent choice behaviour[12]. Since it was impossible to distinguish whether these mistakes result from real inconsistent behaviour or from learning effects, we decided not to exclude these respondents from the sample for the analysis. A comparison between the benchmark sample and the sample excluding these respondents shows no significantly different results.

Finally the impact of a potential free-riding strategy is examined. The survey contained a question about who normally pays the costs of the respondents' car trips. In order to get reliable WTP estimates, it's important that respondents believe they have to pay for the risk or time reduction themselves. While it has been emphasized in the description of the choice scenarios that the travel costs are borne by the respondent, respondents who don't have the experience of paying for car trips themselves might have difficulties empathizing with this hypothetical situation. Excluding these 711 respondents from the sample did not have an impact on the results either.

[8] Every ML model is modelled using 500 Halton draws.

[9] This distribution is bounded by zero on the right and twice the mean on the left, and does not allow to have positive coefficients. This distribution was preferred to the log-normal distribution because, since the latter is bounded on the left but not on the right, the mean coefficients obtained were pulled upwards by the extreme values.

[10] For the time coefficient the normal distribution was preferred over the triangular distribution. The choice of these distributions of the random parameters resulted in a better fit of the model (log-likelihood = −24,362 and AIC = 48,734) compared to models in which the random parameters were all normally distributed (log-likelihood = −27,033 and AIC = 54,079) or models with only triangular distributions bounded by zero (log-likelihood = −24,716 and AIC = 49,439).

[11] Assuming a non-random cost coefficient is common practice when analysing choice data. The reasons therefore are explained by Giergiczny et al. (2012). One of the reasons is that when a random cost coefficient is assumed, this could result in "exploding" WTP values because the share of denominators that approximate zero increases when the variance of the cost coefficient increases.

[12] The scenarios were compared two by two. In some comparisons, alternatives A and B of the first scenario and alternatives C and D of the second scenario were such that: $A < D$ and $B > C$. If individuals declare that they consider that $A > B$ then we can deduce that: $A > B > C$ and as $D > A$ we have $D > A > B > C$ which makes it inconsistent to simultaneously prefer A to B and C to D.

7. Discussion

By applying two different econometric models (MNL and ML), two sets of WTP values are estimated. While the results of both models are close to each other, the ML model is preferred since the panel dimension of the data is taken into account and it has a better fit than the MNL model. The resulting estimates for the whole sample based on the ML model are a VSL of 6.19 Million EUR, a VSSI of 0.95 Million EUR and a VoT of 16.1 EUR. The cost ratio between the VSL and VSSI is about 1 to 7. By addressing the influence of lexicographic respondents, inconsistent respondents, potential hypothetical bias and potential free-riders, the robustness of the model and the reliability of the results is demonstrated. National WTP values are calculated by applying the ML model to the country-specific samples. The WTP values in France (VSL of 5.35 Mill EUR, VSSI of 0.83 Mill EUR and VoT of 12.9 EUR) appear to be significantly lower than the values for Germany (VSL of 7.35 Mill EUR, VSSI of 1.1 Mill EUR and VoT of 19 EUR). The values for Belgium and the Netherlands lie in between.

Since the same sampling procedure and the same SC design have been applied in all four countries, the differences in WTP values cannot by explained by systematic differences between the samples or differences in methodology. There are several possible explanations for these differences. The lowest WTP values are observed in France, which is the only country in which road users have experience with toll roads. Possibly French respondents attach more importance to the cost attribute because they have real life experience with taking costs into account when choosing a route. In addition there are other observable differences between countries that could explain the differences in WTP values, such as income, travel habits or the share of company cars. Next to that also cultural differences could explain different values, such as different attitudes towards risk-taking, time loss or spending money. An examination of interaction effects between these variables and the preference parameters could provide more explanations for the differences between countries.

For each country the estimates from this study are considerably higher than the values that are currently used by national governments. However, comparing the new estimates of this study with official values of the participating countries is difficult because different methodologies are used. Official values in Germany (Baum et al., 2010) are based on a combination of restitution costs and human capital loss, which are known to yield much lower values. In France (CGSP, 2013) and Belgium (RebelGroup, 2013) a value transfer has been applied, which means that WTP values from other countries have been used. The use of WTP as the most appropriate method for assessing the human costs of road casualties has been widely recommended, for example by COST 313 (Alfaro et al., 1994), since the monetary value of safety should reflect the preferences of those affected (de Blaeij, 2003). Only in the Netherlands the official estimates are based on a national WTP survey using a similar survey design (de Blaeij, 2003). These values are still much lower than the estimates of this study and only a small part of the difference can be explained by inflation and income growth. The results of our study might indicate that people's preferences have changed in the last twenty years and that the importance of road safety has increased.

Earlier academic studies on VSL using WTP show a broad dispersion in estimates. Bahamonde-Birke et al. (2015) give an overview of VSL estimates in the context of road safety from WTP studies in several countries, including 13 SC studies. The VSL values found in SC studies range from 548,000 to 7,3 million EUR (price level 2020)[13]. We can conclude that the VSL estimates in this study are at the higher end of the range of VSL estimates found in similar studies. In the United States, Viscusi (2018) found that estimates of the VSL in wage-risk studies are around 10 million USD (price level 2017), which is in the same order of magnitude of the value found in our study. With regards to the VSSI, Wijnen (2021) provided an overview of WTP estimates of preventing serious injuries in the literature, including six SC studies. This overview shows a wide variation of VSSI estimates, from 1% to 47% of the VSL. The ratio between VSL and VSSI in our study is 16% which lies within this range and is consistent with values recommended in European studies on the external costs of transport (van Essen et al., 2019).

There are some well-known challenges in the practical application of the SC method to estimate the WTP such as the hypothetical nature of the experiment, the ability of individuals to deal with small risk changes, behavioural bias, protest answers and respondents' propensity to ignore income constraints (Bahamonde-Birke et al., 2015). When designing the survey, considerable efforts have been made to take these issues into account. Different *meta*-analyses find an influence of the size of the change in the risk on the VSL, with a smaller risk reduction leading to higher WTP values (Elvik, 2016).The differences between attribute levels in our survey are based on a reference value of VSL that was recommended by OECD (2012). For future research we recommend a study on the sensitivity of the results with regards to the attribute levels and their differences.

The survey was conducted in October and November 2020 during the COVID-19 pandemic. It was assumed that the high number of COVID victims and the prevention measures that resulted in a decrease in mobility, could affect the preferences of individuals regarding risk and their perception of road safety. On the one hand we can hypothesize that individuals give less importance to road safety because of the perception that crash risks decreased due to the decrease in traffic. On the other hand, we can hypothesize that road safety became more important during the pandemic because of the overall increased attention to health risks. To investigate the potential influence of the pandemic on the response behaviour in our study, two additional questions were included in the survey. The results are presented in Table 6 and show that the majority of respondents (53%) don't think road safety has increased or decreased since the pandemic. While there is still an important share of respondents that perceive a change in road safety, there is no clear trend with regards to the direction of the change. The same is observed for the importance that respondents attach to road safety: only a moderate share of respondents (15%) reports a change in importance, without a clear trend. This analysis does not support the hypothesis that the pandemic had a big influence on the response behaviour in our survey.

The results of this study can be used for two main purposes in policy-oriented research and applications. Firstly, VSL and VSSI values are an important input for estimating the socioeconomic costs of road crashes since they reflect the human costs related to fatalities and serious injuries as well as consumption loss. Usually, the loss of consumption resulting from road fatalities is included in the calculation of production loss (known as 'gross production loss'), and therefore consumption loss should be deducted from the VSL to arrive at the human costs (Evans, 2001; Wijnen et al., 2009). Reviews show that human costs take up a major share of road crash costs. Wijnen et al. (2019) estimated that human costs amount to 54–94% of the costs per fatality and 51–91% of the costs per serious injury in European countries that use the WTP approach. Future research could be dedicated to estimate the human costs of slight injuries and the economic costs of road crashes in order to have a complete estimate of the socioeconomic costs of road crashes.

Secondly, the VSL and VSSI are needed for cost-benefit analysis (CBA) of road safety measures or broader transport projects with road safety impacts. In CBA, estimates of the socioeconomic costs per casualty or per crash are used to translate casualty reductions into monetary benefits (which are equal to the road crash cost savings). The outcomes of CBAs of road safety measures show whether the socioeconomic return is positive or negative, which can support decision-making about road safety investments and prioritizing road safety measures. Since human costs, and thereby the VSL and VSSI, are essential elements of the socioeconomic cost, the VSL and VSSI have an important impact on the results of CBAs. Higher VSL and VSSI estimates translate into greater

[13] A VSL of 32 million EUR (2020) that was found in a Spanish study was considered implausible.

safety benefits, and thus into more favourable benefit-cost ratios (if there is a positive impact on road safety). In general, most road safety investments are found to be (very) cost-beneficial (Daniels et al., 2019). This is partly explained by the fact that the socioeconomic costs of road crashes, and therefore the benefits of safety improvements, are usually high, as confirmed by results of this study.

More specifically, the results of this study can be incorporated in international studies on the external costs of transport, aiming to update the currently used values of road crash costs. The latest study (van Essen et al., 2019) uses a VSL found in a *meta*-analysis of OECD countries (including non-European countries) and a VSSI which is based on a study from the late 1990 s (ECMT, 1998). Additionally, the results of this research may be useful for other countries which do not have their own country-specific estimates of the VSL and VSSI based on a WTP method. These results can serve as a good source for value transfer (Freeman et al., 2014) applying GDP and price level adjustments.

8. Conclusions

In the present study a common methodology was applied to estimate the human costs of road crashes in four European countries: Belgium, France, Germany and the Netherlands. By means of SC survey that was simultaneously conducted in the four countries, the study provides common VSL, VSSI and VoT values for the four participating countries together, and updated national estimates for each country separately (Table 7).

The national estimates can be used to update the socioeconomic costs of road crashes in the four countries and they can be included in national guidelines for conducting CBAs of transport projects. The average values of the four countries together can be used for international purposes such as European or bilateral transport investments.

Further research can be dedicated to explain the differences between countries by examining the interaction effects between individuals' observable characteristics and the marginal disutilities of the attributes, or by examining the latent variables that are based on attitudes and opinions included in the questionnaire. Also the impact of the estimates on the outcomes of CBAs of transport projects can be investigated.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This paper is published in the framework of the VALOR project that was made possible by the financial support of the Federal Public Service Mobility and Transport (Belgium), the Délégation à la Sécurité Routière (France), the Federal Highway Research Institute (Germany) and KiM Netherlands Institute for Transport Policy Analysis (The Netherlands). The information and views set out in this paper are those of the authors and may not reflect the viewpoint of the funding authorities.

Appendix

Table A1

Table A1
Experimental design

		Alternative	Cost	Time	Fatalities	Serious injuries
BLOCK 1	S1	A	10.0	52	1	3
		B	5.5	28	22	93
	S2	A	7.0	28	22	138
		B	10.0	52	1	3
	S3	A	5.5	52	22	3
		B	10.0	28	1	138
	S4	A	8.5	28	1	138
		B	7.0	44	22	3
	S5	A	5.5	52	1	138
		B	10.0	28	22	48
	S6	A	10.0	52	15	3
		B	5.5	44	22	138
	S7	A	10.0	28	8	138
		B	5.5	52	15	3
BLOCK 2	S1	A	5.5	52	22	93
		B	10.0	36	1	138
	S2	A	10.0	28	15	3
		B	7.0	52	1	138
	S3	A	5.5	52	1	138
		B	10.0	28	22	3
	S4	A	10.0	52	1	48
		B	5.5	28	15	138
	S5	A	10.0	44	1	138
		B	5.5	52	22	3
	S6	A	5.5	52	1	138
		B	10.0	28	22	93
	S7	A	5.5	44	22	138
		B	10.0	52	8	3

References

AAAM. (2020, 12 6). *Abbreviated Injury Scale (AIS)*. Retrieved from Association for the Advancement of Automotive Medicine: https://www.aaam.org/education-resource-center/public-position-statements/abbreviated-injury-scale-ais-position-statement/.

Alfaro, J., Chapuis, M., Fabre, F., 1994. Socio-economic cost of road accidents: final report of action COST 313. Commission of the European Community, Brussels.

ANWB. (2020, 10 1). *Brandstofprijzen Europa*. Retrieved from ANWB: https://www.anwb.nl/vakantie/reisvoorbereiding/brandstofprijzen-europa.

Bahamonde-Birke, F.J., Kunert, U., Link, H., 2015. The Value of a Statistical Life in a Road Safety Context - A Review of the Current Literature. Transport Reviews 35 (4), 488–511. https://doi.org/10.1080/01441647.2015.1025454.

Baum, H., Kranz, T., Westerkamp, U., 2010. Volkswirtschaftliche Kosten durch Straßenverkehrsunfälle in Deutschland. Bundesanstalt für Straßenwesen, Bergisch Gladbach.

Bickel, P. et al. (2006). *Proposal for harmonised guidelines. EU project HEATCO. Deliverable 5*. Stuttgart: University of Stuttgart.

Blincoe, L., Miller, T., Zaloshnja, E., Lawrence, B., 2015. The economic and societal impact of motor vehicle crashes, 2010 (revised). National Highway Traffic Safety Administration, Washington.

Boardman, A.E., Greenberg, D.H., Vining, A.R., Weimer, D.L., 2017. Cost-benefit analysis: concepts and practice. Cambridge University Press.

Burge, P., Rohr, C., 2004. DATIV: SP Design: Proposed Approach for Pilot Survey. Tetra-plan in Cooperation with RAND Europe and Gallup A/S.

CBS Statline. (2019). *Regionale kerncijfers Nederland*. Retrieved from Statline. Nederland in cijfers: https://opendata.cbs.nl/.

CGSP. (2013). *L'évaluation socioéconomique des investissements publics. Rapport de la mission présidée par Emile Quinet*. Paris: Commissariat Général à la Stratégie et à la Prospective.

Croissant, Y. (2020). *mlogit: Multinomial Logit Models*. Retrieved from R package version 1.1-0: https://CRAN.R-project.org/package=mlogit.

Daniels, S., Martensen, H., Schoeters, A., Van den Berghe, W., Papadimitriou, E., Ziakopoulos, A., Perez, O., 2019. A systematic cost-benefit analysis for 29 road safety measures. Accident Analysis and Prevention 133, 105292. https://doi.org/10.1016/j.aap.2019.105292.

de Blaeij, A. T. (2003). *The value of a statistical life in road safety. Stated preference methodologies and empirical estimates for the Netherlands*. Tinbergen Institute Research Series. Amsterdam: Vrije Universiteit Amsterdam.

de Blaeij, A., Florax, R., Rietveld, P., Verhoef, E., 2003. The value of statistical life in road safety: a meta-analysis. Accident Analysis and Prevention 35, 973–986. https://doi.org/10.1016/S0001-4575(02)00105-7.

De Brabander, B. (2006). *Valuing the reduced risk of road accidents. Empirical estimates for Flanders based on stated preference methods*. Universiteit Hasselt, Faculteit Toegepaste Economische Wetenschappen. Hasselt: Universiteit Hasselt.

ECMT. (1998). *Efficient transport for Europe; Policies for internalisation of external costs*. Paris: Organisation for Economic Co-operation and Development OECD.

Elvik, R., 2016. The Value of Life - The Rise and Fall of a Scientific Research Programme. Institute of Transport Economics, Oslo, Norway.

European Commission. (2013). Commission staff working document on the implementation of objective 6 of the European Commission's policy orientations on road safety 2011–2020: First milestone towards and injury strategy. Brussels: European Commission.

European Commission. (2020, 01 31). *CARE database*. [Data file]. Retrieved from CARE database: https://ec.europa.eu/transport/road_safety/specialist/statistics_en.

Eurostat. Total length of motorways [Data file]. Retrieved from Eurostat 2020.

Eurostat. (2020b). *Purchasing Power Parities (PPPs), price level indices and real expenditures*. [Data file]. Retrieved from Eurostat: https://ec.europa.eu/eurostat/databrowser/view/prc_ppp_ind/default/table?lang=en.

Evans, A., 2001. The economic appraisal of road traffic safety measures in Great Brittain. Paper for ECMT Round Table 117 "Economic evaluation of road traffic safety measures". ECMT, Paris.

Flügel, S., Veisten, K., Rizzi, L.I., Ortùzar, J.D., Elvik, R., 2019. A comparison of bus passengers' and car drivers' valuation of casualty risk reductions in their routes. Accident Analysis and Prevention 112, 63–75. https://doi.org/10.1016/j.aap.2018.09.028.

Freeman, A.M., Herriges, J.A., Kling, C.L., 2014. The measurement of environmental and resource values: theory and methods. Routledge.

Census, G., 2011. *Genesis Online Database*. [Data file]. Retrieved from DSTATIS. www-genesis.destatis.de, Statistisches Bundesamt.

Giergiczny, M., Valasiuk, S., Czajkowski, M., De Salvo, M., Signorello, G., 2012. Including cost income ratio into utility function as a way of dealing with 'exploding'implicit prices in mixed logit models. Journal of Forest Economics 18 (4), 370–380. https://doi.org/10.1016/j.jfe.2012.07.002.

González, R.M., Román, C., Amador, F.J., Rizzi, L.I., Ortúzar, J.d.D., Espino, R., Martín, J.C., Cherchi, E., 2018. Estimating the value of risk reductions for car drivers when pedestrians are involved: a case study in Spain. Transportation 45 (2), 499–521.

Henscher, D.A., Rose, J.M., Ortúzar, J.D., Rizzi, L.I., 2009. Estimating the willingness to pay and value of risk reduction for car occupants in the road environment. Transportation Research Part A 43, 692–707. https://doi.org/10.1016/j.tra.2009.06.001.

Hess, S., Rose, J.M., 2009. Should reference alternatives in pivot design SC surveys be treated differently? Environmental and Resource Economics 42 (3), 297–317. https://doi.org/10.1007/s10640-008-9244-6.

Hess, S., Rose, J.M., Polak, J., 2010. Non-trading, lexicographic and inconsistent behaviour in stated choice data. Transportation Research Part D Transport and Environment 15 (7), 405–417. https://doi.org/10.1016/j.trd.2010.04.008.

Hojman, P., Ortúzar, J.D., Rizzi, L.I., 2005. On the joint valuation of averting fatal and severe injuries in highway accidents. Journal of Safety Research 36 (4), 377–386. https://doi.org/10.1016/j.jsr.2005.07.003.

INSEE. (2019). *Statistiques et études*. [Data file]. Retrieved from Institut national de la statistique et des études économiques: insee.fr.

Iraguën, P., Ortúzar, J., d., 2004. Willingness-to-pay for reducing fatal accident risk in urban areas: an Internet-based Web page stated preference survey. Accident Analysis and Prevention 36, 513–524. https://doi.org/10.1016/S0001-4575(03)00057-5.

Johnson, R., Orme, B., 2003. Getting the most from CBC. Sawtooth Software Research Paper Series, Sawtooth Software, Sequim.

Kahneman, D., Tversky, A., 1972. Subjective probability: A judgment of representativeness. Cognitive Psychology 3 (3), 430–545. https://doi.org/10.1016/0010-0285(72)90016-3.

Korzhenevych, A., Dehnen, N., Bröcker, J., Holtkamp, M., Meier, H., Gibson, G., Cox, V., 2014. Update of the handbook on external costs of transport: final report for the European Commission: DG-MOVE. European Commission, Brussels.

Lindhjem, H., Navrud, S., Braathen, N.A., 2010. Valuing lives saved from environmental, transport and health policies: a meta-analysis of stated preference studies. Organization for Economic Co-operation and Development, Paris.

Loftis, K.L., Price, J.P., Gillich, P.J., Cookman, K.J., Brammer, A.L., St. Germain, T., Barnes, J.o., Graymire, V., Nayduch, D.A., Read-Allsopp, C., Baus, K., Stanley, P.A., Brennan, M., 2016. Development of an expert based ICD-9-CM and ICD-10-CM map to AIS 2005 update 2008. Traffic Injury Prevention 17 (sup1), 1–5.

Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated Choice Methods. Cambridge University Press, Analysis and Application.

MacKenzie, E., Steinwachs, D.M., Shankar, B., 1988. April). Classifying Trauma Severity Based on Hospital Discharge Diagnoses. Validation of an ICD-9CM to AIS-85 Conversion Table. Medical Care 27 (4), 412–422. https://doi.org/10.1097/00005650-198904000-00008.

Maibach, M., Schreyer, C., Sutter, D., van Essen, H., Boon, B., Smokers, R., Bak, M., 2007. Handbook on estimation of external cost in the transport sector: produced within the study Internalisation Measures and Policies for All external Cost of Transport (IMPACT). CE Delft, Delft.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembkam, P. (Ed.), Frontiers in econometrics. Academic Press, New York, pp. 105–142.

Milieu Centraal. (2020, 10 1). *Zuinig rijden*. Retrieved from milieu centraal: https://www.milieucentraal.nl/duurzaam-vervoer/autokeuze-en-gebruik/zuinig-rijden/.

Niroomand, N., Jenkins, G.P., 2016. Estimating the Value of Life, Injury, and Travel Time Saved Using a Stated Preference Framework. Accident Analysis and Prevention 91, 216–225. https://doi.org/10.1016/j.aap.2016.03.004.

Nuyttens, N., & Van Belleghem, G. (2014). *Hoe ernstig zijn de verwondingen van verkeersslachtoffers? Analyse van de MAIS-ernstscore van verkeersslachtoffers opgenomen in de Belgische ziekenhuizen in de periode 2004-2011*. Brussels: Belgisch Instituut voor de Verkeersveiligheid - Kenniscentrum Verkeersveiligheid & Vrije Universiteit Brussel - Interuniversity Centre for Health Economics Research.

O'Reilly, D., Hopkin, J., Loomes, G., Jones-Lee, M., Philips, P., McMahon, K., Kemp, R., 1994. The Value of Road Safety: UK Research on the Valuation of Preventing Non-Fatal Injuries. Journal of Transport Economics And Policy 28 (1), 45–59.

A. Obermeyer G. Hirte Willingness to pay for road safety: A conceptual study and pilot survey for Germany 2020 Lyon, France.

OECD. (2020). *Road motor vehicle traffic by road type*. [Data file]. Retrieved from IRTAD database: https://stats.oecd.org/index.aspx?r=930626&errorCode=403&lastaction=login_submit.

Orme, B., 2010. Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research, (2nd ed.). Research Publishers LLC., Madison, Wis.

Pearce, D., Özdemiroglu, E., 2002. Economic Valuation with Stated Preference Techniques Summary Guide. Department for Transport, Local Government and the Regions, London.

Persson, U., 2004. *Valuing reductions in the risk of traffic accidents based on empirical studies in Sweden* (Doctoral dissertation). Lund Institute of Technology, Lund.

RebelGroup, 2013. Standaardmethodiek voor MKBA van transportinfrastructuurprojecten - Kengetallenboek. RebelGroup Advisory Belgium, Antwerp.

Rizzi, L.I., Ortúzar, J.D., 2003. Stated preference in the valuation of interurban road safety. Accident Analysis and Prevention 35 (1), 9–22. https://doi.org/10.1016/s0001-4575(01)00082-3.

Rizzi, L.I., Ortúzar, J.D., 2006. Estimating the Willingness-to-Pay for Road Safety Improvements. Transport Reviews 26 (4), 471–485. https://doi.org/10.1080/01441640600602302.

Sælensminde, K., 2002. The Impact of Choice Inconsistencies in Stated Choice Studies. Environmental and Resource Economics 23 (4), 403–420. https://doi.org/10.1023/A:1021358826808.

Sælensminde, K., 2006. Causes and consequences of lexicographic choices in stated choice studies. Ecological economics 59 (3), 331–340. https://doi.org/10.1016/j.ecolecon.2005.11.001.

Schoeters, A., Wijnen, W., Carnis, L., Wijermars, W., Elvik, R., Daniels, S., Johannsen, H., 2020. Costs related to serious road injuries: a European perspective. European Transport Research Review 12 (1). https://doi.org/10.1186/s12544-020-00448-0.

Sillano, M., Ortúzar, J.D., 2005. Willingness-to-pay estimation with mixed logit models: some new evidence. Environment and Planning 37 (3), 525–550. https://doi.org/10.1068/a36137.

Statbel. (2019). *Structure of the Population*. [Data file]. Retrieved from Statbel. Belgium in figures: https://statbel.fgov.be/.

Stewart, A.E., Lord, J.H., 2002. Motor Vehicle Crash versus Accident: A Change in Terminology Is Necessary. Journal of Traumatic Stress 15, 333–335. https://doi.org/10.1023/A:1016260130224.

Thurstone, L.L., 1927. A law of comparative judgment. Psychological Review 34 (4), 273–286. https://content.apa.org/doi/10.1037/h0070288.

F. Traets D.G. Sanchez M. Vandebroek November). Generating Optimal Designs for Discrete Choice Experiments in R: The idefix Package Journal of Statistical Software 96 3 2020 10.18637/jss.v096.i03.

Train, K., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.

United Nations. (2018). *Demographic Statistics Database*. [Data file]. Retrieved from UNdata: http://data.un.org/.

van Essen, H., van Wijngaarden, L., Schroten, A., Sutter, D., Bieler, C., Maffii, S., El Beyrouty, K., 2019. Handbook on the external costs of transport, version 2019. CE Delft, Delft.

Veisten, K., Flügel, S., Rizzi, L.I., Ortúzar, J.D., Elvik, R., 2013. Valuing casualty risk reductions from estimated baseline risk. Research in Transportation Economics 43 (1), 50–61. https://doi.org/10.1016/j.retrec.2012.12.009.

Viscusi, W.K., 2018. Pricing Lives: Guideposts for a Safer Society. Princeton University Press.

Wijnen, W. (2021). *Economic valuation of preventing non-fatal road injuries: a literature review*. Presentation at the Annual Conference of the Society for Benefit-Cost Analysis 2021.

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.

Wijnen, W., Wesemann, P., de Blaeij, A., 2009. Valuation of road safety effects in cost-benefit analysis. Evaluation and Program Planning 32 (4), 326–331. https://doi.org/10.1016/j.evalprogplan.2009.06.015.

Wiktor, A., Louviere, J., Swait, J., 1998. Introduction to Attribute-Based Stated Choice Methods. NOAA - National Oceanic and Atmospheric Administration. US Department of Commerce, Alberta.