Private and public willingness to pay for safety: A validity test

Henrik Andersson[a,*], Elodie Levivier[a], Gunnar Lindberg[b]

[a] *Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France*
[b] *Institute of Transport Economics (TØI), Oslo, Norway*

ABSTRACT

Stated preference (SP) methods are often used to elicit an affected population's preferences for, e.g., increased safety or better environmental quality. SP methods are based on hypothetical market scenarios which have advantages, since decision alternatives are known to the analysis, but also necessitate thorough validity tests of the results, since decisions are hypothetical. This study suggests a validity test based on theoretical predictions and empirical findings for private and public safety measures. According to the test, willingness to pay (WTP) for a public safety measure should exceed or be equal to the private one. Based on a rich data set eliciting both private and public WTP the results show that private WTP exceeds public WTP. Hence, the findings in this study highlight the importance of validity tests of preference estimates for safety, and suggest that WTP also for a private safety measure should be elicited in studies eliciting WTP for public safety measures, to allow for the validity test.

1. Introduction

To mitigate adverse health effects policy makers can, broadly speaking, choose between private and public safety measures. For instance, whereas legislation on seat-belt usage can be considered a private good for which the individual bears the costs and is the sole person who benefits from the usage, public investments to improve the standard of a road that increase safety will benefit all who travel on that road. To secure an efficient resource allocation, policy makers often rely on benefit-cost analysis (BCA) which evaluates both benefits and costs with a common metric in the form of monetary values (European Commission, 2014). However, many of the effects induced by safety policies do not have easily obtainable monetary values and for those effects non-market valuation methods such as revealed (RP) and stated preference (SP) methods are used.[1] The latter does not rely on the existence of actual market data like in the RP method. Hence, the SP approach is especially useful when eliciting preferences for public goods (or bads) and/or non-use values considering the lack of market information in those situations.

In SP methods, people are asked to state their preferences based on hypothetical scenarios in a survey, e.g. in a face-to-face interview or on the web, and the objective is to estimate the respondents' willingness to pay (WTP) for, e.g., safety measures. Since the seminal paper of Bohm (1972) the possible difference between a hypothetical response and an actual value has been discussed in the literature. Carson and Groves (2007) summarize the discussion of the hypothetical nature of the question in SP studies, strategic behavior, or preferences that may be ill-defined or inconsistent with economic theory and/or intuition. The valuation of safety measure with SP methods may lead to several biases: (i) hypothetical bias, which is related to the fact that the decisions made in the survey have no real consequences, and hence there is a risk that respondents exaggerate their WTP (Blumenschein et al., 2008), (ii) strategic bias, a problem in public economics (Samuelson, 1954) which can be mitigated by adequate mechanism design (Carson and Groves, 2007), and (iii) scope bias, sometimes also referred to as the "embedding" effect, which has been expressed by finding much the same WTP for different level of benefits (Baker et al., 2008) and for risk reductions more specifically as the WTP not being adequately sensitive to the size of the risk reduction (Robinson and Hammitt, 2015). Regarding the estimation of WTP for risk reductions much focus has been on the difficulty for respondents to understand small probabilities, and means to help the respondents to understand the scenarios by for instance using visual aids (Corso et al., 2001; Andersson and Svensson, 2008). Therefore, to ensure the validity of estimates in SP studies eliciting WTP for risk reductions validity tests should systematically be conducted. Bishop and Boyle (2018) describe the three types of validity tests

---

commonly used: (i) *content*, focusing on whether the valuation method and procedure chosen are "conducive to measuring the true value", (ii) *construct*, prior expectations based on theory, intuition, or empirical evidence, of how true value is related to other variables, and (iii) *criterion*, test by comparing different methods with one accepted as having higher level of validity.

Since preferences are expected to be context dependent policy makers may need different values when evaluating private and public safety policies. The empirical evidence suggests that WTP for private and public traffic safety measures differ with private being higher than public (de Blaeij et al., 2003), which suggests that they should be estimated separately. Other evidence suggests that hypothetical bias is a bigger concern in studies eliciting WTP for public compared to private measures (List and Gallet, 2001; Murphy et al., 2005). Hence, the difference between the two may be larger if one considers the issue of the hypothetical nature of SP studies. A concern with much empirical evidence is that comparisons are confounded, since results from different studies are compared where context, design, etc., may differ. It is therefore important to develop reliable validity tests to understand what the predictions are for WTP for public and private safety measure to improve the accuracy of policy evaluations, both for public and private safety measures.

The aim of this study is to suggest a (construct) validity test for private and public WTP for safety. We use a framework with a mortality risk and show the predicted relationship between private and public WTP for safety based on theoretical models and empirical findings. Thus, our test can be defined as a construct validity test and to show the importance of this test we use data from three Swedish contingent valuation (CV) studies (Hultkrantz et al., 2006; Andersson and Lindberg, 2009; Svensson and Vredin Johansson, 2010).[2]

The following section contains the theoretical model and the predictions based on both the theoretical model and empirical findings. Then in Section 3 the surveys used are briefly described, and in Section 4 the empirical model is presented. The results are shown in Section 5 and we end the article with a discussion of the findings in Section 6.

2. The validity test of private and public WTP for safety

For simplicity, our theoretical model is a single-period model with two individuals who face two possible outcomes; staying alive or being dead (Jones-Lee, 1991, 1992; Johannesson et al., 1996). Let V_{ij} (.), π_{ij}, and y_{ij} denote a well-behaved cardinal utility function, survival probabilities, and wealth, respectively. The first subscript $i = \{1,2\}$ refers to the individuals with 1 defining the considered individual, and the second subscript $j = \{0,1\}$ refers to before (0) and after (1) implementation of the safety project, with $\pi_{i0} < \pi_{i1}$. The utility function of individual 1 can now be written as:

$$V_{10} = V_{10}(\pi_{10}, y_{10}, \pi_{20}, y_{20}),$$

which is assumed to be strictly increasing in π_1 and y_1, and non-decreasing in π_2 and y_2. Individual 1 is: (i) purely selfish if $\partial V_1/\partial \pi_2 = 0$ and $\partial V_1/\partial y_2 = 0$, (ii) a pure altruist or paternalist if both are strictly positive, and (iii) a safety paternalist if $\partial V_1/\partial \pi_2 > 0$ and $\partial V_1/\partial y_2 = 0$.[3] A safety paternalist, thus, is only concerned about the safety dimension of the others' well-being. Table 1 summaries the predictions of interest to this study.

For our theoretical prediction of private and public WTP for safety, we follow the analysis in Johannesson et al. (1996). We assume that the private and public safety measures reduce individuals' risk by the same amount, the difference being the characteristic of the good and the way

Table 1
Individual 1's utility, related to wealth (y) and survival probability (π) of individual 2.

	$\partial V_1/\partial \pi_2 = 0$	$\partial V_1/\partial \pi_2 > 0$
$\partial V_1/\partial y_2 = 0$	Purely selfish	Safety paternalist
$\partial V_1/\partial y_2 > 0$	–	Pure altruist or paternalist

it is financed. Whereas the private good is paid for by the individual, the public good is financed through a flat tax. The optimization problem for the private good is:

$$V_{10} = V_{11}(\pi_{11}, y_{10} - p_1, \pi_{20}, y_{20}),$$

where p_1 refers to WTP for the private risk reduction. Since only individual 1 experiences an increased safety level, individual 2 remains at his initial utility level, and p_1 reveals 1's WTP regardless of the type of altruistic preferences (including selfishness).

For the public safety measure, from which both individuals experience increased survival probabilities, 1's WTP is defined by t_1. Since t_1 is assumed identical to all, we now have the following optimization problem:

$$V_{10} = V_{11}(\pi_{11}, y_{10} - t_1, \pi_{21}, y_{20} - t_1),$$

Since a selfish individual would not care how a second individual is affected by the public safety measure, his WTP is the same for both measures, i.e. $t_1 = p_1$. For a pure altruist, his WTP depends on how he believes that the second individual is affected by the public measure. If he believes that t_1 approximates the WTP of the other individual, he would report $t_1 = p_1$ (since $i = 2$ remains at his initial utility). On the other hand, if he believes that the project will increase (reduce) the other person's overall wellbeing he/she would state $t_1 > (<) p_1$. For a pure paternalist $t_1 = p_1$, since 1 imposes his preferences on 2. Finally, a safety paternalistic individual would report $t_1 > p_1$ since he ignores the wealth effect (i.e. forgone consumption opportunities through t_1) the safety project has on individual 2.

Since the theoretical prediction on how private WTP is related to public WTP depends on our assumptions about the considered individual's altruistic preferences, we turn to the empirical evidence to complete the picture. There is strong empirical support that individuals are non-selfish regarding safety (health) of others. For instance, there is empirical evidence that mothers and parents are willing to pay more for the safety of their children than for their own safety (Liu et al., 2000; Dickie and Messman, 2004; Chanel et al., 2005; Andersson and Lindberg, 2009), and that WTP for the entire household is larger than individual WTP (Chanel et al., 2005; Bateman and Brouwer, 2006). Moreover, there is also evidence suggesting that individuals are safety-paternalistic toward others (Holmes, 1990; Vázquez Rodríguez and León, 2004; Jacobsson et al., 2007). Hence, overall, the empirical evidence regarding individuals' preferences for others' safety is that they are safety paternalists.

The empirical evidence on altruistic preference therefore would suggest that $t_1 \geq p_1$. However, the empirical findings from SP studies eliciting WTP for road safety often reveal the opposite, i.e. that respondents state a lower WTP for a public than for a private good (de Blaeij et al., 2003). The empirical findings on private and public WTP are, thus, not consistent with the predictions of the theoretical model combined with the empirical results on safety paternalism. However, between studies differences between private and public WTP may not be very informative since estimates may be confounded by many factors such as survey design differences, but also context, populations, etc. Hence, the suggested validity test here is for studies that elicit the WTP for both a private and a public safety measure with the same risk-reduction and payment scenario.

---

[2] For a description of the CV method, see, e.g., Freeman et al. (2014).

[3] Let m_{kl} denote individual k's marginal rate of substitution of l's y for π. Then k is a pure altruist if $m_{kl} = m_{ll}$, i.e. k respects l's preferences, and a pure paternalist if $m_{kl} = m_{kk}$, i.e. k imposes his preference on l (Jones-Lee, 1992).

3. Data

3.1. Surveys' description

This paper uses three CV surveys conducted in Örebro in Sweden at different points in time, 1998, 2004 and 2006. The pooling relies on the assumption that preferences are stable over time (Liebe et al., 2016). The first two surveys used single-bounded (SB) questions, whereas the last one used an open-ended (OE) format (see, e.g., Johnston et al., 2017, for a discussion of SP methods). In the former respondents are asked to answer yes or no to a question like *"Are you willing to pay SEK X?"* for the good in the hypothetical scenario of the survey, where *X* is the bid level that varies between different subsamples within the same survey. In contrast, with the latter format respondents are asked *"How much are you willing to pay?"* without any suggested level to them to which they may anchor their answer. A brief description of each survey is provided in the following paragraphs.

Before giving an overview of each survey, it is essential to describe the magnitude of the risk scenario of all hypothetical settings. The surveys used the "Vision Zero" context as a basis (Trafikverket, 2017). This road-safety program, adopted by the Swedish Parliament in 1997 (Hultkrantz et al., 2006), has a vision of zero accidents that lead to fatalities or severe injuries. Citizens of Örebro were familiar with the program at the time of the surveys and this brought credibility to the hypothetical risk scenarios and proposed safety programs. In each study, respondents are first informed about the number of people severely injured and killed at a given year and then about safety measures and the risk reduction they will lead to. The size of the risk reductions for the first two studies correspond to 100%, i.e. an elimination of severe and fatal accidents in line with the vision, whereas the risk reduction for the same level of severity is 50% for the last one.

The first questionnaire (Andersson and Lindberg, 2009, subsequently referred to as AAP 2009) was sent by post in 1998 randomly to individuals between 18 and 76 years old. 1040 questionnaires, representing a 55% response rate, were returned. The sample was divided into five predetermined clusters: four about a private safety device and one about a public safety program. The former was differentiated in term of targeted population that would benefit from the device, i.e. the individual himself, a child, the household, or a relative of the respondent. After considering the context conditions of the survey, respondents were asked whether they were willing to rent a private device (private question) or pay the annual fee (public program question) at a predetermined bid level, allocated randomly among the respondents.[4] In our analysis, we kept the individuals who answered the questions about the own private and the public safety device. Our sample for this dataset contains 440 individuals. For the analysis of the other subsamples, see Andersson and Lindberg (2009).

For the second questionnaire (Hultkrantz et al., 2006, subsequently referred to as JRU 2006), a postal survey was sent in the spring of 2004 to 1435 individuals between 18 and 75 years old. The total response rate was 61%. Here as well, the sample was divided in five: one group was asked about the WTP for a private safety device, and the other four about their WTP for a public safety programme. The latter were different for sensitive analysis in term of magnitude of risk as well as presence of a provision point mechanism (PPM). Three of the public safety subsamples used this PPM and were excluded from the analysis since it was not used in the other surveys. Thus, one subsample out of the four on public WTP has been kept for the analysis in this study. As previously, respondents were asked if they are willing to pay SEK *X* for a public (private) safety program (device) defined in the hypothetical scenario of the CV survey. However, an added value of this study is the use of the preference-certainty scale (Blomquist et al., 2009; Svensson, 2009). That is, at the end of the questionnaire, people had to express

their level of certainty on a scale from 1, *"very uncertain"*, to 10, *"very certain"*.[5] We use this additional information for some robustness tests of our results.

The last questionnaire (Svensson and Vredin Johansson, 2010, subsequently referred to as AAP 2010), sent in the spring of 2006 by mail to 1500 individuals, was divided into four sub-samples. The first two asked two questions regarding the respondents' private and public WTP where the order was reversed in the two subsamples "[T]o control for potential anchoring effects of the first question." (p. 1207), whereas in the two other respondents were asked about their private or public WTP. After a description of both the private safety device and the public safety program, and a context explanation in the hypothetical scenario, people had to answer the amount of money they were willing to pay in either one or two OE questions, depending of the subsample they belonged to. The response rate was 59%. The same certainty scale was used as in the previous survey.

3.2. Pooled dataset

For our analysis, we pooled the three previously described datasets. Because they were not constructed based on the same information about the individuals, we kept only the common explanatory variables of interest. Table 2 summarizes the variables used for the empirical analysis. We want to highlight two differences in information collected between the surveys: (i) respondents were in AAP 2010 asked about their gross income, whereas they in the other two surveys were asked about disposable income, and hence income from AAP 2010 has been converted to disposable income, and (ii) there is a subtle difference in how the question on accident experience was asked between JRU 2006 and the other two surveys. However, based on the results in Table 2 we assume that for the JRU 2006 dataset respondents interpreted the question as narrower regarding what constitutes accident experience; hospitalized in that survey compared to traffic accidents or personal injury in AAP 2009 and AAP2010, respectively.

As explained in the previous section, two of our dataset use the SB format and one (AAP 2010) is based on OE questions. Therefore, to treat the data within the same model we need to convert the latter one to a binary yes and no answer before the pooling. To achieve that, we used the same bid levels as in the two other surveys (see Table 3) and distributed randomly a bid level at each observation and created a *Yes* variable such that for each respondent *i*:

$$Yes_i = \begin{cases} 1 \text{ if } WTP_i \geq Bid\ level_i \\ 0 \text{ if } WTP_i < Bid\ level_i \end{cases}$$

which is the same definition of the *Yes* variable as in the two other datasets. Thus, we could do the pooling and proceed to an analysis of the resulting dataset using *Yes* defined previously as a dependent variable (see, e.g., Cameron and Huppert, 1991; Ready et al., 1996).

In Table 3, the bid levels are described and the distributions of the acceptance of paying the attributed bid level are shown. This overview shows that the proportion of yes answer is decreasing with the bid level in each survey (with some exceptions) and a greater acceptance for a private safety device than for a public safety program at the same bid level.

---

[4] Bid levels used provided in Table 3.

[5] As a mean to mitigate hypothetical bias in CV studies questions on how certain respondents are about their answers have been used. Broadly, two formats exist: (i) a qualitative binary question (very certain or rather certain), and (ii) a scale as in JRU 2006. Only "very certain" responses are considered as "true" yes answers, i.e. a "real decision" would have been the same (yes), whereas those interpreted as not sure are treated as "no answers". For a description and evidence of preference-certainty questions mitigating/removing hypothetical bias, see Blumenschein et al. (2008) and Blomquist et al. (2009).

Table 2
Variable descriptions and descriptive statistics.

Variables	Description	JRU 2006 (SB)	AAP 2009 (SB)	AAP 2010 (OE)	Pooled Data (SB)
Private	= 1 if the question asked is about the private safety device	0.48	0.49	0.54	0.51
		(0.500)	(0.501)	(0.499)	(0.500)
Female	= 1 if the respondent is a female	0.54	0.48	0.53	0.52
		(0.499)	(0.500)	(0.500)	(0.500)
Age	Age of the respondent	45.20	44.32	42.96	43.98
		(15.96)	(15.69)	(15.56)	(15.73)
Inc_k	Monthly disposable income of the respondents' household, in SEK 1000 2017 price level	23.58	22.07	25.14	23.85
		(14.14)	(10.51)	(16.35)	(14.36)
Children	Number of children in the respondents' household	0.49	0.62	0.63	0.59
		(0.922)	(0.947)	(0.930)	(0.934)
Acc_exp	= 1 if the respondent or someone else close to him had any traffic accident experience	0.28	0.54	0.56	0.47
		(0.449)	(0.499)	(0.497)	(0.499)
JRU_2006	= 1 if obs. belongs to 2004 dataset				0.29
					(0.452)
AAP_2009	= 1 if obs. belongs to 1998 dataset				0.29
					(0.452)
AAP_2010	= 1 if obs. belongs to 2006 dataset				0.43
					(0.495)
N		438	438	656	1532

Means with standard deviations in parentheses.

Table 3
Proportion of yes answer per category of Bid for each subsample by Type.

Bid level[a]	JRU 2006		AAP 2009		AAP 2010	
	Private	Public	Private	Public	Private	Public
200	0.85	0.44	0.66	0.53	0.81	0.58
1000	0.48	0.25	0.68	0.31	0.64	0.31
2000	0.25	0.13	0.29	0.17	0.29	0.14
5000	0.24	0.05	0.16	0.13	0.15	0.07
10000	0.14	0.06	0.08	0.03	0.13	0.00
20000	0.09	0.05	0.15	0.00	0.02	0.00

[a] For each dataset, expressed in the year price level of money used in each survey.

4. The empirical model

The analysis of our data has been made in a parametric way using a logistic model to estimate the acceptance probability, π, such that:

$$\pi = \frac{1}{1 + \exp(-\Delta \wp)}$$

with $\Delta \wp$ the variation of the utility level. The variables presented in Table 1 were included to examine how different respondent or survey characteristics influence the probability accepting the bid presented to the respondents.[6]

We assume a non-negative distribution of the WTP, both for the private and public WTP. The mean WTP ($\bar{w}$) is therefore estimated as the area under the survival function for positive WTP values (Johansson, 1995):

$$\bar{w} = \int_0^\infty \frac{1}{1 + \exp(-\Delta \wp)} dw$$

Using the multivariate case, we can finally get $\bar{w}$ such that:

$$\bar{w} = -\left(\frac{1}{\beta}\right) \ln(1 + \exp(\alpha + \boldsymbol{x}\delta))$$

with β, α, $\boldsymbol{x}$ and δ the coefficient of the *Bid* variable, the intercept, the vector of covariates and the vector of associated coefficients. We estimate the mean WTP for both public and private safety based on mean

estimates of the different covariates depending on samples used in the regressions.

5. Results

Table 4 shows the results of the logit regressions of the *Yes* variable on our previously defined covariates. When using the data from all three surveys we find that the coefficient for *Bid* is highly statistically significant with the expected negative sign. Moreover, *Private* has a significant positive effect on the probability of accepting the bid, which is contrary to the predictions from section 2. These two effects are persistent when using the information about preference certainty available in AAP 2010 and JRU 2006. We include control variables for the different surveys, i.e. *JRU 2006* and *AAP 2009*. Neither of these variables are statistically significant, which is not surprising since many confounding factors can influence the respondents' answers between surveys.

Regarding individual and household characteristics, we do not find any effect from gender or number of children on the probability of accepting the bid in any of the regressions. We find that age and income have a negative and positive effect on the probability of saying yes except in the regression with Certainty = 10. Based on previous WTP studies on road safety, we further tested the age and income effect by also creating group dummies for these two variables (e.g., Andersson, 2007; Svensson, 2009; Haddak, 2016). After having created three group dummy variables each for age and income, we found statistically significance of the oldest group, i.e. respondents that are more than 50 years old, and for the richest group, i.e. respondents with a monthly income after taxes higher than SEK 30,000 for the full sample (not in the table, available from the authors upon request). Finally, we find weak evidence that accident experience also has a significant effect on the probability of saying yes to the bid.

Table 5 provides mean WTP estimates for private and public safety, and for different levels of certainty. The results confirm the findings from the regressions in Table 4 that for the same risk reduction the WTP for the private safety devise exceeds the one for the public safety measure. Hence, findings are robust and do not disappear if we restrict the sample to more certain respondents. Our findings will hence fail the validity test $\text{WTP}_{\text{public}} \geq \text{WTP}_{\text{private}}$.

6. Discussion

This study analyses the issue of the validity of the estimation of the

---

[6] We assume that readers are familiar with the logistic model, but for those who are not we recommend any textbook on econometrics or statistics.

Table 4
Results of logit regression of Yes on our explanatory variables.

	Full sample	AAP 2010 & JRU 2006		
		Full subsample	Certainty ≥ 8	Certainty = 10
Private	0.942***	1.073***	1.254***	1.285***
	(0.138)	(0.165)	(0.231)	(0.334)
Bid_k[a]	−0.217***	−0.235***	−0.218***	−0.252***
	(0.019)	(0.024)	(0.031)	(0.051)
Female	−0.136	0.077	0.049	−0.360
	(0.134)	(0.161)	(0.226)	(0.334)
Age	−0.019**	−0.018***	−0.018*	−0.019
	(0.005)	(0.005)	(0.008)	(0.012)
Inc_k	0.011*	0.011*	0.017*	0.001
	(0.005)	(0.005)	(0.007)	(0.011)
Children	0.025	0.025	0.015	−0.049
	(0.074)	(0.088)	(0.130)	(0.198)
Acc_exp	0.387**	0.404*	0.282	0.0636
	(0.139)	(0.168)	(0.234)	(0.344)
JRU_2006	0.096	0.116	0.124	0.245
	(0.169)	(0.174)	(0.239)	(0.343)
AAP_2009	0.149			
	(0.162)			
Intercept	−0.443	−0.377	−0.695	0.091
	(0.286)	(0.331)	(0.490)	(0.790)
N	1480	1068	586	307
pseudo R2	0.214	0.234	0.241	0.259

Standard errors in parentheses, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
a: Bid_k, bid in SEK 1000.
AAP 2010 reference group.

Table 5
Mean WTP based on regression results.

	Private safety device	Public safety program
Full sample	4557 [3693, 5420]	2218 [1684, 2753]
AAP 2010 & JRU 2006:	4270 [3506, 5034]	1927 [1451, 2403]
Full subsample		
Certainty ≥ 5	4309 [3427, 5190]	1910 [1373, 2447]
Certainty ≥ 6	4200 [3333, 5067]	1788 [1272, 2304]
Certainty ≥ 7	4528 [3352, 5703]	1752 [1119, 2386]
Certainty ≥ 8	4392 [3253, 5532]	1615 [1018, 2213]
Certainty ≥ 9	4545 [2871, 6219]	1480 [705, 2255]
Certainty = 10	3599 [2201, 4998]	1269 [579, 1958]

SEK 2017 price level (EUR 1 = SEK 9.60, www.riksbank.se, 2017-11-30).
95% confidence intervals in brackets.

WTP for road safety using the SP method. We provide a (construct) validity test based on theoretical and empirical findings that suggests that we expect to find $t_1 ≥ p_1$, i.e. the WTP for the public safety measure should be at least as high as for the private safety devise. However, the analysis of the three surveys in the same city in Sweden, with close to identical scenarios, suggests the opposite. i.e. WTP is higher for the private compared to the public safety measure.

It is hard to pin-point the reason why the validity test suggested in this study is rejected in the empirical analysis. One potential explanation could be a more severe lack of trust in the provision of the public good, in line with the discussion in Abu-Baker and Rouleau (2007). Moreover, the difficulty for people to conceptualize what is the real impact of their response in term of provision of goods must be considered. Both these effects could be more common for the public good measure, since the provision of the good may be less clear and/or a more widespread distrust in actual implementation by government policies. Respondents will then for the public safety measure tend to under-report their WTP since the valuation context often would involve a situation that is complex and unfamiliar to them. In JRU 2006 the test with a PPM (not in our sample) was not conclusive and it is unclear if the respondent found the condition credible in the local context. Moreover, there is also a risk that respondents in the public safety

scenario act more strategically, i.e. do not report their true preferences. This could be due to two reasons: (i) they see a chance of free-riding, i.e. they will still benefit from a public safety measure even if they do not contribute if others will, or (ii) they may state a lower WTP since they are concerned that funding collected for the safety program may be used for other purposes, or that taxes in general will be increased, so they understate their WTP.

One hypothesis that was tested was whether preference uncertainty could help explain our findings (Blomquist et al., 2009). However, the empirical findings suggest that the magnitude of the hypothetical bias is statistically less so for private compared to public WTP (List and Gallet, 2001; Murphy et al., 2005). This would suggest that, *ceteris paribus*, public WTP should exceed private WTP, which is the opposite to what we find when not controlling for preference certainty. When we consider preference certainty our findings can be interpreted as addressing hypothetical bias (mean WTP is reduced) but the relationship between private and public WTP is robust. Hence, hypothetical bias is not the reason why our findings do not pass the validity test.

The debate why estimated WTP does not always pass validity tests in SP is ongoing in the literature. In this paper we have focused on a question of high policy relevance, i.e. private vs. public WTP. Often WTP is framed as a private safety measure even if the findings are going to be used for public safety measures. In this study we have shown that this may be problematic and that thorough validity tests should be conducted. Eliciting WTP for road safety using SP studies is a challenge considering the known difficulties for individuals to understand small probabilities, behavioral aspects among road users/respondents, and the need to consider altruistic motives, which we showed can take different forms, when eliciting preferences for public safety measures. For instance, if some individuals seem to have altruistic preferences, it can be problematic for the SP studies because of a potential imperfect knowledge of others' benefits or if costs to others are not (or only partially) considered in the valuation (Gyrd-Hansen et al., 2016). Above we provided some suggestions on why the validity tests were not met in this study, but the data did not allow for any strong conclusions. We believe that we have raised a critical issue concerning the validity of elicited WTP for public safety measures. Future studies should aim at explaining the findings that private exceeds public WTP despite the evidence on altruistic preferences by eliciting information about altruistic preferences, attitudes towards public and private safety measure, and perception about effectiveness of measures and their effect on others.

References

Abu-Baker, A., Rouleau, G., 2007. Oculopharyngeal muscular dystrophy: recent advances in the understanding of the molecular pathogenic mechanisms and treatment strategies. Biochem. Biophys. Acta 1772, 173–185.

Andersson, H., 2007. Willingness to pay for road safety and estimates of the risk of death: evidence from a Swedish contingent valuation study. Accid. Anal. Prev. 39, 853–865.

Andersson, H., Lindberg, G., 2009. Benevolence and the value of road safety. Accid. Anal. Prev. 41, 286–293.

Andersson, H., Svensson, M., 2008. Cognitive ability and scale Bias in the contingent valuation method. Environ. Resour. Econ. (Dordr) 39, 481–495.

Baker, R., Robinson, A., Smith, R., 2008. How do respondents explain WTP responses? A review of the qualitative evidence. J. Socio. 37, 1427–1442.

Bateman, I., Brouwer, R., 2006. Consistency and construction in stated WTP for health risk reductions: a novel scope-sensitivity test. Resour. Energy Econ. 28, 199–214.

Bishop, R.C., Boyle, K.J., 2018. Reliability and validity in nonmarket valuation. Environ. Resour. Econ. (Dordr). https://doi.org/10.1007/s10640-017-0215-7.

Blomquist, G.C., Blumenschein, K., Johannesson, M., 2009. Eliciting willingness to pay

without Bias using follow-up certainty statements: comparisons between Probably/ Definitely and a 10-point certainty scale. Environ. Resour. Econ. (Dordr) 43, 473–502.

Blumenschein, K., Blomquist, G.C., Johannesson, M., Horn, N., Freeman, P., 2008. Eliciting willingness to pay without Bias: evidence from a field experiment. Econ. J. 118, 114–137.

Bohm, P., 1972. Estimating demand for public goods: an experiment. Eur. Econ. Rev. 3 (2), 111–130.

Cameron, T.A., Huppert, D.D., 1991. Referendum contingent valuation estimates: sensitivity to the assignment of offered values. J. Am. Stat. Assoc. 86 (416), 910–918.

Carson, R.T., Groves, T., 2007. Incentive and informational properties of preference questions. Environ. Resour. Econ. (Dordr) 37, 181–210.

Chanel, O., S. Luchini and J. Shogren. 2005 Does Charity Begin at Home for Pollution Reduction?, Groupement de Recherce en Economie Quantitative d'Aix Marseille.

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

de Blaeij, A., Florax, R.J.G.M., Rietvald, P., Verhoef, E., 2003. The value of statistical life in road safety: a meta-analysis. Accid. Anal. Prev. 35 (6), 973–986.

Dickie, M., Messman, V.L., 2004. Parental altruism and the value of avoiding acute illness: are kids worth more than parents? J. Environ. Econ. Manage. 48 (3), 1146–1174.

European Commission, 2014. Guide to Cost-benefit Analysis of Investment Projects: Economic Appraisal Tool for Cohesion Policy 2014-2020. Brussels, Belgium European Commission, Directorate-General for Regional and Urban policy.

Freeman, A.M., Herriges, J.A., Kling, C.L., 2014. The Measurement of Environmental and Resource Values. RFF Press.

Gyrd-Hansen, D., Kjær, T., Seested Nielsen, J., 2016. The value of mortality risk reductions. Pure altruism – a confounder? J. Health Econ. 49 (Supplement C), 184–192.

Haddak, M.M., 2016. Estimating the willingness-to-pay for road safety improvement. Transp. Res. Procedia 14, 293–302.

Holmes, T.P., 1990. Self-interest, altruism, and health-risk reduction: an economic analysis of voting behavior. Land Econ. 66 (2), 140–149.

Hultkrantz, L., Lindberg, G., Andersson, C., 2006. The value of improved road safety. J. Risk Uncertain. 32 (2), 151–170.

Jacobsson, F., Johannesson, M., Borgquist, L., 2007. Is Altruism Paternalistic? Econ. J. 117, 761–781.

Johannesson, M., Johansson, P.-O., O'Conor, R.M., 1996. The value of private safety versus the value of public safety. J. Risk Uncertain. 13 (3), 263–275.

Johansson, P.-O., 1995. Evaluating Health Risks: An Economic Approach. Cambridge University Press, Cambridge, UK.

Johnston, R.J., Boyle, K.J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., Vossler, C.A., 2017. Contemporary guidance for stated preference studies. J. Assoc. Environ. Resour. Econ. 4 (2), 319–405.

Jones-Lee, M.W., 1991. Altruism and the value of other people's safety. J. Risk Uncertain. 4, 213–219.

Jones-Lee, M.W., 1992. Paternalistic altruism and the value of a statistical life. Econ. J. 102, 80–90.

Liebe, U., Hundeshagen, C., Beyer, H., von Cramon-Taubadel, S., 2016. Context effects and the temporal stability of stated preferences. Soc. Sci. Res. 60, 135–147.

List, J.A., Gallet, C.A., 2001. What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? Environ. Resour. Econ. (Dordr) 20 (3), 241–254.

Liu, J.-T., Hammitt, J.K., Wang, J.-D., Liu, J.-L., 2000. Mother's willingness to pay for her own and her child's health: a contingent valuation study in Taiwan. Health Econ. 9 (4), 319–326.

Murphy, J., Allen, G.P., Stevens, T.H., Weatherhead, D., 2005. A meta-analysis of hypothetical Bias in stated preference valuation. Environ. Resour. Econ. (Dordr) 30 (3), 313–325.

Ready, R.C., Buzby, J.C., Hu, D., 1996. Differences between continuous and discrete contingent valuation estimates. Land Econ. 72 (3), 397–411.

Robinson, L.A., Hammitt, J.K., 2015. Research synthesis and the value per statistical life. Risk Anal. 35 (6), 1086–1100.

Samuelson, P.A., 1954. The pure theory of public expenditure. Rev. Econ. Stat. 36 (4), 387–389.

Svensson, M., 2009. The value of a statistical life in sweden estimates from two studies using the "Cerainty approach" calibration. Accid. Anal. Prev. 41, 430–437.

Svensson, M., Vredin Johansson, M., 2010. Willingness to pay for private and public road safety in stated preference studies: Why the difference? Accid. Anal. Prev. 42, 1205–1212.

Trafikverket. (2017). "Vision Zero." Retrieved November 30, 2017.

Vázquez Rodríguez, M.X., León, C.J., 2004. Altruism and the economic values of environmental and social policies. Environ. Resour. Econ. (Dordr) 28 (2), 233–249.