An assessment of the nonmarket benefits of the Water Framework Directive for households in England and Wales

Paul J. Metcalfe,[1] William Baker,[2] Kevin Andrews,[3] Giles Atkinson,[1]
Ian J. Bateman,[4] Sarah Butler,[5] Richard T. Carson,[6] Jo East,[7] Yves Gueron,[2,8]
Rob Sheldon,[7] and Kenneth Train[5,9]

Received 16 June 2010; revised 1 February 2012; accepted 9 February 2012; published 28 March 2012.

[1] Results are presented from a large-scale stated preference study designed to estimate the nonmarket benefits for households in England and Wales arising from the European Union Water Framework Directive (WFD). Multiple elicitation methods (a discrete choice experiment and two forms of contingent valuation) are employed, with the order in which they are asked randomly varied across respondents, to obtain a robust model for valuing specified WFD implementation programs applied to all of the lakes, reservoirs, rivers, canals, transitional, and coastal waters of England and Wales. The potential for subsequent policy incorporation and value transfer was enhanced by generating area-based values. These were found to vary from £2,263 to £39,168 per km^2 depending on the population density around the location of the improvement, the ecological scope of that improvement, and the value elicitation method employed. While the former factors are consistent with expectations, the latter suggests that decision makers need to be aware of such methodological effects when employing derived values.

Citation: Metcalfe, P. J., et al. (2012), An assessment of the nonmarket benefits of the Water Framework Directive for households in England and Wales, *Water Resour. Res.*, *48*, W03526, doi:10.1029/2010WR009592.

1. Introduction

[2] The European Community (EC) Water Framework Directive (WFD) [*European Parliament*, 2000] requires that all natural water bodies should reach the common minimum European standard of "good ecological status" (GES) by 2015, except where to do so would entail disproportionate cost. This requirement is widely considered to be stringent and substantively different from most water quality standards that are based either on chemical assessments or the ability to support specific types of use. Achieving GES by 2015 will be technically demanding and expensive. It will require member states to restore many natural habitats for plants, fish, and other wildlife by reducing pressures from over-abstraction, point, and diffuse sources of pollution, nonnative species, and from physical modifications such as dams, weirs, and engineered channeling. The cost of achieving full compliance in England and Wales has been estimated to be £2.4 billion per year over a 43-yr term [*Department for Environment, Food and Rural Affairs* (*DEFRA*), 2008; available at http://archive.defra.gov.uk/environment/quality/water/wfd/documents/RIA-river-basin.pdf], an amount that far surpasses the cost of any previous EC water policy directive.

[3] Benefits estimates are valuable to policy makers in this context for two related purposes. First, they can be used to appraise whole programs of improvements at regional or national levels, as a means to help decision makers set the overall scale of implementation of the directive. In addition, they can be used in assessments, on cost-benefit grounds, of whether achieving GES will be disproportionately costly for individual water bodies. In such cases, applications for derogations can be made to allow for a longer time to achieve compliance or for a less-stringent environmental objective to be adopted. The present study was designed to address both purposes simultaneously. In this regard, it departs from most previous studies of water quality improvements which have sought to value either a whole program of improvements [*Carson and Mitchell*, 1993; *Brouwer*, 2008] or improvements to a localized area [e.g., *Alam and Marinova*, 2003; *Bateman et al.*, 2011; *Hanley et al.*, 2003, 2006; *Kontogianni et al.*, 2003; *Kramer and Eisen-Hecht*, 2002; *Loomis et al.*, 2000].

[4] At the core of our study is the development of a model for valuing national and regional programs of WFD improvements as a function of key attributes relevant to strategy setting at these levels. These attributes include measures of the geographic scale of the implementation program, the balance between improvements to the worst

[1]Department of Geography and Environment, London School of Economics, UK.
[2]NERA Economic Consulting, London, UK.
[3]Department of Environment, Food and Rural Affairs, London, UK.
[4]School of Environmental Sciences, University of East Anglia, Norwich, UK.
[5]NERA Economic Consulting, San Francisco, California, USA.
[6]Department of Economics, University of California, San Diego, California, USA.
[7]Accent Market Research, London, UK.
[8]Now at Department of Economics, University College London, London, UK.
[9]Department of Economics, University of California, Berkeley, California, USA.

0043-1397/12/2010WR009592

areas and improvements to raise the number of high-quality sites, and the balance between improvements in densely populated areas and improvements in more remote locations. A key feature of the model is that it can also be used to value individual localized improvements as a function of the size of the area improved, the qualitative range of improvement, and the population density of the area surrounding the water body. An advantage of this approach over typical benefits transfer methods for WFD disproportionate cost assessment [e.g., *Bateman et al.*, 2011], is that the values obtained are fully consistent with the context of a nationwide program of simultaneous improvements. As a consequence, they are not biased because of income and substitution effects, which are liable to cause a discrepancy between the summed values from individual benefits transfers and whole program valuations [*Hoehn and Randall*, 1989; *Bateman et al.*, 1997; *Hoehn*, 1991].

[5] Data to estimate the model come from a large-scale nationwide stated preference survey employing three elicitation methods: a discrete choice experiment (DCE) and two contingent valuation (CV) questions. The discrete choice experiment (DCE) framework [*Louviere et al.*, 2000] is naturally suited to the development of multi-attribute valuation models of the kind required in our study. A number of studies have found, however, that the DCE approach focused on multiple policy changes often elicits higher values for the same package of improvements than a contingent valuation (CV) study focused on a single policy change [*Cameron et al.*, 2002; *Foster and Mourato*, 2003; *Hanley et al.*, 1998]. A number of reasons have been put forth for this finding ranging from strategic behavior to placing less weight on the cost attribute when it is varied simultaneously with other attributes to various types of learning behavior. Most of these suggest that the relative values of noncost attributes derived from a DCE can be considered reliable, but that total values, which depend on cost, may be biased upward. The rationale for this is straightforward. While respondents may want to strategically influence whether a good is provided or how much they have to pay for it, they have no incentive to induce the government to provide a good with a set of attributes that they do not want. A large number of "scaling" tests [e.g., *Carlsson and Martinsson*, 2001] have found a reasonably close correspondence in marginal (relative) values between survey and market comparisons. As *Carson and Groves* [2011] note, "scaling" tests [*Louviere et al.*, 2000] allow deviations in one direction (e.g., price sensitivity) and may be best seen as tests against random behavior in surveys versus markets. These scaling tests have generally accepted the hypothesis that stated and revealed preference data can be combined after allowing for differences in scale suggesting the validity of relative comparisons. As *Lusk and Schroeder* [2004] demonstrate, it is possible for marginal willingness to pay (WTP) for an attribute to be statistically equivalent in survey versus market contexts without total value estimates being the same.

[6] It therefore seems plausible that reliable estimates could be obtained by using relative values from a DCE and then calibrating their scale using CV estimates. We adopt this approach, and hence employ a CV component in addition to the DCE to obtain estimates of the value of a single large-scale program of WFD improvements with which to calibrate the DCE-derived estimates of individual attribute values.

[7] Willingness to pay estimates are known to be sensitive to elicitation methods and question order effects. Comparing across CV questions, many studies have found that dichotomous choice contingent valuation (DCCV) values exceed those obtained by open-ended formats such as the payment card contingent valuation (PCCV) approach [*Venkatachalam*, 2004; *Welsh and Poe*, 1998] to the extent that this is considered a "stylized fact" of the CV approach [*Carson and Groves*, 2007]. Alternative lines of explanation for this divergence have been proposed in the CV literature, from a strategic behavior perspective [*Carson and Groves*, 2007] and from a cognitive psychology perspective [*Green et al.*, 1998]. To test the sensitivity of our findings to elicitation method effects, we utilize both types of CV questions and examine responses to them in a joint model. Additionally, several studies have demonstrated the importance of the order in which elicitation questions are presented [*Day et al.*, 2012] and again alternative lines of explanation have been proposed in the CV literature with differing implications. We therefore vary the order of the elicitation questions asked using split samples to be able to isolate and test sensitivities to these effects.

[8] The contributions of the paper are thus threefold. We obtain a model via a robust large-scale stated preference survey for valuing national programs of improvements as a function of key attributes relevant to strategy setting at this level. Additionally, we derive a transferable value function that can be used for disproportionate cost assessment at the level of individual sites, and which can validly be summed over sites so as to obtain values for regional programs of water quality improvements. Finally, we explore the sensitivity of our estimates to elicitation treatment effects.

[9] The remainder of this paper is organized as follows. Section 2 outlines the design of the survey instrument, describes the survey administration, and characterizes the sample obtained. Section 3 provides an overview of our approach to analysis. Section 4 describes the econometric models we estimate. Section 5 presents our main findings: first, on the CV-derived benefits for a benchmark implementation of the WFD, then on the DCE-derived benefit function for valuing varied implementations of the WFD, and finally, on our preferred model which combines DCE and CV results. Section 6 discusses our findings with respect to elicitation treatment effects. Section 7 concludes the paper.

2. Survey Design, Administration, and Data

[10] The survey design and development for the present study broadly conform to best practices as set out by *Arrow et al.* [1993], *Bateman et al.* [2002], and *Mitchell and Carson* [1989]. The "description of the good" was informed by a stakeholder survey, close work with a team of scientists, and a series of 12 focus groups involving members of the public. The survey instrument was extensively pretested (reports available upon request) with members of the public, via two phases of focus groups (eight groups in total), one phase of 30 cognitive interviews in which respondents were encouraged to "think aloud" and give feedback on the questionnaire as they worked their way through it, and two pilot surveys of 50 and 100 respondents, respectively.

2.1. Attributes and Levels

[11] Policy scenarios for WFD improvements are characterized in the survey by the proportions of a respondent's local area (within 30 miles), and of the national area (England and Wales), that will be high, medium, and low quality in 2015 (8 yr from the survey date), and in 2027 (20 yr from the survey date). Table 1 gives definitions of the attributes used as they appear in our models, and the levels they take in the design.

[12] Note that by including separate attributes for local and national water bodies we are able to obtain values as a function of population density around water bodies. Prior work [*Bateman et al.*, 2006] shows that individual valuations for spatially fixed environmental goods (such as those for water quality improvements) exhibit "distance decay" in that they fall as the distance between that individual and the improvement increases. Given this, individual values within the "local area" should exceed corresponding per hectare values in the "national area." This implies that per hectare values will be higher for improvements within densely populated areas than for those located in sparsely populated areas.

[13] For the DCE, the levels of the future environmental status attributes are based on a "pivot design" methodology [*Rose et al.*, 2008]. Pivot designs, which are common in transportation applications, take the respondent's baseline attributes levels as given and "pivot" off by assigning an increment to those levels to form new attribute levels for the DCE. The variables *LowL8*, *MediumL8*, and *HighL8* in the present case are generated from corresponding baseline conditions *LowL0*, *MediumL0*, *HighL0*, which are known (fixed) quantities and which vary according to respondent location. The generating functions for each level of each environmental attribute used for the main survey are shown in Table 1.

[14] For the DCCV and PCCV questions, one half of the sample was offered a more extensive policy package than the other half in order to allow for analysis and testing of the sensitivity to scope of the CV values. In both scenarios, 95% of all water bodies are brought to high quality within 20 yr with the remainder at medium quality. The scenarios differ only with respect to the extent of improvement that occurs within the first 8 yr: in the "95% scenario," the full 95% is achieved within 8 yr; in the "75% scenario," 75% are brought to high

Table 1. Attributes and Levels

		Levels[c]		
Attribute[a]	Definition[b]	Current[d,e]	CV	DCE[f,g]
HighL8	Proportion at high quality in local area at time = 8 (in 2015)	9.0%	95%	$HighL0 + 0.75\ (MediumL0 - \Delta LowL8)$
			75%	$HighL0 + 0.5\ (MediumL0 - \Delta LowL8)$
				$HighL0 + 0.25\ (MediumL0 - \Delta LowL8)$
				$HighL0 + 0.1\ (MediumL0 - \Delta LowL8)$
LowL8	Proportion at low quality in local area at time = 8 (in 2015)	58.6%	0	0
				$0.25 LowL0$
				$0.5 LowL0$
				$0.75 LowL0$
HighN8	Proportion at high quality in national area at time = 8 (in 2015)	15.0%	95%	$HighN0 + 0.75\ (MediumN0 - \Delta LowN8)$
			75%	$HighN0 + 0.5\ (MediumN0 - \Delta LowN8)$
				$HighN0 + 0.25\ (MediumN0 - \Delta LowN8)$
				$HighN0 + 0.1\ (MediumN0 - \Delta LowN8)$
LowN8	Proportion at low quality in national area at time = 8 (in 2015)	44.0%	0	0
				$0.25 LowN0$
				$0.5 LowN0$
				$0.75 LowN0$
High20	Proportion at high quality in local and national areas at time = 20 (2027)	As now[h]	95%	95%
				75%
Cost	Permanent increase in water bill and other household payments (£ $hh^{-1}\ yr^{-1}$)	N/A	£5	£5
			£10	£10
			£20	£20
			£30	£30
			£50	£50
			£100	£100
			£200	£200

[a]The quantities of high, medium, and low quality always sum to 1, so medium quality is omitted.
[b]"Local area" refers to the area within 30 miles of the location of the respondent's interview and "National area" refers to the whole of England and Wales.
[c]All environmental status levels were rounded to the nearest whole percentage point in the choice sets used.
[d]Current condition levels shown here are those based on data used for the survey itself, rounded to one decimal place. Data are weighted for age, sex, and region based on the 2001 UK Census. Further details on the weights used are available from the authors on request. More recent data may suggest a different picture of current conditions in the water environment.
[e]For attributes *HighL8* and *LowL8*, current levels are sample mean values.
[f]Terms ending in 0 refer to quality levels at time = 0, i.e., current levels.
[g]$\Delta LowL8 \equiv LowL8 - LowL0$, and $\Delta LowN8 \equiv LowN8 - LowN0$.
[h]Although "as now" was how the survey presented current conditions to respondents, a numeric value was needed to enter this attribute into the DCE choice models. This essentially involved a choice between *HighL0* and *HighN0*. We chose to use *HighL0* for statistical reasons.

quality in the first 8 yr, with the remaining improvement up to 95% high quality occurring between the 8- and 20-yr horizons.

[15] The levels of the payment vehicle, *Cost*, for both the DCCV and the DCE questions were £5, £10, £20, £30, £50, £100, and £200, per household per year in extra water bills and other household payments. The amounts shown in the payment card for the PCCV question ranged from £0 to £1000 spread across 28 points distributed on an approximately logarithmic scale.

2.2. Survey Presentation

[16] First, introductory questions on attitudes and use of the water environment were asked; then respondents were shown, in succession, two cards containing carefully developed descriptions of water quality at three color-coded status levels (Figures A1 and A2 in Appendix A). The three status levels were assigned the labels "high quality," "medium quality," and "low quality," and the colors assigned to them were dark blue, midblue, and light blue, respectively. The three adopted status levels were linked to the WFD as follows: "high quality" corresponded to high or good ecological status; "medium quality" corresponded to moderate or poor ecological status; and "low quality" corresponded to bad ecological status. The first card contained generic descriptions of water quality at each status level while the second card gave illustrated descriptions specific to one of four water body types: rural river, urban river, lake, or estuary/coastal. Survey time constraints precluded the presentation of more than one type of water body per respondent. By randomly assigning respondents to different water body types, it was possible to test for any effects caused by the particular water-body type example shown. Statistical tests suggest no effect from the particular example water body the respondent saw so this issue is not discussed further.

[17] Following the status descriptions, respondents were presented with two maps showing current water quality levels, color-coded to match the descriptions just shown. The first map showed the respondent's local area (within 30 miles of the location of the survey interview), and the second showed the whole of England and Wales. A pie chart was included on each map showing the proportions of the water environment in each status category. An example of the maps shown is reproduced in the Appendix Figure A3.

[18] The questionnaire then presented each respondent with a valuation exercise comprising: seven DCE questions each offering a choice between the status quo and two improvement alternatives, one DCCV question offering a choice between the status quo and one large-scale improvement alternative, and one PCCV question asking what amount on the card shown to them, or any amount in between, is the most they would be willing to pay, through increased water bills and other household payments every year to have the improvements shown. Included in Appendix A is the valuation scenario, including statements to enhance consequentiality and the household's budget constraint, read out to respondents prior to their facing the valuation questions. Specific examples of the PCCV, DCCV, and DCE questions are shown in Figures A4, A5, and A6, respectively.

2.3. Experimental Design

[19] The experimental design for the survey was necessarily fairly complex in order to be able to test the range of treatments being considered, which included among other things: the CV scenario presented (75% or 95%), the DCCV cost amount offered (£5, £10, £20, £30, £50, £100, or £200), the combination of DCE choice profiles shown, and the order of elicitation questions (PCCV before or after the DCE, DCCV at the beginning middle or end of the DCE). In addition, survey instruments varied across sampling locations due to differences in current water status levels in the local area.

[20] The design for allocating these treatments aimed to minimize the correlation between them and to achieve a good degree of balance across the sample. For the DCE design problem (i.e., the selection of combinations of choice profiles), this involved drawing choice sets (status quo plus two improvement alternatives) randomly, without replacement from the full factorial of every possible combination of attribute levels, excluding strictly dominated and practically impossible combinations, so that each choice card for each respondent was unique. Because of the large sample of unique option profiles, an experimental design created in this way should, with a large sample, provide a reasonable approximation to the full factorial itself, and so thereby be internally near orthogonal. Compared to a main effects design, it is possibly less statistically efficient, but it has the advantage of allowing estimation of lower-order interaction terms which, given the variety of design issues under consideration, was considered a key requirement for the present study.

[21] The remaining treatments were allocated independently from the above procedure, and were structured to ensure an even spread of treatments across each location sampled. To this end, by location, each instrument type (defined by its unique combination of water-body type example, CV scenario, DCCV cost amount, and order of elicitation questions) was drawn with equal probability from the set of all instrument types, without replacement, so that no combination of treatments was allocated to more than one respondent in any one location. This procedure ensured that each instrument type was given an equal probability of selection overall, that sufficient numbers of certain key combinations would be present in the sample, that there would not be any clustering of treatments by location, and that orthogonality, with respect to the DCE design, would be preserved.

2.4. Survey Administration and Data

[22] The study's target sample was developed as a set of 50 locations, with a target of 30 respondents for each. Locations were sampled in proportion to their population size, and respondents were recruited off the street from the busiest places in the area, with quotas set for age, gender, and socio-economic characteristics. Additionally, in order to be in scope, recruits had to be responsible, solely or jointly, for paying the water bill and they had to live within 15 miles of the survey location so that the 30-mile radius map presented to them adequately represented what they would call their local area. An £8 incentive was offered to encourage participation. Although consideration of the range of treatments offered might suggest a need for a larger sample size, the emphasis on orthogonality in the experimental design ensured that all relevant comparisons, e.g., between question order treatments, could be tested without the need to control for all interactions with other

treatment types. The target sample size of 1500 was therefore expected to be more than adequate to estimate the models desired with reasonable precision.

[23] In July 2007, 1487 respondents were interviewed across the 50 sampling areas. Interviews were conducted face-to-face in a designated location by experienced professionals under the supervision of Accent Market Research using the computer-aided personal interviewing (CAPI) technique. The interviews lasted an average of 32 min and the interviewers found good levels of understanding and attention were given to the questions.

[24] A total of 165 respondents stated a PCCV WTP of £0 for the scenario, of whom 58 respondents were removed due to giving an invalid protest response. A further 23 were removed for giving no response at all to the PCCV question, and 17 were removed as outliers. Protest cases were identified by examining the verbatim follow-up responses to the elicitation questions; outliers were defined as those in the top 1% of the distribution of PCCV responses, which corresponds to all WTP amounts greater than or equal to £350 per household per year. The DCCV data show roughly 90% are in favor at the lowest amounts (with one small monotonicity violation at £10) dropping to about 40% in favor at £200, the highest amount asked. Interestingly, there were no nonresponses to the DCCV or DCE questions, a result that may reflect the higher cognitive load of the more open-ended PCCV format. The total number of respondents removed amounts to 6.6% of the full sample. Additional analysis reported in our technical report [*NERA-Accent*, 2007] examined the sensitivities of our main results to more liberal and more conservative approaches for identifying and excluding protests and outliers. Our results are robust in a qualitative sense to the specific approach used.

[25] Since believability of the scenario is crucial for obtaining valid estimates of WTP, we examined the verbatim responses to the PCCV follow-up question for evidence of any disbelief in the scenario presented. We expect that had respondents doubted that the improvements would take place, and had expressed this doubt by lowering their stated WTP, then they would have articulated this doubt when asked for the reasons underlying their stated response. From the verbatim follow-up responses, we identified only eight people out of 1389 in the analysis sample who indicated that they did not believe the improvements would occur. This constitutes only 0.6% of the sample, which we take as evidence that disbelief in the scenario was not widely held. In addition to this evidence, we found that not a single person during the extensive pretesting process expressed any doubt that the improvements would take place as described.

[26] Table 2 presents population and sample characteristics, for both the raw sample and the analysis sample, which excludes protests, outliers, and nonresponses. The analysis sample characteristics are almost identical to the raw sample characteristics. The raw sample appears to match the population reasonably well although clearly not perfectly. The sample contains a somewhat lower proportion of men than the population, and contains more people out of work, and a lower range of incomes. The sample also appears to be better educated than the population at large. In all the analysis that follows, the data are weighted using a three-way table of survey weights to match sample to population by age, sex, and region, based on the 2001 UK Census.

Table 2. Sample and Population Characteristics

	England and Wales Population[a]	Raw Sample[b]	Analysis Sample[c]
Age[d]			
18–29	19%	14%	14%
30–64	60%	65%	65%
65+	21%	21%	21%
N (=100%)	40,246,981	1486	1388
Sex[d]			
Male	48%	43%	43%
Female	52%	57%	57%
N (=100%)	40,246,680	1487	1389
Children?[e]			
Yes	29%	27%	27%
No	71%	73%	73%
N (=100%)	21,660,682	1487	1389
Education[f]			
Basic	31%	19%	19%
Medium	39%	42%	42%
High	30%	39%	39%
N (=100%)	34,998,226	1373	1285
Economic activity[f]			
Working	64%	53%	55%
Not working	29%	44%	43%
Student	7%	2%	2%
N (=100%)	37,606,305	1373	1285
Income (weekly)[g]			
Low (<£300)	30%	42%	42%
Med (£300–£1000)	53%	46%	46%
High (£1000+)	17%	12%	13%
N (=100%)	18,823	1060	1009

[a]Stats are drawn from the 2001 UK Census except where indicated otherwise.
[b]Base includes all respondents who answered the relevant question in the survey, unless indicated otherwise.
[c]Base excludes from the raw sample the 98 respondents who failed to answer the WTP questions, or who were identified as protestors or outliers.
[d]Base for population statistics equals all individuals.
[e]Base for population statistics equals all households.
[f]Base for sample statistics equals respondents aged 18–74; base for population statistics equals individuals aged 16–74.
[g]Population statistics are drawn from the 2007/8 UK Family Resources Survey (DWP, 2008, 172 pp.; available at http://research.dwp.gov.uk/asd/frs/2007_08/frs_2007_08_report.pdf); Base equals all households.

3. Overview of Analysis

[27] We analyze the data obtained from the survey as follows. First, we combine the DCCV and PCCV responses using a single estimation technique: interval censored regression. This yields estimates of the value of the benchmark "95% scenario" for each question type, the effects of the question order on these estimates, and the effects of respondent covariates. Interval frameworks are well suited to representing both DCCV and PCCV responses. *Cameron and Huppert* [1989, 1991] have argued that the language of a payment card question lends itself to an interval interpretation, with WTP lying between the amount indicated and the next highest amount labeled on the card. Interval frameworks have also long been used to represent DCCV responses [*Carson and Hanemann*, 2005], with a no

response indicating that WTP lies between zero and the amount asked and a "yes" response indicating that WTP lies between the amount asked and an upper bound reflecting financial resources. To be conservative and ensure consistency with a key assumption made about the PCCV data, we use an upper bound of £350 for the interval when a respondent said "yes" to the DCCV question, which is substantially higher than the largest amount used (£200). This does not rule out the possibility that larger WTP values are held by respondents, only that they were not observed in either our PCCV or DCCV data. One weakness of this study is that our largest DCCV bid used was not high enough to clearly pin down the right tail.

[28] Our next step is to analyze the DCE data, and we adopt the standard conditional logit model for this purpose [*McFadden*, 1973]. This model obtains distinct marginal, i.e., per percentage point, values of improvements from low to medium, and from medium to high, in local and national areas. Again, we also examine question order effects and the effects of covariates as a test of the validity of the results, and report the range of values we obtain. The DCE-derived marginal values give rise to estimates of the value of the benchmark 95% scenario via inputting the degrees of improvement in each attribute that correspond to this scenario. We compare the results for the 95% scenario obtained from the DCE model with those from the CV model, and test the null hypothesis that they are the same. In line with findings from some other studies [e.g., *Foster and Mourato*, 2003], we find that the DCE results significantly exceed those derived from the CV model. The DCE results may be biased upward, either because the act of presenting multiple packages to a respondent causes them to behave strategically, rather than accepting a choice at face value [*Carson and Groves*, 2007], or because the act of answering multiple questions where many attributes vary encourages respondents to place less focus on the cost [*Kahneman et al.*, 2006; *Schkade and Kahneman*, 1998]. In the spirit of adopting a conservative approach to analysis, we scale the marginal values derived from the DCE, so that the estimated value of the benchmark 95% scenario derived using the scaled DCE marginal values, is equal to the CV value for this scenario.

[29] The final step in our analysis is to invert the scaled DCE-based valuation function so that instead of measuring the value of national policies to households it measures the value of individual water body improvements as a function of the size of the area improved, the qualitative scope of improvement, low to medium, medium to high, or low to high, and the density of the population surrounding the water body improved. Regional water body improvement programs such as River Basin Management Plans, and also water utility investment plans that impact on water quality in their area, can then be valued by summing the values over the water bodies improved. By way of example, we present average values for improvements in a low population density region and a high population density region in England and Wales.

4. Econometric Models

4.1. Contingent Valuation Models

[30] The interval-censored framework is straightforward to implement in a maximum likelihood context. Let y_n be our interval-censored variable, which we model as a linear function of explanatory variables x_n plus an independent and identically distributed (i.i.d.) error term ε_n with mean zero and variance σ^2. Then we have:

$$\text{Prob}(y_n) = F\left(\frac{y_n^U - x_n\beta}{\sigma}\right) - F\left(\frac{y_n^L - x_n\beta}{\sigma}\right),$$

which implies the following log likelihood:

$$LL = \sum_n \log\left[\text{Prob}(y_n)\right].$$

[31] A distributional assumption is required for $F(.)$ to implement the estimation. We chose the lognormal because it ensures that WTP is non-negative (a problem with the normal) and it is straightforward to implement. Since the lower bound for some intervals is zero, the number "1" was added to all lower and upper bound values before taking logs because the log of zero is undefined. This "1" was then subtracted in obtaining later estimates for mean and median WTP. In the panel context, where we have two responses per person, indexed by t, we thus let $y_{nt} = \log(1 + \text{WTP}_{nt})$ and define lower and upper bounds accordingly, where WTP_{nt} is the willingness to pay by respondent n, as elicited by question type t ($t \in$ {PCCV, DCCV}). $F(.)$ is then simply the standard normal cumulative distribution.

[32] The above log likelihood is based on the assumption that error terms are independent of one another. Independence is unlikely, however, when responses to both PCCV and DCCV questions are combined. To take account of the within-person correlation between responses, we also estimate a random effects panel version of the above model, which involves decomposing the error term into an individual specific effect, u_n, assumed to be normally distributed with mean zero and variance σ_u^2 and an i.i.d. normal variate with mean zero and variance σ_e^2.

[33] We present results for two models following this approach:

$$\text{(CV1)}\log\left(1 + \text{WTP}_{nt}\right) = f(\boldsymbol{Scope}_{nt}, \boldsymbol{Treat}_{nt}; \alpha^{CV1}) + u_n + e_{nt},$$

$$\text{(CV2)}\log\left(1 + \text{WTP}_{nt}\right) = f(\boldsymbol{Scope}_{nt}, \boldsymbol{Treat}_{nt}, \boldsymbol{Covariate}_{nt}; \alpha^{CV2}) + u_n + e_{nt}.$$

[34] The first set of variables to enter the models is ***Scope***, which captures the degree of environmental improvement represented by the scenario presented for valuation. Two variables are entered from this group, *Log %Change*, the log of the percentage change in high quality locally that occurs within 8 years and *T95* × *Log %Change*, a variable which interacts *Log %Change* with an indicator for the 95% scenario treatment, *T95*, equal to 1 if the CV scenario results in an improvement to 95% in 8 years, and equal to zero otherwise. Standard economic theory suggests the larger the change the more respondents should be willing to pay. Since initial water quality levels vary over the sample, however, two different respondents could be shown the same size change, and for one respondent it would represent 95% of the water in the local area being of high quality, while for another respondent it would represent 75%. If

there is declining marginal utility in the spatial extent of the improvement, then a scenario resulting in 95% high quality locally should be worth less than the improvement to 75% high quality locally if the absolute size of the improvement is the same in each case. It is possible, however, that respondents only care about the magnitude of the actual change in which no effect should be found.

[35] The second group to enter the CV models, ***Treat***, contains three variables. The first is *Payment Card*, an indicator for whether the observation relates to a response to the payment card question as opposed to a DCCV response. Given past empirical comparisons and the theoretical rationale put forward by *Carson and Groves* [2007], we expect to see this variable enter with a negative coefficient. The other two variables in this set are *PC* × *PC First*, a dummy equal to 1 if the observation is a PCCV response and the PCCV question was asked first, and *DC* × *DC First*, a dummy equal to 1 if the observation is a DCCV response and the DCCV question was asked first. These two variables do not have clear-cut predictions; however, we present some possible interpretations of the findings in our discussion of the results in section 6.

[36] The third group of variables, which enter model CV2 only, is $\boldsymbol{Covariates_{nt}}$, a vector of respondent characteristics, such as income, education, use of the water environment, and membership of an environmental club. Some of these covariates have a theoretical expectation, such as that frequent users of the water environment should be willing to pay more for its improvement than nonusers. Consistency of the results with such theoretical expectations is an important test of the validity of the results. We discuss our findings in relation to this test in section 5.

[37] The terms $\boldsymbol{\alpha}^{CV1}$ and $\boldsymbol{\alpha}^{CV2}$ are the parameter vectors to be estimated for models CV1 and CV2, respectively. The error term includes u_n, an individual specific effect assumed to be i.i.d. normally distributed over respondents with mean zero and variance σ_u^2, and e_{nt}, a normal i.i.d. variate with mean zero and variance σ_e^2. This allows the response errors to be correlated within the respondent.

4.2. Discrete Choice Experiment Models

[38] We analyze the data obtained from the DCE using the conditional logit model [*McFadden*, 1973]. Let choice_{nit} be a dummy variable equal to 1 if respondent n chose option i in choice situation t, and equal to zero otherwise. Respondent utility u_{nit} is composed of a deterministic component $v_{nit} = f(\boldsymbol{X};\ \beta)$ plus an i.i.d. standard Gumbel error term $\varepsilon_{nit.}$ Then we have the conditional logit probability expression:

$$\text{Prob}(\text{choice}_{nit}) = \frac{e^{v_{nit}}}{\sum_j e^{v_{njt}}},$$

where j indexes the alternatives in choice situation t. The above probability implies the following log likelihood:

$$LL = \sum\nolimits_n \sum\nolimits_j \sum\nolimits_t \text{choice}_{njt} \log\left[\text{Prob}(\text{choice}_{njt})\right].$$

We estimate the following two DCE models within this framework:

$$(\text{DCE1})\ v_{nit} = g(Scope_{nit}, SQ_{nit}, Treat_{nit}, Cost_{nit}; \beta^{DCE1}),$$

$$(\text{DCE2})\ v_{nit} = g(Scope_{nit}, SQ_{nit}, Treat_{nit}, Cost_{nit}, Covariates_{nit}; \beta^{DCE2}).$$

The two models differ only insofar as DCE2 includes respondent covariates but DCE1 does not. The utility functions corresponding to the DCE models do not map neatly onto the willingness to pay functions specified for the CV models despite the appearance of the same variable sets *Scope*, *Treat*, and *Covariates*. The richness of the data obtained via the DCE allows a richer specification of the value of environmental improvements than does the CV data. In particular, within the *Scope* group, we are able to include separate variables for each of the attributes *HighL8*, *LowL8*, *HighN8*, *LowN8*, and *High20*. For the CV data by contrast, there was insufficient variation to identify each of these scope variables separately. All observations on low quality nationally, for example, were identical in our data set. A single scale, high quality locally, was therefore used to capture the degree of improvement.

[39] The DCE models also include an alternative specific constant, labeled *SQ*, which indicates the status quo, or "no change" alternative, which is present in each choice set. This variable captures the average preference for the status quo after allowing for the influence of the attribute level differences, modeled linearly. When such a variable is included in a choice model and enters the model with a positive coefficient, it is typically interpreted as a status quo bias, an excessive preference for the status quo given the levels of its attributes in comparison with change alternatives. The opposite interpretation holds for a negative coefficient [*Hartman et al.*, 1991].

[40] The *Treat* group in the DCE models contains two variables, *PC First*, an indicator for whether the PCCV question was asked before the DCE questions, and *DC First*, an indicator for whether the DCCV question was asked before the DCE questions (in which case the PCCV question would have been asked after the DCE questions). These variables are entered into the model as interactions with *SQ*.

[41] The payment vehicle variable *Cost* enters both DCE models linearly with coefficient β^{Cost}, which we interpret as minus the marginal utility of income. The final group of variables, *Covariates*, appears in model DCE2 only. Respondent characteristics enter the DCE model via interactions with the *SQ* and *Cost* variables. When interacted with *Cost*, respondent characteristics impact on their willingness to pay via their effect on the marginal utility of income; when they enter via an interaction with *SQ* they impact the probability of choosing an improvement scenario at all.

5. Findings

5.1. Contingent Valuation Estimates of Benchmark WFD Implementation Scenarios

[42] Results from the interval censored regression models combining DCCV and PCCV responses are presented in Table 3. In model CV1, which includes no respondent covariates, the coefficient on *Log %Change* is of its expected positive sign and significant at the $p < 0.05$ level. It is an elasticity, so it implies that a 1% improvement in the proportion of high quality improved, e.g., the difference between an improvement of 50%, such as from 25% to 75%, and an

Table 3. Interval Censored Models Combining DCCV and PCCV Responses

Variable	Mean[a]	CV1[a,b,c,d] Coefficient	CV1 Standard Error	CV2[a,b,c,d] Coefficient	CV2 Standard Error
Constant	1.000	1.378	(1.151)	−0.471	(1.298)
Log %Change	4.315	0.676	(0.275)**	0.509	(0.305)*
T95 *X Log %Change*	2.188	−0.080	(0.024)***	−0.069	(0.025)***
PC X T95 *X Log %Change*	1.094	0.042	(0.016)***	0.043	(0.016)***
Payment card (PC)	0.500	−0.562	(0.061)***	−0.817	(0.079)***
DC *X* DC first	0.092	0.189	(0.088)**	0.167	(0.086)*
PC *X* PC first	0.245	−0.361	(0.055)***	−0.346	(0.053)***
Log income	4.325			0.227	(0.037)***
Missing income	0.265			1.237	(0.220)***
Male	0.470			0.155	(0.054)***
Child at home	0.262			0.120	(0.063)*
Wales	0.099			0.077	(0.108)
Water user	0.850			0.685	(0.201)***
Pollution control	0.859			0.796	(0.199)***
P_Con. *X* water user	0.730			−0.581	(0.217)***
Environmental club member	0.271			0.230	(0.063)***
Understood	0.865			0.284	(0.084)***
Under *X* not concentrate	0.051			−0.209	(0.124)*
PC *X* Edu_High	0.197			0.399	(0.074)***
PC *X* Edu_Med	0.194			0.193	(0.071)***
σ					
σ_u		0.885	(0.027)***	0.791	(0.026)***
σ_e		0.715	(0.019)***	0.706	(0.018)***
ρ		0.605		0.556	
Observations		2778		2778	
Log likelihood		−5101.1		−4983.4	
Pseudo R^2		0.055		0.076	

[a]Results are weighted for age, sex, and region based on the 2001 UK Census. Further details on the weights used are available from the authors on request.
[b]Models 1 and 2 are interval censored regression, with no assumed within-person correlation; Models 3 and 4 are interval censored regressions which do allow for within person correlation. The left-hand side for each model is the pair {ly1,ly2},where ly1 is the log of one plus the lower bound of WTP and ly2 is the log of one plus the upper bound of WTP.
[c]Standard errors are robust, calculated using the Huber-White estimator.
[d]Stars indicate p-value for two-sided t-test: $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

improvement of 50%*(1 + 1%), such as from 25% to 75.5%, results in a 0.68% increase in WTP. The interaction of the T95 indicator with *Log %Change* has a negative sign ($p < 0.01$) suggesting that respondents are less willing to pay for a given change if it takes them all the way to 95% of the local area water being of the highest quality than if it takes them to 75%. The effect is just over 10%, the magnitude of the main *Log %Change* coefficient. The interaction of this variable with the *Payment Card* indicator is positive ($p < 0.01$) and about half the magnitude of the original T95 *X Log %Change*, suggesting that PCCV WTP for increased amounts of a high quality water environment diminishes less than DCCV WTP over the range of spatial high quality density.

[43] Turning to treatment effects, consistent with prior expectations, *Payment Card* has a very significant ($p < 0.01$) downward effect on the WTP estimate. The coefficient of −0.562 implies that the PCCV treatment leads to an ∼43% lower WTP estimate, all else equal. The order effects are also significant. *DC X DC First* enters with a positive coefficient indicating that when the DCCV question comes first it results in higher DCCV estimates. By contrast, *PC X PC First* enters negatively, implying that PCCV WTP is lower when it is the first of the elicitation questions to appear. Taken together these findings suggest that PCCV WTP is substantially below DCCV WTP, and particularly when it occupies its typical first position.

[44] The statistic ρ is equal to the fraction of total variance accounted for by the random individual effects. This takes a value of 0.605 in model CV1, which indicates the importance of these effects to the fit of the model, and hence shows strongly that response errors are correlated across choice.

[45] Model CV2 in Table 3 includes the same scope and treatment variables as model CV1 plus a number of respondent covariates. The first thing to note about these results is that the addition of these new variables does not change the signs of any of the experiment treatment variables. The most noticeable changes are in the magnitude of the *Payment Card* indicator, which has jumped up substantially (for reasons noted below), and in the *Log %Change* variable, which has fallen ∼25% in magnitude and lost some significance (although it is still significant at the $p < 0.10$ level). The addition of the 14 respondent-related covariates results in a large improvement in the log likelihood that is significant at the $p < 0.01$ level.

[46] The first respondent covariate considered in model CV2 is income. Since some of the sample refused to provide income information (27%), as is typical in surveys, we include two variables to model the income effect. The first, *Log Income*, is equal to the log of household income (£/week) for those that answered the income question and equal to zero otherwise. The second is an indicator variable, *Missing Income*, equal to 1 if income was not recorded for

the respondent, and equal to zero otherwise. In combination, the coefficient on *Log Income* can be interpreted as the income elasticity of WTP for those who answered the income question, and the coefficient on *Missing Income* can be interpreted as the mean income effect of those who did not provide their income. The magnitude of the income elasticity of WTP is 0.23 in model CV2 and is significant at the $p < 0.01$ level. This elasticity tends to be smaller than its ordinary income elasticity of demand for theoretical reasons [*Flores and Carson*, 1997] and because measurement error in income tends to attenuate the coefficient toward zero.

[47] Females tend to give lower WTP estimates ($p < 0.01$), which is, to some degree, offset by those with children at home tending to give higher WTP estimates ($p < 0.10$). Last, with respect to demographic variables, residents of Wales are WTP slightly more than those of England but this effect is not statistically significant at conventional levels suggesting that responses from England and Wales can be combined.

[48] As expected, water users are willing to pay substantially more ($p < 0.01$) than those who do not use water under our broad definition of using water outdoors in England and Wales in the previous year. Likewise, those who express a pro-environmental view with respect to pollution control are WTP substantially more ($p < 0.01$) than those who did not. An interaction between *Pollution Control* and *Water User* is negative and significant ($p < 0.01$). This term suggests that while the joint effect of these different variables is positive, it is subadditive. Finally, being a member of an environmental club or organization (broadly defined) is associated with a moderate size increase in WTP which is significant at the $p < 0.01$ level.

[49] Two variables related to interviewer assessment of the respondent during the interview are included. The first of these is an indicator of whether the respondent was seen as having understood the valuation questions. Those rated as understanding (86.5% of respondents) are WTP more than those who did not ($p < 0.01$). The other variable is an interaction of the understood indicator with the interviewer rating the respondent as not concentrating. The 4.8% of respondents classified as understanding but not concentrating are willing to pay less (about the same as those not understanding) with this effect being significant at $p < 0.1$ in both models. This pair of variables worked better than inclusion of both understanding and concentration indicators because of the high correlation between them.

[50] The final pair of variables in the model is a set of interactions between *Payment Card* and indicators for the middle and high education groups in our sample. The high education interaction is large, offsetting almost half of the negative payment card coefficient, and highly significant ($p < 0.01$). The interaction of the middle education group with the *Payment Card* indicator is substantially smaller, although still significant at $p < 0.01$. There were two somewhat surprising aspects of these two interaction terms. At first, we included indicators for the middle and high education groups in our original modeling effort and they were significant predictors of WTP. We then added a number of interactions of the respondent covariates with the *Payment Card* indicator. Only two of the education interactions turned out to be strong predictors and when they were included, the indictors for high and medium education levels were no longer significant on their own. This suggests that those with different education levels may be responding differently to a payment card with the response of higher education levels being much closer to that of the DCCV treatment.

[51] Table 4 presents estimates of median and mean WTP from model CV1, our preferred model since it only includes the experimental design variables and accounts for the within-respondent correlation, by question type and order. Under the commonly used assumption that WTP is log-normally distributed, mean PCCV WTP is either £50.5 or £72.9, and mean DCCV is either £106.5 or £128.9 depending on the order of elicitation questions asked. The difference between PCCV is greatest when the PCCV question comes first and smallest when both PCCV and DCCV questions are preceded by the DCE.

[52] A useful comparison can be drawn between DCCV estimates from the model, and those based on the Turnbull nonparametric method [*Turnbull*, 1976], which imposes weak monotonicity on the percent "yes" as a function of the bid amount and allows a calculation of a lower bound on mean WTP by assuming all of the density for each set of interval observations is at the lower bound of the interval. This effectively assumes the most conservative distribution that is consistent with the observed choices. The Turnbull method allows for the possibility of a spike at zero, a likely feature of our data that is not well approximated by a lognormal and which is likely responsible for a lack of fit in both the left and right tails. For our data, the Turnbull lower bound on mean WTP is £127.4 when the DCCV question is asked first, and £106.7 when the DCCV question is not asked first. These estimates are approximately the same as those shown in Table 4. The reason why our model estimates are not higher than the Turnbull estimates is due to the fact that we have assumed a lognormal distribution, and have applied the conservative assumption that none of our observed respondents who said "yes" would have paid more than £350 for consistency. It is something of a coincidence that both of these assumptions result in model estimates that are almost identical to the Turnbull estimates. The comparison suggests, however, that the DCCV estimates shown in Table 4 can be

Table 4. CV WTP Estimates for Benchmark 95% Scenario, by Question Type and Order

Question Order	PCCV WTP £ hh^{-1} yr^{-1} [a,b,c,d]			DCCV WTP £ hh^{-1} yr^{-1} [a,b,c,d]		
	Median	Mean	95% C.I.	Median	Mean	95% C.I.
PCCV first	26.0	50.5	(47.4, 53.6)	55.3	106.5	(100, 113)
DCCV first	37.7	72.9	(68.4, 77.4)	67.0	128.9	(121, 136.8)
DCE first	37.7	72.9	(68.4, 77.4)	67.0	128.9	(121, 136.8)

Median WTP is calculated as $\exp(X\beta) - 1$ and mean WTP is calculated as $\exp(X\beta + 0.5[\sigma_u^2 + \sigma_e^2]) - 1$.

[a] Figures are calculated based on improvements from current (2007) water environment status levels, as presented in Table 1, to 95% high quality in local and national areas by 2015, with the remainder at medium quality. No further improvement occurs beyond this date.

[b] WTP results are based on CV model 3 coefficients presented in Table 3.

[c] Estimates are based on £ (July 2007).

[d] Results are weighted for age, sex, and region based on the 2001 UK Census. Further details on the weights used are available from the authors on request.

considered conservative. Higher estimates of mean WTP can be derived from the DCCV data with reasonable alternative assumptions.

5.2. Discrete Choice Experiment Estimates

[53] Table 5 presents DCE results estimation results for the two models described in section 4 above. Model DCE1 includes no covariate effects, except for a treatment effect to control for whether or not the PCCV question was asked before or after the DCE. The model is a reasonably good fit for the data. The (McFadden) pseudo-R^2 is 0.18, and for the coefficients all are of the expected sign and statistically significant at least at the 5% level. The attributes *HighL8*, *HighN8*, and *High20* enter positively, and *LowL8* and *LowN8* enter negatively, as expected. Furthermore, *Cost* is also negative and significant at $p < 0.01$.

[54] Both models are linear in environmental improvement attributes. This functional form implies that the value of improving a water body depends on its current status and the status following the improvement, but not on the status of surrounding water bodies or the state of the national water environment generally. Nonlinear forms were tested which allowed for, and found, diminishing marginal utility with respect to water environmental improvement, but these models did not significantly outperform the linear model in a statistical sense, and so the simpler linear model was adopted.

[55] With respect to the nature and location of environmental improvements, the results imply that respondents prefer percentage point improvements from medium to high quality over percentage point improvements from low to medium quality (since the coefficients on *HighL8* and *HighN8* are greater, in absolute value terms, than the coefficients on *LowL8* and *LowN8*, respectively). This finding of an increasing marginal utility over the qualitative range of improvement does not contradict our previous CV result that increases in the spatial extent of high quality, measured by the *Log %Change* variable, give rise to a diminishing marginal utility. It is easy to imagine a person attaching a higher value to improvements from medium to high quality than from low to medium quality, for example, if only high quality is sufficient to use the site for a valued recreational pursuit. Yet, once there is a sufficient number of high quality sites, adding to this number might add little improvement to the same person's recreational opportunities, and so his marginal utility for increases in the spatial extent of high quality would be diminishing.

[56] Additionally, respondents prefer a percentage improvement in the national environment more than a percentage improvement in their local area (the coefficients on *HighN8* and *LowN8* are greater, in absolute value terms, than the coefficients on *HighL8* and *LowL8*, respectively). With regard to the latter finding, however, the size ratio of national to local areas is $\sim$20:1, and so the coefficients on *HighL8* and *LowL8* should be multiplied by 20 to draw a comparison with the coefficients on *HighN8* and *LowN8* in equivalent spatial terms. If this is done, local improvements are seen to be valued very much higher per hectare than nonlocal improvements. Thus, the results show that the typical person values local improvements substantially more than nonlocal improvements per hectare, which is as expected.

[57] The *StatusQuo* (*SQ*) indicator variable enters model DCE1 with a negative coefficient, indicating that people would prefer an improvement alternative to the status quo after taking account the utility effects of the associated environmental improvements.

[58] Thus, rather than the more commonly cited "status quo bias," we find a general reluctance to stick with the status quo. The variable *SQ X PC First* enters the model positively, however, and with a coefficient 85% of the size of the *Status Quo* coefficient. This implies that the *SQ* effect

Table 5. DCE Estimation Results

Variable	Mean[a]	DCE1[a,b] Coefficient	DCE1[a,b] Standard Error[c,d]	DCE2[a,b] Coefficient	DCE2[a,b] Standard Error[c,d]
HighL8	0.340	0.915	(0.100)***	0.934	(0.102)***
LowL8	0.346	−0.615	(0.123)***	−0.658	(0.121)***
HighN8	0.399	1.128	(0.110)***	1.151	(0.111)***
LowN8	0.293	−0.918	(0.171)***	−0.944	(0.171)***
High20	0.605	0.423	(0.189)**	0.439	(0.186)**
Status quo (SQ)	0.333	−0.364	(0.180)**	3.361	(0.560)***
SQ *X* PC question first	0.163	0.311	(0.130)**	0.331	(0.134)**
SQ *X* log income	1.442			−0.459	(0.091)***
SQ *X* missing income	0.088			−2.208	(0.528)***
SQ *X* water user	0.283			−0.590	(0.168)***
SQ *X* pollution control	0.286			−0.681	(0.164)***
SQ *X* Edu_High	0.131			−0.573	(0.159)***
Cost (£ hh^{-1} yr^{-1})	0.398	−1.185	(0.048)***	−1.474	(0.082)***
Cost *X* male	0.188			0.201	(0.095)**
Cost *X* Edu_High	0.158			0.203	(0.098)**
Cost *X* Wales	0.041			0.595	(0.131)***
Observations		29,169		29,169	
Log likelihood		−8769.83		−8440.64	
Pseudo-R^2		0.18		0.21	

[a]Results are weighted for age, sex, and region based on the 2001 UK Census. Further details on the weights used are available from the authors on request.
[b]The model is conditional logit; dependent variable is choice, a dummy equal to 1 if the option was chosen.
[c]Standard errors are robust, calculated allowing for within-person correlation.
[d]Stars indicate p-value for two-sided t-test: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

is almost wiped out if PCCV was the first elicitation question asked. That is, respondents are more likely to choose the status quo if the PCCV question had already been asked, than if it had not.

[59] The coefficients on the scope and treatment variables in model DCE2 are qualitatively the same, and quantitatively almost identical, to those found for model DCE1. The only substantial differences are for the *Status Quo* and *Cost* variables, and this is because these variables enter with interaction terms in DCE2. The interactions with *SQ* all enter negatively, and indicate that people were more likely to choose an improvement alternative if they had high incomes, were water users, held attitudes supporting pollution control efforts, and had a higher level of education. All of these findings are consistent with expectation and so are supportive of the construct validity of the survey. The interactions with *Cost* indicate that, all else equal, men are willing to pay more than women for environmental improvements, those with a higher level of education are willing to pay more than others, and those living in Wales are willing to pay more than those living in England.

[60] Table 6 shows marginal WTP figures for each of the environmental attributes, and corresponding WTP for the benchmark 95% scenario by question order. As anticipated, given the literature on DCE-CV comparisons [e.g., *Cameron et al.*, 2002; *Foster and Mourato*, 2003; *Hanley et al.*, 1998], the estimated values from our DCE model for the 95% scenario are substantially and significantly ($p < 0.01$) higher than those from the PCCV model and the DCCV model for all question orders, based on two-sided *t*-tests.

5.3. Scaled WTP Estimates

[61] We derive low (PCCV)-scaled and high (DCCV)-scaled values for percentage point changes in local high and low quality, and national high and low quality, by applying the formula below:

$$s_{k0}^{CV} = s_{k0}^{DCE} \frac{s^{CV}}{\sum_k s_{k8}^{DCE} \Delta_{k8}^{95\%}} \delta_8.$$

In this expression, s_{k0}^{CV} is CV-scaled WTP for an instantaneous 1% change in dimension $k \in \{HighL, LowL, HighN, LowN\}$; s_{k8}^{DCE} is the corresponding DCE estimate from Table 6, measuring the value of an 8-yr improvement path to an ultimate 1% change, s^{CV} is the CV estimate of WTP for the 95% scenario, as drawn from Table 4; $\sum_x s_{x8}^{DCE} \Delta_{x8}^{95\%}$ is the sum of the marginal DCE WTP estimates for *HighL8*, *LowL8*, *HighN8*, and *LowN8* multiplied by the corresponding changes in those variables under the 95% scenario, as drawn from Table 6. The final term in the expression, δ_8, is the discount factor necessary to equate the value of an 8-yr improvement path with an instantaneous change, i.e., $\delta_8 = \frac{1}{8}\sum_{t=1}^{8} (1+r)^{-t}$, for discount rate *r*.

[62] The formula embeds a crucial step, which is to treat the s_{k8}^{DCE} parameters as representing relative values of *HighL*, *LowL*, *HighN*, and *LowN*, into which the 95% scenario can be exhaustively decomposed. That is, in applying the formula we interpret the derived s_{k0}^{CV} values as estimates of WTP for a 1% improvement in the *k*-th value dimension with no deterioration thereafter. This step is innocuous if one is willing to impose, as is the case here, an exogenous discount rate.

[63] Based on the expression above, Table 7 presents PCCV- and DCCV-scaled values for percentage improvements in local high and low quality, and national high and low quality, for two discount rate assumptions, 3.5%, which is the the "Green Book" rate used for U.K. public policy, and 7.0%. The figures for low (PCCV)-scaled values are derived using $s^{CV} = £50.5$ per household per year, the lower of the PCCV estimates for the 95% scenario, which corresponds to the PCCV question having been asked first. For high (DCCV)-scaled values, $s^{CV} = £128.9$ per household per year, which is the higher of the DCCV estimates corresponding to the DCCV question having been asked first. We use the furthest apart estimates from each elicitation method in order to capture the full range of possible values, although we note that the range of values reported could be justifiably extended to incorporate a sampling variation, as measured by the 95% confidence intervals reported in Table 4.

[64] The estimates presented in Table 7 allow valuation of programs of water environment improvements as a function of the geographic scale of the improvements, the

Table 6. DCE WTP Estimates for Marginal Changes in Variables, and for Benchmark 95% Scenario

Parameter	Marginal Effect ($\Delta X = 1\%$) WTP (£ hh^{-1} yr^{-1})[b,c]		95% Scenario[a] Δ_X[d]	WTP (£ hh^{-1} yr^{-1})[b,c]	
$S_{HighL8}^{DCE}\Delta_{HighL8}$	0.77	(0.6, 0.95)	86.0%	66.4	(51.3, 81.5)
$S_{LowL8}^{DCE}\Delta_{LowL8}$	−0.52	(−0.72, −0.32)	−58.6%	30.4	(18.5, 42.3)
$S_{HighN8}^{DCE}\Delta_{HighN8}$	0.95	(0.76, 1.14)	80.0%	76.2	(60.8, 91.6)
$S_{LowN8}^{DCE}\Delta_{LowN8}$	−0.77	(−1.07, −0.48)	−44.0%	34.1	(21.3, 46.9)
$S_{High20}^{DCE}\Delta_{High20}$	0.36	(0.05, 0.67)	86.0%	30.7	(3.9, 57.5)
$S_{SQ}^{DCE}\Delta_{SQ}$			−1	30.7	(1, 60.4)
$S_{SQPC\,First}^{DCE}\Delta_{SQPC\,First}$			−1	−26.2	(−47.8, −4.7)
Total WTP (PCCV first)				242.3	(216.5, 268.1)
Total WTP (DCE or DCCV first)				268.5	(241.8, 295.3)

[a]Under the "95% scenario," 95% local and national areas are brought to high quality by 2015, with the remainder at medium quality. No further improvement occurs beyond this date.
[b]WTP results are based on DCE model 1 coefficients presented in Table 5.
[c]Estimates are based on £ (July 2007).
[d]Based on improvements from current (2007) water environment status levels, as presented in Table 1.

Table 7. Scaled WTP Estimates for Marginal Changes in Current Status

	Low (PCCV)-Scaled WTP (£ hh^{-1} yr^{-1})[a]		High (DCCV)-Scaled WTP (£ hh^{-1} yr^{-1})[a]	
Parameter	d.r.[b] = 3.5%	d.r. = 7.0%	d.r. = 3.5%	d.r. = 7.0%
S^{CV}_{HighL0}	0.16	0.14	0.41	0.36
S^{CV}_{LowL0}	−0.11	−0.09	−0.28	−0.24
S^{CV}_{HighN0}	0.20	0.17	0.51	0.44
S^{CV}_{LowN0}	−0.16	−0.14	−0.41	−0.36

[a]Estimates derived as discussed in the text of the paper, based on £(July 2007).
[a]d.r., discount rate.

extent of population around the area improved, and the change in quality afforded by the improvements. The final step in our analysis is now to derive the inverted valuation function so that instead of measuring the value of national policies to households it measures the value of individual catchment and water body improvements as a function of the size of the area improved, the qualitative scope of improvement, low to medium, medium to high, or low to high, and the density of the population surrounding the area improved.

[65] Table 8 presents the inverted function, which makes use of the s^{CV}_{k0} parameters presented in Table 7, plus two additional parameters, p and q. The parameter p is a local scalar equal to the population living within 30 miles of the water body in question divided by 1% of the area of a 30-mile radius circle. Similarly, q is a national scalar equal to the national population, divided by 1% of the area of the country, including coastal areas.

[66] Making use of the valuation function in Table 8 requires the use of geographic information systems (GIS) to obtain local population data for each water body. As an example, suppose there is a lake measuring 1 km^2 in a region with 1000 surrounding households living within 30 miles. Suppose further that the lake is currently at medium quality, and we wish to know the value of improving it to high quality, assuming a discount rate of 3.5%. We would apply the formula in Table 8 to give, as a low (PCCV)-scaled WTP estimate, $(1000/73.23)*0.16 + 12{,}883*0.20 =$ £2,579; and as a high (DCCV)-scaled WTP estimate, $(1000/73.23)*0.41 + 12{,}883*0.20 =$ £6,576. These values represent total WTP, relative to a medium quality base, for each year at which the lake is at high quality.

[67] The average value of an improvement from low to medium quality in the Solway-Tweed River Basin district, the lowest density district in England and Wales, is £2,870 per km^2 yr^{-1} for the low (PCCV)-scaled estimate and £7,321 per km^2 yr^{-1} for the high (DCCV)-scaled estimate. A similar improvement in the Thames River Basin district, the highest-density district in England and Wales, is valued at £8,911 per km^2 yr^{-1} for the low (PCCV)-scaled estimate and £22,802 per km^2 yr^{-1} for the high (DCCV)-scaled estimate. These examples demonstrate the importance of local population to the values obtained.

Table 8. Water Body Valuation Function

Qualitative Scope of Improvement	WTP (£ km^2 yr^{-1})[a]
Low quality to medium quality	$-(ps^{CV}_{LowL0} + qs^{CV}_{LowN0})$
Medium quality to high quality	$(ps^{CV}_{HighL0} + qs^{CV}_{HighN0})$
Low quality to high quality	$-(ps^{CV}_{LowL0} + qs^{CV}_{LowN0}) + (ps^{CV}_{HighL0} + qs^{CV}_{HighN0})$

[a]$p = (1/73.23)*$(households within 30 miles) (NB 73.23 = 1% of a 30-mile radius circle, measured in km^2; $q =$ (households in nation/national area [including coastal area, in km^2]/100). For England and Wales, $q =$ 12,883; s^{CV}_{k0} parameters to be drawn from Table 7.

6. Discussion of Treatment Effects

[68] The effects of question type and order have been found to be both economically and statistically significant in this study, although the bounds they form are still likely to be useful for many policy purposes. The value of the benchmark 95% scenario, for example, was found to vary from £50.5 to £72.9 per household per year via the PCCV responses, from £106.5 to £128.9 per household per year via the DCCV responses, and from £242.3 to £268.5 via the DCE responses. All six of these values are significantly different from one another ($p < 0.01$).

[69] Comparing across CV questions, the finding that DCCV values are higher than PCCV values is consistent with many previous findings [*Venkatachalam*, 2004; *Welsh and Poe*, 1998]. From a strategic behavioral perspective [*Carson and Groves*, 2007], the PCCV method is considered less robust than the DCCV method because it allows respondents more discretion to attempt to bring about the result they most want by giving answers that do not truthfully reflect their actual valuations. Under reasonable assumptions, this leads the PCCV approach to result in downwardly biased estimates. The DCCV approach, by contrast, is argued to be compatible with truth-telling due to the "take-it-or-leave-it" nature of the question, provided that a stringent set of auxiliary conditions are met. An analysis of DCCV responses is more sensitive, however, to distributional assumptions and outliers. Further, some cognitive psychologists have argued that the DCCV method may signal a value for the good, which "anchors" respondents' perceptions of what they would be willing to pay when unsure of their true valuations. Open-end approaches like the PCCV elicitation method are thought to be less susceptible to this sort of anchoring effect because respondents select their own WTP amount [*Green et al.*, 1998; *Jacowitz and Kahneman*, 1995; *Johnson and Schkade*, 1989]. (Although it should be noted that earlier suggestions were made that the particular range of amounts shown in the PCCV question can influence respondent answers.) Since these two perspectives have potential offsetting issues, the DCCV and PCCV approaches were both used to allow us to estimate a reasonable range of WTP for benchmark WFD implementation programs. There is always some possibility that the responses to either the DCCV or the PCCV questions are not indicative of what would happen if the government gave citizens an opportunity to vote on the policy scenario. However, comparisons of valuation survey responses with actual voting have generally shown reasonably close comparisons [e.g., *Vossler and Kerkvielt*, 2003; *Johnson*, 2006], even though laboratory experiments using purely hypothetical questions often suggest an upward bias. Success in obtaining an accurate response to our survey questions is likely dependent on the quality of the scenario description and convincing participants that the survey was

consequential in the sense that the government was going to seriously consider its results.

[70] We also find that values are sensitive to the order in which the questions were asked, a result that is also consistent with many previous studies [e.g., *Day et al.*, 2012] and behavior in actual markets. In the present case, DCCV WTP is found to be higher if the DCCV question came first, and PCCV WTP is found to be lower if the PCCV question came first. From the strategic behavioral perspective, the first scenario presented has special status since only the first scenario is free from the influence of prior scenarios. In contrast, various types of (nonstrategic) hypothesized learning [e.g., *Braga and Starmer*, 2005; *Plott*, 1996] suggest that answers to later questions are likely to be more reliable than answers to earlier questions. In the present analysis we have not attempted to distinguish between the strategic and anchoring hypotheses, instead, we have simply controlled for the order effects and reported the range of estimates we obtained.

7. Concluding Remarks

[71] The research presented here on WTP for potential water quality changes meets a new substantive policy need in England and Wales. Results are based on a carefully designed and well-tested stated preference survey that was implemented using a large in-person sample. The principal goal of the study was to develop a robust statistical valuation function capable of providing benefits estimates for national and regional programs of water quality improvements to meet the requirements of the Water Framework Directive (WFD), and to support the development of these programs by quantifying household's priorities with respect to the location of improvements and the types of improvements to be made. The results suggest that households in England and Wales value local improvements much higher than national improvements per km^2 of catchment, lake, or coastal water improved, as expected, and value improvements from medium quality (poor/moderate ecological status) to high quality (good/high ecological status) substantially more than improvements from low quality (bad ecological status) to medium quality (poor/moderate ecological status). Regionally averaged values for WFD improvements are found to vary from £2,263 to £39,168 per km^2 improved depending on where the improvement is made, the ecological scope of the improvement, and the source of the valuation estimate from within the range of treatments modeled.

[72] The results are limited in three important ways. First, the decision to focus on programs of improvements rather than on individually specified improvements meant that no information was given to respondents regarding which areas were to be improved except insofar as they were to be made in the local area, i.e., within 30 miles, or elsewhere. It is not hard to imagine that the range of values for individual water body improvements is likely to be substantial within these broad categories. For broad enough programs, errors in the values attributed to individual improvements are likely to cancel each other out. Considerable care should be taken, however, if using these results to make valuation estimates for one-off improvements. A second limitation of the results is that they only provide values for broad ranges of improvement. It is not strictly possible, for example, to use the results to value an improvement from poor to moderate ecological status because both status categories are embedded within the medium quality level. For some purposes, this may be a significant restriction on applicability. The final limitation of the results is that the range of estimates reported with respect to elicitation treatments may be too wide for some policy purposes. Narrowing this range is likely to require taking a stance on the most preferred elicitation method and question ordering.

[73] Despite these limitations, initial benefits estimates obtained from this study (as presented in the project's technical report [*NERA-Accent*, 2007]) have already been successfully used in several applications, including a national impact assessment, impact assessments for all 11 regional River Basin Management Plans in England and Wales, and the appraisal of water utility investment programs in support of the 2009 water price review in England and Wales. The results presented in this paper, which have been revised since the initial policy applications, might usefully be applied in the future to the second phase of River Basin Management Plans in 2015, and to the 2014 water company price review in England and Wales. Finally, with suitable adaptation, the results would serve as a cross-check on the values of water quality improvement programs in other countries.

Appendix A: Selected Show Card Materials

[74] This valuation context statement is an extract from the survey questionnaire:

[75] "Water quality is affected by pollution from households, farms, and businesses, and climate change. Some works are needed just to prevent water sites from getting worse. The government's policy is that the polluter will have to pay for these works. This will make some every day products more expensive and will increase household water and sewerage bills too.

[76] "The government has estimated that these extra costs to each household, including yours, will be £10 per year, in terms of higher water and sewerage bills and higher prices on everyday products.

[77] "Improving the environment requires more cutting of pollution, which will make products more expensive and will further increase household water and sewerage bills.

[78] "I am now going to show you cards which have two or three options for water environment improvements. For all the options, steps will be taken so there will be no worsening of the water environment at any site, the most cost-effective works will be used, the money will be ring-fenced to make the improvements, and information will be made available to the public on progress toward the improvements.

[79] "It is important for us to get realistic choices from you regarding the values of these programs, so before you make some real choices, please consider your household budget and all of the things that you and your household need or would prefer to spend your money on before you decide. Please also bear in mind that your water bill and other household expenses may change in future for other reasons not related to the water environment, and your income may also change in future. Your choices will influence how far to go with improvements, so will influence everyone's payment for improvements."

Dark Blue – quality is "High".

- There will be a diverse and natural range of plants, insects, fish, birds and other animals.
- Water will generally have the right degree of clarity and there will be no noticeable pollution.
- Water will generally be suitable for contact activities, such as rowing or wind surfing.

Mid-Blue – quality is "Medium".

- There will be plants, insects, fish, birds and other animals, but there will be some fish and other wildlife missing.
- Water will be slightly murky or discoloured in parts, and there will sometimes be visible pollution in some places, and some algal blooms.
- Water will be suitable for contact activities in some areas but not others.

Light Blue – quality is "Low".

- There may be limited or no plants or wildlife, or the water may be dominated by a single plant species.
- Water will generally be murky or discoloured, and may sometimes be bad-smelling in some places. There may also regularly be visible pollution in some places, and frequent algal blooms.
- Water will be unsuitable for contact activities.

Figure A1. CARD 4a: quality levels.

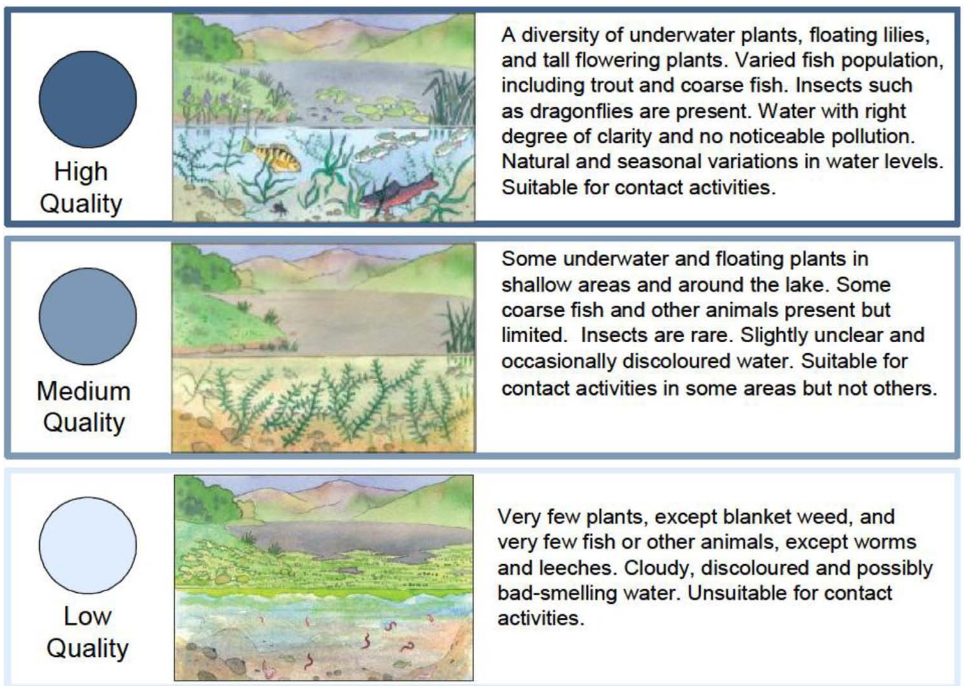


Figure A2. CARD 4b: lake.

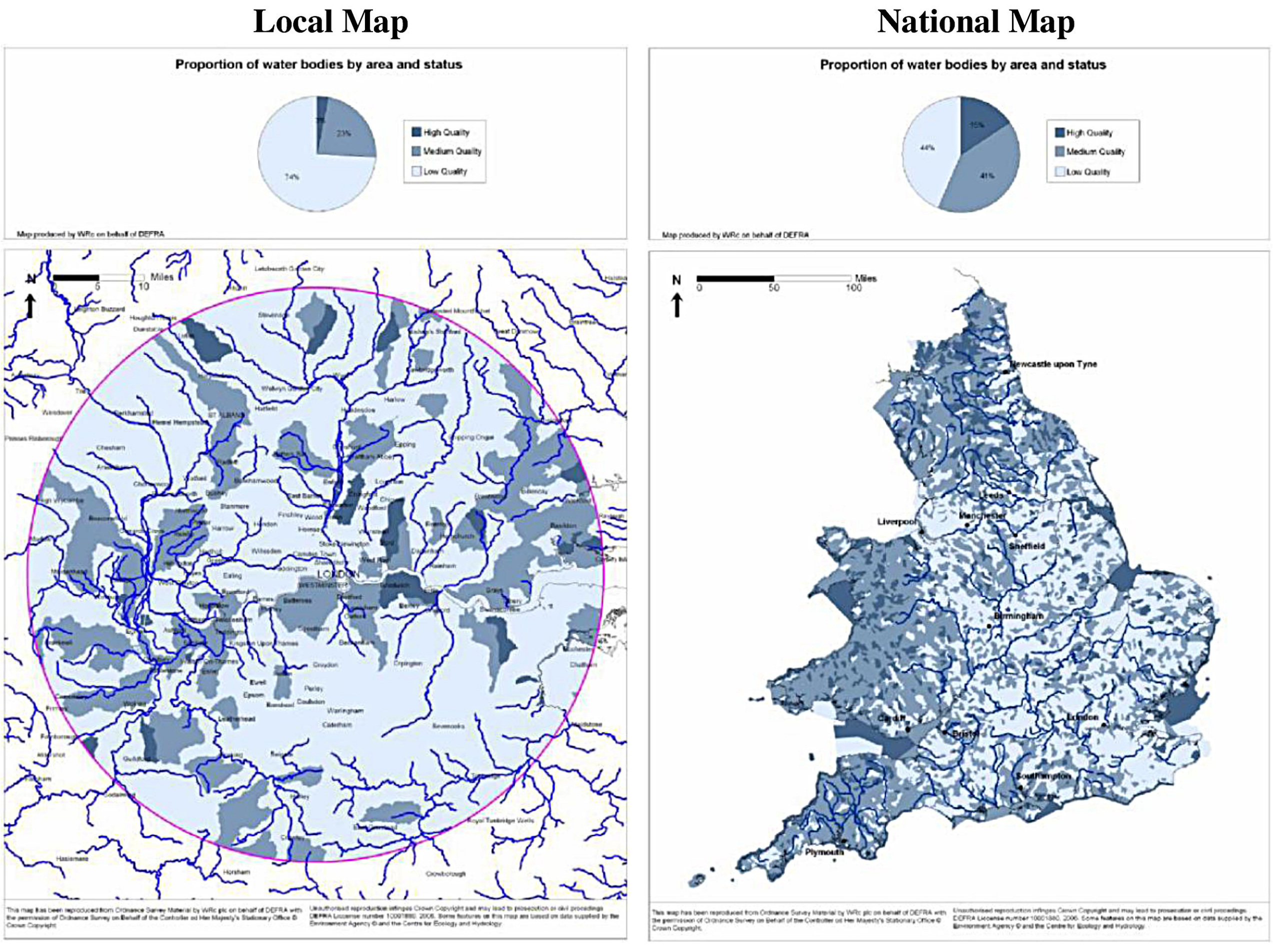


Figure A3. Example maps.

High Quality	Medium Quality	Low Quality

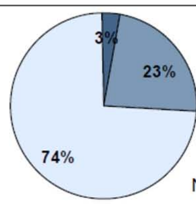
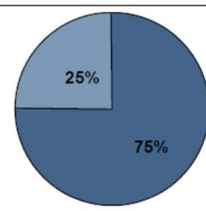
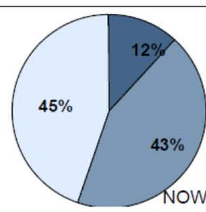
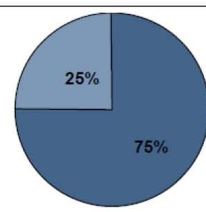
WILLINGNESS TO PAY CARD	Option A - No Change	Option B
Status of Local Area in 8 years time	 NOW AND 2015	 IN 2015
Status of England and Wales in 8 years time	 NOW AND 2015	 IN 2015
Status of England and Wales and Local Area in 20 years time IN 2027	Same as Now	95%
Increase in your water bill and other household payments. Note: this payment will be added to the cost of avoiding any worsening of the water environment.	**No increase in water bills or other household payments**	**£ --------- per year** (continuing indefinitely)

Figure A4. Example of a PCCV card.

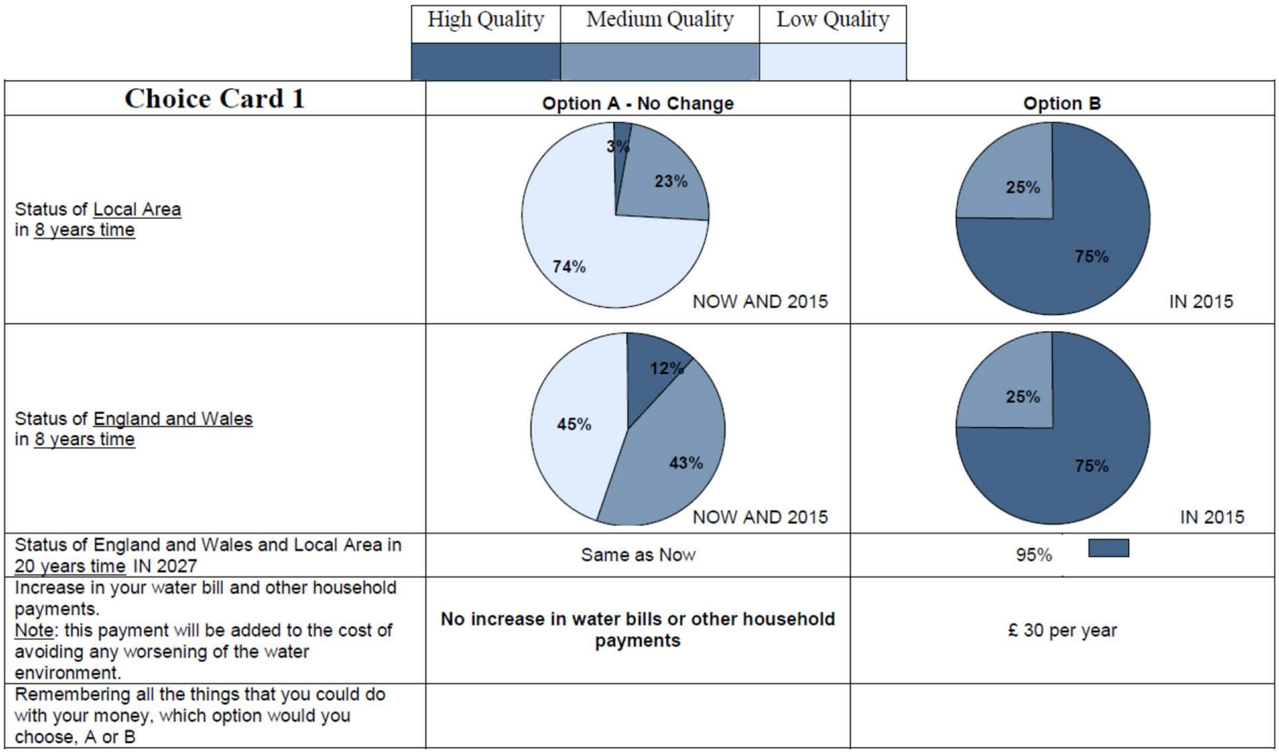


Figure A5. Example of a DCCV card.

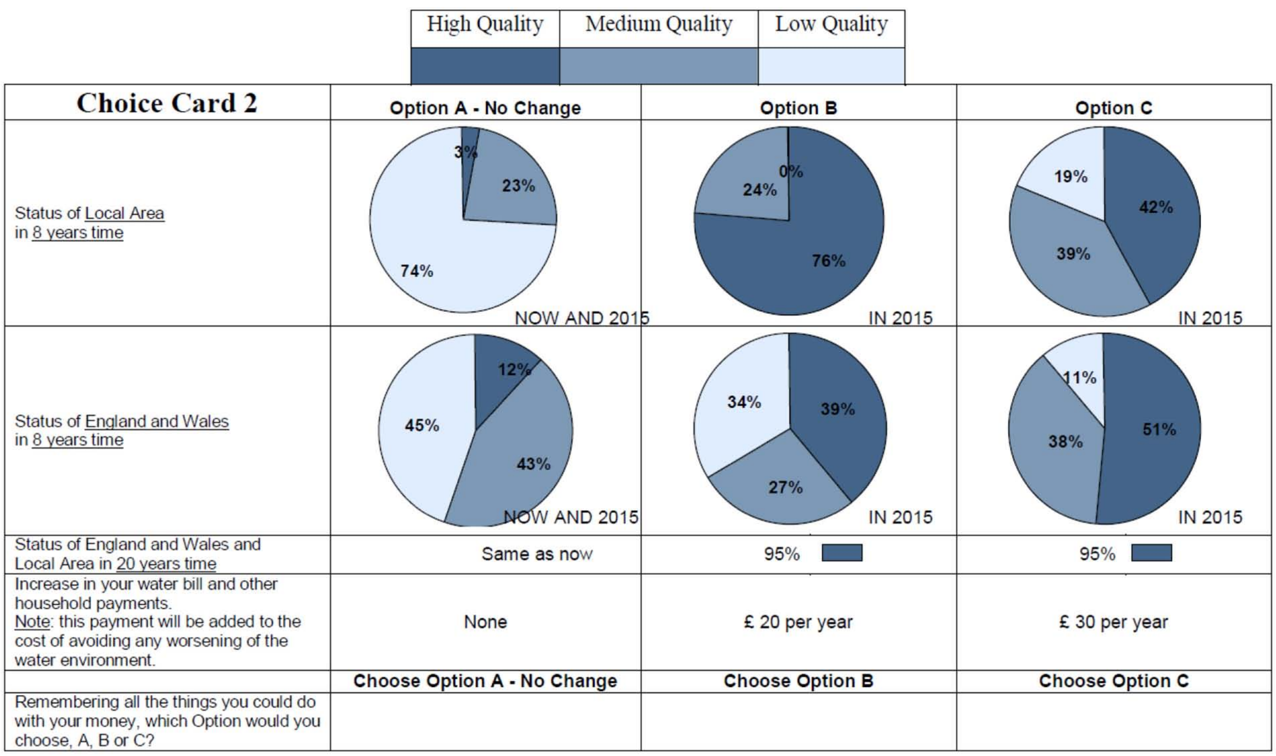


Figure A6. Example of a DCE card.

[80] **Acknowledgments.** This study was jointly funded and steered by Water Framework Directive stakeholders in the UK as part of the DEFRA-led Collaborative Research Programme, in a project commissioned from NERA Economic Consulting. Financial support to Paul Metcalfe from the ESRC is also gratefully acknowledged. The authors would like to thank Rob Curry (Environment Agency), Iain McGuffog (South West Water), John Joyce (Independent Consultant), Camilla Lundbak (DEFRA), and George Hutchinson (Queen's University, Belfast) for their inputs throughout the study. Any errors contained herein are our own.

References

Alam, M. K., and D. Marinova (2003), Measuring the total value of a river cleanup, *Water Sci. Technol.*, *48*(7), 149–156.

Arrow, K., R. Solow, E. Leamer, R. Radner, and H. Schuman (1993), Report of the NOAA panel on contingent valuation, *Federal Register*, *58*(10), 4601–4614.

Bateman, I. J., A. Munro, B. Rhodes, C. Starmer, and R. Sugden (1997), Does part-whole bias exist? An experimental investigation, *Economic J.*, *107*(441), 322–332.

Bateman, I. J., et al. (2002), *Economic Valuation with Stated Preference Techniques: A Manual*, 458 pp., Edward Elgar, Cheltenham, U.K.

Bateman, I. J., B. Day, S. Georgiou, and I. Lake (2006), The aggregation of environmental benefit values: Welfare measures, distance decay and total WTP, *Ecol. Econ.*, *60*(2), 450–460.

Bateman, I. J., et al. (2011), Making benefit transfers work: Deriving and testing principles for value transfers for similar and dissimilar sites using a case study of the non-market benefits of water quality improvements across Europe, *Environ. Resour. Econ.*, *50*(3), 365–387, doi:10.1007/s10640-011-9476-8.

Braga, J., and C. Starmer (2005), Preference anomalies, preference elicitation and the discovered preference hypothesis, *Environ. Resour. Econ.*, *32*, 55–89.

Brouwer, R. (2008), The potential role of stated preference methods in the Water Framework Directive to assess disproportionate costs, *J. Environ. Plan. Manage.*, *51*(5), 597–614.

Cameron, T. A., and D. D. Huppert (1989), OLS versus ML estimation of non-market resource values wth payment card data interval data, *J. Environ. Econ. Manage.*, *17*(3), 230–246.

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

Cameron, T. A., G. L. Poe, R. G. Ethier, and W. D. Schulze (2002), Alternative non-market value-elicitation methods: Are the underlying preferences the same?, *J. Environ. Econ. Manage.*, *44*(3), 391–425.

Carlsson, F., and P. Martinsson (2001), Do hypothetical and actual marginal willingness to pay differ in choice experiments?: Application to the valuation of the environment, *J. Environ. Econ. Manage.*, *41*(2), 179–192.

Carson, R. T., and T. Groves (2007), Incentive and informational properties of preference questions, *Environ. Resour. Econ.*, *37*(1), 181–210.

Carson, R. T., and T. Groves (2011), Incentive and informational properties of preference questions: Commentary and extensions, in *International Handbook of Non-Market Environmental Valuation*, edited by J. Bennett, ch. 15, pp. 300–321, Edward Elgar, Northampton, Mass.

Carson, R. T., and W. M. Hanemann (2005), Contingent valuation, in *Handbook of Enviromental Economics*, edited by K. M. Maler and J. R. Vincent, pp. 822–920, Elsevier, Amsterdam, Netherlands.

Carson, R. T., and R. C. Mitchell (1993), The value of clean water: The publics willingness-to-pay for boatable, fishable, and swimmable quality water, *Water Resour. Res.*, *29*(7), 2445–2454.

Day, B. H., I. J. Bateman, R. T. Carson, D. Dupont, J. J. Louviere, S. Morimoto, R. Scarpa, and P. Wang (2012), Ordering effects and choice set awareness in repeat-response stated preference studies, *J. Environ. Econ. Manage.*, *63*, 73–91, doi:10.1016/j.jeem.2011.09.001.

Department for Environment, Food and Rural Affairs (DEFRA) (2008), Overall Impact assessment for the Water Framework Directive (EC 2000/60/EC), *Rep.*, U.K., 65 pp., [Available at http://archive.defra.gov.uk/environment/quality/water/wfd/documents/RIA-river-basin.pdf.]

European Parliament (2000), Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action, *Official J. European Communities*, *L327*, 1–72.

Flores, N. E., and R. T. Carson (1997), The relationship between income elasticities of demand and willingness to pay, *J. Environ. Econ. Manage.*, *33*(3), 287–295.

Foster, V., and S. Mourato (2003), Elicitation format and sensitivity to scope: Do contingent valuation and choice experiments give the same results?, *Environ. Resour. Econ.*, *24* 141–160.

Green, D., K. E. Jacowitz, D. Kahneman, and D. McFadden (1998), Referendum contingent valuation, anchoring, and willingness to pay for public goods, *Resour. Energy Econ.*, *20*(2), 85–116.

Hanley, N., R. E. Wright, and W. L. Adamowicz (1998), Using choice experiments to value the environment: Design issues, current experience and future prospects, *Environ. Resour. Econ.*, *11*(3–4), 413–428.

Hanley, N., D. Bell, and B. Alvarez-Farizo (2003), Valuing the benefits of coastal water quality improvements using contingent and real behaviour, *Environ. Resour. Econ.*, *24*, 273–285.

Hanley, N., R. E. Wright, and B. Alvarez-Farizo (2006), Estimating the economic value of improvements in river ecology using choice experiments: An application to the water framework directive, *J. Environ. Manage.*, *78*(2), 183–193.

Hartman, R. S., M. J. Doane, and C. Woo (1991), Consumer rationality and the status quo, *Q. J. Econ.*, *106*(1), 141–162.

Hoehn, J. P. (1991), Valuing the multidimensional impacts of environmental: Policy, theory and methods, *Am. J. Agric. Econ.*, *73*(2), 289–299.

Hoehn, J. P., and A. Randall (1989), Too many proposals pass the benefit cost test, *Am. Econ. Rev.*, *79*(3), 544–551.

Jacowitz, K. E., and D. Kahneman (1995), Measures of anchoring in estimation tasks, *Personality Social Psychol. Bull.*, *21*(11), 1161–1166.

Johnson, E. J., and D. A. Schkade (1989), Bias in utility assessments: Further evidence and explanations, *Manage. Science*, *35*(4), 406–424.

Johnson, R. J. (2006), Is hypothetical bias universal?: Validating contingent valuation responses with a binding referendum, *J. Environ. Econ. Manage.*, *52*(1), 469–481.

Kahneman, D., A. B. Krueger, D. Schkade, N. Schwarz, and A. A. Stone (2006), Would you be happier if you were richer? A focusing illusion, *Science*, *312*(5782), 1908–1910.

Kontogianni, A., I. H. Langford, A. Papandreou, and M. S. Skourtos (2003), Social preferences for improving water quality: An economic analysis of benefits from wastewater treatment, *Water Resour. Manage.*, *17*(5), 317–336.

Kramer, R. A., and J. I. Eisen-Hecht (2002), Estimating the economic value of water quality protection in the Catawba River basin, *Water Resour. Res.*, *38*(9), 1182, doi:10.1029/2001WR000755.

Loomis, J., P. Kent, L. Strange, K. Fausch, and A. Covich (2000), Measuring the total economic value of restoring ecosystem services in an impaired river basin: Results from a contingent valuation survey, *Ecol. Econ.*, *33*(1), 103–117.

Louviere, J. J., D. A. Hensher, and J. Swait (2000), *Stated Choice Methods: Analysis and Application*, Cambridge Univ. Press, N. Y., 402 pp.

Lusk, J. L., and T. C. Schroeder (2004), Are choice experiments incentive compatible? A test with quality differentiated beefsteaks, *Am. J. Agric. Econ.*, *86*, 467–482.

McFadden, D. (1973), Conditional logit analysis of qualitative choice behavior, in *Frontiers in Econometrics*, edited by P. Zarembka, pp. 105–142, Academic, N.Y.

Mitchell, R. C., and R. T. Carson (1989), *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Resources for the Future, Washington, D. C., 463 pp.

NERA-Accent (2007), The benefits of Water Framework Directive programmes of measures in England and Wales, *Rep. CRP 4b/c*, Dept. for Environment, Food and Rural Affairs, London, U.K.

Plott, C. R. (1996), Rational individual behavior in markets and social choice processes: The discovered preference hypothesis, in *Rational Foundations of Economic Behavior*, edited by K. Arrow, M. Perleman, and C. Schmidt, pp. 225–250, Macmillan, London, U.K.

Rose, J. M., M. C. A. Bliemer, D. A. Hensher, and A. T. Collins (2008), Designing efficient stated preference experiments in the presence of reference alternatives, *Trans. Res. B*, *42*(4), 395–406.

Schkade, D. A., and D. Kahneman (1998), Does living in California make people happy? A focusing illusion in judgments of life satisfaction, *Psychological Science*, *9*(5), 340–346.

Turnbull, B. W. (1976), The empirical distribution function with arbitrarily grouped, censored, and truncated data, *J. R. Stat. Soc.*, *38*, 290–295.

Venkatachalam, L. (2004), The contingent valuation method: A review, *Environ. Impact Assess. Rev.*, *24*(1), 89–124.

Vossler, C., and J. Kerkvliet (2003), A criterion validity test of the contingent valuation method: Comparing hypothetical and actual voting behavior for a public referendum, *J. Environ. Econ. Manag.*, *45*(3), 631–649.

Welsh, M. P., and G. L. Poe (1998), Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach, *J. Environ. Econ. Manage.*, *36*(2), 170–185.

K. Andrews, Department of Environment, Food and Rural Affairs (DEFRA), Nobel House, 17 Smith Sq., London, SW1P 3JR, UK.

G. Atkinson and P. J. Metcalfe, Department of Geography and Environment, London School of Economics, Houghton St., London WC2A 2AE, UK. (p.j.metcalfe@lse.ac.uk)

W. Baker, NERA Economic Consulting, 15 Stratford Pl., London W1C 1BE, UK.

I. J. Bateman, School of Environmental Science, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK.

S. Butler, NERA Economic Consulting, One Front St., Ste. 2600, San Francisco, CA 94111, USA.

R. T. Carson, Department of Economics, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 94720-3880, USA.

J. East and R. Sheldon, Accent Market Research, Chiswick Gate, 598-608 Chiswick High Road, London, W4 5RT, UK.

Y. Gueron, Department of Economics, University College London, Gower St., London, WC1E 6BT, UK.

K. Train, Department of Economics, University of California, Berkeley, 508-1 Evans Hall, 3880, Berkeley, CA 94720-3880, USA.