

Bateman, Ian J.; Covey, Judith; Loomes, Graham

Working Paper

Valuing risk reductions: Testing for range biases in payment card and random card sorting methods

CSERGE Working Paper EDM, No. 05-02

Provided in Cooperation with:

The Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia

Suggested Citation: Bateman, Ian J.; Covey, Judith; Loomes, Graham (2005) : Valuing risk reductions: Testing for range biases in payment card and random card sorting methods, CSERGE Working Paper EDM, No. 05-02, University of East Anglia, The Centre for Social and Economic Research on the Global Environment (CSERGE), Norwich

This Version is available at:

<https://hdl.handle.net/10419/80296>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

**Valuing risk reductions: Testing for range biases
in payment card and random card sorting
methods**

by

**Ian J. Bateman, Judith Covey and
Graham Loomes**

CSERGE Working Paper EDM 05-02

**Valuing risk reductions: Testing for range biases in
payment card and random card sorting methods**

by

**Ian J. Bateman¹, Judith Covey² and
Graham Loomes¹**

¹ **Centre for Social and Economic Research on the Global Environment,
University of East Anglia, Norwich NR4 7TJ, UK**

² **University of Durham**

Author contact details:

**Ian Bateman - email: i.bateman@uea.ac.uk
Tel: +44 (0) 1603 593125
Fax: +44 (0) 1603 593739**

Acknowledgements

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. This work was part of the interdisciplinary research programme of the ESRC Research Centre for Social and Economic Research on the Global Environment (CSERGE).

The authors wish to thank Sven Baaz for collecting the survey data reported in this study.

ISSN 0967-8875

Abstract

Concerns have been raised that the payment card (PC) format widely used in contingent valuation surveys of health treatments and risk reductions is subject to range bias. In response recent surveys have adopted an alternative random card sorting (RCS) approach – though this approach's susceptibility to range bias has not yet been formally tested. This study addressed this gap and showed, somewhat unexpectedly, that the RCS procedure was no less vulnerable to range bias than the PC method for eliciting both monetary values of health risk reductions and non-monetary estimates of death rates. Conclusions for future research initiatives are drawn.

Keywords: willingness-to-pay; contingent valuation; risk reduction; elicitation format

1. INTRODUCTION

The rise and promulgation of cost-benefit approaches to public sector decision making has brought to prominence the question of how analysts should value preferences for non-market goods such as health treatments, safety improvements or environmental benefits. An apparently attractive solution to this problem is provided by the contingent valuation (CV). For example, the CV method can be used to value risks to life and health by asking individuals for their maximum willingness to pay (WTP) for small hypothetical reductions in risk (for a recent example of a CV study for valuing a risk reduction see Carlsson, Johansson-Stenman & Martinsson, 2004 - for reviews see Jones-Lee, 1989; Viscusi, 1993; and Beattie et al., 1998).

One of the methodological questions raised by CV research is how to elicit such information. Researchers are faced with the choice between open-ended (OE), iterative bidding (IB), dichotomous choice (DC), or payment card (PC) formats. The dilemma is that empirical comparisons across these techniques reveal that they typically fail to yield the same values for a given good (Kealy & Turner, 1993; Kristrom, 1993; Bateman et al., 1995; Boyle et al., 1996; Bateman, Langford & Rashbash, 1999; Ryan, Scott & Donaldson, 2004), but rather reveal regular patterns of difference (Bateman & Jones, 2003).

Whilst the DC and PC formats both compare favourably to the OE and IB methodsⁱ at this time there is no universal consensus about which method is the preferred alternative (Smith, 2000; Bateman et al. 2002). Both approaches are being extensively used in CV studies though it is notable that the PC method has emerged in recent years as the popular choice for valuing health benefits. While the number of health valuation studies almost tripled between 1995-1999 and 2000-2004 the number of studies using the PC increased five-fold (Smith, 2003; 2004a).

One of the reasons for the PC method's popularity is its ease of use. Respondents are simply presented with a range of monetary amounts and asked to identify the maximum amount they would be willing-to-pay (WTP). Because this approach allows the analyst to identify maximum WTP for each respondent (or at least narrow it down to some tractable range), the PC approach avoids the statistical inefficiency of the DC approach while involving lower cognitive loads than the OE method (Mitchell & Carson, 1989; Bateman et al., 2002). Moreover, by asking respondents to identify all the amounts they are sure they would pay and all the amounts they are sure they would not pay the method can also provide information about the range of values over which people are uncertain (Dubourg, Jones-Lee & Loomes, 1997).

The PC approach is not however without its problems. One particular concern is that the method may be subject to range and centring biases (Rowe, Schulze & Breffle, 1996; Dubourg et al., 1997; Bateman, Lovett & Brainard, 2003; Whynes, Wolstenhulme & Frew, 2004). For example, both Dubourg et al. (1997) and Bateman et al. (2003) show that payment cards with higher upper values may elicit higher mean *and* median values – the latter suggesting that this is not simply the result of higher upper values allowing bigger outliers, but that the whole distribution of responses is influenced in an upward direction.

In response to this problem some recent CV studies have used an alternative elicitation method which involves a randomized card sorting (RCS) procedure (Carthy et al., 1999; Chilton et al., 2004; Smith, 2004b). Rather than the PC procedure of presenting all amounts together on one sheet, the RCS approach is to write individual amounts on separate cards. These are then shuffled in the presence of respondents and cards are drawn one at a time in a random order. As each card is presented the respondent is asked to sort them into one of three categories: amounts they are sure they would pay; amounts they are sure they would

not pay; and amounts they are unsure about. This innovative procedure offers a number of potential advantages over the typical PC.

By reducing the PC format to a series of “take it or leave it” choices, it simplifies the valuation task further and increases the respondent’s engagement with the exercise. Furthermore, by avoiding explicit presentation of the full range of values to the respondents, the RCS may hope to attenuate the range bias described above. Finally, by overtly shuffling the value cards in front of the respondent the analyst reinforces the message that the first value seen is randomly selected and therefore should not be interpreted as having any particular significance. The intention is to weaken the anchoring heuristic evident in so many contexts (Tversky & Kahneman, 1974; Chapman & Johnson, 2002).

To-date however the only empirical evidence in support of these suggested advantages is a recent study by Smith (2004b) which compared three versions of the PC administered either over the phone or in face-to-face interviews. The version of the PC where the values were listed from high-to-low produced significantly higher values than versions where the values were listed from low-to-high or when presented in random order from a pack of shuffled cards. This result suggests that reading the list from high-to-low had a tendency to inflate the valuations relative to the other versions. Moreover, the fact that the random order version produced valuations which were similar to the low-to-high version suggests that the starting card value did not have an undue influence on the estimates – otherwise one might expect the results from the random version to lie between the high-to-low and low-to-high formats since the average starting card value would typically lie somewhere between. The study also found that significantly fewer respondents who were presented with random order version had values that they were unsure about paying and fewer changed their initial maximum WTP when challenged whether that amount was indeed their maximum. With supporting qualitative data suggesting that the random format encouraged people to think more closely about their values, Smith (2004b) drew the conclusion that the estimates derived from the RCS version were most likely to reflect true WTP.

Whilst these conclusions are clearly encouraging for the RCS procedure it is notable that Smith’s study did not test for range biases. This paper presents such a test by comparing the extent to which the PC and RCS methods are affected by the range of amounts presented on the list or cards. In the next section we describe the method used. Section 2 presents results and Section 3 discusses their implications and concludes.

2. METHOD

2.1 Design

The design centred around questions which asked people to value reductions in the risks of road injuries. To enable comparability with previous research which has demonstrated PC range effects we used two of the scenarios used by Dubourg et al. (1997) where significantly higher WTP responses were found when respondents were presented with a PC list of amounts from 0-£1500 than when a PC range of 0-£500 was used. Indeed for these two scenarios the difference was so marked that the mean confidence intervals did not overlap at all: that is to say, the average amount the 0-£1500 sub-sample were sure they *would* pay was strictly greater than the average amount the 0-£500 sub-sample said they were sure they would *not* pay.

For each of the scenarios, respondents valued safety features which might be added to their car. They were told that this feature would halve their risk of a particular type of injury. The first scenario elicited willingness to pay to halve a 24 in 100,000 annual risk of sustaining an injury labelled **S** (1-4 weeks hospitalization with moderate to severe pain, followed by some permanent restrictions to work and leisure activities); the second to halve a 36 in 100,000 annual risk of sustaining an injury labelled **X** (1-4 weeks hospitalization with slight to moderate pain, temporary restrictions to work and leisure activities, 1-3 year recovery to normal health).

Survey respondents were randomly allocated to one of two elicitation procedures – the typical PC procedure or the RCS approach. Within each of these procedures respondents were randomly allocated to be presented with amounts on the lists or cards ranging from 0-£500 (LOW RANGE) or 0-£1500 (HIGH RANGE). The experiment therefore took a 2 (elicitation method: PC vs. RCS) x 2 (range: LOW v HIGH) between-participants factorial design, with respondents' best estimates of their maximum WTP as the dependent variable. Because the RCS procedure does not explicitly reveal the range of values in advance we hypothesised that there would be a smaller difference between the LOW and HIGH responses in the RCS condition than the PC condition.

We also tested this hypothesis for two non-monetary quantitative estimation tasks which differed from the valuation task and from each other in ways which we considered might influence people's susceptibility to positional cues provided by the range. First, we included an estimation task for which respondents might have expected the investigator to know the correct answer: respondents were asked to provide an estimate of the number of people killed in road accidents each year in the UK. Second, we asked respondents to reveal a personal value for which they would not expect the investigator to know to correct answer (although they might expect them to have some idea about the typical or average for the population): namely, the number of miles the respondent expected to travel by car in the coming 12 months. Here our hypothesis was that in the case where the researchers might be expected to know the correct answer, respondents might perceive that there was greater informational content in the ranges presented to them than in the case where the respondents were likely to be more knowledgeable than the researchers. Hence our hypothesis was that responses would be more influenced by range in the first case than the second.

2.2 Pilot Study

In order to assist in the setting of response ranges, a pilot sample of 60 respondents (recruited from the same population as the main survey) were asked the following target questions, using an open-ended (OE) response format:

- "What is the most you would be willing to pay to halve your risk of sustaining Injury S?" (WTP-S)

- “What is the most you would be willing to pay to halve your risk of sustaining Injury X?” (WTP-X)
- “How many people do you believe are killed in road accidents each year in the UK?” (ROAD-DEATHS)
- “How many miles would you estimate that you will travel as a vehicle driver or passenger in the next 12 months?” (MILEAGE)

On the basis of these 60 responses, we established that the largest value we wanted to use for the LOW range for the CV questions (i.e., £500), was located at the 90th percentile of the OE responses for WTP-X and the 86th percentile for WTP-S. The largest value we wanted to use for the HIGH range (i.e., £1500) was between 7 and 8.5 standard deviations higher. We therefore selected largest values for the LOW and HIGH ranges for the ROAD-DEATHS and MILEAGE questions which matched those percentile positions, and the other values used in those scales were also selected in such a way that the values closest to the median and mean for each task were located at similar points on the scale. The resulting scales are shown in Figure 1. The values in normal type were common to both the LOW and HIGH range scales, while the values in bold type are specific to each scale.

2.3 Main Survey

The respondents for the main survey consisted of a convenience sample of 240 adults resident in the Aberdeen area of the UK who were interviewed at their own homes on a one-to-one basis. Respondents were randomly assigned to receive one of four versions of the interview – denoted by PC-LOW, PC-HIGH, RCS-LOW, RCS-HIGHⁱⁱ.

The ROAD-DEATHS question was the first question in the interview in all four versions. Those respondents who were randomly assigned to the PC versions were shown either the LOW or HIGH range for this question and asked to:

- Put the letter L next to each number they believed was certainly lower than the number of people killed in road accidents each year;
- Put the letter H next to each number they believed was certainly higher than the number of people killed in road accidents each year;
- And put a question mark next to each number they were unsure about.

The interviewer then reminded the respondent of the largest number they had put L against and the smallest number they had put H against and went on to ask them for their best estimate of the number of people killed in road accidents each year in the UK. Respondents were prompted that their answer didn’t have to be a number shown on the list.

The respondents allocated to the RCS procedure were asked to sort a shuffled pack of cards, with each card showing one of the numbers from the range on the corresponding PC. The interviewer instructed the respondent to sort the cards one at a time onto a template to indicate whether they believed the number on the card was ‘TOO LOW’ or ‘TOO HIGH’ for the number of people killed in road accidents each year, or whether they were ‘UNSURE’ about the number. When all cards were sorted, the interviewer noted the largest-L and the smallest-H values, and then elicited the best estimate in the same way that was done in the PC version.

Following the ROAD-DEATHS question, the interview continued in all versions with a number of general questions including one in which respondents were asked to rate how they believed their risk of being killed in a road accident compared with the average for people in this country. These general questions were followed by the two CV scenarios in which respondents were presented with the descriptions of Injury S and Injury X and told that they were going to be asked how much they would value specified reductions in the risks of these different kinds of injuries. The injury descriptions and CV scenario (which was

common to all four versions) is reproduced in Appendix A. The first CV question asked respondents about their maximum WTP to halve their risk of sustaining injury S during the next 12 months, while the second CV question asked for their maximum WTP to halve their risk of sustaining injury X during the next 12 months.

The PC and RCS procedures used were very similar to the procedures used for the ROAD-DEATHS question apart from the following small differences. The PC procedure asked respondents to put ticks next to amounts they were certain they would pay, and crosses next to amounts they were certain they would not pay. Similarly the RCS procedure asked respondents to sort the cards onto a template with spaces headed 'CERTAIN WOULD PAY' 'CERTAIN WOULD NOT PAY' and 'UNSURE'.

The last section of the interview included some questions about respondents' car usage. The final MILEAGE question then asked them to estimate their coming year's mileage as a car driver or passenger, with respondents being required to make decisions about whether the numbers shown on the PC or RCS shuffle cards were too low or too high.

3. RESULTS

Table 1 presents summary statistics of best estimate responses for each of the four target questions as a function of both the METHOD of elicitation (PC vs. RCS) and RANGE of the PC or RCS shuffle-cards (LOW vs. HIGH).

Although we retested the WTP-S and WTP-X scenarios of the Dubourg et al. (1997) exercise, we did not replicate the magnitude of range effect with the PC format observed in that study; in the present case, the ratios between the HIGH and LOW range mean responses were 1.18 and 1.17 as opposed to 2.65 and 2.33 in the earlier study.

Contrary to our expectations, the RCS procedure did not reduce the size of the range effects on the WTP responses. In fact the RCS procedure produced a larger effect, with ratios of 2.01 and 1.94 for WTP-S and WTP-X. This result is confirmed by the results of 2 (METHOD: PC vs. RCS) x 2 (RANGE: LOW vs. HIGH) analyses of variance (ANOVAs) reported in Table 1. The two-way ANOVAs show significant RANGE x METHOD interactions, with subsequent one-way ANOVAs indicating that the range only had a significant effect on the WTP responses elicited using the RCS method. Nor does this appear to be a result driven by a few outliers: although (as is usual in value elicitation studies) the distribution of responses is right-skewed, the ratios of medians – 2.0 for WTP-S and 2.33 for WTP-X – are no less pronounced than the ratios of means.

The results for the ROAD-DEATHS and MILEAGE questions also fail to support the supposed superiority of the RCS method over the PC method in combating range effects: indeed in both cases the RCS range effects are more significant than those generated by the PC approach, the latter being insignificant for the MILEAGE question (the METHOD x RANGE interaction effects were not significant). Additional analyses reported below aim to provide further insight into these findings.

One possible explanation of why the RCS procedure produced a range effect at least as big or bigger than that generated by the PC method might be an anchoring of respondents' best estimates on the first card shown in the RCS treatment. As Table 2 shows, these starting values were typically much lower in the LOW range version than the HIGH range version for all four questions (e.g., medians £25 vs. £300 for the LOW vs. HIGH versions for the WTP-S questionⁱⁱⁱ). Generally, the proportions of respondents placing their first cards into the CERTAIN WOULD PAY/ TOO LOW category was not particularly sensitive to the magnitude of the starting value. However, although there is some evidence of respondents' best estimates being correlated with either the first card values presented or the median of the first five cards, the pattern is uneven, with correlations being insignificant for a number of tests (detailed analyses shown in Table 2)^{iv}. So despite the signs of "yea-saying" (which is also suggested by the fact that between 2 and 3 of the first five cards were placed in the CERTAIN WOULD PAY/ TOO LOW categories in both the LOW and HIGH range conditions), anchoring on the first card values cannot fully explain the stronger range effects shown in the RCS condition. This conclusion is further supported by the ANCOVA (analysis of covariance) results reported in Table 1 which show that the inclusion of the first card value as a covariate does not change the statistical significance of any of the RANGE effects in the RCS sub-sample.

Table 1: Summary statistics of ‘best estimates’ as a function of METHOD of elicitation (PC vs. RCS) and RANGE of values presented (LOW vs. HIGH)

	WTP-S		WTP-X		ROAD-DEATHS		MILEAGE	
LOW RANGE	PC	RCS	PC	RCS	PC	RCS	PC	RCS
Mean	£142	£162	£130	£126	6,929	8,691	6,333	6,946
Median	£100	£100	£78	£75	5,000	5,500	5,250	6,500
Standard Deviation	£198	£195	£185	£163	6,509	8,851	5,753	4,855
HIGH RANGE	PC	RCS	PC	RCS	PC	RCS	PC	RCS
Mean	£168	£327	£152	£245	10,086	12,104	7,188	9,470
Median	£100	£200	£90	£175	7,250	7,000	5,500	7,000
Standard Deviation	£186	£361	£183	£290	11,471	14,500	4,413	7,406
M_{HIGH}/M_{LOW}	1.18	2.01	1.17	1.94	1.46	1.39	1.14	1.36
TWO-WAY ANOVAs								
	F(p)	Effect size (d)	F(p)	Effect size (d)	F(p)	Effect size (d)	F(p)	Effect size (d)
RANGE	17.3 (<.001)	0.53	13.9 (<.001)	0.49	6.50 (.011)	0.34	6.91 (.009)	0.35
METHOD	11.9 (.001)	0.44	4.26 (.040)	0.27	2.06 (.15)	0.19	8.78 (.003)	0.39
RANGExMETHOD	6.2 (.014)	N/A	4.15 (.043)	N/A	.01 (.95)	N/A	2.14 (.15)	N/A
ONE-WAY ANOVAs								
	F(p)	Effect size (d)	F(p)	Effect size (d)	F(p)	Effect size (d)	F(p)	Effect size (d)
RANGE (PC SUB-SAMPLE)	2.47 (.12)	0.33	1.59 (.21)	0.24	3.18 (.077)	0.34	0.97 (.33)	0.17
RANGE (RCS SUB-SAMPLE)	15.5 (<.001)	0.71	15.2 (<.001)	0.71	3.32 (.071)	0.35	6.48 (.012)	0.40
ANCOVAs (FIRST CARD COVARIATE)								
	F(p)	Effect size (d)	F(p)	Effect size (d)	F(p)	Effect size (d)	F(p)	Effect size (d)
RANGE (RCS SUB-SAMPLE)	8.99 (.003)	0.50	5.76 (.018)	0.42	2.71 (.103)	0.30	5.62 (.020)	0.43

Table 2: Summary statistics of the first cards presented in the random card sorting (RCS) procedure

	<u>WTP-S</u>			<u>WTP-X</u>			<u>ROAD-DEATHS</u>			<u>MILEAGE</u>		
	LOW	HIGH	(pooled)	LOW	HIGH	(pooled)	LOW	HIGH	(pooled)	LOW	HIGH	(pooled)
FIRST CARD PRESENTED												
Mean	83	499	(291)	89	427	(266)	7704	16,253	(11,979)	5467	10,333	(7900)
Median	25	300	(100)	30	250	(75)	5000	10,000	(8750)	4000	7000	(5000)
Rank (1-14)	5.5 th	9 th	--	6 th	8.5 th	--	8 th	7 th	--	7 th	7 th	--
Spearman's correlation with best estimate	.223*	.054	(.230**)	.027	.143	(.195**)	.070	-.056	(.030)	.220*	0.031	(.133)
% respondents placing 1 st card into CERTAIN WOULD PAY/ TOO LOW category	71.7%	46.7%	(59.2%)	61.8%	43.3%	(52.2%)	46.7%	33.3%	(40.0%)	46.7%	41.7%	(44.2%)
MEDIAN OF FIRST FIVE CARDS												
Mean	79	249	(164)	94	208	(153)	5362	10,525	(7944)	5200	8867	(7033)
Median	50	200	(100)	50	175	(100)	3000	10,000	(6250)	5000	7000	(6000)
Rank (1-14)	7 th	8 th	--	7 th	7.5 th	--	7 th	7 th	--	8 th	7 th	--
Spearman's correlation with best estimate	.042	-.069	(.149)	.162	.067	(.224**)	.185	.070	(.169*)	.182	.245*	(.243***)
Number of cards (1-5) placed into CERTAIN WOULD PAY/ TOO LOW category												
Mean	2.98	2.23	(2.61)	2.64	2.45	(2.54)	2.58	1.8	(2.19)	2.47	2.03	(2.25)
Median	3	2	(3)	3	3	(3)	3	2	(2)	2.5	2	2

Note: *p<.10, **p<.05, ***p<.01 (two-tailed)

Table 3: Mean numbers of responses in the CERTAIN WOULD PAY/ TOO LOW, UNSURE and CERTAIN WOULD NOT PAY/ TOO HIGH categories.

		WTP-S		WTP-X		ROAD-DEATHS		MILEAGE	
		PC	RCS	PC	RCS	PC	RCS	PC	RCS
CERTAIN WOULD PAY/ TOO LOW									
	LOW	7.73	7.97	7.23	7.11	6.5	6.85	6.73	7.03
	HIGH	5.70	7.03	5.47	6.47	4.5	4.82	4.80	5.53
TWO-WAY RANGE x METHOD ANOVAs		F (p)		F (p)		F (p)		F (p)	
	RANGE	12.98 (<.001)		7.86 (.005)		35.2 (<.001)		13.1 (<.001)	
	METHOD	3.63 (.058)		1.03 (.31)		0.96 (.33)		1.19 (.28)	
	RANGExMETHOD	1.79 (.18)		1.70 (.19)		<0.01 (.96)		0.21 (.65)	
UNSURE									
	LOW	0.92	1.93	0.83	1.2	2.02	2.67	1.27	2.17
	HIGH	1.03	2.02	0.64	1.58	2.47	2.87	1.23	2.08
TWO-WAY RANGE x METHOD ANOVAs		F (p)		F (p)		F (p)		F (p)	
	RANGE	0.26 (.61)		0.28 (.60)		2.21 (.14)		0.12 (.72)	
	METHOD	25.4 (<.001)		12.8 (<.001)		5.75 (.017)		30.0 (<.001)	
	RANGExMETHOD	0.01 (.93)		2.45 (.12)		0.33 (.57)		0.03 (.87)	
CERTAIN WOULD NOT PAY/ TOO HIGH									
	LOW	5.35	4.10	5.93	5.69	5.48	4.48	6.0	4.8
	HIGH	7.27	4.95	7.88	5.95	7.03	6.32	7.97	6.38
TWO-WAY RANGE x METHOD ANOVAs		F (p)		F (p)		F (p)		F (p)	
	RANGE	11.2 (.001)		6.46 (.012)		19.6 (<.001)		14.3 (<.001)	
	METHOD	18.5 (<.001)		6.26 (.013)		5.05 (.026)		8.81 (.003)	
	RANGExMETHOD	1.67 (.19)		3.78 (.053)		0.14 (.71)		0.17 (.68)	

We then explored how the process of placing shuffle cards into the CERTAIN WOULD PAY/ TOO LOW, CERTAIN WOULD NOT PAY/ TOO HIGH, and UNSURE categories compared to the process of completing the PCs. The mean numbers of responses in the three categories are presented in Table 3 along with the results of 2 (RANGE: LOW vs. HIGH) x 2 (METHOD: PC vs. RCS) ANOVAs for each of the four questions. The first notable feature about these data was the tendency of the RCS method to produce significantly more UNSURE responses than the PC. However the larger bands of uncertainty in the RCS condition were not symmetrical. Whereas the RCS method resulted in significantly fewer values placed in the CERTAIN WOULD NOT PAY/ TOO HIGH category than the PC method no significant differences were found between the methods in the numbers of values placed in the CERTAIN WOULD PAY/ TOO LOW categories.

Notably the mean numbers of UNSURE responses were not significantly different between the LOW and HIGH range conditions. However, the non-linear nature of the scale (i.e., the increments between values at the low end were smaller than the increments at the high end) had the effect of increasing the differences between the respondents' largest CERTAIN WOULD PAY/ TOO LOW and smallest CERTAIN WOULD NOT PAY/ TOO HIGH responses to a greater degree in the HIGH range condition than the LOW range condition. This is illustrated in Figure 1 which presents the mean CERTAIN WOULD PAY/ TOO LOW, BEST ESTIMATE and CERTAIN WOULD NOT PAY/ TOO HIGH values in graphical form for each question and method of elicitation. Taking for example the responses to the WTP-S question, the difference between the mean CERTAIN WOULD PAY and CERTAIN WOULD NOT PAY values was £154 (£271-£117) in the RCS-L condition and £374 (£610-£236) in the RCS-H condition. Hence, any tendency for respondents either just to split the difference between the CERTAIN WOULD PAY-WOULD NOT PAY values when they provided their best estimates, or to take positional cues from the information provided by the full range of cards which was now exposed, would tend to have a bigger effect on increasing the best estimate provided in the HIGH range condition than the LOW range condition. Although this explanation implies that the biasing effects of RANGE occurred to some extent after the process of card sorting the effect was of course supplemented by the fact that the RCS method tended to produce a larger number of UNSURE responses.

Figure 1: Low and high range scales for monetary valuation and non-monetary valuation questions

LOW range scales			HIGH range scales		
WTP-S, WTP-X	ROAD- DEATHS	MILEAGE	WTP-S, WTP-X	ROAD- DEATHS	MILEAGE
£1	100	1000	£1	100	1000
£5	300	1500	£10	500	2000
£10	500	2000	£20	1000	3000
£15	750	2500	£50	3000	4000
£20	1000	3000	£75	5000	5000
£30	2000	3500	£100	7500	6000
£50	3000	4000	£150	10,000	7000
£75	5000	5000	£200	12,500	8000
£100	7500	6000	£300	15,000	10,000
£150	10,000	7000	£400	20,000	12,000
£200	12,500	8000	£500	25,000	15,000
£300	15,000	10,000	£750	30,000	20,000
£400	20,000	12,000	£1000	40,000	30,000
£500	25,000	15,000	£1500	60,000	40,000

4. DISCUSSION

In this study our main aim was to examine and compare the extent to which the PC and RCS methods were susceptible to range bias. Our hypothesis that the RCS procedure would be less susceptible to such bias because it disguises the range of values being presented was not supported by our results. If anything we found range bias to be as much, if not more, of a problem with the RCS method than the PC method. In light of the fact that this method has already been used in CV studies (Carthy et al., 1999; Chilton et al., 2004; Smith, 2004b) and has been identified as a promising alternative procedure for use in CV studies (Bateman et al., 2002), these results suggest the need for some caution.

However, while it is important for the research community who are interested in CV methodology to be alerted to these results, it would be somewhat premature to write the RCS method off on the basis of this study alone. After all, Smith's (2004b) results painted a somewhat more positive picture and we cannot be sure that the poor performance that we found with the RCS procedure would necessarily apply across all types of CV surveys which can vary not only in terms of the types of goods they ask their respondents to value but also by their mode of administration.

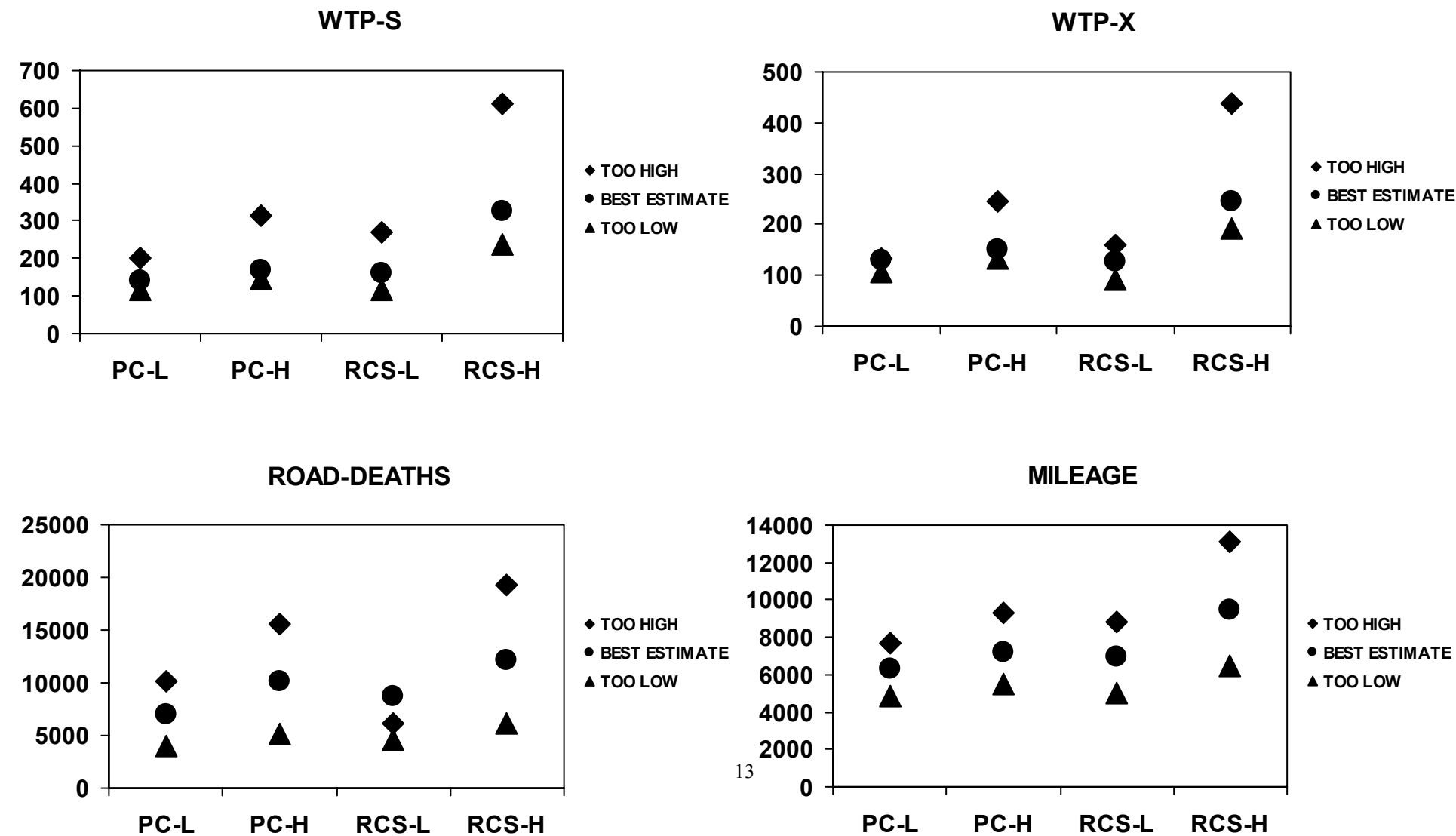
The results of the present study might be quite specific to the way in which it was administered; respondents were recruited by canvassing people in the street or on the doorstep, and interviewed them there and then in their own homes. This is a rather different approach to theCarthy et al. (1999) and Chilton et al. (2004) surveys which recruited respondents by letter or telephone and interviewed them by appointment in hotel meeting rooms or in their own homes. Interviewing people on-the-spot rather than by pre-arranged interview is likely to not only have placed different cognitive demands on the respondents but also, as noted by Tourangeau, Rips & Rasinski (2000) in their comprehensive review of the psychological aspects of survey methodology, have had different impacts on the perceived legitimacy and importance of the study and the depth of processing of survey questions.

We might conjecture that the on-the-spot method is likely to not only reduce the respondents' perceptions of the legitimacy and importance of the study relative to the pre-arranged interview method but also the depth of processing of the questions because of a feeling of time pressure. Both a failure to take the interview sufficiently seriously and lack of time might have led respondents to rush the sorting of their cards which may in turn have resulted in rather too many values being sorted into the unsure category (which as noted in the previous section provides part of the explanation why the range had such an influence). It could be that the RCS method works better in the pre-arranged interview situation where respondents are not so likely to feel rushed into giving their answers – and although we cannot be sure that the Chilton et al. (2004) data was not subject to a range bias, it is worth noting that the majority of respondents in that study placed no values at all into the unsure category. It would seem therefore that the choice of elicitation method might be dependent on the mode of survey administration though supporting evidence is clearly needed to establish which methods work best under which circumstances.

As well as comparing the PC and RCS methods we also compared the extent of range biases across different types of survey questions. Bearing in mind that the end-points of the LOW and HIGH range scales were set at equivalent points in the distribution of open-ended (OE) responses obtained in the pilot study, the most notable finding here was that the range had a bigger impact on the ROAD-DEATHS question than either the MILEAGE or WTP questions. This result is consistent with the hypothesis that the values presented by the researchers are more likely to be perceived as informative under conditions where respondents believe that the person who has designed the survey knows the correct answer to the question. Because people expect questions to meet standards of conversational informativeness by conveying information that they should take into account in determining

their answer (Grice, 1975; Sperber & Wilson, 1986), they may well infer that the scale has been constructed with the correct value in mind. Of course the robustness of this finding needs to be established for other types of factual questions before definitive conclusions can be drawn, though it highlights the importance of understanding not only the cognitive processes associated with different elicitation methods, but also the social psychological

Figure 2: Mean CERTAIN WOULD PAY/ TOO LOW, BEST ESTIMATE, and CERTAIN WOULD NOT PAY/ TOO HIGH responses for each question and method of elicitation



References

- Bateman, I. J., Carson, R. T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., Pearce, D. W., Sugden, R., & Swanson, J. (2002). *Economic Valuation with Stated Preference Techniques: A Manual*. Cheltenham: Edward Elgar.
- Bateman, I.J. & Jones, A.P., (2003) Contrasting conventional with multi-level modelling approaches to meta-analysis: An illustration using UK woodland recreation values, *Land Economics*, 79(2): 235-258.
- Bateman, I.J., Langford, I.H. and Rasbash, J. (1999) Elicitation effects in contingent valuation studies, in Bateman, I.J. and Willis, K.G. (eds.) *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford University Press, p511-539.
- Bateman, I. J., Langford, I. H., Turner, K., Willis, K. G., & Garrod, G. D. (1995). Elicitation and truncation effects in contingent valuation studies. *Ecological Economics*, 12, 161-179.
- Bateman, I. J., Lovett, A.A. and Brainard, J.S. (2003) *Applied Environmental Economics: a GIS Approach to Cost-Benefit Analysis*, Cambridge University Press, Cambridge
- Beattie, J., Chilton, S., Cookson, R., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Robinson, A., & Spencer, A. (1998). *Valuing health and safety controls: A literature review*. Norwich: HSE Books.
- Boyle, Kevin J., Richard C. Bishop, and Michael P. Welsh. 1985. "Starting Point Bias in Contingent Valuation Bidding Games." *Land Economics*, 61 (May):188-194.
- Boyle, K. J., Johnson, F. R., McCollom, D. W., Desvousges, W. H., Dunford, R. W., & Hudson, S. P. (1996). Valuing public goods: Discrete versus continuous contingent valuation responses. *Land Economics*, 72, 381-396.
- Carlsson, F., Johansson-Stenman, O., & Martinsson, P. (2004). Is transport safety more valuable in the air? *Journal of Risk and Uncertainty*, 28(2), 147-163.
- Carthy, T., Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., & Spencer, A. (1999). On the contingent valuation of safety and the safety of contingent valuation: Part 2 - The CV/SG "chained" approach. *Journal of Risk and Uncertainty*, 17, 187-213.
- Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 120-138). Cambridge: Cambridge University Press.
- Chilton, S., Covey, J., Jones-Lee, M., Loomes, G., & Metcalf, H. (2004). Valuation of health benefits associated with reductions in air pollution. London: DEFRA.
- Dubourg, W. B., Jones-Lee, M. W., & Loomes, G. (1997). Imprecise preferences and survey design in contingent valuation. *Economica*, 64, 681-702.
- Grice, H. (1975). Logic and conversation. In P. Cole & T. Morgan (Eds.), *Syntax and Semantics: Vol 3, Speech Acts* (pp. 41-58). New York: Seminar Press.
- Jones-Lee, M. W. (1989). *The Economics of Safety and Physical Risk*. Oxford: Blackwells.
- Kealy, M. J., & Turner, R. W. (1993). A test of the equality of closed-ended and open-ended contingent valuations. *American Journal of Agricultural Economics*, 75, 321-331.
- Kristrom, B. (1993). Comparing continuous and discrete contingent valuation questions. *Environmental and Resource Economics*, 3, 63-71.
- Mitchell, R. C., & Carson, R. T. (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington DC: Resources for the Future.
- Rowe, R. D., Schulze, W. D., & Breffle, W. S. (1996). A test for payment card biases. *Journal of Environmental Economics and Management*, 31, 178-185.
- Ryan, M., Scott, D. A., & Donaldson, C. (2004). Valuing health care using willingness to pay: A comparison of the payment card and dichotomous choice methods. *Journal of Health Economics*, 23, 237-258.

- Smith, R. D. (2000). The discrete-choice willingness to pay question format in health economics: Should we adopt environmental guidelines? *Medical Decision Making*, 20, 194-206.
- Smith, R. D. (2003). Construction of the contingent valuation market in health care: A critical assessment. *Health Economics*, 12, 609-628.
- Smith, R.D. (2004a) Development of CV in health over the past 20 years, presented at the Valuing the Public Sector conference, Business School, Stephenson Centre, University of Newcastle upon Tyne, 29th September 2004.
- Smith, R. D. (2004b). It's not what you do it's the way that you do it: The effect of different payment card format and survey administration on willingness to pay for health gain. Unpublished Manuscript. University of East Anglia.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Cambridge MA: Harvard University Press.
- Stalhammar, N.-O. (1996). An empirical note on willingness to pay and starting-point bias. *Medical Decision Making*, 16, 242-247.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Viscusi, W. K. (1993). The value of risks to life and health. *Journal of Economic Literature*, 31, 1912-1946.
- Whynes, D. K., Wolstenhulme, J. L., & Frew, E. (2004). Evidence of range bias in contingent valuation payment scales. *Health Economics*, 13, 183-190.

Appendix A

Injury Descriptions and CV Scenario

The next two questions concern the two types of injury described in this card [*PRESENT INJURY SHOWCARD*]. What I am interested in is how much you would value reductions in the risks of these different kinds of injuries. [*READ OUT INJURY SHOWCARD*]

INJURY S	INJURY X
<p>In hospital</p> <ul style="list-style-type: none">• 1-4 weeks• Moderate to severe pain <p>After hospital</p> <ul style="list-style-type: none">• Some pain gradually reducing, but may recur when you take part in some activities• Some restrictions to leisure and possibly some work activities for rest of your life	<p>In hospital</p> <ul style="list-style-type: none">• 1-4 weeks• Slight to moderate pain <p>After hospital</p> <ul style="list-style-type: none">• Some pain/ discomfort, gradually reducing• Some restrictions to work and leisure activities, steadily reducing• After 1-3 years, return to normal health with no permanent disability

In each case we should like you to imagine a situation which is somewhat artificial, but which will help us to get a clearer picture of the importance you attach to reducing the risks of these different kinds of injuries.

You will be asked to imagine that you can choose to have an extra safety feature fitted to the vehicle you travel in. This feature is quickly and easily fitted, but to work effectively it must be renewed every 12 months. If you have it fitted, then for the next 12 months:

- it will reduce your risks of a particular kind of injury by half;
- but it will not change the risks of anyone else in the vehicle, and it will not affect your risk of any other kind of injury.

Of course, it is difficult to imagine an actual safety device that would work exactly like this. But what we are asking you to do is to think about each kind of injury on its own, and to think about an annual payment for an annual reduction in the risk of each particular kind of injury. We are also asking you to think just about your own risk - we know that you may be concerned about other people's risks, but we shall take account of them by asking a large number of other people for their values too.

So please focus on the most you would be willing to pay for a device which reduces your risk of each kind of injury. When answering, please ignore the possible financial effects of the injury such as loss of income, vehicle repair costs etc., and concentrate on the personal, non-financial effects such as distress, pain and temporary or permanent disability or damage to your health.

Notes

ⁱThe OE method is generally discounted by researchers because of its high non-response rate and tendency to exhibit free-riding and strategic overbidding tendencies (e.g., Bateman et al., 1995) whereas the IB method is particularly susceptible to starting point bias (e.g., Boyle, Bishop & Welsh, 1985; Stalhammar, 1996).

ⁱⁱ Copies of the stimulus materials and interview protocols used by the investigators are available from the corresponding author.

ⁱⁱⁱ It is worth noting that random chance has resulted in higher than expected differences. With 14 cards in each pack, we might have expected the median to be nearer £50-£75 for the LOW range and £150-£200 for the HIGH range.

^{iv} It is worth noting that the correlations found in this study between the first card value and the best estimates were of a similar magnitude to the correlation of .142 reported in Chilton et al. (2004).