Journal of Environmental Management 79 (2006) 221–231
Journal of Environmental Management
www.elsevier.com/locate/jenvman

Comparing contingent valuation and contingent ranking: A case study considering the benefits of urban river water quality improvements

I.J. Bateman[a,b], M.A. Cole[c,*], S. Georgiou[a,b], D.J. Hadley[b]

[a]Programme on Environmental Decision Making (PEDM), School of Environmental Sciences, University of East Anglia, Norwich, UK
[b]Centre for Social and Economic Research on the Global Environment (CSERGE), School of Environmental Sciences, University of East Anglia, Norwich, UK
[c]Department of Economics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Received 16 September 2003; accepted 22 June 2005
Available online 22 September 2005

Abstract

This paper contrasts applications of both the contingent valuation (CV) and contingent ranking (CR) methods as applied to a common issue, the valuation of improvements to the water quality of an urban river (the River Tame, running through the city of Birmingham, UK). Building upon earlier experimental work, the CV design used ensures that respondents are fully aware of all impending valuation tasks prior to undertaking any one of those tasks. Such an approach is directly comparable to the CR design for which full awareness of all options is a pre-requisite. Findings indicate that the CV responses exhibit strong internal consistency with expected relationships observed between values and theoretically expected parameters. External comparisons show that CR valuations are substantially larger than those elicited through CV (with protest votes excluded), and that the response rate for the CR survey is significantly higher than that for the CV survey.

1. Introduction

The past 30 years have witnessed the institutional use of cost–benefit analysis (CBA) move from a minority sport to a central preoccupation. Early one-off analyses have been replaced with systematic guidelines mandating use of CBA principles as the bedrock of decision making (DETR (Department of the Environment, Transport and the Regions), 1998). Yet in practice this application remains patchy with one of the major stumbling blocks concerning methods for valuing the benefits or costs of environmental change (Pearce, 1998; Hanley, 2001).

Interest in the evaluation of such methods has for some considerable time focussed upon the contingent valuation (CV) method (Mitchell and Carson, 1989; Bateman and Willis, 1999), wherein survey respondents are directly questioned about their willingness to pay (WTP) or willingness to accept (WTA) compensation for increments or decrements of, typically, public goods.[1] However, the CV method has been the focus of considerable academic debate, focussing on issues such as the sensitivity of CV derived values to the magnitude or 'scope' of goods under consideration (Boyle et al., 1994; Diamond and Hausman, 1994; Kahneman and Knetsch, 1992; Schkade and Payne, 1994). While other commentators have mounted a robust defence of the method (Hanemann, 1994, 1996; Carson, 1997; Carson et al., 1998), this debate has led to increased interest in alternatives to CV. Notably the UK Environment Agency has recently commissioned research into the viability of applying the Contingent Ranking (CR) method (Smith and Desvousges, 1986) to assessments of river water management options (GIBB Ltd, 2001).

Like the CV approach, the CR method also relies upon survey responses. However, in a CR study respondents are asked to rank or rate a series of 'product profiles' that describe products having specific attribute levels. The expressed trade-offs between respondents' assessments and product attributes can then be used to estimate the

* Corresponding author. Tel.: +121 414 6639; fax: +121 414 7377.
E-mail address: m.a.cole.1@bham.ac.uk (M.A. Cole).

0301-4797/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2005.06.010

[1] For a CV analysis of all four Hicksian welfare change measures see Bateman et al. (2000).

marginal utility of each attribute. Since price is usually one of the attributes, it is possible to rescale the ranking (utility) index in monetary terms and so derive WTP estimates for particular attribute bundles. Such ranking activities are especially useful for valuing environmental programs, which often have several components (see, e.g. Lareau and Rae, 1985; Garrod and Willis, 1998; and, Foster and Mourato, 2000).

Whilst CR methods are attractive since they avoid the explicit elicitation of WTP from respondents, the process of ranking imposes a significant cognitive burden upon respondents that is positively related to the number of attributes included and the number of alternatives presented (Foster and Mourato, 2002). The level of this burden is such that the reliability of ranked data has been called into question. In particular the results of Ben-Akiva et al. (1991) provide evidence to suggest that ranked data are unreliable, in the sense that choice models for different ranks were found to differ significantly. Foster and Mourato (2002) note that possible explanations for this may either be due to responses being determined by differing decision protocols according to level of rank; or, as a result of increasing noise corresponding to the depth of the ranking task. The same authors develop tests of consistency of CR results and provide empirical evidence that whilst a significant proportion of CR responses are not fully coherent, only a small minority are systematically inconsistent, and further that the presence of inconsistent rankings does not have a statistically significant effect on estimates of WTP.

Despite the interest of institutions such as the Environment Agency there have been few studies comparing CV and CR formats. Those that have made such a comparison suggest that systematic differences may arise with inferred WTP from a CR exercise typically being greater than that elicited through CV (see for instance Desvousges and Smith, 1983). Given the absence of theoretical drivers for such differences,[2] most commentators have fallen back upon psychological critiques of the methods. Stevens et al. (2000) provide three reasons why CR and CV results may differ. Firstly, the fact that substitutes are often made more explicit in the ranking format may encourage respondents to explore their preferences and trade-offs in greater depth (see also Boxall et al., 1996). Secondly, the psychological process of ranking in the CR format is somewhat different to the process underlying decisions about WTP in the CV format (Tversky et al., 1988). Irwin et al. (1993) argue that the relative prominence of money in a CV format may increase the weight accorded to money losses (in a WTP exercise) relative to that in a CR study where money is just one of a number of attributes of any given product profile. Thirdly, because it is easier in a CR experiment to express indifference to the choices (by ranking them equally), then non-response and protest zero-bidding behaviour may be less of a problem than in CV exercises.

Building upon the experimental work of Bateman et al. (2001), a further reason for differences between CV and CR estimates may be that, while respondents facing a ranking exercise necessarily have to be shown all options prior to undertaking any ranking task, in CV studies respondents may or may not be made aware of the entire choice set of all valuation opportunities prior to commencing valuation tasks. Bateman et al., show that advance disclosure of the full choice set results in valuations which exhibit significantly greater internal consistency (e.g. with respect to scope sensitivity) and robustness to question ordering effects (i.e. invariance to the order in which valuation tasks are presented; problems regarding which are noted by Boyle et al., 1994) than do valuations where the full set of options is revealed to respondents in a stepwise manner with valuation tasks interspersed between each expansion of the choice set. Accordingly this paper presents the first comparison of such an advance disclosure CV format with a CR design. We believe this provides a fairer comparison of these two methods given that, in both cases respondents are aware of all options prior to undertaking any valuation or ranking task.

Given the interest of the Environment Agency in applying CR techniques to assessment of river water quality improvements, we choose such an example to provide the empirical motivation for our comparison of CR and CV techniques, this being the first such comparison of its kind to be conducted in the UK to date. The chosen case study concerns potential schemes to improve the water quality of the River Tame which passes through the city of Birmingham, UK. The river flows through predominantly urban areas and is only one of a handful of UK rivers to be classified by the Environment Agency as having very poor water quality. However, the river does have ecological and recreational potential and passes through residential areas, playing fields and a country park. Clearly, any realisation of this potential will involve expense which is likely to fall largely on local authorities and hence ultimately on the local population. It is therefore important to examine the public's WTP for such clean-up schemes.

Alongside the comparison of the CR and CV techniques we also conduct a number of internal consistency checks. Within the CV study we assess the robustness of estimates to changes in the ordering of valuation tasks and examine whether derived values vary in relation to the distance of a respondent's residence to the river. Both the CV and CR estimates are also examined for evidence of scope sensitivity.

The remainder of the paper is organised as follows: Section 2 describes study design in terms of the various levels of river water quality improvement being considered and the CV and CR designs applied. This also describes the various internal consistency tests used within this study. Section 3 presents findings from both these internal checks and

[2] Further information on the economic principles underlying contingent valuation and contingent ranking can be found in Georgiou et al. (2000).

the results obtained from our CV and CR exercises. Finally Section 4 presents conclusions derived from this study.

2. Study design

As mentioned above, the case study used for this exercise concerned improvements to water quality in the River Tame passing through the City of Birmingham, UK. While it was once a high quality river providing habitat for trout and similar game fish, the Tame is now classed as one of the UK's most polluted rivers with fish stocks virtually non-existent and plant growth, insects, birds and animal life all severely limited. Direct human use is consequently extremely limited with the river classified as being unsuitable even for boating.

The survey questionnaire used to elicit the CV and CR responses, opened by asking respondents for details regarding their residency in Birmingham, together with their patterns of use of the River Tame (how often they visit the river, for what purpose, etc.). The survey then proceeded to ask a number of attitudinal questions employing Likert scales to assess respondents' perceptions of the importance of environmental issues relative to other social issues and, of those environmental issues, which they believe to be the most important.[3] Respondents were then asked for their perceptions of the water quality in the river before being provided with some detailed information regarding past and present state of the water quality in the River Tame.

In order to undertake a viable CR exercise and to assess the internal consistency of CV findings, three nested levels of river quality improvement were used for both exercises, which for simplicity we shall refer to as a small (S), medium (M) and large (L) improvement. It was decided to present respondents with the CV questions prior to the CR questions. We do not claim that derived results would be invariant to changes in this order, however such concerns were not central to our objectives and the within-person approach adopted here should enhance the likelihood of consistency between derived valuations. Consequently any differences between CV and CR values derived from such a design should be treated as a cause for significant concern.

As noted both the CV and CR exercises concerned three nested improvements to water quality. In the CV part of the exercise these were described textually making reference to a table shown to respondents (described subsequently) detailing the impact of each improvement upon three attributes of water quality: fishing; plants and wildlife; and boating and swimming (full details of these descriptions together with complete questionnaires are provided in Georgiou et al., 2000a).[4] Respondents were told that these three improvements were alternatives to each other, of which only one would be implemented, and therefore should be evaluated relative to a common baseline of the current situation, which was also described. It was made clear to respondents that values for all three improvements would be elicited, and this was then undertaken using a direct open-ended question asking the respondent to state their maximum WTP for the good in question. The open-ended approach has been criticised as liable to result in downwardly biased estimates of WTP (partly as a result of free-riding behaviour) by Carson et al. (1999) who advocate use of a single dichotomous choice (DC) format. However, as demonstrated by Giraud et al. (1999),[5] the DC approach faces its own problems in nested good, scope sensitivity situations such as the present study. Given this, the statistical efficiency of open-ended estimates and the knowledge that, if anything, derived estimates are liable to be lower than those derived using other formats, provides a clear expectation for our CV study.

The specified payment vehicle was an annual increase in council tax paid by respondents. This was felt to be realistic in the sense that any improvements would indeed be partly funded by local authorities and as such result in higher council tax levels.[6] Furthermore, it was felt that the other possible mode of payment, namely an increase in water rates, was inappropriate due to the degree of public distrust and opposition to private water institutions which exists, stemming from the politics of water privatisation in the 1980s (see Georgiou et al., 2000b for some discussion on this).

A simple test of the consistency of CV-derived values was undertaken by varying the order in which the three goods were presented to and valued by each respondent. Roughly half the sample respondents faced a 'bottom–up' (B) treatment in which the S good was valued first followed by goods M and then L. Conversely the remainder of the sample faced a 'top–down' (T) ordering where the order of

[3] These questions are elicited using a Likert type scale, where 1 = not important and 5 = very important.

[4] Georgiou et al. (2000a) is available for download at: http://www.uea.ac.uk/env/cserge/pub/wp/gec/gec_2000_18.pdf.

[5] The Giraud et al. study fails to find significant internal (within respondent) scope sensitivity to substantial changes in the magnitude of an environmental good. The setting of DC bid amounts between nested goods is problematic, risking accusations of induced scope sensitivity if bid amounts increase with the scope of goods, while risking implausibility if amounts decline as the scope of goods increases. It is also debateable whether the strong incentive compatibility properties of a single DC question are retained across multiple valuation tasks.

[6] Using taxes as a payment vehicle has well known disadvantages, particularly with regard to incentives (Chilton and Hutchinson, 1999). However, we feel the use of taxes as a payment vehicle in this case is defensible on two fronts. Firstly, and as we note in the text, taxes are the only realistic option for this scenario. Secondly, there is evidence that respondents prefer taxes as a payment vehicle for certain public goods as they provide some form of certainty of provision that voluntary contributions do not. Bateman et al. (2003, p. 28) test three payment vehicles including taxes and a donation to a targeted charitable trust fund. They found that while the tax vehicle may have given a stated willingness to pay lower than true willingness to pay this was still considerably higher than stated willingness to pay from donations to the trust find.

Table 1
Water quality improvement scenarios, together with payment levels for the contingent ranking exercise

Water quality level	Characteristics of water quality level			Council tax increase	
	Fishing	Plants and wildlife	Boating and swimming	Per year	Per month
Large improvement in quality (level L)	Trout and Salmon return Good game fishing possible	Increase in plants and wildlife Possible for Otters to survive	Suitable for boating and swimming	£30	£2.50
Medium improvement in quality (level M)	Some game fish species return (e.g. perch) Good enough for fishing	Increase in number and types of insects Greater numbers of birds and wildlife	Suitable for boating, but not swimming	£15	£1.25
Small improvement in quality (level S)	A few fish species return (e.g. roach)	More plants would grow, waterfowl can use river	Suitable for boating, but not swimming	£5	£0.42
Current situation (level C)	Fish stocks virtually non-existent	Plant growth, insects, birds and animal life limited	Unsuitable for boating or swimming	£0	£0

valuation tasks was for good L, then M and finally S. We can develop a simple notation to differentiate the derived values in terms of treatment and ordering as follows:

$$\text{Sample T}: \quad L_T^1, M_T^2, S_T^3$$

$$\text{Sample B}: \quad S_B^1, M_B^2, L_B^3$$

here subscripts denote presentation treatment and superscripts denote the order in which any given good was presented and valuation elicited. Given that goods are presented as alternatives to each other we do not expect substitution effects between valuations and therefore (under certain assumptions[7]) anticipate that the value for any given good will be invariant to the order in which it is presented and the degree of scope sensitivity between goods should not vary between the T and B presentation modes. Note that we have no prior expectations regarding scope sensitivity other than invariance across treatments; the presence or absence of scope sensitivity across nested goods is a purely empirical question reflecting the underlying preferences of respondents.

Following the CV valuation tasks all respondents were asked open-ended questions concerning the motives underlying their responses. Such questions permit analysis both of the factors underpinning higher levels of WTP but also to disentangle true zero from protest responses and understand the motives behind such responses.

The subsequent CR exercise again referred to the same three water quality improvements, however here the table defining these goods was supplemented with information on a pricing attribute. The resultant description of the schemes is as detailed in Table 1. Here, all columns other than that concerning the relevant increase in council tax are identical to those shown to respondents in the course of the CV exercise. The increases in council tax were determined via a pilot exercise (Georgiou et al., 2000a) which looked at the relationship between the ranking of goods and their attribute levels including council tax increases and set the latter to ensure that this permitted sufficient variability in ranking (setting payments too high or low would simply result in a lack of variation across observed ordering). The CR exercise asked respondents to rank the four options (S, M, L and the current situation C) in order of preference. This provides an ordinal equivalent of the prior CV exercise, with CR valuations being inferred from rank orderings.

The final section of the questionnaire elicited information concerning the socio-economic and demographic characteristics of the respondents including gender, age, education, household numbers, marital status, employment status, income and the council tax charge band of the respondents' property.[8]

3. Results

A sample of 675 useable observations was gathered through face-to-face interviews conducted during August and September 1999 at respondents' places of residence in the Birmingham area,[9] Table 2 provides some general characteristics of this sample, including a dissaggregation into the type T and B treatments used to elicit CV responses, with full information supplied by Georgiou et al. (2000a). Note that the majority of the sample (54%) had not visited the River Tame whereas a minority (11%) visited every

[7] These assumptions concern the extent to which respondents engage in strategic behaviour for the purchase of public goods. Importantly the advance disclosure of all valuation tasks means that any strategies engaged in by respondents should be constant across all valuations irrespective of presentation ordering. For further discussion see Bateman et al. (2001).

[8] Interviewers were also asked to subjectively assess, post-interview, respondents' understanding of, and consideration given to, the valuation questions.

[9] Note that the focus of the survey was on the comparison of CV and CR methods and as such the sampling frame was nor specifically designed to capture a representative sample of the population of Birmingham households and hence produce WTP estimates for use in policy decisions. However, the sample was randomised by the random selection of streets throughout the Birmingham area within which households were approached.

Table 2
General characteristics of respondents

Characteristic	Full sample	Type T subsample	Type B subsample
Sample size and % of total	675 (100.0)	346 (51.3)	329 (48.7)
Mean income (£) and % don't know/refused	19023 (50.0)	18872 (51.5)	19172 (48.6)
Age (mean)	48.4	48.9	47.8
Sex (% of men)	45.9	43.6	48.3
Employment (% employed)[a]	51.1	49.2	53.1

[a] Full or part-time.

week. Therefore, we expect a high variability in the use value element of any stated values.

3.1. CV results

Of the full sample,[10] 23.1% were unable to state a WTP, the most commonly cited reasons being that the exercise was too hypothetical or too difficult (20.5% of those unable to state a WTP), or that more information on costs was needed (15.3%). The sample also contains a high level of zero bids, with 39.0% of respondents stating a zero WTP for all three water quality levels. The most commonly cited reason for a zero WTP was that the respondent felt they already paid enough tax (35.7% of those stating zero WTP), followed by the claim that existing funds should be used (16.3%). Others felt that the water companies (5.7%) or the polluters (3.4%) should pay. Only 9.8% of those stating a zero WTP claimed not to care about the river, indicating that most who expressed a zero WTP were registering a protest vote. Overall, these findings suggest considerable resistance to the CV format, perhaps deriving from the open-ended format and/or the use of tax as a payment vehicle.

Table 3 presents univariate WTP statistics for the three goods valued in the CV exercise. The table presents the results according to whether protest zero responses are included or excluded from the analysis. In this respect, protests are defined as zero WTP responses that are given in protest against some aspect of the contingent commodity, even though the improvement under consideration is actually valued. It should be noted that there is some debate regarding the legitimacy of including or excluding zero responses from any analysis, which can be somewhat resolved if one considers what is actually being valued under each point of view. Those who advocate removal of protest votes seek to measure the value of the good or service as their objective, whilst those advocating inclusion of protests do so on the grounds that they are interested in estimating the value which theoretically coincides with the amount that would be approved for spending under a referendum - the manner in which public policy is most often decided (see Jorgensen and Syme, 2000; Randall et al., 1981; Randall et al., 2001; Alvarez-Farizo et al., 1999 for further discussion on the issue of protest responses). The table further disaggregates the values according to the T and B presentation orderings used in the valuation exercise. The significance of differences between WTP values can be assessed by inspection of confidence intervals that were derived using bootstrap techniques (Efron and Tibshirani, 1993). It can be seen that although the exclusion of protest zero responses increased the valuation estimates found for each improvement as expected, the differences in means were nevertheless not statistically significant ($P<0.05$) across any of the corresponding samples.

As noted previously, we have no prior expectations regarding the presence or absence of scope across the three goods valued (other than, assuming these are normal goods, non-declining WTP as scope increases). However, the univariate results presented in Table 3 indicate clear scope sensitivity at the full sample level (irrespective of whether protests are included or excluded). Internal consistency of the CV findings is explored by formal non-parametric testing of differences across goods - but within presentation treatments - with results provided in Table 4. Here all of these 'internal' *tests* (i.e. within subject tests, as opposed to 'external' tests, which relate to between subject comparisons) prove significant ($P<0.01$) for comparisons between each of the water quality improvements considered (again irrespective of whether protests are included or excluded).

Some testing of external consistency is also possible by comparing between the goods valued first across presentation treatments. For the T and B treatments these are the large and small good respectively, and hence this represents a test of external scope. As shown in Table 5, the external test proves significant ($P<0.05$).

In addition, formal external testing of possible ordering effects was undertaken by comparing values for a given good across the T and B treatments. Table 6 reports tests of the null hypothesis of equality between values for each good. Results indicate that we cannot reject this hypothesis i.e. within each pair the difference in WTP is not statistically significant. Thus, we find no statistically significant evidence of ordering effects.

[10] One observation was omitted on the grounds that the stated WTP was 10 SD from the mean, and stemmed from a respondent who was a full-time student. It was therefore felt that the stated figure was unlikely to be a true reflection of the respondent's WTP.

Table 3
Univariate analysis of WTP (£ per household per annum) for River Tame water quality improvements

	Water quality improvement, mean WTP (CI)[a]		
	Small (S)	Medium (M)	Large (L)
Protests included			
Full sample ($n=518$)	7.60 (6.14–9.05)	12.07 (9.83–14.33)	18.12 (15.13–21.48)
Sample T ($n=270$) Sequence L, M, S	7.08 (5.03–9.42)	11.43 (8.70–14.86)	17.43 (13.40–22.39)
Sample B ($n=248$) Sequence S, M, L	8.17 (6.24–10.60)	12.75 (9.60–16.47)	18.86 (14.46–23.50)
Protests excluded			
Full sample ($n=410$)	9.60 (7.89–11.50)	15.24 (12.67–18.18)	22.89 (19.28–27.26)
Sample T ($n=209$) Sequence L, M, S	9.15 (6.55–12.08)	14.77 (10.99–18.93)	22.52 (17.45–28.77)
Sample B ($n=201$) Sequence S, M, L	10.07 (7.74–12.71)	15.73 (12.11–19.86)	23.28 (18.45–29.04)

The importance of protecting the environment was the main reason given for stating a positive WTP (as stated by 55.4% of those giving a positive WTP).
[a] Bootstrapped 95% confidence intervals.

Table 4
Internal tests of scope sensitivity

Internal scope test	WTP sample including protests			WTP sample excluding protests		
	% difference in Mean WTP for levels considered	Paired difference *t*-statistic ($P>\|t\|$)	Wilcoxon signed-rank *Z* statistic ($P>\|t\|$)	% difference in mean WTP for levels considered	Paired difference *t*-statistic ($P>\|t\|$)	Wilcoxon signed-rank *Z* statistic ($P>\|t\|$)
1. $L_t^1 > M_t^2$	52.5	6.514 (0.0000)	9.821 (0.0000)	52.5	6.665 (0.0000)	9.747 (0.000)
2. $M_t^2 > S_t^3$	61.44	6.189 (0.0000)	9.477 (0.0000)	61.42	6.318 (0.0000)	9.199 (0.000)
3. $L_t^1 > S_t^3$	146.19	6.6484 (0.0000)	9.715 (0.0000)	146.12	6.810 (0.0000)	9.645 (0.000)
4. $L_b^3 > M_b^2$	47.92	6.228 (0.0000)	9.262 (0.0000)	49.99	6.3420 (0.0000)	9.376 (0.000)
5. $M_b^2 > S_b^1$	56.06	6.402 (0.0000)	8.882 (0.0000)	56.20	6.5265 (0.0000)	8.818 (0.000)
6. $L_b^3 > S_b^1$	130.84	7.0603 (0.0000)	9.381 (0.0000)	131.18	7.2296 (0.0000)	9.306 (0.000)

Table 5
External test of scope sensitivity

External scope test	WTP sample including protests			WTP sample excluding protests		
	% difference in mean WTP for levels considered	Paired difference *t*-statistic ($P>\|t\|$)	Wilcoxon signed-rank *Z* statistic ($P>\|t\|$)	% difference in mean WTP for levels considered	Paired difference *t*-statistic ($P>\|t\|$)	Wilcoxon signed-rank *Z* statistic ($P>\|t\|$)
$L_t^1 = S_b^1$	113.3	3.6625 (0.0003)	2.075 (0.0380)	123.6	3.9522 (0.0001)	2.835 (0.0046)

Following conventional norms and recognised guidelines (Mitchell and Carson, 1989; Bateman et al., 2002), overall validity testing was undertaken through the estimation of a multivariate bid function (Table 7).[11] This permits inspection of the consistency of findings with prior expectations derived from theory and empirical regularities observed in the literature.

The bid function was modelled using a maximum likelihood random effects approach (Greene, 1990) in order to allow for possible intra-respondent correlation between responses given for the three water quality improvements. Since the responses are also censored at zero (negative WTP amounts are not permitted), Tobit analysis is appropriate. The explanatory factors include dummy variables for two of the three water quality improvements (the omitted one being used as the base case), as well as a dummy for the order used to elicit values (top down or bottom up). A variety of other potential explanatory variables were investigated with the best fitting model of WTP responses being reported in Table 7.[12]

The bid function described in Table 7 takes WTP responses for the largest good (L) as the base case. In line with the findings of Table 3 we find that dummy variables for responses to the M and S goods are

[11] The bid function analysis is undertaken on the data set that excludes protest responses since the factors determining protests may be quite different from those underlying the variation in WTP values of the good itself. It is this latter variation that the bid function analysis seeks to explain.

[12] Results obtained from other model specifications are available from the authors.

Table 6
Ordering effect tests

Ordering test	WTP sample including protests		WTP sample excluding protests	
	Two sample *t*-statistic ($P>\|t\|$)	Mann–Whitney Z statistic ($P>\|t\|$)	Two sample *t*-statistic ($P>\|t\|$)	Mann–Whitney *Z* statistic ($P>\|t\|$)
$L_t^1 = L_b^3$	0.4443 (0.6570)	0.795 (0.4264)	0.1920 (0.8479)	0.404 (0.6863)
$M_t^2 = M_b^2$	0.5897 (0.5556)	1.184 (0.2362)	0.3516 (0.7253)	0.816 (0.4142)
$S_t^3 = S_b^1$	0.7037 (0.4819)	1.546 (0.1220)	0.4877 (0.6261)	1.283 (0.1993)

Table 7
Random effects Tobit regression of bid function describing responses to the CV WTP questions

Explanatory variable	Coefficient	Standard error
Constant	44.77*	4.64
Medium (=1 if WTP refers to M (medium) good; 0 otherwise)	−12.19*	1.51
Small (1 if WTP refers to S (small) good; 0 otherwise)	−21.71*	1.54
Order (=1 if order is top–down (T); 0 if order is bottom–up (B))	−2.72	2.11
Distance from respondent's residence to river (miles)	−0.27	0.46
Age of respondent (years)	−0.71*	0.07
Knowledge (=1 if respondent has prior knowledge of River Tame water quality, 0 otherwise)	9.52*	2.24
Work (=1 if respondent is in full time/part time work, 0 otherwise)	−10.95*	2.42
σ_e	44.66*	1.12
σ_u	17.58*	0.50
Log-likelihood	−3545.50	–
Wald statistic	310.03*	–

N=409; σ_e and σ_u relate to the random parameters of the model, where *u* is the random disturbance characterising each individual, and *e* is the conventional error term. *Significant at the 1% level.

significant and that estimated coefficients have expected signs (i.e. are negative, indicating lower WTP compared to the baseline case of a large improvement in quality). Similarly, in line with Table 6 we find that an indicator for whether the T or B orderings are used to elicit values proves clearly insignificant (i.e. there is no evidence of any ordering effects). For the most part other relationships also accord with expectations. WTP is higher for those who had prior knowledge (and thereby, arguably, higher concern) regarding the water quality problems affecting the River Tame. The negative sign of the coefficient estimated for the age variable indicates that WTP decreases with increasing age (as might be expected since the more elderly, retired, members of the sample are likely to have lower incomes than those younger and in work). A further negative relation was found with increasing distance between the respondent's permanent address and the River Tame. Whilst this again seems highly plausible and accords with previous empirical findings (Sutherland and Walsh, 1985; Bateman et al., 2000a,b), the coefficient is insignificant. Finally, and contrary to expectations (where we might expect that those in employment will have higher incomes and hence higher WTP than those not in employment), the sign of the estimated coefficient on the employment variable is negative. One possible explanation for this result might arise from a trade-off between working and leisure hours;

Table 8
Frequency distribution for the rankings of water quality and payment

Ranking[a]	Frequency	Percent
LMSC	208	31.37
LMCS	2	0.30
LSMC	1	0.15
LSCM	0	0
LCMS	7	1.06
LCSM	1	0.15
MLSC	36	5.43
MLCS	2	0.30
MSLC	61	9.20
MSCL	51	7.69
MCLS	0	0
MCSL	0	0
SLMC	0	0
SLCM	0	0
SMLC	71	10.71
SMCL	11	1.66
SCLM	0	0
SCML	32	4.83
CLMS	4	0.60
CLSM	0	0
CMLS	1	0.15
CMSL	0	0
CSLM	0	0
CSML	175	26.40
Total	663	100.00

[a] The rankings are from highest to lowest level, with the alternatives defined as follows: C=status quo, S=small improvement, M=medium improvement and L=large improvement.

those in work may simply be less familiar with, and make less use of, this environmental feature.

3.2. CR results

Following the contingent valuation questions, respondents were asked to complete a contingent ranking exercise in which they had to rank combinations of water quality improvements and payment levels (as defined in Table 1). In contrast to the CV exercise, respondents seemed to find the CR format substantially easier to respond to as reflected in a non-response rate of less than 2%.

Table 8 provides the frequency distributions for the different permutations of rankings. The results indicate that 31% of the sample provided rankings which were in line with maximising changes in water quality (ranking = L, M, S, C), whilst 26% provided ranking in line with minimising increases in costs (ranking = C, S, M, L). Thus, just over 40% of the sample chose intermediate rankings in which some trade off between increasing costs and water improvements was clearly visible (note that other respondents may be induced to trade at other attribute levels; i.e. it should not be inferred that those ranking consistently with maximising water quality improvements have infinite WTP for such changes).[13]

The rankings of water quality and WTP were analysed using a rank ordered logit model (McKelvey and Zaviona, 1975) in order to derive the indirect utility function, specified as a linear function of alternative scenario attributes.[14] This, as we shall see below, enables us to estimate the marginal rates of substitution between scheme attributes (characteristics). As such it is also possible to obtain valuations of individual scheme attributes (in this case these are a composite index of environmental quality—which is described further below—and cost) as well as for the scheme as a whole. The model was estimated for all observations.[15] The rank-ordered maximum likelihood estimation procedure (StataCorp, 2003), finds the coefficients that maximise the likelihood that a randomly selected individual ranks the alternatives in the order that they were chosen. Negative coefficient estimates imply that an increase in the respective attribute will reduce utility (i.e. the probability of higher ranking for alternatives with higher levels of the respective attribute is decreased).

Table 9
Water quality improvements and RFF water quality index

Water quality level	RFF water quality index	Dissolved oxygen[a] (% saturation)	BOD[a] (mg/l)	Total Ammonia[a] (mgN/l)
Large Improvement—L	7.0	80	2.5	0.25
Medium improvement—M	5.0	65	5	0.95
Small improvement—S	2.5	50	8	2.5
Current situation—C	0.8	20	15	9

[a] The correspondence of these technical measures to the RFF index scores and water quality improvement levels considered is derived from the UK Environment Agency's River Ecosystem Classification scheme, which has been used to describe rivers according to their suitability for fish in terms of these technical measures. Given that both the RFF index and the water quality improvement levels considered in this study are also characterised by their suitability for fish it has thus been possible to match up the various technical measures to the water quality improvement levels and the RFF index values.

The attributes of the scenarios that are considered in the modelling procedure are those for water quality and cost of scenario, as outlined in Table 1. The water quality attributes include fishing, plants and wildlife, boating and swimming and although categorical attribute level parameters could have been estimated for each of these (as is commonly done in the literature, e.g. Beggs et al. (1981)), due to their correlation and the consequences of this in terms of non-orthogonality, a proxy composite variable was used, in the form of a water quality level scale. One such scale that has been previously used (Smith and Desvousges, 1986) is the Resources for the Future (RFF) Water Quality Index (Vaughan, 1981). This is a 10-point index of technical water quality measures and associated recreational activities. Alternatively, one might use technical water quality measures such as levels of dissolved oxygen, biological oxygen demand, and total ammonium. Table 9 shows the water quality level improvements used in the survey along with their associated RFF water quality index values and other technical measure values.

Whilst the RFF index does not specifically incorporate plant and wildlife characteristics in defining the water quality, these are considered to be consistent with the index

[13] It is possible that our results display some degree of lexicographic ordering. However, it is not possible to rigorously test for lexicographic orderings given the limited ranking tasks respondents were asked to undertake (Spash and Hanley, 1995; Foster and Mourato, 2002). Hence it is not possible to say definitively whether respondents are truly following a lexicographic ordering pattern or not. However, as a reviewer noted, nontrading of any form distorts valuation inferences when nontrading patterns are not randomly distributed across attributes, and inflates variances when nontrading patterns are randomly distributed. Thus the results of our analysis should be considered with regard to this issue and its implications.

[14] The assumption that the utility function is linear results in the unattractive implication that the marginal utility associated with attributes does not vary at different levels of the attributes, and hence is at odds with economic intuition. An alternative approach is thus to specify a log-linear relationship between utility and attribute levels, which has the desired effect of making the marginal utility dependent on the attribute level. Whilst this was undertaken within our study, comparison of the linear and log-linear specifications indicated that under the logarithmic transformation of the data, coefficient estimates had uniformly smaller *t*-ratios as well as a lower value of the maximised log-likelihood function, indicating a worse statistical performance. As such there was no statistical case for preferring the logarithmic functional form to the linear one.

[15] There were no tied rankings.

Table 10
Contingent ranking ordered logit results

Variable	Coefficient	SE	95% CI	Z	P
Payment (£)	−0.2415	.0164	−.2737: −.2092	−14.680	0.000
Water quality (RFF score)	1.2256	.0797	1.0693: 1.3820	15.364	0.000

$n=2649$, log-likelihood $=-3547.9$, LR $\chi^2(2)=248.8$ ($P<0.0001$).

Table 11
Estimates of WTP (£ per household per annum) for River Tame water quality improvements derived from the contingent ranking exercise

Water quality improvement, mean WTP (CI)		
Small (S)	Medium (M)	Large (L)
8.64 (8.40–8.86)	21.34 (20.75–21.88)	31.50 (30.63–32.30)

value pertaining to each of the recreational activity levels (boating, fishing and swimming—see Table 1). Such an interpretation of the RFF index is subject to a certain degree of qualification. Others may interpret it somewhat differently, which may have some effect on the results. Nevertheless we feel that our interpretation is on balance defensible. Similarly, the extent to which the values of dissolved oxygen, BOD and ammonia correspond with the small, medium and large water quality improvement levels considered here can be questioned.

Table 10 provides the results of the modelling procedure. The dependent variable is the ranked position of the policy and the independent variables are the policy attributes. The coefficient on 'payment' is negative as expected, indicating that as the payment required for a scheme increases, utility decreases and so schemes requiring higher payment are more likely to receive lower ranked positions. Conversely the coefficient on 'water quality' is positive (as expected) indicating that utility increases as the water quality index increases. Schemes with higher water quality are more likely to receive higher ranked positions. The coefficients are both highly significant. More complicated specifications of the model involving income, demographic and socio-economic variables can also be developed, though this is beyond the scope of the present study.[16]

In order to calculate WTP estimates from the ranking exercise we follow the approach of Lareau and Rae (1985). The trade off between attribute levels and disposable income is found by firstly assuming an indirect utility function in which the deterministic portion is specified as a linear function of the attribute levels, as shown by the following form:

$$V = \alpha c + \beta q$$

where c is the vector of payment associated with the scheme in question, q is the vector of the level of the other attributes associated with the schemes (in this case river water quality), and α and β are the respective coefficient parameters which represent the relative importance of each attribute in determining a respondent's ranking (or the marginal (dis)utility associated with a one-unit change in the attribute). In accordance with welfare theory, a respondent's maximum WTP for an increase in river water quality generated by a unit increase in q is such that his/her overall level of utility is constant. The change in payment relative to the change in river water quality necessary to keep welfare constant is $\Delta c/\Delta q$, which is given by the ratio—β/α. This is the marginal rate of substitution between the water quality and payment attributes (ratio of marginal utilities). Calculating this for the specification shown in Table 10 gives an estimated WTP trade-off of £5.08 per household per annum (for a unit increase in RFF water quality index) which translates into the WTP estimates for water quality improvements shown in Table 11 (and which can be directly compared to those figures given for the CV exercise in Table 3).

Inspection of bootstrapped confidence intervals around the estimates of marginal values reported in Table 11 indicates that values obtained from the CR exercise are significantly higher than those derived from CV questions when protests are included in the CV values. However this is only the case for the medium and large improvements when protests are excluded from the CV values. Possible implications of these results are considered within our concluding remarks.

4. Summary and concluding remarks

The policy background to this paper is provided by the UK Environment Agency's decision to test the CR method for valuing changes in the provision of publicly funded environmental goods. Specifically the Agency is testing the method within the context of river water management issues but has chosen to test the CR technique in isolation of other methods such as the more commonly applied CV method. Our application compares both the CR and CV method within the same application of river water management schemes.

Our findings suggest strong internal and external consistency within the CV element of our study. As per our previous experimental findings (Bateman et al., 2001) the advance disclosure of all valuation tasks prior to their

[16] These variables can be incorporated in two ways. Firstly, the variables of interest can be interacted with the payment or water quality attribute, though this can add to the burden of estimation and make interpretation of resulting coefficients complicated. Alternatively one can divide the data up according to the particular socio-economic variable of interest and estimate the model separately for each subset. The effect of the variable can then be deduced by comparing the magnitude and sign of the coefficients for each subset.

execution appears to underlie our rejection of the ordering effects found by previous, more conventional studies (e.g. Boyle et al., 1994). We find significant, but more importantly, consistent scope sensitivity across the two presentation orderings used. Furthermore, the estimated bid function shows that an array of factors affect WTP responses in a manner that is consistent with theoretically or empirically derived expectations. However, this exercise can be criticised in terms of the high proportion of respondents (23%) who were unable to state WTP responses to the CV valuation question, which may have been due to the open-ended response format of the survey.

With regard to the CR element of our study, while just under one-third of respondents gave rankings which were consistent with maximisation of water quality improvements, just over one-quarter gave rankings consistent with minimising payments, leaving the remainder of the sample giving intermediate rankings. This allowed the estimation of a bid function in which both water quality and payment levels exhibited expected effects. The rate of non-response in the CR exercise (2%) was far lower than for the CV format. However one can question some of the assumptions upon which the results of the CR model rest, e.g. the assumptions it imposes on ranking behaviour (all rankings follow a logistic distribution), and use of a linear functional form (resulting in non-diminishing marginal utility).

In isolation therefore, we can see positive and negative aspects to both the CV and CR elements of this study. In comparison to each other we find that although the CR values tended to be higher than those recorded via CV responses, from both a theoretical and empirical perspective this finding is not surprising. The open-ended WTP format adopted in the CV element is known to produce lower bound value estimates (compared to those obtained by other elicitation formats, see Carson et al., 1999) while previous empirical comparisons of CV and CR formats have also found that values derived from the latter exceed those obtained from the former (Stevens et al., 2000). Nevertheless, the finding of higher CR values was not unambiguous and was to some extent dependent on the inclusion of protest zero responses encountered in the CV exercise. Furthermore, the presence of nontrading patterns in our CR data casts some doubt on reliability of the estimates of WTP we derive.

The most striking result of this comparison between the two methods, and the result of most relevance for practical policy, is in terms of the response rates to the CV and CR surveys. Responses to the CR survey were significantly higher than they were to the CV survey, with a 98% response rate to the former compared to a response of approximately 77% to the latter. A large proportion of non-responders to the CV survey cited the difficulty, or hypothetical nature, of the exercise as the reason for their inability to state a WTP. However, the majority of these same respondents were able to complete the CR exercise. This suggests that the ranking exercise is perhaps conceptually easier for respondents than is the open-ended valuation format - as long as the number of alternatives and attributes presented as part of the exercise are kept relatively small (as they are in this study) so that the cognitive burden that respondents face is minimised. Hence, in public policy situations, where response to valuation surveys needs to be maximised, and where the environmental scenario being valued is conceptually difficult to understand, the CR format may be a more appropriate choice of valuation tool than an equivalent CV survey with an open-ended format. Of course, the question then arises whether the same result would have arisen if a CV survey using a dichotomous choice format had been used, however this question will need to be explored in future research.

Acknowledgements

The PEDM is funded by the UK Economic and Social Research Council (ESRC). Additional funding for this study was provided by the University of Birmingham's Interdisciplinary Research Fund.

References

Alvarez-Farizo, B., Hanley, N., Wright, R.E., Macmillan, D., 1999. Estimating the benefits of agri-environmental policy: econometric issues in open-ended contingent valuation studies. J. Environ. Plan. Manage. 42 (1), 23–43.

Bateman, I.J., Willis, K.G., 1999. Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US EU, and Developing Countries. Oxford University Press, Oxford.

Bateman, I.J., Langford, I.H., Munro, A., Starmer, C., Sugden, R., 2000a. Estimating the four Hicksian measures for a public good: a contingent valuation investigation. Land Econ. 76 (3), 355–373.

Bateman, I.J., Langford, I.H., Nishikawa, N., Lake, I., 2000b. The axford debate revisited: a case study illustrating different approaches to the aggregation of benefits data. J. Environ. Plan. Manage. 43 (2), 291–302.

Bateman, I.J., Cole, M., Cooper, P., Georgiou, S., Hadley, D., Poe, G.L., 2001. Visible choice sets and scope sensitivity: An experimental and field test of study design effects upon contingent values, CSERGE Environmental Decision Making Working Paper EDM01-01. Centre for Social and Economic Research on the Global Environment, University of East Anglia p. 76.

Bateman, I.J., Carson, R.T., Day, B., Hanemann, W.M., Hanley, N., Hett, T., et al., 2002. Economic Valuation with Stated Preference Techniques: A Manual. Edward Elgar, Cheltenham.

Bateman, I.J., Lovett, A.A., Brainard, J.S., 2003. Applied Environmental Economics: A GIS approach to Cost-Benefit Analysis. Cambridge University Press, Cambridge.

Beggs, S., Cardell, S., Hausman, J., 1981. Assessing the potential demand for electric cars. J. Econometrics 16, 1–19.

Ben-Akiva, M., Morikawa, T., Shiroishi, F., 1991. Analysis of the reliability of preference ranking data. J. Business Res. 23 (3), 253–268.

Boxall, P.C., Adamowicz, W.L., Swait, J., Williams, M., Louviere, J., 1996. A comparison of stated preference methods for environmental valuation. Ecol. Econ. 18 (3), 243–253.

Boyle, K.J., Desvousges, W.H., Johnson, F.R., Dunford, R.W., Hudson, S.P., 1994. An investigation of part-whole biases in contingent valuation studies. J. Environ. Econ. Manage. 27 (1), 64–83.

Carson, R.T., 1997. In: Kopp, R.J., Pommerehne, W.W., Schwarz, N. (Eds.), Contingent valuation surveys and tests of insensitivity to scope Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods. Kluwer Academic Publishers, Boston.

Carson, R.T., Flores, N.E., Hanemann, W.M., 1998. Sequencing and valuing public goods. J. Environ. Econ. Manage. 36, 314–323.

Carson, R.T., Groves, T., Machina, M.J., 1999. Incentive and informational properties of preference questions, Plenary Address, Ninth Annual Conference of the European Association of Environmental and Resource Economics (EAERE), Olso, Norway, June 1999

Chilton, S.M., Hutchinson, W.G., et al., 1999. Some Further Implications of Incorporating the Warm Glow of Giving into welfare measures: a comment on the use of donation mechanisms by champ et al.. J. Environ. Econ. Manage. 37, 202–209.

Desvousges, W.H., Smith, V.K., 1983. Option price estimates for water quality improvements: a contingent valuation study for the monogahela river. J. Environ. Econ. Manage. 14, 248–267.

DETR (Department of the Environment, Transport and the Regions), 1998.Anon., 1998. Technical Guidance on Policy Appraisal: A Report to the Department of the Environment, Transport and the Regions by the Economics for the Environment Consultancy (EFTEC). DETR, London.

Diamond, P.A., Hausman, J.A., 1994. Contingent valuation: is some number better than no number? J. Econ. Perspect. 8 (4), 45–64.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.

Foster, V., Mourato, S., 2000. Valuing the multiple impacts of pesticide use in the UK: A contingent ranking approach. J. Agric. Econ. 49, 1–15.

Foster, V., Mourato, S., 2002. Testing for consistency in contingent ranking experiments. J. Environ. Econ. Manage. 44, 309–328.

Garrod, G., Willis, K., 1998. Using contingent ranking to estimate the loss of amenity value for inland waterways from public utility service structures. Environ. Resour. Econ. 12, 241–247.

Georgiou, S., Bateman, I.J., Cole, M.A., Hadley, D., 2000a. Contingent Ranking and Valuation of River Water Quality Improvements: Testing for Scope Sensitivity, Ordering and Distance Decay Effects. CSERGE Working Paper GEC 2000-18.

Georgiou, S., Bateman, I.J., Langford, I.L., Day, R.J., 2000b. Coastal bathing water health risks: developing means of assessing the adequacy of proposals to amend the 1976 directive. Risk Decision and Policy 5, 49–68.

GIBB Ltd, 2001.Anon., 2001. River Mimram Low Flow Public Preferences Study: Phase 1 Final Initial Report to the Environment Agency, Thames Region. GIBB Ltd..

Giraud, K.L., Loomis, J.B., Johnson, R.L., 1999. Internal and external scope in willingness-to-pay estimates for threatened and endangered wildlife. J. Environ. Manage. 56, 221–229.

Hanemann, W.M., 1994. Valuing the environment through contingent valuation. J. Econ. Perspect. 8 (4), 19–43.

Hanemann, W.M., 1996. Theory versus data in the contingent valuation debate. In: Bjornstad, D.J., Kahn, J.R. (Eds.), The Contingent Valuation of Environmental Resources: Methodological Issues and Research Needs. Edward Elgar, Cheltenham, pp. 38–60.

Hanley, N., 2001. Cost-benefit analysis and environmental policy making. Environ. Plan. C 19, 103–118.

Irwin, J.R., Slovic, P., Lichtenstein, S., McClelland, G., 1993. Preference reversals and the measurement of environmental values. J. Risk and Uncertainty 6 (1), 5–18.

Jorgensen, B.A., Syme, G.J., 2000. “Protest responses and willingness to pay: attitude toward paying for stormwater pollution abatement”. Ecol. Econ. 33, 251–265.

Kahneman, D., Knetsch, J.L., 1992. Valuing public goods: the purchase of moral satisfaction. J. Environ. Econ. Manage. 22, 57–70.

Lareau, T.J., Rae, D.A., 1985. Valuing willingness to pay for diesel odor reduction: an application of the contingent ranking technique. Southern Econ. J. 55 (3), 728–742.

McKelvey, R., Zaviona, W., 1975. A statistical model for the analysis of ordinal level dependent variables. J. Math. Sociol. 4, 103–120.

Mitchell, R.C., Carson, R.T., 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method. Resources for the Future, Washington, DC.

Pearce, D.W., 1998. Cost-benefit analysis and environmental policy. Oxford Rev. Econ. Pol. 14 (4), 84–100.

Randall, A., Hoehn, J.P., Tolley, G.S., 1981. The Structure of Contingent Markets: Some Results of a Recent Experiment, Paper Presented at the American Economic Association Annual Meeting 1981.

Randall, A., DeZoysa, D., Yu, S., 2001. In: Bergstrom, J.C., Boyle, K.J., Poe, G.L. (Eds.), Ground Water, Surface Water, and Wetlands Valuation in Ohio The Economic Value of Water Quality. Edward Elgar, Cheltenham, UK.

Schkade, D.A., Payne, J.W., 1994. How people respond to contingent valuation questions—a verbal protocol analysis of willingness-to-pay for an environmental regulation. J. Environ. Econ. Manage. 26 (1), 88–109.

Smith, V.K., Desvousges, W.H., 1986. Measuring Water Quality Benefits. Kluwer, Boston.

Spash, C.L., Hanley, N., 1995. Preferences, information and biodiversity preservation. Ecol. Econ. 12, 191–208.

StataCorp, 2003.Anon., 2003. Stata Statistical Software: Release 8.0. Stata Corporation, College Station, Texas.

Stevens, T.H., Belkner, R., Dennis, D., Kittredge, D., Willis, C., 2000. Comparison of contingent valuation and conjoint analysis in ecosystem management. Ecol. Econ. 32, 63–74.

Sutherland, R.J., Walsh, R.G., 1985. Effect of distance on the preservation value of water quality. Land Econ. 61 (3), 281–291.

Tversky, A., Sattath, S., Slovic, P., 1988. Contingent weighting in judgement and choice. Psychol. Rev. 95, 371–384.

Vaughan, W.J., 1981. In: Mitchell, R.C., Carson, R.T. (Eds.), The Water Quality Ladder, Appendix II An Experiment in Determining Willingness to Pay for National Water Quality Improvements. Draft Report, Resources for the Future, Washington, DC.