Collaborative Research Programme

On River Basin Management Planning Economics

Report on The Benefits of Water Framework Directive Programmes of Measures in England and Wales

Nera & Accent

6 November 2007

JEP

6 November 2007

The Benefits of Water Framework Directive Programmes of Measures in England and Wales

A Final Report to DEFRA re CRP Project 4b/c

NERA
Economic Consulting

Accent

MMC Marsh & McLennan Companies

NERA-Accent Project Team

Bill Baker	(NERA, Project Director)
Paul Metcalfe	(NERA, Project Manager)
Sarah Butler	(NERA)
Yves Gueron	(NERA)
Rob Sheldon	(Accent)
Jo East	(Accent)

NERA's Expert Advisors

Giles Atkinson	(London School of Economics)
Richard Carson	(University of California, San Diego)
Kenneth Train	(NERA and University of California, Berkeley)

Project Steering Group

John Joyce	(Independent Consultant)
Kevin Andrews	(DEFRA)
Camilla Lundbak	(DEFRA)
Rob Curry	(Environment Agency)
Iain McGuffog	(South West Water and UK Water Industry Research)

Client's Peer Reviewers

Ian Bateman	(University of East Anglia)
George Hutchinson	(Queen's University, Belfast)

NERA Economic Consulting
15 Stratford Place
London W1C 1BE
United Kingdom
Tel: +44 20 7659 8500
Fax: +44 20 7659 8501
www.nera.com

Acknowledgements

The authors would like to thank a number of people who have contributed in varied ways to the design and implementation of the study: Deryck Hall (Consumer Council for Water); Huw Brain (Welsh Assembly Government); Ian Dickie (Royal Society for the Protection of Birds); Yvette de Garis (Thames Water); Paul Logan, Dave Martin and Jonathan Fisher (Environment Agency); Michael Payne (Environmental Consultants); Steve Nixon, Francois Ford, Ruth Cullingford and Yvonne Walker (WRc); the staff at Oxford Designers and Illustrators; Michael Spackman, David Harrison and Kent Van Liere (NERA); and the Accent team of programmers, field controllers, supervisors and interviewers.

Contents

Appendices (In Separate Documents)

Appendix E. Show Cards 2 - CV, Then CE Order

Appendix F. Demographics Showcards

Appendix G. Potential Biases

Appendix H. Variable Definitions and Summary Statistics

Appendix I. Sample Characteristics

Appendix J. Inconsistencies

Appendix K. Validity Assessment

Appendix L. Verbatim Responses

Appendix M. Supplementary Analysis Tables

Appendix N. Comparator Benefit Valuation Studies

Appendix O. CRP Caveats Regarding Usage of this Report

List of Tables

Executive Summary

This report presents research undertaken by NERA Economic Consulting and Accent Market Research for the UK Collaborative Research Programme into the Water Framework Directive (WFD). We use survey methods to estimate, in monetary terms, the value placed by households in England and Wales on improvements to the water environment brought about by the WFD.

Approach and implementation

This research is designed to estimate the total value placed on water environment improvements by households only. The development of the approach and its implementation were peer reviewed step by step:

- work with scientists developed a method for representing WFD improvements through the changing proportions of water sites nationally and locally (30 mile radius) falling into defined categories of high quality, medium quality and low quality;
- a representation of the quality status categories, and of the proportional changes in status over time, that household survey respondents would understand, was developed through a substantial series of focus and deliberative group enquiries informed by a literature search and a stakeholder survey;
- a survey instrument was developed to ascertain household willingness to pay (wtp) for water environment improvements using three elicitation methods in combination for every respondent: a payment card question (PCCV), a dichotomous choice question (DCCV), and a choice experiment (CE), which were presented in various orders. It was expected that the PCCV and DCCV estimates would bracket the wtp range, and that the CE would provide better estimates of marginal values. The instrument was piloted twice and adjusted accordingly;
- the target sample of 1500 household respondents for interview was developed as a set of 50 hall locations randomly drawn from a list of districts, stratified by population size and River Basin District, across England and Wales, with quotas for the demographic characteristics of 30 target respondents in each hall location.

In July 2007 1487 interviews were undertaken in 50 locations:

- the full sample matches England and Wales demographic characteristics quite well, though more women were interviewed. Aggregated results are re-weighted for this;
- the interviews lasted an average of 32 minutes and the interviewers found good levels of understanding and attention were given to the questions;
- across the sample the wtp answers varied with income, environmental attitudes, and many other factors, very much as predicted. We formed a reduced sample for econometric valuation analysis conservatively, dropping 23 non-response cases, plus 58 “protesting” and 17 “outlier” respondents. We retained a substantial number of respondents who gave apparently different answers to the different wtp elicitation methods, but we also derive results for a reduced sample dropping these respondents as a sensitivity;

- the responses performed satisfactorily against a suite of tests of internal and external validity, leading us to conclude that the wtp measures are meaningful estimates of the value of water environment improvements as we have represented them, though there are some reasons for caution.

Survey-derived water-environment benefit estimates

The table below shows wtp estimates from the three different elicitation methods (PCCV, DCCV, CE), for the best models NERA was able to find through an intensive econometric exercise in the short time available for this analysis. To place the wtp values from the different formats on a comparable basis, they have all been presented as the permanent increase in real annual payments that the household is willing to pay to enjoy the scenario of 95% improvement to High Quality status nationally by 2015.

We consider that the 'true' mean wtp is very likely to lie in the range between the mean wtp from the PCCV model (£44.5 per household per year) / and the mean wtp from the DCCV model (167.9 £/hh/yr). The PCCV method is known theoretically and empirically as likely to lead to conservative estimates, possibly downwardly biased, while the DCCV method used in this research may yield upwardly biased estimates. We consider a conservative approach appropriate and this would imply use of estimates from the lower end of the range as the main benefit figures for policy-making purposes, as well as sensitivity analysis.

Table 1.1
WTP for Water Environment Improvement (95% by 2015) by Elicitation Method*

Elicitation Method / Model	England		Wales		England and Wales	
	Mean WTP £/hh/yr	Median WTP £/hh/yr	Mean WTP £/hh/yr	Median WTP £/hh/yr	Mean WTP £/hh/yr	Median WTP £/hh/yr
PCCV Sample Statistics	49.2	30.0	62.6	50.0	50.4	30.0
PCCV OLS Model	44.8	25.3	40.1	22.7	44.5	25.1
DCCV Turnbull Statistics*	101.6	100-200	129.9	>200	103.0	100-200
DCCV Logit Model	167.0	167.0	181.4	181.4	167.9	167.9
CE Logit Model	293.7	293.7	508.0	508.0	299.9	299.9

Source: NERA
** Note: DCCV Turnbull Statistics for Wales are calculated using choices made for both the 95% and 75% improvement scenarios. There were insufficient respondents in Wales that were shown the 95% scenario to calculate these statistics reliably using only these observations.*

The table shows that the estimated willingness to pay for water environment improvements is sensitive to the elicitation method used, in the way expected. The PCCV method, where the respondent must freely select a monetary sum from a wide range presented on a card, produces lower answers than the DCCV and CE methods where the respondent accepts or rejects choices including a given monetary sum.

The table also shows that the median household wtp is lower than the mean or average household wtp, another relationship familiar from most environmental wtp studies (the DCCV and CE models force the mean to equal the median). Consistent with this in our data:

- there is a large variation across households in wtp for water environment improvements, and this is evident in the responses to each of the three elicitation methods;

- the majority of households are willing to pay a notable sum for water environment improvements;
- about 8% of households have zero wtp, and a further 14% have wtp less than £10 for water environment improvements;
- some households are willing to pay many times the median amount for water environment improvements; and
- there is evidence that households likely to fall into groups where 'affordability' is a concern (based in income, household size, socio-economic group, etc) have substantially lower wtp for water environment improvements. However there is also a substantial degree of heterogeneity across households in wtp values that is not explained by any of the wide set of measures in our household characteristics data. Associated with this, numbers of low income households have high wtp, and numbers of high income households have zero wtp for water environment improvements.

Regarding different timings and types of water environment improvement:

- there is evidence in our data that for many people, when improvements by 2027 are expected, delivering improvements earlier adds little to their willingness to pay. However some households value earlier improvement very highly. As a result the mean willingness to pay for earlier improvements is substantial;
- percentage improvement applying to the whole nation appears to be valued only a little more than the same percentage improvement applying only in the local area (30 mile radius). This implies that 'per household per site' wtp values would be much higher for improvements at local sites than at distant ones. This finding is supported by responses to another question in our survey showing that the majority of respondents consider that making improvements near where people live should be a priority; and
- the general pattern is for increases in good quality to be valued somewhat more than decreases in low quality, though not at all models exhibit this, and the qualitative research found differently.

While estimates of the mean willingness to pay are generally regarded as the appropriate main foundation for national-level cost-benefit analysis, relying on a hypothetical-compensation test, we consider that the distribution of willingness to pay should in this case be fully considered in policy-making. As sated above, we consider that the true WFD household mean benefit value for the 95% WFD improvement scenario as we have represented it is very likely to lie between the mean estimate from the PCCV model and the mean estimate from the DCCV model. The PCCV estimate, under reasonable assumptions, is known to be biased downward as respondents with a low wtp have an incentive to give a zero value (in an attempt to avoid eventually being forced to pay for the program), and respondents with a high wtp have an incentive to give a wtp amount which is biased downward in the direction of what they think the expected cost of the program will be (in an attempt not to be forced to pay more than this if the program proceeds). Our survey data appear to be consistent with this behaviour, as there are substantially more PCCV respondents expressing zero wtp than are suggested by the DCCV responses, and there are also considerably fewer respondents giving high PCCV values than are suggested by the DCCV responses. A single DCCV question (offering the status quo or one alternative including a specified payment sum, Yes or No) is generally regarded as incentive compatible in the sense that truthful preference revelation is a dominant strategy if respondents believe they will have to pay the specified sum. Unfortunately, in order to value as many different aspects of the WFD as we did, it was impossible to field a large survey that asked only one DCCV valuation question. When respondents see multiple payment figures for a program, it is common to see them engage in a kind of averaging so that (relative to asking a single binary DCCV question) the proportion of respondents who are willing to pay the lower payment figures falls (because they think the actual amount they will be forced to pay will be some type of weighted average between the amount stated and the earlier and generally higher amounts seen), while the proportion of respondents willing to pay the higher payment amounts increases (because they think the amount they will actually be forced to pay will be some type of weighted average between the amount stated and the other amounts seen earlier which are generally lower). The net effect of this sort of behaviour in a repeated discrete choice model is usually to increase the estimated mean wtp above its true value. There is an active academic debate over the nature of these response effects.

The marginal values for different types of environmental improvements are estimated much more precisely in the CE models, due to the substantial additional information available on preferences obtained by this method. Because of this, we consider that where relative marginal estimates of benefit values are required these are best obtained by scaling the CE relative figures to match the total PCCV benefit value estimate.

Estimates of WFD scenario benefit values

DEFRA provided the following scenarios of WFD implementation by key dates.

Table 1.2
Environmental improvement Scenarios

Scenario Name	Description
Maximum benefits	full improvement (100%) to High Quality achieved by 2015
Front loaded	50% of improvements by 2015, followed by 30% in 2021, and 20% in 2027
Even loaded	33% of improvements achieved by each of 2015, 2021, 2027
Back loaded	20% start in 2015 followed by a further 30% in 2021 and 50% in 2027
Less stringent objectives	25% by each of 2015, 2021, 2027, then no more (i.e. assumes less stringent ultimate objectives, amounting to the last 25% of improvement)
Nature assimilation lag	constraints from natural conditions, such as stocks of pollutants in sediment, mean that 50% of the improvement will not occur until 50+ years

Source: DEFRA

We translate each of these scenarios into year by year proportions of quality status over water environment sites. Separately, for each wtp model and each case considered (e.g. for England, or Wales, or both; for income classes separately or together; re-weighting the gender proportions), we derive an associated annual monetary value applying to the quality proportions of water environment sites (£ per percentage point, per household, per year). This is applied to the scenario's quality proportions to form a year by year series of monetary values. This is discounted at the UK Treasury Green Book social discount rate (3.5% per year) and summed to form the present value of the scenario benefits, per household. Multiplying by the number of households gives the total present value of the scenario benefits. This procedure makes allowance for differences between countries in the explanatory factors applied in the models (e.g. income) and in the numbers of households.

The first table below presents aggregate values for WFD benefits in England, Wales, and in England and Wales by scenario, based on the mean PCCV wtp values. We consider these values conservative. Across the range of scenarios, the total present value of WFD benefits based on the mean PCCV figures ranges from £18bn to £29bn, for England and Wales.

As sensitivities, the second table presents aggregate values for the WFD scenario benefits calculated from the median PCCV wtp, and the third shows aggregate benefit values based on the DCCV wtp estimates.

Table 1.3
Aggregate PCCV Estimates of WFD Benefits by Scenario

Policy Scenario	England			Wales			England and Wales		
	Mean WTP	Annual WTP (for 20.5m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 1.2m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 21.7m hh)	PV WTP (@3.5% discount rate)
	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million
Scenario 1 - Maximum	47.4	968.63	27,675.15	42.9	51.89	1,482.43	47.1	1,020.15	29,147.10
Scenario 2 - Front Loaded	40.9	836.10	23,888.55	37.0	44.79	1,279.60	40.7	880.57	25,159.10
Scenario 3 - Even Loaded	38.2	782.25	22,349.93	34.7	41.90	1,197.19	38.0	823.85	23,538.65
Scenario 4 - Back Loaded	35.7	731.06	20,887.37	32.4	39.16	1,118.84	35.5	769.94	21,998.30
Scenario 5 - Less Stringent Objectives	28.7	586.85	16,767.20	26.0	31.44	898.14	28.5	618.06	17,658.99
Scenario 6 - Nature Assimilation Lag	31.3	639.68	18,276.45	28.3	34.26	978.99	31.1	673.70	19,248.51

Table 1.4
Sensitivity 1: Median PCCV Estimates of WFD Benefits by Scenario

Policy Scenario	England			Wales			England and Wales		
	Mean WTP	Annual WTP (for 20.5m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 1.2m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 21.7m hh)	PV WTP (@3.5% discount rate)
	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million
Scenario 1 - Maximum	26.8	547.47	15,641.87	24.3	29.33	837.87	26.6	576.58	16,473.81
Scenario 2 - Front Loaded	23.1	472.56	13,501.70	20.9	25.31	723.23	23.0	497.69	14,219.81
Scenario 3 - Even Loaded	21.6	442.12	12,632.08	19.6	23.68	676.64	21.5	465.64	13,303.94
Scenario 4 - Back Loaded	20.2	413.19	11,805.45	18.3	22.13	632.37	20.1	435.17	12,433.34
Scenario 5 - Less Stringent Objectives	16.2	331.69	9,476.75	14.7	17.77	507.63	16.1	349.33	9,980.78
Scenario 6 - Nature Assimilation Lag	18.9	385.97	11,027.62	17.1	20.67	590.70	18.8	406.49	11,614.14

Table 1.5
Sensitivity 2: Aggregate DCCV Estimates of WFD Benefits by Scenario

Policy Scenario	England			Wales			England and Wales		
	Mean WTP	Annual WTP (for 20.5m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 1.2m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 21.7m hh)	PV WTP (@3.5% discount rate)
	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million
Scenario 1 - Maximum	176.7	3,614.42	106,883.67	193.9	234.46	6,933.26	177.7	3,849.03	113,821.44
Scenario 2 - Front Loaded	152.6	3,119.89	92,259.50	167.4	202.38	5,984.63	153.4	3,322.40	98,248.03
Scenario 3 - Even Loaded	142.7	2,918.94	86,317.23	156.6	189.34	5,599.17	143.5	3,108.41	91,920.05
Scenario 4 - Back Loaded	133.4	2,727.93	80,668.70	146.4	176.95	5,232.76	134.1	2,905.00	85,904.87
Scenario 5 - Less Stringent Objectives	107.1	2,189.83	64,756.28	117.5	142.05	4,200.57	107.7	2,331.97	68,959.58
Scenario 6 - Nature Assimilation Lag	116.7	2,386.94	70,585.13	128.1	154.83	4,578.67	117.4	2,541.87	75,166.78

1 Introduction

This report for DEFRA is the final report from CRP Project 4b and Project 4c. The report outlines the development and implementation of a stated preference survey instrument to value the benefits to households of Water Framework Directive (WFD) Programmes of Measures in England and Wales, and presents the results for WFD scenarios.

The remainder of this section is as follows:

- Section 1.1 provides background to the project.
- Section 1.2 sets out the objectives for the stated preference research.
- Section 1.3 outlines the stages of the project.
- Finally, Section 1.4 outlines the remainder of this report.

1.1 Background

The WFD is a far-reaching piece of European Community (EC) legislation. It covers lakes, rivers, transitional and coastal waters, artificial and heavily modified water bodies, and groundwater. Its first requirement is for member states to protect all water bodies from ecological deterioration. It then requires member states to specify Programmes of Measures to improve the ecological status levels of its water bodies over time-limited periods. Natural water bodies will usually be required to meet the standard of "good ecological status", commonly defined according to physical, chemical and biological criteria. For artificial water bodies, such as canals and reservoirs, and in cases where water bodies are designated as "Heavily Modified", the less stringent objective of good ecological potential will be applied. This objective allows for the retention of human modifications, e.g. concrete channelling, dams, etc, but still requires the chemical quality of the water to satisfy similar criteria to natural water bodies. Measures to improve the status levels of water bodies will potentially be broad ranging, from command and control-type standard setting, to economic instruments and information schemes.

Estimates of the benefits of WFD Programmes of Measures are needed at the present time for two reasons. First, the WFD itself allows for derogations from the general requirement of member states to reach good ecological status in all water bodies by 2015 in cases where the costs of doing so can be shown to be disproportionate. In such cases, the WFD allows for a longer time frame to achieve good ecological status or for a less stringent environmental objective to be met. In England and Wales, the assessment of disproportionate costs has been interpreted as requiring a comparison of the costs and the benefits of alternative Programmes of Measures. Second, benefits estimates are needed for a national Impact Assessment (IA) for the WFD, and for individual River Basin District RIAs.

The present study has been jointly funded and steered by WFD stakeholders in the UK as part of the UK Collaborative Research Programme (CRP). The CRP was set up early in the life of the WFD to develop the methodologies needed to undertake the WFD economic analysis.

The CRP involves fourteen parties and is chaired by DEFRA[1]. Following a strategic steer by the UK Economics Steering Group, CRP members decided to take a long term approach to benefits analysis for the WFD. This approach comprises an initial expert workshop to refine the strategic approach (4a) followed by a series of interrelated projects, 4b-4h, which together are intended to provide the high level estimates of the total benefits of the WFD needed to inform formation of the first River Basin Management Plan (RBMP).

The present stated preference study, comprising the development Project 4b and its direct implementation follow-on Project 4c, is to estimate the non-market benefits of WFD programmes of measures. Project 4d will investigate the public's preferences between alternative types of measures. Project 4e is to provide complementary estimates of the direct market benefits. Project 4f is to provide additional evidence on use and access to water bodies. Project 4g, covering the interpretation of the science of potential WFD improvements in terms useful for economic studies, has linked closely with the present study by informing descriptions and materials representing the good we value. The final project within this sub-programme of the CRP is Project 4h, which is to focus exclusively on the non-use benefits arising from the WFD. For future planning periods, additional valuation work is expected to be commissioned to provide further guidance on WFD benefits.

1.2 Objectives

The objectives of this research are to use stated preference (SP) methods:

- primarily to establish a figure for the potential non-market benefit of the WFD impacts as a whole nationally to use in the Regulatory Impact Assessments on implementation which are to be filed later this year; this must be broken down between the nations of England and Wales; and
- secondarily, and to the extent possible, to establish figures for the potential benefit of the WFD impacts broken down by RBD (or finer locations), by water body type, by quality improvement type, by timing, by degree of certainty of effects, or by other helpful dis-aggregations. These figures on "relative priorities" are to be useful in the first pass studies later in 2007 which will select initial WFD measures for adoption.

The research needs as far as possible to characterise WFD benefit functions whose arguments reflect the science and the policy choices. It also needs to provide a reasonable, coherent and even-handed context for survey respondents in making decisions, and must do this in a non-threatening, intelligible way, encouraging the respondent to engage with the information and make reliable preference-reflecting choices.[2]

---

1 Parties to the CRP are: Department of Environment, Food and Rural Affairs (DEFRA), Scottish Executive, the Environment Agency, Scottish Environment Protection Agency (SEPA), Natural England, Department of Trade and Industry (DTI) now the Department for Business, Enterprise and Regulatory Reform (BERR), UK Water Industry Research (UKWIR), Royal Society for the Protection of Birds (RSPB), Welsh Assembly Government (WAG), Department of Environment Northern Ireland (DOENI), British Ports/UK Major Ports Group (UKMPG), Countryside Landowners and Business Association (CLBA), National Farmers Union (NFU), and Joint Environment Programme (JEP). Involved groups include the Consumer Council for Water (CCW).

2 See M. Kaplowtiz, F. Lupi and J. P. Hoehn, "Multiple Methods for Developing and Evaluating a Stated-Choice Questionnaire to Value Wetlands," in *Methods for Testing and Evaluating Survey Questionnaires,* (2004) S. Presser et al.

Practically, the SP design answers the following questions:

- What is to be valued?
- Which value elicitation method is to be used?
- What visual materials?
- What contextual information?
- Which attributes to describe baseline and improvements?
- What payment vehicle?
- What initial and follow-on questions?
- What sampling, survey mode and fieldwork method?
- What analysis and aggregation method?

Our approach results from staged development and testing with continued peer review, as described in the next subsection.

1.3 Development Stages

The SP design and the accompanying stated preference survey questionnaire were developed in three distinct stages.

1.3.1 Stage I - Preliminary Research

The first stage comprised three preliminary research components: a literature survey, a series of deliberative research groups, and a series of stakeholder interviews. This created an evidence base for understanding how change in the water environment should be represented for the purposes of valuation. In detail, the elements comprised:

- a review and synthesis of the relevant information from previous literature relevant to valuation of national improvements to the water environment, specifically considering:
  - the previous science work on the dimensions of the WFD and the risks it would address;
  - how context is derived and presented;
  - how valuing overall programmes rather than specific schemes is dealt with;
  - if and how use values and non-use values are separately derived;
  - how studies deal with the issue of risk to the delivery of benefits;
  - how studies deal with the issue of latency in the delivery of benefits; and
  - what approaches studies have taken to deal with common validity issues, including: protest responses; different potential biases.
- new deliberative qualitative research of members of the public, investigating:
  - the language and concepts used by members of the public to understand and value the water environment;

  - how people process various visual and textual materials, and how people cope with exercises of varying types and degrees of difficulty;
  - baseline perceptions of current status and drivers of change in the water environment;
  - priorities in relation to types of value (e.g. use+non-use), types of environmental change, types of site, distance / locations of sites; and
  - the contexts in which value is derived and on which it is dependent, e.g. timing of environmental change, substitutes and complements, and attitudes to uncertainties and responsibility.
- new interviews with stakeholders with expert knowledge or experience of the water environment, exploring:
  - organisational and individual roles in relation to the water environment and WFD implementation;
  - priorities in relation to WFD aims and implementation options, including attitudes to timing of environmental change, uncertainty, and responsibility; and
  - the factors regarded as critical to valuation, drawing on experiences with previous valuation or prioritisation work.

Outputs included reports on the three components, all of which were reviewed by NERA's expert panel and the client's peer reviewers. In addition, a memorandum outlining the rationale for Stage I, and the topic guides for the deliberative research groups and the stakeholder interviews, were all also reviewed by NERA's expert panel and the client's peer reviewers.

1.3.2 Stage II - Recommendations

Drawing on the evidence base from Stage I, the second stage developed recommendations for the design and testing of a stated preference survey instrument, including a draft questionnaire and show materials, and an associating sampling and aggregation approach. At this stage, the project team worked closely with WRc and the Environment Agency to create and refine the description and representation of the good so that it is both a fair description of the WFD effects in a scientific sense, and understandable for members of the public. Comments and feedback were also received from a number of stakeholders during this period, including Natural England, Thames Water, Three Valleys Water, and Consumer Council for Water.

The principal output from this stage from NERA was a peer reviewed report entitled "Recommendations for the Design and Testing of a Stated Preference Instrument for CRP Project 4c". Appendices to this report included a draft main survey stated preference questionnaire, a topic guide for focus group testing, and a questionnaire designed for cognitive interview testing. In addition, WRc produced a separate self-standing report (WRc, 2007) documenting the development of the descriptions and visual materials provided used for the survey.

1.3.3 Stage III - Testing and Refinement

The third stage of development tested and refined successive versions of the SP survey instrument. The main elements of testing included 10 focus groups located in five locations around England and Wales, 30 cognitive interviews, two pilot tests on 50 and 98 households respectively, and six further focus group discussions. Over the course of this stage, aspects of the survey instrument were successively tested and refined, including the visual and textual representation of the baseline and WFD improvements, the design and presentation of the value elicitation mechanisms, and the number and style of the supporting questions and materials. The draft survey instrument was reduced in length at each stage, cutting areas which were found to be less useful, whilst retaining sufficient material to ensure a valid and reliable survey instrument. As before peer reviewers commented at each successive step.

1.3.4 Stage IV – Implementation and Analysis

The SP survey was undertaken in the last week of July and first week of August 2007 and the results were immediately analysed. A preliminary and a final report on the results were peer reviewed and drawn on for Ministerial Guidance on the WFD. This over-arching project report concludes the research. A separate tool to allow valuation of WFD scenarios is also to be delivered.

1.4 Structure of Report

Section 2 below reviews relevant literature.

Section 3 and 4 outline the SP Design and wider elements of the survey instrument.

Section 5 discusses the sample, Section 6 reports on implementation of the survey, and Section 7 outlines our derivation of samples for analysis.

Sections 8, 9 and 10 present general responses and then our approach to valuation analysis, and finally our wtp results.

Section 11 presents WFD benefit results aggregated for England and Wales and for WFD scenarios.

Section 12 concludes.

2 Literature Review

2.1.1 Scope and Methodology

The literature review focussed on national, or wide-area, level empirical valuation studies of generic types of changes to the environment, or of measures or policies to bring about such changes. It covered air and land-use impacts in addition to water-related environmental changes and covered all countries. It did not cover individual site-specific valuation studies in detail, nor did it cover the theoretical literature.

The studies were reviewed with a focus on the following range of issues pertinent to the design of the stated preference instrument:

- how context is derived and presented
- the manner in which valuing overall programmes rather than specific schemes is dealt with
- if, and how use values and non-use values are separately derived
- how studies deal with the issue of risk to the delivery of benefits
- how studies deal with the issue of latency in the delivery of benefits
- what approaches studies have taken to deal with common validity issues, including: protest responses; hypothetical bias; information bias; scope, embedding and part-whole effects; and anchoring bias.

2.1.2 Studies Identified

Our survey of the literature found useful predecessor studies; in particular, in the US EPA National Freshwater Benefits study (Mitchell and Carson, 1981,1984,1986; also Carson and Mitchell, 1993; and US EPA, 1994); and the US EPA Valuing Inland Water Quality study (Huber and Viscusi, 2006). A number of other studies reviewed contained useful examples of how to deal with particular issues, including developing and presenting the contextual information necessary for valuing a complex policy.

In addition to the SP valuation studies, a number of supplementary studies were identified that could provide supporting quantitative or qualitative evidence to support and test the findings that emerge from the Project 4c SP survey. These studies included the first WFD Regulatory Impact Assessment (WRc et al, 1999), the Environment Agency's PR04 Benefits Assessment (Fisher et al, 2002), US EPA Iowa Lakes Valuation (Azevedo et al, 2003), and two opinion and attitude surveys: Scottish Executive Survey (Ipsos-MORI, 2006) and the PR04 joint water customer research study (MVA, 2003).

Table 2.1 below lists the names and references for all relevant studies identified in the literature review by study type.

Table 2.1
Relevant Studies Identified in Literature Review

Type of Study	Study Name (Reference)
Multi-Site Water SP	Defra Bathing Water Directive Revisions (Eftec, 2002)
	British Waterways Canals (Adamowicz et al, 1995)
	US EPA National Freshwater Benefits (Mitchell and Carson, 1981, 1984, 1986)
	US EPA Groundwater Cleanup (McClelland et al, 1992)
	US EPA Valuing Inland Water Quality (Huber and Viscusi, 2006)
	YW LEADA (Willis et al, 2005)
Supplementary Multi-Site Water Non-SP	WFD RIA (WRc, 1999)
	PR04 Benefits Assessment (EA, 2002)
	US EPA Iowa Lakes Valuation (Azevedo et al, 2003)
Multi-Site Non-Water SP	UK Forestry Commission Valuing Forest Recreation (Scarpa, 2003)
	Defra Health Benefits From Reductions in Air Pollution (Chilton et al, 2004)
	Defra Valuing Changes in Biodiversity (Christie et al, 2004)
	Ireland Rural Environment Protection Scheme (Campbell et al, 2006a, b, c and d)
	Finland Natura 2000 Conservation Program (Li et al, 2004)
Opinion and Attitude Surveys	Scottish Executive Survey (Ipsos-MORI, 2006)
	PR04 Joint Water Customer Research (MVA, 2003)

2.1.3 Deriving and Presenting the Context

The valuation scenario in a survey instrument should be explained to respondents as accurately and fully as possible without over-burdening the respondent with complex and unnecessary details (Mitchell and Carson, 1989). A particularly methodical approach to this challenge was displayed in the US EPA groundwater cleanup study by McClelland et al (1992). This study began with a “full-information” instrument, and reduced this down to a concise version, while retaining all the necessary information, using successive instruments with cognitive debriefing at each stage. A key lesson from the wider literature is that the survey instrument should be tested and re-tested on small samples to draw out any potential problems and should be well defined through use of photographs, maps, and detailed descriptions of impacts, etc.

2.1.4 Spatial Sampling, Referencing and Aggregating

Two main alternatives were distinguished in the literature for valuing improvements to multiple sites. These were a "top-down" approach and a "bottom-up" approach. In the top-down approach, the valuation scenario depicts the whole policy; in the present case, this would be all the benefits brought about under the WFD. The alternative, bottom-up, approach characterises the benefits to an individual site, or collection of sites, and develops a benefits transfer function to value the benefits of all other site improvements.

It is well known in the broad willingness-to-pay literature that a simple bottom-up summation of the benefits arising from site-specific improvements brought about by a policy is likely to give a seriously misleading estimate of the benefits of the whole policy (Hoehn and Randall, 1989). This is due to substitution and/or complementarity effects between sites. Thus, if a bottom-up valuation method is adopted, then one needs to carefully take into account the availability of substitutes and complements, which mean that the value of an improvement in one place depends on which other sites, in which condition, are available. Without this the sum will likely overstate the tone value. This ideally requires a consideration of the effects of distance on cross elasticities of demand between sites. Besides this challenge there are few relevant site-specific values.

In contrast to this, a top-down approach is relatively simple to perform. However, even with the top-down study design, one still needs to take care over questionnaire design to ensure that survey respondents correctly comprehend the policy under valuation, and do not, for example, only consider a single site, or at the other extreme, consider a broader environmental clean-up programme than is actually on offer. Thus, the survey instrument needs to be carefully designed to minimize the possibility of scope sensitivity bias, and / or to test for the presence of this bias.

The problems involved in the bottom-up approach appear to us to be practically insurmountable for the WFD stated preference study design. On the basis of the literature reviewed, a sensible way forward for the stated preference study design is to adopt, a top-down approach to valuing national and/or regional water quality improvements directly.

2.1.5 Use / Non-Use

In our literature review, only one study, the US EPA Groundwater Cleanup study, seriously attempted the separate estimation of non-use values. The Environmental Economics Advisory Committee reviewed this study, and noted that the need for separate estimate of non-use value for groundwater cleanup, imposed by the EPA, was an onerous requirement. A lessons learnt from this study is that attempting to separate use values from non-use values within a single study is liable to be extremely difficult, due to the fact that use and non-use values may be jointly produced and not straightforward to decompose in a valid way. The attempt to derive a non-use only value may in some valuation contexts even be fundamentally misconceived.

2.1.6 Risk / Uncertainty

In the vast majority of the reviewed studies, either no mention was made of any risks associated with non-delivery of benefits, or in one case, respondents were specifically asked to treat the delivery of benefits as being certain. The one study that did take account of risk in some way, was the Defra Health Benefits From Reductions in Air Pollution project. This study informed respondents about the impact of air pollution on health in a series of showcards. From these it could be inferred that this was not an exact science. Some uncertainty as to extra life expectancy in good and bad health. Sub-samples valued either 1, 3 or 6 months of extra life expectancy and wtp was derived as a function of extra life expectancy.

Notwithstanding the Defra Health Benefits study, the fact that few studies have modelled the attitudes towards the uncertainty of benefits being achieved as described is at least partly testament to the profound difficulties commonly found by experimental and survey researchers when studying people's approaches to decision-making under conditions of uncertainty.

2.1.7 Latency

In all the studies examined in this review, only two took measures to control for delays to the timing of benefits in their survey instruments. These are the US EPA Groundwater Cleanup study, and the US EPA Valuing Inland Water Quality study.

The approaches presented in these two studies were similar and provide useful templates. Furthermore, both studies showed highly consistent evidence of declining implicit discount rates in the valuation of benefits that occur in the future.

2.1.8 Overall Findings

Our survey of the literature found useful template instruments; in particular, the US EPA National Freshwater Benefits study (Carson and Mitchell, 1993); and the US EPA Valuing Inland Water Quality study (Huber and Viscusi, 2006). A number of other studies reviewed contained useful examples of how to deal with particular issues, including developing and presenting the contextual information necessary for valuing a complex policy.

We distinguished two broad approaches in the literature for representing the spatial location of environmental changes for a wide-area policy. These are: (1) a "top-down" approach, and (2) a "bottom-up" approach. In the first methodology, changes are described for a wide area; in the latter, they are described for individual sites and then values are aggregated to the whole using a benefits transfer function. We conclude from our literature review that the top-down approach is to be preferred for the present study. This is because it avoids difficult issues relating to the presence of substitutes and complementary water and non-water sites.

Our review of the literature found no straightforward way to estimate separate values for use and non-use values for a wide area policy. A robust dichotomy of values is impracticable.

It is equally not straightforward to estimate values that vary by uncertainty or latency in the delivery of WFD outcomes in a sophisticated way. At best, simple descriptors of these attributes may be adopted, but achieving a refined representation of preferences over these attributes is likely to be impracticable.

In addition to the studies covered by this review, our recommendations for the design of the stated preference instrument for Project 4c also draw on best practice guidance and surveys of the wider stated preference literature as set out in, eg Mitchell and Carson (1989), Arrow et al

(1993), Bateman et al (2002), Eftec (2006), Jacobs (2006) and Pearce, Atkinson and Mourato (2006).

3 Design of the Valuation Framework

Stated preference valuation of national water environmental improvements requires: deciding what is to be valued; selecting an appropriate value elicitation technique; representing the good to be valued in a simple, accurate, meaningful and non-leading manner; developing questions to elicit valid and reliable values; and creating an experimental design with adequate statistical precision. This section describes our approach to these issues. It is structured as follows:

- Section 3.1 outlines which WFD benefits are valued, and which are not.
- Section 3.2 addresses the choice of stated preference technique.
- Section 3.3 discusses the baseline scenario specification;
- Section 3.4 explains how the baseline and WFD improvements have been represented;
- Section 3.5 covers the choice of payment vehicle;
- Section 3.6 presents the experimental design, including the attributes and levels, and the value elicitation formats.

3.1 Types of Benefits Valued

The objectives of the stated preference survey were restricted at an early stage to valuation of the non-market benefits of WFD improvements. While it would be possible to use the SP survey to value the full range of benefits, including direct and indirect market benefits, SP techniques are mostly appropriate for measuring non-market benefits (Bateman et al, 2002). Our understanding is that where benefits from WFD projects and programmes can be accounted for using market prices, then these will be valued as part of CRP Project 4e (Direct market benefits).

Household recreational benefits are a boundary case, where many benefits will be non-market but others will be derived from activities with a market price. The survey includes WFD-induced changes in household values placed on the latter activities. Introductory material states that access to the water environment will not change. In adding benefits estimates from the survey to market-based benefit estimates care should be taken not to double-count changes in market recreational benefits.

Such double counting would occur if prices for recreational activities changed relative to peoples' expectations at the time of the survey and the market benefit of the improvement was calculated based on the new higher price.

It was also decided that the SP study should not value improvements to groundwater. The human benefits of groundwater improvements derive from their effects on its status as a resource for abstraction (or preservation as a storage arrangement to assist with withdrawal in future), and from their indirect effects on surface water status. In the first case, in England and Wales, there are very few households that abstract groundwater. The vast majority buy water from the water companies and so will benefit only via a potential reduction in water bills, which is a market benefit. The non-market benefits are therefore likely to be very small. In the latter case, the surface water quality benefits of groundwater improvements should be calculated by tracing the impact of groundwater improvements through to their impact on surface water ecological quality status. These indirect improvements to surface water quality may then be valued using results from the SP survey. For these reasons, the SP survey instrument excludes any direct reference to groundwater.

Figure 3.1 shows a typology of total economic benefits and points to those that are considered in this study.

Figure 3.1
A Typology of Potential Water Framework Directive Benefits

Potential Water Quality Benefits	Current User Benefits	Direct Use	In Stream	Recreational* - fishing, swimming, boating, rafting etc.
				Commercial - fishing, navigation
			Withdraw	Municipal - drinking water, waste disposal
				Agriculture - irrigation
				Industrial / Commercial - cooling, process treatment, waste disposal, steam generation
		Indirect Use	Near Stream	Recreational *- hiking, picnicking, bird watching, photography, etc
				Relaxation *- viewing
				Aesthetic *- enhancement of adjoining site amenities
	Intrinsic Benefits	Potential Use	Option	Near-term potential use*
				Long-term potential use*
		No Use	Existence	Stewardship - maintaining a good environment for everyone to enjoy (including future family use-bequest)*
				Vicarious consumption - enjoyment from the* knowledge that others are using the resource.

* Considerad in this project

Source: Desvouges, Smith and McGivney (1983)

The benefits of WFD improvements covered by the stated preference study include improved conditions for recreation in or around the water, improved aesthetic appearance of the water environment, and non-use benefits deriving from the provision of improved ecosystems.

Benefits arising that relate to the use of water bodies as a resource for abstraction have been excluded from the present study. This is because benefits that arise in the form of, for example, lower water treatment costs for drinking water, due to cleaner water abstracted, can be valued directly from the cost functions of water companies and other abstractors and CRP Project 4e will address this. Likewise, increases in production costs that arise due to the curtailment of abstraction licences as a result of the WFD will best be valued directly from the cost functions of water companies and other abstractors.

3.2 Stated Preference Valuation Technique

The two principal techniques for stated preference valuation that are consistent with welfare economics are Contingent Valuation (CV) and Choice Experiments (CE). Both techniques are well established in the academic literature and in policy applications, and have been successfully employed to value national water quality improvements (Mitchell and Carson, 1984, and Huber and Viscusi, 2006) as well as hundreds of other non-market benefits. Both approaches have as their objective the estimation of maximum willingness-to-pay for improved policy alternatives relative to some baseline. The maximum that people are willing to pay for a good is an economic measure of its value. We used both approaches within the SP survey for Project 4c.

The CV technique is focused on valuing one scenario so is suited primarily to situations where estimates of total benefits of an environmental programme are needed. By contrast, CE questions value marginal changes, as well as valuing whole environmental programmes, and so are useful for valuing elements of policies and programmes. Adopting both techniques provides the ability to cross-validate the total value estimates, though differences are expected.

The key elements of CE and CV questionnaires are similar. Both require the specification and presentation of baseline and improvement scenarios, and the selection of a payment vehicle, e.g. taxes or water bills. Both require the selection of an elicitation format and an experimental design.

3.3 Baseline Specification

A number of existing EU Directives require measures to be taken in the period leading to 2015 which will affect the ecological status of the water environment across England and Wales even without the WFD. For the WFD IA, costs and benefits are to be measured against a baseline incorporating the implications of these existing directives. Despite the baseline for the IA being an improvement relative to the current state of the environment, following discussion with the Steering Group it was decided that the baseline for the valuation framework should be the assumption of no deterioration and no improvement at any site. This approach was necessary to simplify the SP instrument, to simplify presentation to respondents and in view of the current uncertainty over the actual IA baseline. This does mean that care is needed in interpreting the benefit values: some will be attributable to existing planned measures. Also the likely increase in payments for existing planned measures needs to be presented to allow WFD benefit sensitivity to this to be explored.

3.4 Representation of the Baseline and WFD Improvements

One of the key design challenges of the study was to characterise the good – to select a combination of attributes and a form of presentation to characterise the differences between policy alternatives in a way that was: closely attuned to the policy-levers available; understandable, meaningful, and helpful for respondents; scientifically defensible; and unbiased by use of language, symbols or omission. These design choices were made jointly and over the three stages of development as described in above with substantial input from peer reviewers and Steering Group. It was desired to value a large number of possible variations on the extent, nature and timing of improvements, and this influenced selection of elicitation methods.

At an early stage in the study design process, the policy alternatives to be valued were conceptualised, quite generally, by their effect on the risks that different water bodies would fail to achieve good ecological status/potential, in different qualitative ways and at different times (eg by the key WFD dates of 2015, 2021 or 2027). We selected attributes and visual materials in light of the relative importance of the spatial, quality, time and risk dimensions, and the interplay between them.

3.4.1 Spatial Referencing: Use of Maps

Results from the deliberative group research confirmed our expectation that the location of water bodies is a key determinant of how they are valued. In addition, individuals responded very well to the use of maps in the deliberative groups. The maps provided individuals with a simple way to review the quality of water bodies in their region, and they were able to match the information in the maps to their own knowledge and perceptions of the areas shown. Moreover, participants found it helpful to see the number of water bodies is in various quality states.

From a design perspective, the use of maps is desirable because this allows a large amount of information to be conveyed concisely. Since the maps were well received and understood by respondents, the recommended SP survey instrument was built around the use of maps to represent baseline water environment status. As part of the science support project (4g), WRc developed successive versions of maps according to jointly developed specifications during Stages II and III. A summary of the final design of the maps is given below.

3.4.2 Use of Three-Level Status Classification

A strong finding from the deliberative group research was that participants generally display only a very coarse appreciation of the ecological status of water bodies. Virtually everyone in all the locations felt unable to judge the water status except at some superficial level by how it looks and smells, and unable to value fine differences in status or the reasons why status had changed. On the basis of this evidence, it was decided at an early stage that a very simple metric of ecological status was needed, which captured the most policy-relevant improvements brought about by the WFD, namely movements between WFD-defined status levels, but which did not attempt more detailed qualitative separation of improvement categories. The consequence of this decision was that to changes in WFD status level would be valued, but the design would not provide for different values on the basis of types of pressure to be addressed (flow, alien species, etc.)

3.4.3 Matching the Five WFD Levels to the Three SP Survey Status Levels

The WFD contains five normative levels for the classification of surface water bodies, namely: High, Good, Moderate, Poor and Bad ecological status. For Artificial Water Bodies (AWB) and Heavily Modified Water Bodies (HMWB), the relevant categories are Maximum, Good, Moderate, Poor and Bad ecological potential. These levels were mapped into three as shown in Table 3.1.

Table 3.1
Status Levels in the WFD and the SP Survey Instrument

Status Level for SP survey	WFD Status Level
1 High Quality	High or Good Ecological Status (natural water bodies); Maximum or Good Ecological Potential (artificial or heavily modified water bodies)
2 Medium Quality	Moderate or Poor Ecological Status (natural water bodies); Moderate or Poor Ecological Potential (artificial or heavily modified water bodies)
3 Low Quality	Bad Ecological Status (natural water bodies); Bad Ecological Potential (artificial or heavily modified water bodies)

Source: NERA

This mapping allows benefit values to be estimated for improvements from below good status to good status or above, a key policy-relevant distinction for disproportionate cost assessment. It also allows separate values to be estimated for improving the worst areas, which links to a finding from the deliberative group research that participants favour improving the worst areas first (we note that the main survey did not find values matching this qualitative finding).

3.4.4 Wording of Status Descriptions

In Stage I of the survey development, views were sought from stakeholders on the expected outcomes of the WFD, and on the expected benefits that would result. The deliberative group research also sought views from respondents on priorities and sources of value from the water environment.

During Stage II, as part of the science support project (4g), WRc drafted 200 to 500-word layman's descriptions of each of the three status levels for each of lakes, urban rivers, rural rivers, and transitional and coastal waters. These full descriptions are contained in WRc's interim report. The WRc descriptions remained too long for the survey questionnaire and so were successively reduced in length to fit onto an example showcard for each water body type, and a further showcard containing generic descriptions covering all water body types.

The showcards were tested and adjusted in the Stage III focus groups and cognitive interviews. In addition, in Stage III, the full descriptions were broken down into 18 attributes, which focus group participants and cognitive interview respondents were asked to rank in order of importance to them. The results of the attribute rankings, combined with the findings from the prior deliberative group research, and the reactions to the cut-down versions included in the example showcards, were drawn on to revise the wording prior to the pilot survey.

Comments from the Programme Steering Group, the Environment Agency and WRc were sought and taken account of prior to the focus groups and cognitive interviews, and again prior to the pilot survey to refine the wording of the descriptions. The final versions of the descriptions included in the SP survey questionnaire were reached following extensive iteration and refinement following successive comments from sources including English Nature, Thames Water, Three Valleys Water, DEFRA and the Environment Agency, and proved to be meaningful to members the public in the pilot survey.

3.4.5 Use of Examples and Illustrations

The SP instrument has four variants, each of which includes one generic show card showing status level descriptors plus one show card describing the three status levels for one water body type (see Appendix D). Card 3a shows a rural river, Card 3b shows an urban river, Card 3c shows a lake, and Card 3d shows an estuary / coastal area. The descriptions are accompanied by illustrations produced by Oxford Designers & Illustrators. One of the four versions is selected at random for each respondent.

The use of the illustration was proposed as a way of helping respondents construct a value for an unfamiliar and complex good. An alternative that was considered at an early stage, and rejected, was to use photographs of water bodies of different types at different status levels. While photographs proved useful in the Stage I deliberative research groups for generating discussion, they did not appear to aid participants in establishing priority frameworks and many expressed a desire to know where the photographs were taken. Moreover, certain important aspects relating to ecological status are not visible in photographs, while other things that aren't related to the WFD appear in photos and lead to a biased interpretation, e.g. the presence of buildings, or the nature of the weather in the photos.

The illustrations were tested in the Stage III through many steps of focus groups and cognitive interviews for coherence with other materials presented and for their overall usefulness. At each step successive improvements were made. The descriptions were also successively improved in terms of their scientific accuracy through several iterations with the Steering Group, WRc and Environment Agency. They were retested in the pilot survey, at which they proved valuable and acceptable.

A single example, rather than all four, is presented in order to keep the interview to a reasonable length.

3.4.6 Assignment of Normative Labels and Colours

The three status levels in the SP survey instrument are labelled "High", "Medium" and "Low" quality, and colour coded using shades of blue. These labels and colours help respondents arrive at a value for a complex unfamiliar good without leading them excessively. Notwithstanding the desirable property of helpfulness to respondents, it was essential to test that the labels and colours were not leading. Extensive tests were therefore conducted in the Stage III focus groups, cognitive interviews and pilot surveys.

The normative labels were tested by checking with pilot that they matched the descriptions. The colour codes were tested by asking focus group participants and cognitive interview respondents "What do you think of the colours selected?", and, "Do the colours fit the description?". In addition, in the focus groups, many alternative colour coding schemes were considered and tested. The variants included red, amber, green, which participants liked but which was ultimately discarded for fear it may be leading despite its clarity. A dozen or so other colourways were tested before settling on shades of blue. There was no evidence that respondents felt that the normative labels or colours were inappropriate given the descriptions, and strong evidence that they were helpful, in forming a value for an unfamiliar and complex good. This finding should not be surprising, since it seems clear that the WFD status levels are likely to be normatively ordered by almost everyone.

The possibility remains that people might choose differently given a different colour and labelling schema. However, removal of helpful labels and colours altogether would leave respondents with a more difficult task, reducing the validity of the resulting value estimates.

3.4.7 Status Level Mapping

The mapping of the baseline status of water environment sites according to our status levels by WRc uses GQA data primarily, and the Environment Agency's Risk Assessment data secondarily where the GQA data is missing. WRc (2007) contains details of the approach taken to ensure the scientific accuracy of the mapping. We do not discuss the approach further here.

The mapping was tested in the pilot stage, in conjunction with the descriptions, normative schema and illustrations. The key question at this stage was whether respondents accepted the information presented to them as representing the truth, or whether they disbelieved and disregarded certain aspects due to misalignment with their own experience. This is because we want respondents to value the situations presented, not their own alternative versions, so we can compare and aggregate their answers. The final mapping passed this test well.

3.4.8 Catchments, Rather than Lengths of River, Drawn and Coloured

The maps are drawn with catchments coloured, rather than with lengths of river coloured. The former provides much improved clarity. During the Stage I deliberative group research, maps were used as show materials, with rivers as lines on the maps. While finding the maps useful, respondents found the coloured river lines difficult, and in many cases impossible, to discern. By contrast, in the later focus group and cognitive interview research, respondents almost uniformly understood and appreciated the form of presentation. Main survey questions test respondents' reactions to the maps.

3.4.9 Spatial Metrics for Representing Baseline and Improvements

The unit adopted to measure the "quantity" of each status level is hectares of catchment area for rivers, and hectares of surface water area for lakes, estuaries and coastal areas. Alternatives considered included the number of sites, as defined by the WFD, and a metric combining the surface water area for lakes, estuaries and coastal areas, with an approximation to river water surface area, based on kilometres of river.

Within the valuation framework, the improvement amount does not distinguish the relative amounts or values of improvements between different water body types. This approach was chosen during Stage II after consideration of the options at a roundtable discussion with scientists from the Environment Agency and WRc. The rationale was two-fold. First, and most importantly, the catchment unit gives the closest match to the cost structure of maintaining and improving water sites, given a catchment-based approach to water resource management. Second, the proportions of status levels derived as summary statistics using this metric, and presented in a pie chart, most closely matched the visual appearance of the maps and were best understood by respondents. The metric was tested against an alternative based on number of sites in the Stage III focus groups, and the majority of participants expressed a weak preference for the catchment unit.

3.4.10 Spatial Referencing: Local and National Areas

The SP instrument includes two maps, one of which covers the whole of England and Wales and is shown to every respondent. The other is a local area map, which is unique for each interview location. This map covers the 30-mile surrounding area. In the choice cards (see Appendices C and D), there is a corresponding separation between local and national area information on environmental status levels. The presentation of the maps mirrors the information presented in the choice cards: the area proportions by status level in the "No-Change" option are the same as those presented on the maps.

The rationale for separating local and national areas is to take account of the strong finding from the Stage I deliberative group research, and many other studies, that proximity is a key determinant of the value households place on environmental improvements..

Alternative degrees of detail in representing the location of improvements in the CE were considered and tested during Stage II of the survey design process, including a single national set of status proportions, and separate status proportions by water body type (lake, river, estuary, coast). A number of focus group participants and cognitive interview respondents expressed a preference for the simplest version, and by contrast, the version including separate water body type information on water status levels was widely rejected as not understood.

The selection of a 30-mile radius as the boundary of the local area for the maps was made after testing both 10 mile and 50 mile radius maps in the Stage III focus groups and cognitive interviews. Both versions tended to be accepted in this Stage as representing local areas, with a mixture of fairly weak and unsubstantiated preferences one way or the other. The 30-mile radius was selected, as a compromise. (Note that over recruitment protocol required respondents to live within 15 miles of the centre of each local area, to be sure the map is relevant to them as their local area.)

3.4.11 Format of Representation: Pie-Charts vs Bar-Charts

The choice cards and maps in the survey each use pie charts to represent information on the status proportions. During the Stage II focus group testing, alternative versions of the choice experiment and maps were presented, half of which used pie-charts, and the other half of which used bar-charts to present the information. The vast majority of focus group participants found pie charts to be easier to understand and interpret, hence this format was chosen.

3.4.12 Timing of WFD Improvements

For each option in the CE, respondents are told the proportions of status levels after 8 years (2015), and after 20 years (2027). This specification of the timing of improvements draws on evidence from the deliberative groups, consistent with prior expectation, that people do not expect to see change immediately but that they would like to see measures undertaken sooner rather than later. The variation in timing of improvements afforded by the CE design allows the analysis of the survey data to estimate a discount rate and thereby understand households' values more fully. We followed the Green Book in using the social rate of time preference in aggregation.

Regarding the representation of the timing of improvements, given the lower sensitivity of respondents to the 20-year situation, relative to the 8-year situation, it was decided that the representation would be simpler, and still effective, if in the 20-year case the amount of "low" quality were restricted to always be zero, and the amount of medium and high quality were always the same for the local as for the national environment. These restrictions enable the choice card to include only a single line, and number, to represent the 20-year situation, rather than needing to present the full series of pie charts as for the 8-year environmental situation. This does not prevent scenarios involving lower improvement in 20 years from being valued, using the step from the status-quo to the 20-year level, and by extrapolating the intermediate values obtained from the 8-year scenarios.

3.4.13 Uncertainty

The SP survey instrument does not include a separate attribute to capture the extent of uncertainty that policy alternatives will achieve the improvements specified. Such an attribute was considered in early versions of the CE to enable support for decision-making amongst policy-mixes that vary in the degree of risk associated with the achievement of the intended outcomes. Versions including an uncertainty attribute were tested in the Stage III focus groups and cognitive interviews. Analysis of the cognitive interview results revealed, as expected, that people strongly preferred more certainty to less, but that people seemed to respond qualitatively to the labels of the levels of the uncertainty attribute ("Almost certain" and "Likely"), and not to the probability values themselves, which varied in split samples. This finding cast some doubt on using the probability numbers quantitatively in subsequent valuation work, due to concerns that they would inevitably be inconsistently interpreted by respondents rendering the values invalid. The uncertainty attribute was therefore excluded from the CE for the pilot surveys, and, consequentially, from the instrument for the main survey. The benefit values must therefore be interpreted as applying to scenarios which will occur.

3.5 Payment Vehicle

It is critical that the payment vehicle be something respondents think they would actually have to pay and could not avoid. This is so that they give answers which reveal their true valuations. At the same time, the payment vehicle must be realistic in the sense that respondents will see that the valued-goods could lead to costs being recovered through the payment vehicle suggested. Also, the payment vehicle needs to allow that other groups will also (if true) be paying, so that household respondents do not give answers based on fairness concerns instead of their own benefit values.

The payment vehicle adopted for the CE and CV is a combination of water bills plus higher prices on everyday products. Adopting water bills on their own was considered first due to its desirable properties of universal coverage and clear necessity of payment. However, it was desired that the valuation scenario introduce the fact that many actors are involved in improving the waters including farms and industry, rather than just water companies. In light of this, there was concern that using water bills as the sole payment vehicle was not credible.

Approaches using higher prices on everyday products as a (partial) payment vehicle have been used in past studies successfully, e.g. Mitchell and Carson (1984). While some negativity was expressed towards water companies in the deliberative research, this was perhaps less than might have been anticipated based on experiences in past studies. In the Stage III focus group, cognitive interview and pilot testing, the payment vehicle (i.e. an increase in “water bills and other household payments”) worked without significant problems. For example, only two respondents out of 50 in the first pilot survey gave zero willingness to pay amounts in the CV, and only a small number, 4 out of 50, cited protest-type reasons in explanation of their willingness to pay bids.

Respondents are asked to react to or provide amounts which are the link between the values’ respondents derive from WFD improvements and the values they could derive from the wider alternative ways of using the payment amount. For economic valuation of WFD improvements we want respondents’ to state values that are informed by all the other possible ways the payment amount could be used. To help ensure respondents are in this mindset the introductory material in the survey, and later questions, use the ‘corrective entreaty’ of asking respondents to provide values remembering all the other things they could do with the money.

3.6 Experimental Design

The survey design included stratification of the sample across a range of design variables, and randomisation of a number of other variables. This approach led to each respondent receiving a unique set of survey materials. This subsection describes our approach in respect of the allocation of materials to respondents.

3.6.1 Order of PCCV, DCCV and CE Questions

The survey instrument contained three value elicitation mechanisms for each respondent: a payment card contingent valuation (PCCV) question, a dichotomous choice contingent valuation (DCCV) question, and eight choice experiment (CE) questions, including one example (or practice) choice card. The respondents were talked through each card-type line-by-line when first presented, to be sure they could interpret the material.

In our initial design we employed a payment card contingent value question (PCCV) and a set of choice experiment questions (CE). This was because CE is required to examine marginal values of change locally, nationally and at different times, but CV methods are regarded as better for valuing packages of improvements such as the WFD effects. The PCCV was initially adopted over the single bound dichotomous choice contingent value elicitation method (DCCV) because the latter format is very similar to the CE format, and because we wished to understand the effect on valuation of the different focus of the PCCV format with its presented payment card money focus (the DCCV format is very like the CE format). In the pilot the PCCV stated total values were substantially below those derived from simple CE models, so to provide a further check we introduced the DCCV format as well for the main survey, by varying one of the CE choices to take DCCV form. This was to see whether the difference seemed to be due more to the format or to the port-whole differences.

The sample was split into two, with half receiving the PCCV question before the CE, and half receiving the PCCV question after the CE, so that ordering effects could be checked. These could include anchoring of later responses on values presented earlier.

The DCCV question was embedded within the CE choice questions at either the beginning, in the middle, or at the end of the series. The sample was thus split into three for both "CE, then PCCV", and "PCCV, then CE" question orderings, with each third completing a survey questionnaire with one of the three positions for the DCCV question within the CE.

3.6.2 CV Scenario Levels (Payment Card and Dichotomous Choice)

In all cases the baseline scenario is for environmental quality to remain as in 2007.

For each respondent, the PCCV and DCCV scenarios are identical in terms of environmental improvements offered.

For both elicitation questions, a respondent received either a "75% high quality in 8 years" scenario, or a "95% high quality in 8 years" scenario. An equal proportion of the sample received each of these two scenarios.

For each improvement scenario, the proportion of the water environment at high quality to be obtained in 20 years time is 95%.

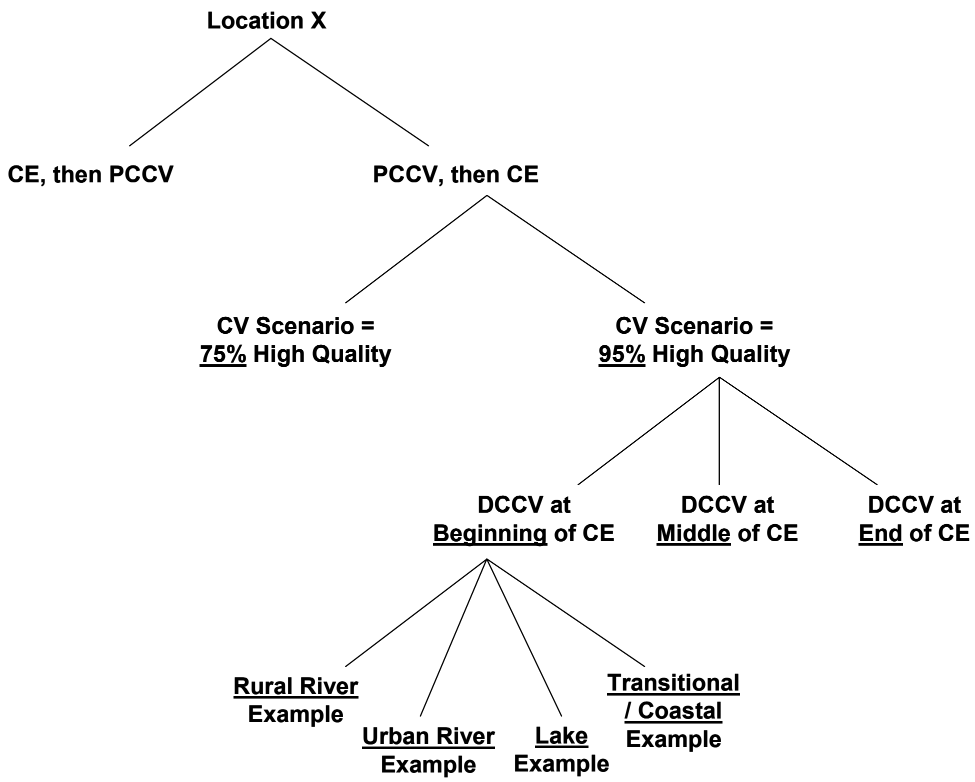
3.6.3 Water Body Type Examples Used

Four different water body type examples were presented, with each respondent receiving one example only. An equal proportion of the sample received each of the four water body type examples.

3.6.4 Summary of Experimental Design Stratification

Figure 3.2 illustrates how this design meant the sample was stratified over the design variables.

Figure 3.2
Illustration of Split Sample Design



For each location, an equal number of instruments were created for each node of the tree in the above figure, resulting in 48 unique questionnaires per location. Since fewer than 48 surveys were conducted in each location, Accent allocated the sets randomly at the time of interviewing from the available collection.

3.6.5 No Deterioration Cost

The cost presented to each respondent for maintaining current conditions in the water environment was drawn randomly from the range {£5, £10, £20}, measured as an annual increment to water bills and other household payments. A single value was drawn for each respondent and applied throughout their interview. This was to allow us to check whether wtp for WFD scenarios is sensitive to other expected changes in bills.

3.6.6 DCCV Scenario Costs

The cost presented to each respondent for the DCCV scenarios was drawn randomly from the range {£5,£10,£20,£30,£50,£100,£200}, stated to be an annual increment to "water bills and other household payments." This range was derived by open questions in focus groups about the extent of wtp, and adjusted after the pilot to reflect the PCCV pilot answers.

3.6.7 Payment Card

The same payment card was used for each respondent in the sample (see Appendix D).

3.6.8 CE Attributes and Levels

Table 3.2 defines the attributes used in the choice experiments.

Table 3.2
CE Attribute Definitions

Attribute	Definition
ll0, ml0, hl0	Percent low, medium and high quality in local area at time=0 (current conditions)
ll8, ml8, hl8	Percent low, medium and high quality in local area at time=8 (in 2015)
ln0, mn0, hn0	Percent low, medium and high quality in national area at time=0 (current conditions)
ln8, mn8, hn8	Percent low, medium and high quality in national area at time=8 (in 2015)
hn20	Percent high quality in local and national areas at time=20 (2027)
bill	Increase in water bill and other household payments (£/hh/yr)

Source: NERA

Possible levels of the future environmental status attributes: *ll8, ml8, hl8, mn8, hn8* , and *hn20* were generated from the levels of the baseline conditions: *ll0,ml0, hl0,ln0 ,mn0* and *hn0* , which are fixed known quantities, and given steps. The generating functions for each level of each environmental attribute, used for the main survey, are presented in Table 3.3 below.

Table 3.3
Environmental Attribute Levels

Attribute	Level 1	Level 2	Level 3	Level 4
ll8	0	¼*ll0*	½*ll0*	¾*ll0*
hl8	hl0+¾(ml0 - Δll8)	hl0+½(ml0 - Δll8)	hl0+¼(ml0 - Δll8)	hl0+0.1(ml0 - Δll8)
ln8	0	¼**ln0*	½*$\%L^N_0$	¾*ln0*
hn8	hn0+¾(mn0- Δln8)	hn0+½(mn0 - Δln8)	hn0+¼(mn0 - Δln8)	hn0+0.1(mn0 - Δln8)
hn20	95	75		

Note: Δ118 = 118-110, etc
All environmental status level numbers were rounded to the nearest whole percentage point for inclusion in the choice sets.

The levels are presented in Table 3.3 in order of presumed value, with the highest valued level included as Level 1 for all attributes. Thus, for *ll8*, the best level is 0, which is where there is no low quality status remaining in 2015. Levels 2 to 4 for this attribute all include some positive quantity of low quality, provided that there is some low quality in the baseline. There is no need to present medium quality in this table since this is always equal to one minus the sum of low and high quality.

Table 3.3 shows how the possible levels of the environmental status attributes in 2015 are generated from the baseline levels. The first attribute to be generated is *ll8*, which in Levels 2 to 4 is equal to a quarter, a half and three quarters respectively of the initial quantity of red in the baseline. The four levels of *hl8* are then calculated by adding to the baseline *hl0* decreasing proportions, for successive levels, of the potential quantity of medium quality in 2015, after the change in low quality from baseline to 2015 has been factored in. Thus, if the initial quantity of high quality is 20%, the initial quantity of medium quality is 25% and the decrease in low quality between baseline and 2015 is 15%, then Level 1 of *hl8* will be (20 + 0.75*(25 + 15))% = 50%; Level 2 of *hl8* will be (20 + 0.5*(25 + 15))% = 40%, and so on.

The levels of the cost attribute *bill* used in the main survey are {5,10,20,30,50,100,200}, measured in pounds per household per year, a range which was adopted to reflect verbal responses in the focus groups and adjusted given the PCCV responses in the pilot, with the intention of allowing the best identification of the frequency curve of value responses.

3.6.9 Allocation of CE Levels to Option Profiles

Random sampling was adopted to generate the combinations of levels presented, on each option and in each choice card. Option A is always to continue the baseline. The combinations of levels assigned to Options B and C were generated by random sampling from the full factorial design. That is, a list of every possible combination of levels from Table 3.3 above and the bill range was created, and then pairs of options were drawn randomly from that database without replacement so that each choice card for each respondent was unique.

To increase the information recovered in the survey all pairs with a strictly dominated combination of attributes were removed, ie where Option C was better than Option B in every attribute, including all environmental status attributes, and the bill attribute, or vice versa. In addition, practically impossible combinations were also removed. This included combinations where *hn20* was less than *hn8*, and cases where *ln8* = 0 and *1l8* > 0.

This experimental design is intended to approximate to the full factorial and so it is possible to estimate all effects and combinations of effects without bias. While more efficient designs (more statistical power for a given sample size) are possible than our simple randomization approach, they complicate both the development of the materials and the later analysis, and may not allow interaction terms to estimated. With samples of our size, efficient designs do not add much power.

4 Questionnaire Elements

4.1 Overview

The questionnaire (see Appendix C) begins with an introductory section, which includes questions on the respondent's use of the water environment. The next section introduces the valuation scenario. It presents the status descriptions, followed by the maps, and describes the potential benefits of the WFD. The next section introduces the valuation tasks, elicits CE and CV values and asks follow-up questions. The final section asks demographic questions. After the interview is finished, there are a number of debriefing questions for the interviewer to complete.

The introduction to the valuation questions sets out the context carefully, mentioning among other things: that under the WFD no site will deteriorate; a base payment may be needed to achieve this; there is no change to access or litter resulting from the WFD; that water bills and household expenses may increase in future for reasons not related to the water environment; that household income may change in future.

4.2 Attitudes, Opinions and Uses of the Water Environment

The questionnaire includes two attitude questions. The first asks respondents to consider five statements about pollution control and indicate which of them they agree with the most. The statements range from: "Protecting the environment is so important that pollution control requirements and standards cannot be too strict, and continuing improvement must be made regardless of cost", to: "Pollution control requirements and standards have gone too far and they already cost more than they are worth." This variable was found to be a good predictor of PCCV willingness-to-pay amounts in the pilot survey. Demonstrating that attitude variables such as this have predictive power for willingness to pay is important in validating the elicited willingness to pay amounts.

The second attitude question asks respondents to choose, from a list of three options, their preferred order of improvement for implementation of the WFD. The alternatives offered are: a) First improve the worst areas, as judged by scientists, regardless of where they are located; b) First improve areas near to where many people live or visit; or c) First improve areas where quick improvements can be achieved at low cost. The answers provide a cross-check on individuals' explanations for their CE and CV responses, and thereby on the validity of the CE and CV responses themselves.

The instrument collects data on uses of the water environment disaggregated into three types, contact use, fishing, and other non-contact uses. One would expect that frequent users of the water environment would be willing to pay more for improvements to the water environment than non-users. The data from this section will be used to test this hypothesis. The questions also encourage the respondent to consider the ways they, and their household, use the water environment, which will serve well to initiate the construction of a value for its maintenance and improvement.

4.3 Follow-up Questions

Debriefing questions are asked of respondents at various stages notably after presentation of the maps and status levels showing status levels for local and national water environments, and after the valuation questions.

Following the presentation of the maps and status levels, the questionnaire asks whether there is more or less low quality than the respondent would have expected. These questions encourage respondents to reflect on prior perception of current local water status and relate to the new information presented. In addition, answers provide evidence of the extent to which the state of the environment differs from prior expectations across the population.

Following the presentation of potential WFD benefits, respondents are asked to choose from a list which is the most important to them. This question helps respondents to mentally review and incorporate the information just presented. The responses provide data on relative priorities accorded to different benefit types, ie user and non user values.

Following the example or practice choice card which the respondent is talked through line-by-line, respondents are asked directly why they chose the option they did. This question encourages respondents to reflect on the reason for their choice and helps them make choices in the remainder of the valuation section. It also provides top of mind responses on how options were understood and how alternatives were viewed, allowing the interviewer to correct any obvious misunderstandings, and allowing validity checking.

Following the full series of questions for the CE, respondents are asked: "In making your choices between options on the cards, what factors did you consider, and which were the most important?" Similarly, following the CV question, respondents are asked: "How did you decide the amount you/your household would be willing to pay?" Pre-coded alternatives to these two follow-up questions were considered, following the pilot surveys, however it was decided that this could lead to the loss of potentially valuable information. The verbatim responses are used to identify protest explanations, to understand the reasons for any differences between CE and CV results, and to test the validity of the valuations by checking reported reasons against the choices people made.

Additionally, following the CV exercise, respondents are asked how sure they are about the wtp answers given. Answers to this question add context to the examination of outliers and provide information to understand the degree of confidence to be held overall in the valuation results.

Respondents are then asked three questions on the extent to which they considered current and future costs and income, and are given an opportunity to revise their bids following these questions should they wish to do so. These questions are important to demonstrate the validity of the value responses obtained.

4.4 Demographic Information

The demographic questions included in the survey instrument are standard. They include age, sex, education, employment status, socio-economic status, income group, and whether or not the respondent is a member of an environmental group. Many of these variables have predictive power over willingness to pay amounts. In addition, they are used to weight the sample using census data to adjust for any degree of unrepresentativeness of the survey sample obtained in any of these dimensions.

5 Target Population, Survey Mode and Sampling Strategy

5.1 Target Population

Household Benefits; Not Business Benefits

The target population for the survey is the set of households in England and Wales. Alternatives considered included individuals, or households and businesses. Households were adopted over individuals as the target population due to the fact that budgetary consumption decisions are often made at a household level rather than an individual level and so aggregating individual's decisions could lead to double counting. Also, water and wastewater bills are paid at household level rather than individual level.

While businesses and other employers may care about the quality of a river nearby to their offices, due to the commercial benefits associated with the attractiveness of their office location as a place of work, and the corresponding impact on ease and cost of recruitment, and the salaries needed to attract employees to work there, these impacts are captured to a certain extent by the amount that households are willing to pay for improvements. It is demanding to ask people to ignore impacts of improvements near their place of work on their salary prospects, and we do not expect people to naturally lower their willingness to pay for improvements to take account of these feedback effects. If businesses or employers were to be included in the population of valuers, there would likely be a degree of double-counting of benefits. If business/employer benefits seem substantial enough to be important to WFD valuations then they should be assessed directly.

Current Generation's Benefits

The target population for the SP survey is restricted to the current generation. It is only the present generation who can express their valuations in today's SP work, though we expect that benefits will also accrue to future generations. If today it is thought that future generations will value the WFD improvements in a different way, and this different way is clear, then today's WFD valuations should be adjusted or overtaken accordingly in forming a "full value" estimate.

5.2 Survey Mode

Given the complexity, length and need to show stimulus material it is necessary that interviews are conducted face-to-face, using computers. Further, given the number of interviews and the need to conduct the fieldwork in around four weeks, a hall test approach was adopted. This involves setting up a hall and recruiting people to come in for a face-to-face interview on the spot or later.

In comparison with telephone or mail surveys, face-to-face interviews have the advantage that

- complex issues are easier to explain.
- show material can be used during the initial contact.
- longer interviews are possible as it is easier to retain the respondent's interest.

The only real disadvantage of face-to-face interviews is that it is less easy to get a good geographical spread as the sample is typically clustered for cost purposes. Given the possible importance of geographic variation to the valuation of water environment improvements, we aimed to ensure that an average of no more than 30 interviews was carried out at each sampling point.

Hall tests have the following advantages relative to in-house interviews:

- interviewers have access to instant support from the resident supervisor. This is very important when as here the survey is complex. In the alternative household location interviewers can feel isolated and continue to make the same mistakes as they are not picked up quickly enough.
- the first day briefing in any location can be immediately followed up by ‘trial’ interviews and then an observed (by the supervisor) first real interview; again very important given the complexity.
- recruitment leads straight into an interview so is much quicker than door-to-door recruitment which can take several ‘call-backs’ if undertaken properly.
- interviews in a hall will provide more geographical dispersion in the local area than a household based survey as recruits to the hall test can have come from any location in the area, whereas a household based survey needs to ‘jump off’ from a number of specific locations and then work within a fairly tightly subscribed area, for example every fifth house.
- the hall approach allows for access to recruitment of a wide range of people making it easier to fit respondents to particular respondent quota groups.
- the hall provides a quiet environment for the interview to take place, without home distractions of family etc, which will allow the respondents to give careful consideration to the show materials which are a vitally important element of this study.
- complex issues can be explained face-to-face in either environment, but usually with less distraction in a hall.
- generally the hall approach makes longer interviews more feasible, as hall interviews have been shown to make it easier to retain respondent interest.

The hall approach requires high standards of recruitment and fieldwork procedure to ensure that the sample obtained is robust. Whereas poorly recruited household samples may suffer from a bias against those not often home, poor recruitment into a hall may suffer from a bias towards those most easily available or self selecting at the recruitment questionnaire stage.

5.3 Sampling Strategy

5.3.1 Sample Size and Stratification

The reliability of all statistics increases with sample size, all else equal, but this increases costs so stratification is used to allow accurate results to be obtained for respondent sub-sets. On the basis of discussions with the CRP Steering Group, about the priorities for the research, informed by calculations of the likely impact on confidence ranges, it was decided that:

- the regional stratification would involve only ensuring a set minimum of two sampling points for the smallest RBDs, with an otherwise population weighted sample size from each RBD. This approach strikes a good compromise between the need for robust National statistics, as the most important figures to generate, and reasonable sample sizes to generate cross-checking estimates for the smallest RBDs; and
- a sample size of 1500 respondents from 50 locations would provide adequate confidence.

5.3.2 Sampling of Areas for Hall Locations

A probability sampling approach was adopted for selecting the areas in which the halls were located. This approach requires the selection of a sampling frame, or list, from which elements are to be drawn with some probability, and an approach for drawing the elements from the list. The frame we used was the list of urban areas in England and Wales with population greater than 1,500, which is available from the UK Census 2001. In this frame, unique elements of the list vary from Greater London to small villages with populations of only 1,500. This sampling frame, rather than a frame listing local authorities or other population enumeration districts, is preferred for the selection of locations for two reasons: first, because the halls will need to be in areas where there is a reasonable amount of foot traffic to enable recruitment within a reasonable time frame, and secondly, because the centre of the location observation, being a village, town or city centre, is a natural centre for the drawing of a map which will be interpreted as a local area for potential recruits.

Locations were sampled from the list as follows:

- First, the frame was ordered by RBD and then by population size, *pop*.
- Then, the cumulative population, *cupop*, of elements in the frame was calculated. That is, for each element, *n*, *cupopn* measures the sum of the populations of all elements above it in the frame.
- Thirdly, the total population of the country resident in urban areas, (settlements > 1500 population, approx 90% of the England and Wales population), was divided by 100 to create a sampling interval, *x*, allowing the selection of locations at regular spacing in the list, and thus with proportional sample sizes by region and population size.
- Finally, a single random number, *e*, was selected from the range 0 to *x*, to provide the random starting point for the interval-based sample selection.

An important point to note is that the first stage of ordering by RBD and size has the effect of ensuring that the number of locations sampled in each region is proportional to the population of the region, and that the sample contains a representative number of locations by size of urban area, ie small and large areas will each be sampled with this approach.

Individual respondents' home post codes were coded and classified according to urban-rural categories in the Joint Government Department Classification system, as Appendix I shows (Table I.8).

5.4 Research Targets and Recruitment

To ensure that the sample was representative of the population as far as possible, in order that sample statistics accurately measure the population parameters of interest, we used quota sampling:

- identifying and selecting measurable variables to use (i) as quota for recruitment purposes, and (ii) to check for sample representativeness post-survey; and,
- adopting a fieldwork approach that mitigates against potential biases within quota due to recruitment location, target recruit selection, and approach / introduction.

The measurable variables upon which the quotas were determined are age, gender, and SEG. Urban / rural split, area population, household composition, and income, were used to check for sample representativeness post-survey. In addition, we collected information in both the 4c survey and a contemporaneous Omnibus survey to allow us to check the representativeness of our sample for general attitudes, such as concern for the environment, against a separate contemporaneous study. Attitudes are key determinants of willingness to pay, but it is difficult to know whether the attitudes of a sample recruited off the street are representative of those of the population. The use of the Omnibus survey aimed to provide such a check. The results from this survey are presented alongside those from the main 4c survey for comparison. The inclusion of attitude questions will also potentially allow wtp results to be updated in future for changing attitudes over time.

Our fieldwork was designed to recruit a sample conforming to the population. Steps taken included:

- recruiting in places where all segments of the local and regional population are likely to visit, e.g. supermarkets, high streets (village, town and city);
- approaching every 1 in nth person to avoid biases towards friendly faces in the sample;
- providing £8 incentives to take part to maximise sample coverage and minimise self-selection for each location;
- recruiting at different times of the day to capture all segments of the population;
- recruiting on different days of the week to ensure a representative spread of the population; and
- allowing for future appointments when recruiting, e.g. recruit for immediately, later that day, or later that week.

The recruitment questionnaire is included as Appendix B to this report. In order to be in scope, recruits need to be responsible, solely or jointly for paying the water bill, and they need to live within 15 miles of the survey location in order to ensure that they are not so close to the edge of the 30-mile radius map presented for that location that their ‘local’ area is largely off the map.

6 Survey Implementation

The current section provides details on the implementation of the survey, including:

- the locations of the halls;
- fieldwork;
- interviews achieved against the target quotas;
- interviews achieved against a range of additional geographic and demographic criteria, including the urban/rural split, income, education, household composition, receipt of state support, and accommodation occupation status; and,
- length of interviews and interviewer feedback.

6.1 Survey Locations

Hall venues were sought for each of the locations identifed in the sample design specification. In some cases, the original locations required substitutions due to our encountering certain problems at the time of booking the venues.

A few did not have a suitable venue, or the only suitable accommodation for the survey could not be used during the fieldwork period.

Two locations did not allow street interviews without prior permission being granted and this could not be gained within the fieldwork timescale.

Finally, for Berwick-upon-Tweed, a large part of the area within 30-miles lies in Scotland. The local water quality map for this location was unhelpful because the whole Scottish part of the map was not colour coded. It was felt that the presentation of materials for this area was liable to cause confusion amongst respondents.

Where substitute locations were needed, the replacement locations used were the areas closest in population size within the same RBD.

In accordance with the design specification, the vast majority of surveys were conducted in halls in the assigned locations. However, there were three locations, where no hall venue was available during the limited fieldwork period. So interviews were conducted house-to-house, and in five further areas some house-to-house interviews were conducted to supplement the hall interviews in order to get closer to quota.

A final list of the locations used for the survey, including their postcodes, is shown in Table 6.1.

Table 6.1
Final Hall Locations

Location	River Basin District	County	Postcode
Wimblington/Doddington	Anglian	Cambridgeshire	Wimblington, PE15 0QT
Stowmarket	Anglian	Suffolk	Stowmarket, IP14 1AD
Peterborough	Anglian	Cambridgeshire	Peterborough, PE1 1XS
Brentwood	Anglian	Essex	Brentwood, CM14 4RR
Neston	Dee	Merseyside	Door-to-door in CH64
Mold	Dee	Clwyd	Mold, CH7 1AP
Dordon/Polesworth	Humber	Staffordshire	Tamworth, B79 7BX
Wigston	Humber	Leicestershire	Wigston, LE18 1DS
Cannock	Humber	Staffordshire	Cannock, WS11 1EB
Walsall	Humber	Staffordshire	Walsall, WS1 1DA
Birmingham	Humber	West Midlands	Birmingham, B2 4NU
Dunnington	Humber	North Yorkshire	Dunnington, YO19 5PW
Goole	Humber	East Yorkshire	Goole , DN14 5AA
Grimsby	Humber	Lincolnshire	Grimsby, DN31 1HG
Kingston Upon Hull[3]	Humber	East Yorkshire	Hull, HU1 3JP
Sheffield	Humber	South Yorkshire	Sheffield, S1 2OB
Wardle	North West	Cheshire	Door-to-door in CW5
Whitefield	North West	Greater Manchester	Prestwich, M25 1AY
Skelmersdale	North West	Merseyside	Door-to-door in WN8
Cheadle and Gatley	North West	Cheshire	Cheadle, Stockport, SK8 1AX
St. Helens	North West	Merseyside	St Helens, WA10 1AF
Manchester	North West	Greater Manchester	Manchester, M2 1FB
Hexham	Northumbria	Northumberland	Hexham, NE46 3LS
Gateshead	Northumbria	Tyne and Wear	Gateshead, NE8 1EP
Hagley	Severn	West Midlands	Stourbridge, DY8 1DW
Coventry	Severn	West Midlands	Coventry, CV1 1LZ
Chippenham	Severn	Wiltshire	Chippenham, SN15 3ER
Bath	Severn	Somerset	Bath, BA1 2EB
Ebbw Vale	Severn	Gwent	Chepstow, NP6 5EP*
Penrith	Solway Tweed	Cumbria	Penrith, CA11 7XX
Carlisle	Solway Tweed	Cumbria	Carlisle, CA3 8JE
Heathfield	South East	East Sussex	Heathfierld, TN21 8LB
Eastbourne	South East	East Sussex	Eastbourne, BN21 3JX
Winchester	South East	Hampshire	Winchester, SO23 9LQ
Shaftesbury	South West	Dorset	Shaftesbury, SP7 8JE
Taunton	South West	Somerset	Taunton, TA1 3XZ
Crowborough	Thames	East Sussex	Crowborough, TN6 1AF
Crawley	Thames	West Sussex	Crawley, RH10 1BS
Bexley	Thames	Kent	Bexleyheath,DA6 7EZ
Southwark	Thames	London	Southwark, SE22 8EP
Ealing	Thames	London	Ealing, W5 2BY*
Goring/Streatley	Thames	Berkshire	Streatley-on-Thames, RG8 9HR*
Wokingham	Thames	Berkshire	Wokingham, RG40 1AS
Aylesbury	Thames	Buckinghamshire	Ayelsbury, HP20 2RW
Barking and Dagenham	Thames	London	Barking, IG11 8HG
Luton	Thames	Bedfordshire	Luton, LU1 3AA
Havering	Thames	Essex	Romford, RM1 3AB
Wembley	Thames	London	Wembley, HA9 7JA*
Llangefni	Western Wales	Isle of Anglesey	Llangefni, LL77 7LR
Bridgend	Western Wales	Glamorgan	Bridgend, CF31 1EB*

Source: NERA-Accent
** some house-to-house interviews undertaken to supplement hall tests*

6.2 Fieldwork

Respondents were recruited on street, using the recruitment questionnaire and invited to undertake an interview in a nearby hall venue. In the small number of locations where house-based interviews were conducted, respondents were recruited by means of a random walk approach.

An £8 Gift voucher was offered as an incentive to participate in the survey in all cases.

In scope respondents had to live within 15 miles of each of the chosen locations. A map which had a 15 mile ring on it was shown to establish whether respondents were in scope. The home postcode was recorded in the CAPI questionnaire and used to classify respondents as urban/rural (see Appendix I).

A target of 30 interviews was set for each of the 50 locations.

Quotas were set for each location as follows:

- Age: equal spread of following categories:
  - 18-34;
  - 35-54;
  - 55+
- Gender:
  - 45%+ male
  - 45%+ female.
- Socio-Economic Grade (SEG):reasonable spread from across the categories[3]
  - AB Higher and intermediate managerial/administrative/professional
  - C1 Supervisory, clerical, junior managerial/administrative/professional
  - C2 Skilled manual workers
  - D Semi-skilled and unskilled manual workers
  - E On state benefit, unemployed, lowest grade workers
  - X Not applicable

  (Not applicable category (X) comprises: people aged 15 or under or aged 75 or over)

Interviews took place between Friday 13 July and Friday 3 August 2007.

6.3 Interviews Achieved Against Quota

On average the strike rate was about 1 in 7 (ie seven potential respondents were approached for every one recruited). Strike rates were typically better at the beginning of shifts and worse at the end of shifts when specific quota targets were being looked for.

[3] National Statistics Socio-Economic Classification (NS-SEC) codes were derived from the SEG data following implementation of the survey using the National Statistics approximation methodology (see http://www.scrol.gov.uk/scrol/metadata/Socio_Econ_Classification.htm). The NS-SEC codes have been used to compare the sample achieved against the 2001 census, which records this category. The reason for including SEG rather than NS-SEC data at the time of recruitment is due to the fact that Accent's market research interviewers were trained to code SEG, but did not have training in NS-SEC coding. Accent therefore decided to code to SEG during the recruitment interview and then code to SEC in the office with suitably trained coders. Interviewers were briefed to write down details of the occupation on the paper recruitment questionnaire.

The target of 30 interviews for each location proved difficult to achieve for a number of reasons. Particular difficulties faced were:

- Low footfall in some of the locations, which meant recruitment of in-scope respondents was difficult
- Heavy rain in some locations during the fieldwork period
- The stated interview length was off putting for some.

Difficulties in recruitment and the specific nature of some areas (ie large numbers of specific types of respondents such as elderly, specific SEG groups) also meant that meeting some elements of quotas was very difficult in many locations.

The numbers of interviews by survey location are shown in Table 6.2.

In total 1,487 interviews were undertaken against a target of 1,500. The numbers of interviews achieved by location ranged from 17 to 36 against the target of 30 in each location.

Table 6.2
Interviews Achieved by Survey Location

Urban Area	River Basin District	Government Office Region	Country	Interviews Achieved
Doddington	Anglian	East of England	England	21
Stowmarket	Anglian	East of England	England	33
Brentwood	Anglian	East of England	England	32
Peterborough	Anglian	East of England	England	24
Neston	Dee	North West	England	36
Mold	Dee	Wales	Wales	36
Dunnington	Humber	Yorkshire and The Humber	England	36
Dordon/Polesworth	Humber	West Midlands	England	31
Goole	Humber	Yorkshire and The Humber	England	27
Wigston	Humber	East Midlands	England	23
Cannock	Humber	West Midlands	England	32
Grimsby	Humber	Yorkshire and The Humber	England	29
Walsall	Humber	West Midlands	England	31
Kingston upon Hull	Humber	Yorkshire and The Humber	England	33
Sheffield	Humber	Yorkshire and The Humber	England	31
Birmingham	Humber	West Midlands	England	31
Wardle	North West	North West	England	30
Whitefield	North West	North West	England	32
Skelmersdale	North West	North West	England	25
Cheadle and Gatley	North West	North West	England	23
St. Helens	North West	North West	England	30
Manchester	North West	North West	England	17
Hexham	Northumbria	North East	England	26
Gateshead	Northumbria	North East	England	30
Stourbridge	Severn	West Midlands	England	28
Chippenham	Severn	South West	England	35
Bath	Severn	South West	England	19
Coventry	Severn	West Midlands	England	30
Ebbw Vale	Severn	Wales	Wales	35
Penrith	Solway Tweed	North West	England	29
Carlisle	Solway Tweed	North West	England	29
Heathfield	South East	South East	England	32
Winchester	South East	South East	England	25
Eastbourne	South East	South East	England	22
Shaftesbury	South West	South West	England	30
Taunton	South West	South West	England	34
Goring/Streatley	Thames	South East	England	36
Crowborough	Thames	South East	England	28
Wokingham	Thames	South East	England	36
Aylesbury	Thames	South East	England	26
Crawley	Thames	South East	England	34
Barking/Dagenham	Thames	London	England	36
Luton	Thames	East of England	England	30
Bexley	Thames	London	England	32
Havering	Thames	London	England	30
Southwark	Thames	London	England	36
Brent	Thames	London	England	22
Ealing	Thames	London	England	36
Llangefni	Western Wales	Wales	Wales	24
Bridgend	Western Wales	Wales	Wales	34
			England	1358
			Wales	129
			England & Wales	1487

Source: NERA-Accent

The numbers of interviews by River Basin District (RBD) are shown in Table 6.3.

Table 6.3
Interviews Achieved by River Basin District

River Basin District	Target Number of Interviews	Interviews Achieved
Anglian	120	110
Dee	60	72
Humber	300	304
North West	180	157
Northumbria	60	56
Severn	150	147
Solway Tweed	60	58
South East	90	79
South West	60	64
Thames	360	382
Western Wales	60	58
Total	1,500	1,487

Source: NERA-Accent

Table 6.3 shows that there is a close correspondence between the target number of interviews and the number achieved within each RBD.

6.4 Length of Interview and Interviewer Feedback

The length of the interview averaged 32 minutes with some lasting over an hour.

Table 6.4, Table 6.5 and Table 6.6 present summaries of interviewers' feedback on respondents' understanding of the questionnaire, degree of consideration given to answering the questions, and signs of fatigue.

Table 6.4
Respondent's Understanding of the Questionnaire

Respondent's Understanding of the Questionnaire, Assessed by Interviewers	Frequency	Percentage of Sample (%)
Understood completely	897	60.32
Understood a great deal	376	25.29
Understood somewhat	144	9.68
Understood a little	42	2.82
Did not understand very much	14	0.94
Did not understand at all	9	0.61
Other	5	0.33
Total	1,487	100

Source: NERA

Table 6.5
Respondent's Degree of Consideration Given to Answering the Questionnaire

Respondent's Degree of Consideration Given to Answering the Questionnaire, Assessed by Interviewers	Frequency	Percentage of Sample (%)
Gave questions careful consideration	1,094	73.6
Gave questions some consideration	350	23.5
Gave the questions very little consideration		2.3
Other	9	0.6
Total	1487	100

Source: NERA

Table 6.6
Respondent's Fatigue

Respondent fatigue, Assessed by Interviewers	Frequency	Percentage of Sample (%)
Maintained concentration throughout	1,287	86.6
Lessened concentration in later stages	187	12.6
Other	13	0.87
Total	1,487	100

Source: NERA

Table 6.4 shows that the interviewers judged that the majority of the sample (85%) understood completely (60%) or a great deal (25%). There were a number of respondents (15%) who were thought to have understood only somewhat, not very much or did not understand at all.

Table 6.5 shows that the interviewers considered that the majority of the sample (74%) gave the questionnaire careful consideration. The majority of the remainder (24%) gave the questions some consideration, and a very small number (2%) gave the questions very little consideration.

Table 6.6 shows that the interviewers believed that the majority of the sample (87%) maintained concentration throughout, while 13% lessened concentration in the later stages of the survey interview.

While we are here reliant on the interviewers' judgments of the respondents' understanding, degree of consideration, and fatigue, these judgemental ratings of understanding (or not) are very useful supplements to the indications of understanding (or not) that are available from the survey responses direct. They are particularly helpful in considering the likely validity of the responses.

7 Analysis of Raw Responses

A part of the full sample must be omitted from analysis due to the respondents having given invalid responses, or no responses at all, to the valuation questions, or having given infeasibly large outlying values for the benefits of the WFD improvements. Our approach has been to identify potential groups of respondents for exclusion from the sample, select our preferred combination of groups to exclude, and examine the effect of excluding alternative combinations of these groups on wtp statistics. Out of the full sample of 1,487, our preferred approach leads to a total of 98 respondents being removed, leaving an analysis sample containing 1,389 respondents. The proportion of the full-sample removed following this approach is equal to 6.6%. This number is relatively low in comparison with some similar studies. For example, Mitchell and Carson (1984) identified and removed 30% of the original sample of 813 for reasons of invalid responses. However, other studies have removed a similar proportion to our own and some have removed fewer.

The first part of this section describes our approach to identifying potential groups of respondents for exclusion, and defining our preferred sample, and alternative samples, from these groups for analysis. The second part of the section then examines descriptive statistics from our preferred analysis sample, which we then use for the majority of the remainder of the analysis. The final part concludes on the characteristics of this core sample of useable respondents.

7.1 Identification and Removal of Non-Useable Responses

7.1.1 Item Non-Responses

A total of 5 respondents refused to answer the contingent valuation willingness to pay question, and 18 said “don’t know”. Thus, 23 respondents are therefore automatically excluded from the sample.

There were no respondents that refused to answer the choice experiment questions, including the dichotomous choice question.

7.1.2 Protests

Protests are considered to be choices or willingness to pay amounts that are in fact underestimations[4] of respondents’ actual willingness to pay. These responses are typically a result of a misunderstanding about the scenario specification, a reaction to the payment vehicle, a moral objection to the question being asked or some combination of all of the above. Zero protest votes or amounts do not reflect a respondent’s true willingness to pay and can cause the overall estimate to be too low. Removing respondents where there is strong evidence that their responses are protests, rather than true willingness to pay values, helps to remove a potential bias in the aggregate estimates.

---

[4] The literature varies in its use of the term “protest”; some studies depict protest as including both under and over estimates. For our purposes, we treat protests as underestimates and outliers (see Section 1.1.3) as overestimates.

In line with best practice for stated preference research, the survey questionnaire included open-ended questions following both the choice experiment and the contingent valuation exercises. These questions asked respondents to articulate the reasons for their responses to the choices presented to them and for their stated willingness to pay amount. An analysis of these responses, alongside additional information such as household demographics, attitudes and use of the water environment, provides a good means to discern whether a valuation given is a protest. Unfortunately, there is no perfect way to identify protests in all cases; some responses will provide some evidence that the response is a protest, but this evidence will not be compelling. We classify these responses as "borderline" protests. The subsections below describe how protests have been identified from the choice experiment and the contingent valuation responses. Borderline cases are also identified, but to be conservative, these cases are retained in our preferred sample and are a part of the aggregate analysis.

7.1.2.1 Choice Experiment and Dichotomous Choice Protests

To examine for protest cases within the choices that respondents made, we first screened for those who selected the "no change" option across all eight choice sets, including the dichotomous choice question because they are presented together. We considered these to be candidate cases which warranted a detailed review. In total, there were 127 respondents identified by this screening process. The next step involved examining the respondents' verbatim answers to the follow-on question on their reasons for making the choices they did, and interpreting these responses using selected criteria. These criteria are described below. To be conservative, if a response indicated one or more of the protest criteria, and additional non-protest factors, the respondent was not considered a protest. While such respondents may have made some of the choices based on protests, some choices may have been driven by true valuations. Since we have no way to parcel out the true choices from the protest, we conservatively treat these respondents as non protests.

The first criterion for protests relates to indications that the WFD should be funded by the government, by the water companies or by polluters. Such responses strongly suggest that respondent choices did not reflect their true valuation of the improvements shown in the choice profiles, but rather were a reaction to the perceived fairness of the payment vehicle. These respondents were treated as protesters and removed from the sample. There were 14 such cases in total.

The second criterion relates to statements that current monies are being misspent or that the WFD is not the correct solution. Verbatim answers of this kind imply that the choice experiment responses are driven by either 1) a reaction to the credibility of the scenario, ie the respondent does not believe improvements would occur, or, 2) a reaction to the payment vehicle ie the respondent does not believe he/she should be paying for an improvement scenario when money is already being wasted that could be used for the WFD. Both cases are forms of protests. Therefore, this set, comprising five respondents, was also removed from the sample.

The third criterion is based on statements that the improvements proposed by the WFD will not actually happen or conversely, that improvements will occur without the WFD. These statements reflect a rejection of the scenario described to respondents, wherein they were asked to assume that the water environment will evolve as described in the choice experiment. Thus respondents were asked to believe that if they choose "no change", there would be no change in the water environment and if they chose an improvement option, the water environment would improve exactly as described. This group of verbatim answers are therefore strong evidence that the responses are protests rather than true valuations, and are therefore removed from the sample. There were four such respondents in total.

The fourth, and final, criterion relates to statements that alternative policies, such as education, water meters, etc. should be implemented as opposed to the initiatives detailed in the WFD. These answers suggests that the respondents care about the WFD improvements, but that their responses to the choice experiment reflect protests regarding the perceived delivery mechanisms of the improvements, rather than the value of the improvements themselves. Aspects of delivery mechanisms can, and will, have genuine value for respondents, e.g. digging up roads to fix leaks to improve flow levels in local rivers carries a genuine negative value to road users. However, the survey instrument did not, and could not, provide sufficient details of how the improvements would be achieved to allow respondents to construct a value over the delivery mechanisms independently of the value of the improved water environment. It was intended that respondents would value only the improvements and not the delivery mechanism, and that when the costs and benefits of future projects were to be considered, the non-market costs and benefits of delivery mechanisms would be estimated separately. Verbatim answers which indicate a protest against the perceived delivery mechanisms are evidence that the choices do not reflect the respondents' true values for the options presented and so should be removed from the sample. Only one respondent provided this type of answer, and was removed from the sample.

In total, 24 respondents were characterised as protesters at the choice experiment stage, and are removed from the sample.

In addition to the criteria described above, a number of borderline cases were identified, but not removed from the sample in the interests of taking a conservative approach overall. Additional, borderline, criteria included the following:

- Answers indicating that the government or water company should pay but also indicating that personal budget constraints contributed to the choice; (14 respondents)
- Answers indicating that they currently pay enough for their water; (20 respondents)
- Answers stating that there is no guarantee that the proposed changes will come about; (3 respondents)
- Answers arguing that rates will increase in the future anyway; (1 respondent)

The borderline cases here include a further 29 respondents in total. These have not been removed from the sample.

Table 7.1 below shows the distribution of respondents classified as protests and borderline cases using the above criteria.[5]

[5] It should be noted that some respondents mentioned more than one of the protest criteria in their responses. The codes were assigned on the basis of the primary or first comment in the verbatim response.

Table 7.1
Protest Cases for DCCV / CE Questions

Verbatim Answer Type	Protest Cases Removed from Sample (Frequency)	Borderline Cases Not Removed from Sample (Frequency)
Government/Water Company/Polluter should pay	**14**	
Current monies are misspent	**5**	
Improvements will not actually occur[6]	**4**	
WFD not the solution	**1**	
Pay enough already		20
Government/Water Company/Polluter should pay and individual concerns		5
No guarantees		3
Rates will increase anyway		1
Total	**24**	**29**

Source: NERA

The 24 protest cases removed from the choice experiment analysis represent 1.6% of the full sample.

7.1.2.2 Payment Card Contingent Valuation Protests

For the contingent valuation responses, we first screened for potential protesters by examining only individuals who stated that their willingness to pay was zero. While some small number of respondents offering willingness to pay amounts more than zero had verbatim responses which could have been classified as protests, in the interests of taking a conservative approach overall, we have not included these individuals here.

The full sample contained 164 respondents who gave a value of £0 when asked about their willingness to pay.[7] Following the same criteria as set out above, 48 cases were classified as protests. There were a further 116 classified as borderline protest cases using the additional criteria as described above.

As above, the first criterion relates to respondent statements that the WFD should be funded by the government, by the water companies or by polluters. Respondents such as these are not providing a true valuation, but rather are reacting the perceived fairness (or lack thereof) of the payment vehicle. There were 23 such respondents and they were removed from the total.

The second type of protest response relates to all statements indicating that money is currently misspent or inappropriately handled. Again, these respondents either do not believe the scenario is credible (ie they feel government and water companies' spending in the past has not aided environmental issues) or they believe that the money to fund the project already exists (in tax revenues, water rates or water company profits). There were 16 such respondents removed from the sample for these reasons.

---

6 One of these four respondents stated that the improvements will occur regardless of the WFD programme.

7 This figure excludes the 23 cases with missing WTP values, which are automatically excluded from the analysis sample.

The third criterion is characterised by respondents who do not feel that the improvements will actually occur. These respondents specifically reject the assumptions of the hypothetical scenario; that is respondents were asked to believe that the science would provide the improvements as depicted. One respondent further rejected these assumptions by indicating that the environment would not only not improve, but would get worse. Given that these responses reflect the lack of credibility these respondents perceived in the scenario as presented, the responses are to be considered as protests rather than true valuations, and therefore are removed from the sample. There were six such respondents in total.

As above, a small number of respondents indicated that alternative policies should be implemented as opposed to the initiatives detailed in the WFD. These answers, given by two respondents, suggest that they have some value for WFD improvements, but that they "protest" or object to the delivery mechanisms of the improvements. These two respondents were removed from the sample.

There was one additional respondent who provided a response indicating a protest to the payment mechanism. This respondent stated that she felt that it was "morally wrong" to ask citizens to fund the WFD programmes. This statement indicates that the valuation was focused on the fairness of the vehicle and not the improvements. This respondent was also removed.

As with the choice experiment, a number of borderline cases were identified, but not removed from the sample. These responses included:

- Answers indicating respondents currently pay enough for their water; (60 respondents)
- Answers indicating that the government or water company should pay but also indicating that personal budget constraints contributed to the choice; (5 respondents)
- Answers stating that water in general should be free; (2 respondents)
- An answer arguing that there are no guarantees that the changes will come about; (1 respondent)

These 68 respondents have not been removed from the sample.

Table 7.2 below summarises the distribution of protest cases and borderline cases for the contingent valuation question.

Table 7.2
Protest Cases for Payment Card CV Question

Protest Type	Protest Cases Removed from Sample (Frequency)	Borderline Cases Not Removed from Sample (Frequency)
Government/Water Company/Polluter should pay	23	
Current monies are misspent	16	
Improvements will not actually occur[8]	6	
WFD not the solution	2	
Morally wrong to ask for payment	1	
Pay enough already	--	60
Government/Water Company/Polluter should pay – and individual concerns	--	5
Water should be free	--	2
No guarantees		1
Total	**48**	**68**

Source: NERA

The 48 protest cases removed from the PCCV analysis represent 3.2% of the full sample.

7.1.2.3 Summary and Comparison of CE/DCCV and PCCV Protests

As would be expected, many of the respondents characterised as protest cases for the choice experiment were similarly characterised in the contingent valuation section. Table 7.3 below summarises the numbers of respondents characterised as protest cases for the CE questions and for the CV question. The table also shows the total number of protest cases, taking into account the fact that some respondents are characterised as protest cases for both sets of questions. In other words, if a respondent is characterized as a protest in the choice experiment and is again a protest in the contingent valuation, he or she is only counted one time in the total shown in the table.

Table 7.3
Summary of Protest Cases for DCCV / CE and PCCV

	Protest Cases Removed from Sample (Frequency)	Borderline Cases Not Removed from Sample (Frequency)
Choice Experiment and DCCV protest cases	24	29
Payment Card CV protest cases	48	68
Total no. of protest cases in sample	58	71

Source: NERA

Table 7.3 shows that there were 58 respondents in total coded as protests, either for the CE section, the CV section or both. These respondents have been removed from the sample. The borderline cases totalled 71 such respondents. These were retained in the sample.

[8] One respondent indicated that the changes would be irrelevant because the environment was going to deteriorate with increased population growth.

The literature on stated preference design supports our finding that protests more frequently occur in response to contingent valuation questions than to choice experiments (see Bateman et al 2002). This is likely due to the fact that the CV method asks for direct payment and thus emphasises the monetary contribution, whereas the CE method asks for a preference rating of which the financial contribution is just one of many attributes.

The number of cases removed due to protests for the DCCV / CE section, the PCCV section or both (58) is equal to 3.9% of the full sample.

This share of protests is reasonable in light of past research; for example, Mitchell and Carson (1984) had 16 percent of the total sample giving protest responses to their contingent valuation questions. In their design, they attempted to "convert" protest respondents and encouraged these individuals to relate their actual willingness to pay. The authors report that their efforts to do so were not particularly successful. More recent studies recognise the difficulties associated with attempts to "force" respondents to reveal their true willingness to pay and instead these studies, once systematically identifying the protest, remove them from the analyses.[9] Huber and Viscuzi (2006) remove five percent of their total sample for "logically inconsistent results" (pg. 6). The respondents in this study who were excluded indicated that they would prefer the dominated choice; that is the respondents indicated a preference for the option that was equal on one level of the good but worse on the other. The authors determined that these responses were economically irrational and were therefore removed from their study. While a number of studies reveal protest values between 3 and 10 percent, a few report much higher rates. A choice experiment study done by Mourato et al (2003) on the UK Bathing Water Directive yielded protest responses equal to approximately 21 percent of the overall sample. These responses were removed. Similarly, a study done by Georgiou et al (2000) identified 35 percent of the total sample as protests, which the authors point out may have been related to the controversy surrounding the privatisation of the water industry in the 1980s (page25).

7.1.3 Outlier Analysis

Just as protests can cause a model to underestimate the benefits of a policy, very high values can skew the estimation and lead to an overestimation of the total benefits. Unfeasibly large outliers in this direction can occur for a number of reasons; respondents can feel as though they are "sending a message" about the importance of the good, respondents can be attempting to provide a socially desirable response or respondents could be forgetting other demands on the household income. Some very high amounts are also likely to be genuine valuations. Unfortunately, there is no perfect way to identify which outliers are genuine high valuations, and which are overestimates. As an approximation, and to help ensure our valuations are conservative, in our preferred sample we have removed respondents from our preferred analysis sample that are in the top 1% of the distribution of PCCV responses. This corresponds to all responses greater than or equal to £350 per household per year, and totals 17 responses that are removed from our preferred analysis sample.

---

9 Mitchell and Carson found a correlation between income level and protests and therefore the removal of any group of respondents must also consider whether recalibration of the weights will be necessary so that the demographics of the sample remain representative.

In subsequent analysis, we investigate an alternative approach to excluding outliers in order to test the sensitivity of the results to our methodology. In this alternative approach, only two outliers are excluded. These are cases where respondents stated they would be willing to pay £1,000 to see improvements. For one of these cases, the interviewer indicated that the respondent, "…was very argumentative and denigrated the whole survey repeatedly, a most difficult interview to conduct. Respondent does not want to be contacted again,". The second respondent indicating a wtp of £1,000 was younger (aged 25 to 29), was not employed nor had an employed partner, would not state income and was on a war pension. While this individual endorsed many pro-environment attitudes, his use of the water environment was somewhat limited and given the lack of steady income it is most unlikely that the £1,000 is a realistic figure.

7.1.4 Summary on Protests and Outliers

There were 23 item non-responses to the PCCV question, including 8 refused, and 15 don't know responses. There were no such responses for the CE or the DC questions.

There were 58 protests in total, including 24 protest respondents identified for the CE/DCCV and 48 for the PCCV question. Some respondents (14) were identified as protesters for both parts. As a sensitivity check, we also identified a further 71 cases that could reasonably be considered as outliers, but which are treated as borderline cases and included in the preferred sample.

A total of 17 respondents, equal to the top 1% of the wtp distribution in the full sample, were removed as outliers from our preferred sample. To use as a sensitivity check, we also proposed an approach which excludes only the two cases where respondents gave £1000 as their PCCV wtp amount.

Table 7.4 presents a summary of the groups identified as potential candidates for exclusion from the analysis sample. Table 7.5 then defines our preferred analysis sample, and comparator samples for sensitivity checking, that we have created by excluding combinations of these groups.

Table 7.4
Potential Exclusion Groups

Variable Name	Frequency	Definition
item	23	Item non-responses to PCCV question (5 refusals and 18 "don't know")
ce_pts1	24	Core CE protests
ce_pts2	29	Borderline CE protests
cv_pts1	48	Core PCCV protests
cv_pts2	68	Borderline PCCV protests
total_pts1	58	Total Core protests (CE+PCCV)
total_pts2	71	Total Borderline protests (CE+PCCV)
outlier1	2	PCCV WTP≥ £1000
outlier2	17	PCCV WTP≥ £350
inconsistent	402	PCCV<DC_bill and DC_choice=1 - or - PCCV>DC_bill and DC_choice=0

Source: NERA

Table 7.5
Samples for Comparison

Sample Name	Frequency	Definition
pref	1,389	Preferred Sample: excludes *item*, *total_pts1* and *outlier2*
trim1	1,462	Exclude *item* and *outlier1*
trim3	1,319	Exclude *item*, *total_pts1*, *total_pts2*, and *outlier2*
pref_cons	1,006	Exclude *item*, *total_pts1*, *outlier2 inconsistent*

Source: NERA

Our preferred core analysis sample contains 1,389 respondents out of the full sample of 1,487. A total of 98 respondents are thus removed in this approach due to giving invalid protest responses, or no responses at all, to the valuation questions, or infeasibly large outlying values for the benefits of the WFD improvements. This amounts to 6.6% of the full-sample. This number is relatively low in comparison with some similar studies. For example, Mitchell and Carson (1984) identified and removed 30% of the original sample of 813 for reasons of invalid responses. However, other studies have removed a similar proportion to our own and some have removed fewer.

7.2 Characteristics of Excluded Respondents

The following tables show the proportions of the full sample excluded from the *pref* sample by sex, age, employment status, household income, and level of education.

Table 7.6
Proportion of Sample Excluded by Sex

Respondent's Sex	Proportion of Category Excluded from Sample (%)
Male	7.8
Female	5.6

Source: NERA

A slightly greater proportion of males are excluded than females.

Table 7.7
Proportion of Sample Excluded by Age

Respondent's Age	Proportion of Category Excluded from Sample (%)
18-29	2.9
30-64	6.7
65+	8.2

Source: NERA

Younger respondents are less likely to have been removed from the *pref* sample than older respondents.

Table 7.8
Proportion of Sample Excluded by Employment Status

Respondent's Employment Status	Proportion of Category Excluded from Sample (%)
Working	4.1
Not Working	8.8
Student	13.3
Other	0.0

Source: NERA

A large proportion of students are removed from the sample (13.3%). More non-workers are removed from the *pref* sample than working people.

Table 7.9
Proportion of Sample Excluded by Total Weekly Income

Respondent's Income	Proportion of Category Excluded from Sample (%)
Low (<£300 per week)	5.6
Med	4.1
High (>£1000 per week)	3.1

Source: NERA

Those on higher incomes are less likely to be removed from the *pref* sample than respondents with low household incomes.

Table 7.10
Proportion of Sample Excluded by Level of Education

Respondent's Level of Education	Proportion of Category Excluded from Sample (%)
Primary	4.9
O levels, GCSE or CSE (1+ passes), NVQ Level 1 or foundation level GNVQ	6.3
5+ O, CSE grade 1's or GCSE A-C; School certificate; 1+ A or As; NVQ level 2 or intermediate GNVQ	7.8
2+ A levels; 4+ As; Higher School Certificate; NVQ Level 3; or Advanced GNVQ	6.4
First Degree, Higher degree, NVQ Level 4/5; HNC; HND; Qualified teacher status; Qualified medical doctor, dentist; nurse; midwife; health visitor	6.2
Other / Don't know	13.6
Refused	11.8

Source: NERA

An even mix of respondents are removed with respect to education level.

7.3 Conclusions on Characteristics of Core Sample of Useable Responses

Following removal of the 98 respondents from the full sample, the full set of demographic characteristics were re-examined based on the *pref* core analysis sample.

We consider that the sample of useable responses provides an adequate representation of the national population, noting that there are fewer men than women, and that there are more low income households and fewer high income households. Accordingly, unweighted sample descriptive statistics regarding values of the water environment, and opinions, attitudes and uses of the water environment, are likely to adequately represent those of the national population. Sub-populations, for example at RBD level, are less well represented, however our analysis does not rely on making inferences from these sub-samples. When deriving aggregate values for the benefits to the population, and to sub-populations, we use a weighting procedure to account for demographic differences between the sample of usable responses and population counterparts.

7.4 Inconsistent Responses

We have also analysed at length inconsistencies between responses to the PCCV, DCCV and CE questions. This analysis, as discussed in Appendix J, has shown there to be a large number of cases where PCCV amounts are on the face of it "too low" given accepted DCCV and / or CE choices, and a small number of cases where PCCV results are "too high" given a rejected DCCV scenario. Our preferred sample retains the inconsistent cases. As a sensitivity check, we compare some of our results from the preferred sample against those from a sample which excludes respondents with inconsistent PCCV and DCCV responses (the *pref_cons* sample).

8 Results on Attitudes, Opinions and Uses of the Water Environment

This section presents descriptive results from the *pref* sample on respondents' attitudes, opinions and uses of the water environment. Omnibus survey results are also reported for the three questions in the 4c survey that were also included in the Omnibus survey carried out at the same time.

Table 8.1 presents results on respondents' opinions regarding the level of government spending on public services. The following table presents the corresponding results from the Omnibus survey.

Table 8.1
Opinions on Levels of National Spending on Public Services

	Pensions (%)	Social Security (%)	Education (%)	Water Pollution (%)	Transport (%)	Police & Criminal Justice (%)	Air pollution (%)	Health
Too much	2.7	29.5	3.5	1.8	9.7	7.9	3.5	3.0
About right	19.1	38.5	35.5	32.8	35.0	40.0	29.6	28.2
Too little	74.3	22.8	57.4	52.7	52.6	49.5	58.8	67.7
Don't know	4.0	9.1	3.5	12.7	2.7	2.6	8.2	1.1
Refused			0.1			0.1		1
N	1389	1389	1389	1389	1389	1389	1389	1389

Source: NERA analysis of 4c survey data

Table 8.2
Opinions on Levels of National Spending on Public Services – Results from Omnibus (2007) Survey

	Pensions (%)	Social Security (%)	Education (%)	Water Pollution (%)	Transport (%)	Police & Criminal Justice (%)	Air pollution (%)	Health
Too much	1.54	39.93	4.28	1.71	7.8	6.52	2.57	5.1
About right	20.74	26.22	34.36	31.02	26.48	24.53	22.90	22.6
Too little	65.98	15.77	53.56	29.82	53.73	59.18	43.40	66.6
Don't know	11.75	18.08	7.80	37.45	12.0	9.78	31.13	5.8
N	1167	1167	1167	1167	1167	1166	1166	1167

Source: NERA analysis of ONS (2007) Omnibus survey data

When asked about current levels of national spending on public services (Table 8.3), most people responded with "too little" or "about right" for all spending categories. The single exception is social security for which there were more people saying "too much" than "too little". Water pollution spending was regarded as "too much" (1.8%) by the fewest people of any spending category. By contrast, 52.7% responded that water pollution spending was "too little". More responded "don't know" re water pollution spending (12.7%) than on any other category.

In comparison with the Omnibus survey results, there is generally a close match. There is a larger "don't know" proportion of responses to the Omnibus survey. However, the results do not indicate any problems with the 4c survey data.

Table 8.3 presents results on respondents' priorities for water company spending. The following table presents the corresponding results from the Omnibus survey.

Table 8.3
Opinions on Water Companies' Expenditure Priorities

	1st Priority (%)	2nd Priority (%)
Reducing the risks of drinking water discolouration	20.1	12.5
Reducing the risks of sewer flooding	30.7	28.9
Fixing leaks in pipes	34.0	22.8
Improving river water quality	5.7	10.7
Reducing the risks of interruptions to supply	3.7	8.0
Improving bathing water quality	0.8	2.5
Protecting animal and plant life around waterways	5.0	14.3
Don't know	0.1	0.3
Refused	-	-
N	1,389	1,389

Source: NERA analysis of 4c survey data

Table 8.4
Opinions on Water Companies' Expenditure Priorities - Results from Omnibus (2007) Survey

	1st Priority (%)	2nd Priority (%)
Reducing the risks of drinking water discolouration	11.58	6.25
Reducing the risks of sewer flooding	26.76	34.18
Fixing leaks in pipes	44.25	22.46
Improving river water quality	9.69	16.60
Reducing the risks of interruptions to supply	2.74	7.32
Improving bathing water quality	.26	1.17
Protecting animal and plant life around waterways	2.57	12.01
Don't know	2.14	0.0
Refused	-	
N	1166	1024

Source: NERA analysis of ONS (2007) Omnibus survey data

Turning to water company expenditure (Table 8.4), the top priorities were most frequently fixing leaks, and sewer flooding, neither of which is directly linked to WFD outcomes. Protection of habitats around waterways, and improving river water quality were second priorities for some respondents, but improving bathing water quality was not seen as a priority by many at all.

In comparison with the Omnibus survey results, there is again generally a close match in terms of priority ordering. There are more respondents in the Omnibus survey that consider improvements to river water quality as the top or second highest priority for water company spending, however, there are more in the 4c survey that protecting animal life around waterways to be one of the top two priorities. Overall, the results do not indicate any problems with the 4c survey data.

Table 8.5 presents results from both the 4c survey and the Omnibus survey on respondents' opinions regarding the level of national spending on environmental protection and improvement.

Table 8.5
Attitudes towards Paying for Environmental Protection and Improvement

	4c (%)	Omnibus (%)
Protecting the environments so important that pollution control requirements and standards cannot be too strict, and continue improvements must be made regardless of cost	46.2	30.2
Protecting the environment is important and continue improvements should be funded, provided that they are not excessively cost	39.4	52.6
We are spending about the right amount on cleaning up the environment and don't need to increase or decrease this spending	6.8	7.1
We have made enough progress on cleaning up the environment and we should now concentrate on holding down costs rather than requiring stricter pollution controls	4.3	3.6
Pollution control requirements and standards have gone too far and they already cost more than they are worth	3.0	1.2
Don't Know	0.4	5.2
N	1389	1148

Source: NERA

The 4c sample showed a strong preference towards paying for improving the environment. There were 46.2% who said environmental improvements should continue regardless of cost and 39.4% who said they should continue if not excessively costly. Few thought that, as a nation, we are spending the right amount or too much. Results from the Omnibus survey are similar in terms of the proportion answering in these first two categories, there are 85.6% in the 4c survey and 82.8% in the Omnibus survey. In comparison with the Omnibus results there are more respondents in the top category and fewer in the second category, however overall, the results do not indicate any problems with the 4c survey data.

Table 8.6 presents results from the 4c survey on uses of the water environment.

Table 8.6
Uses of the Water Environment

	Water Contact Activities (%)	Fishing (%)	Other Activities (%)
Often	10.4	5.3	40.5
Sometimes	14.1	4.3	28.8
Rarely	19.9	6.8	11.7
Never	55.6	83.5	19.0
N	1389	1389	1389

Source: NERA

There are a good number of users of different types in the sample, as well as many non-users.

Table 8.7
Opinions on Most Important WFD Benefits

Most Important WFD Benefit	Proportion of Sample (%)
Direct Use Benefits	
Improved conditions for fishing	2.7
Improved conditions for contact activities	23.0
Improved conditions for other activities	20.7
All Use Benefits	46.4
Non-use and Perhaps Use Benefits	
Knowledge of improved habitats for plants, animals and fish	49.2
Other	1.9
Equal	1.9
None/Don't know	0.6
N	1389

Source: NERA

An approximately equal number of respondents said that improved conditions for direct use benefits were the most important, to those that said that knowledge of improved habitats benefits were the most important.

Table 8.8
Attitudes towards WFD Improvement Priorities

Preferred Order of Improvement	Proportion of Sample (%)
Worst areas, judged by scientists	36.4
Near where people live	44.2
Quick + low cost	18.9
Don't know	0.4
Not Stated	0.1
N	1389

Source: NERA

The most popular order of improvements amongst respondents was for those areas near where people live to be improved first. Quick, low cost solutions were not preferred.

Table 8.9
Reactions to Local and National Water Environment Quality Maps

Reactions to Water Environment Quality Maps, vs Expectations	Local Map vs. Expected Status	National Map vs. Expected Status
Great deal better than expected	3.7	1.2
Somewhat better than expected	10.2	6.7
As expected	29.4	34.8
Somewhat worse than expected	30.8	35.3
Great deal worse than expected	25	21.0
Don't know	0.9	1.1
N	1389	1389

Source: NERA

The majority of the sample thought that the local and national maps were showing them a worse picture than they expected (Table 8.9). Verbatim responses giving reasons for respondents' reactions are presented in Appendix L.

9 Approach to Analysis

9.1 Objectives

The objectives of our analysis were:

1. to derive key summary statistics showing elemental findings from the sample obtained
2. to contribute to an assessment of the validity of PCCV, DCCV and CE responses by analysing the variance of responses in light of prior expectations;
3. to conclude on a suitable model, or models, for aggregation of the WFD benefits valuation results to the populations of England, Wales, and England and Wales, where the populations differ from the sample in some respects.
4. to assess the sensitivity of the WFD benefits valuation results to alternative elicitation methods, estimation approaches, and inclusion/exclusion of protests, outliers and inconsistent responses.

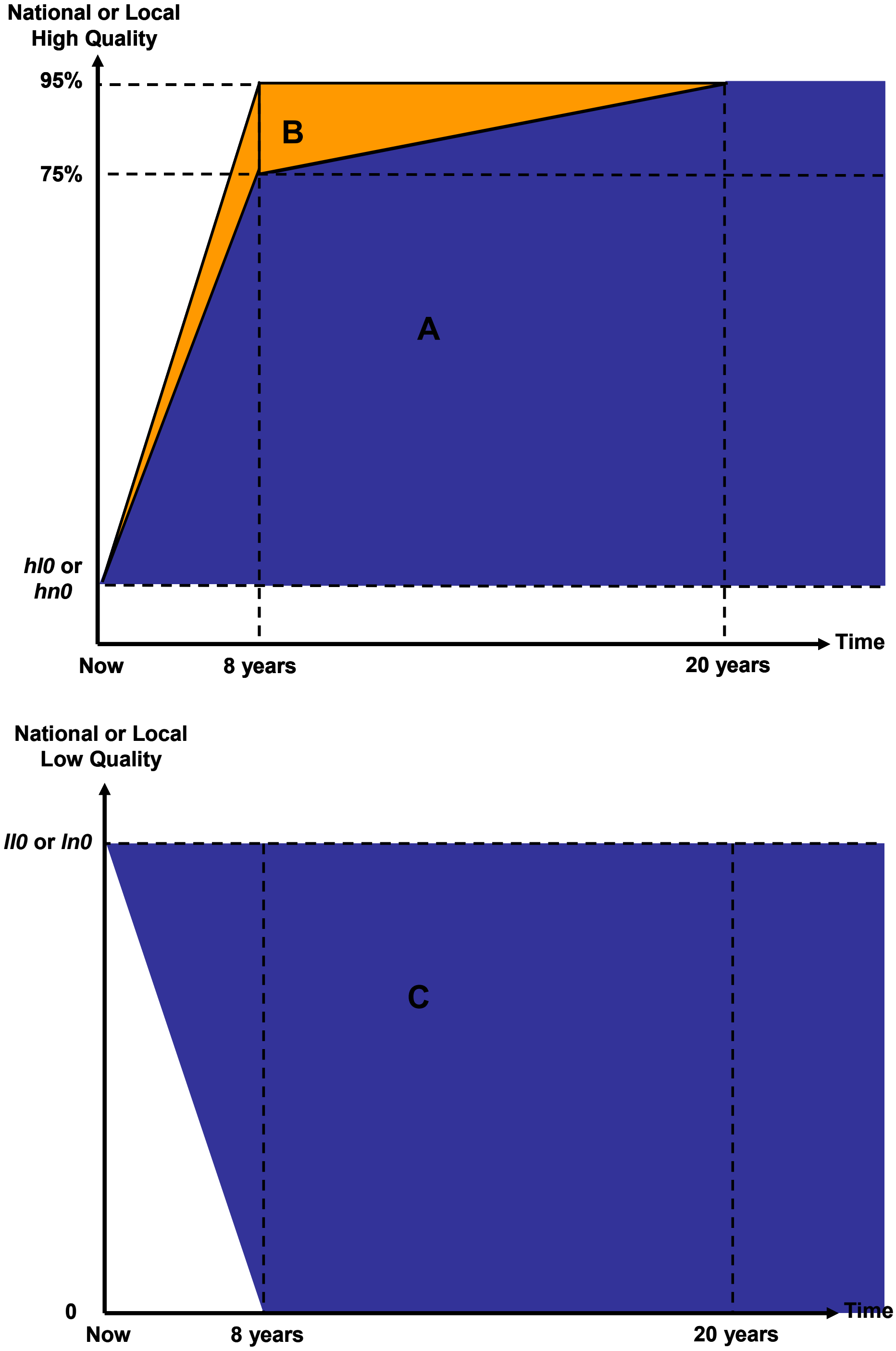
9.2 The Improvements to be Valued

9.2.1 The Improvements to be Valued

CV Valuation of Environmental Improvements

Figure 9.1 illustrates the CV scenario in respect of the scope of environmental improvements. From the baseline level of environmental quality at time Now (i.e. the proportion *hl0* of high quality local sites, *hn0* of high quality national sites, *ll0* of low quality local sites, *ln0* of low quality national sites), the quality improves through an increase in the high quality proportion and a decrease in the low quality proportion over time.

Figure 9.1
Environmental Improvements Valued in the CV Scenarios



The PCCV and DCCV responses to the "75% scenario" are interpreted as values for the scope of improvement indicated by areas **A** (local+national) + **C** (local+national) in Figure 9.1, while responses to the "95% scenario" are interpreted as values for the area indicated by **A** (local+national) + **B** (local+national) + **C** (local+national).

The national-level environmental improvements are the same for all respondents in the sample, except that half are offered the 95% scenario and half are offered the 75% scenario.

The local improvements offered vary across respondents according to the current status of their local area, as well as by whether they are offered the 95% or the 75% scenario.

Differences across respondents in local conditions might potentially allow sufficient variation to estimate separate values for local and national parameters for the values of each of **A**, **B** and **C**, however identification is only obtained if: (i) there is a substantial difference between treatments offered, and, (ii) the treatments are randomly allocated to respondents. In the present case, neither of these conditions is likely to hold. The variation in treatments offered is small relative to the similarities, and is probably in a region where declining marginal utilities apply, so the variation in values is also likely to be small, making it difficult to identify without a very large sample. Also, in the case of local initial conditions, the treatments are not randomly allocated since, for example, people with higher values of the water environment may choose live in currently good local water environments, or, living in currently good water environments may influence peoples' values for improvements. Because of these issues, we did not anticipate being able to obtain the population's separate values for local, national, high quality and low quality changes from the CV scenarios.

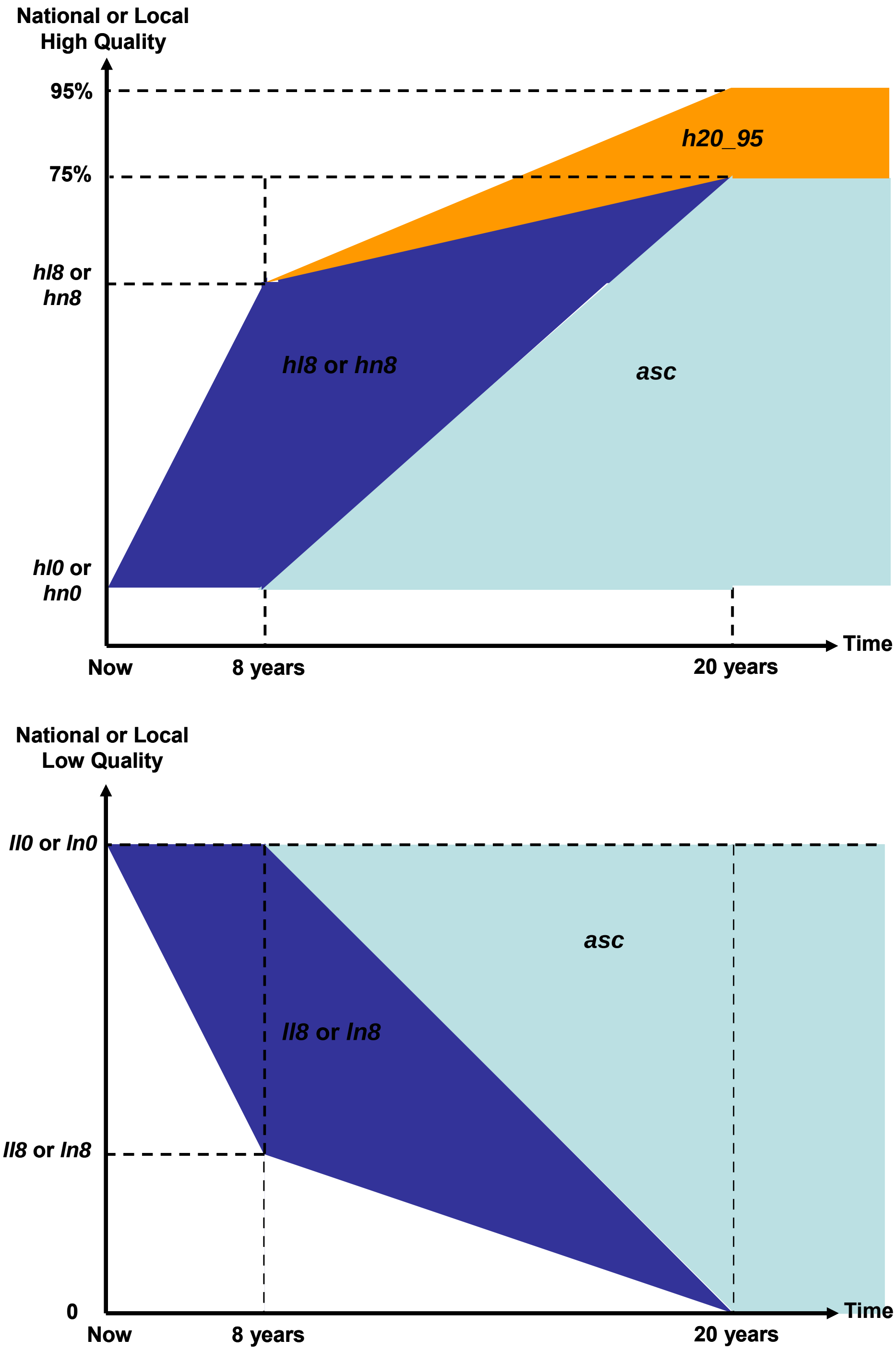
The split sample approach (95% or 75%) with respect to environmental improvement scenario offered has sufficient randomised variation to check whether respondents are sensitive to scope however. We view this as a test on the validity of the instrument, in line with recommendations by, eg Arrow et al (1993), rather than as a precise way of valuing marginal changes in environmental quality. A finer degree of precision was expected to be obtained from the CE analysis, which contains multiple randomised treatments per respondent and thus is less affected by these problems.

Given the design adopted for the study, the dataset was expected to contain a rich source of information on preferences, leaving substantial scope for alternative explanatory models to be constructed. Our approach to analysing the CV data is described in following sections. The aggregation to form national results from the CV household results depends on the final model adopted.

9.2.2 CE Valuation of Environmental Improvements

Figure 9.2 illustrates the way that the environmental change attributes are related to the CE parameters.

Figure 9.2
Environmental Improvements Related to CE Parameters



The main parameters to be estimated from the CE responses are *asc*, *h20_95*, *hl8*, *hn8*, *ll8* and *ln8*. These variables, except for *asc* and *h20_95*, correspond directly to the variables included in the CE design. The variable *asc* is an alternative specific constant which captures differences between the "no-change" option and all of the change options as a group. The size of the area indicated by *asc* varies a little by location as it depends on *ll0* and *hl0*. The variable *h20_95* is a {1,0} dummy variable indicating, if equal to 1, that the option leads to 95% high quality by 2027. It is equal to 0 otherwise, indicating 75% improvement by 2027.

The size of the area captured by this variable does not depend on initial water environment quality conditions or 2015 quality conditions. The size of the effects of *hl8*, *hn8*, *ll8* and *ln8* do depend on the relevant initial conditions *hl0*, *hn0*, *ll0* and *ln0*.

The CE design was expected to provide richer information than the CV on preferences, with substantial scope for alternative models to be constructed. Our approach to analysing the CE data is described below. The form of aggregation to national estimates from the CE design depends on the final model adopted. Depending on the specification of the econometric model, the CE design is also potentially able to derive an implied discount rate from the observed choice behaviour due to the fact that there is variation in the 20-year water environment high quality levels.

9.3 PCCV Analysis Approach

For the PCCV analysis, the key summary statistics derived include the mean, median, and lower and upper quartiles from the distribution of willingness to pay. We also present a table and chart showing the full cumulative distribution of willingness to pay responses

To contribute to an assessment of the validity of PCCV responses, we first examined descriptive cross-tabulations of sample wtp against a large number of variables in the dataset to discover which variables appear to be strongly correlated with wtp. Key tables from this descriptive analysis are reported below; detailed supplementary tables are provided in Appendix L.

We then conducted our econometric analysis as follows: first, we began with a combination of explanatory variables selected from the descriptive results, by picking the seemingly most important variables, and estimated two models using Ordinary Least Squares (OLS) regressions. The first regression was run with *wtp* as the dependent variable; the second with ln_wtp as the dependent variable, where *ln_wtp* is the natural log of (1+*wtp*).10 Our findings from this analysis, reported as supplementary tables in Appendix L, showed that the model with *ln_wtp* as the dependent variable fit the data substantially better than the model with wtp as the dependent variable. Subsequent analysis therefore focussed on models with *ln_wtp* as the dependent variable.

Then, for each of eight different groups of variables, holding the functional form of all the seven other groups constant, we checked a range of models using OLS.

- Group 1 - Environmental changes (level of high quality / low quality, at the national and local level);
- Group 2 - Affordability (income, adjusted for household composition, benefits, etc);
- Group 3 - Use of the environment (contact activities, fishing, etc…);

[10] The choice of log(1+WTP) as the dependent variable, rather than log(WTP), was due to the fact that there were many zero WTP responses in the sample and log(0) is undefined. By contrast, log(1+0)=0. Estimation results using log(1+WTP) as the dependent variable are interpreted in the same way as those using log(WTP), but there are many more observations available for estimation under the adopted approach.

- Group 4 - Attitudes towards the environment (opinion on pollution control, membership to an environmental club, spending priorities, reactions to the maps, most important WFD benefit, etc);
- Group 5 - Other demographics, e.g. sex, age, education, SEC;
- Group 6 - Geography - RBD, country, urban/rural;
- Group 7 - Survey instrument and interview features: ordering of the questionnaire, water body type example, interviewer gender;
- Group 8 - Comprehension of questionnaire: e.g. understanding, concentration, fatigue.

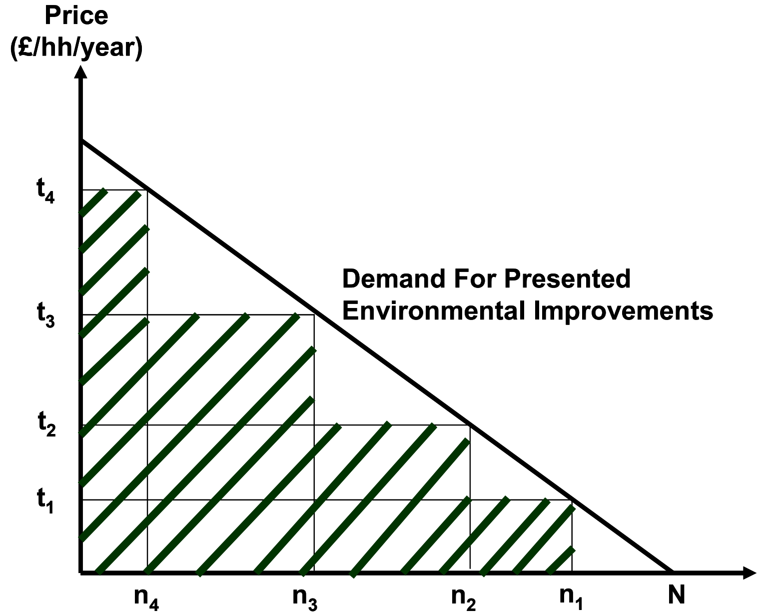
A summary of the findings from this analysis, including a table showing our final PCCV model for aggregation to the population, is reported below.

Further sensitivities examined include sensitivity to sample selection and exclusions. In respect of exclusions, we estimated the final model for aggregation on the *trim1, trim3* and *pref_cons* samples, and compared the results against those from the *pref* sample. In addition, following the draft version of this report, CC Water raised the hypothesis that wtp results for England and Wales from the survey may have been significantly lower had the sample included residents of the South West Water Company area instead of the respondents we interviewed at the sampling points which were randomly selected for the South West RBD. Those South West RBD sampling points were all outside the region served by South West Water, which has the highest average household water and sewerage bills in England and Wales. This hypothesis has been explored in two ways. First, we tested the sensitivity of wtp to the average water and sewerage bill in respondents' company areas for the original sample of responses, by adding the average bill for the region as a new explanatory factor and revisiting the econometric modelling to form a co-efficient for this factor. Secondly, we compared the England and Wales mean wtp for the original sample against the England and Wales mean wtp derived by modelling with each of two amended datasets. The two amendments were as follows: first, we replaced the wtp amounts for all South West RBD respondents with zero, and second, we replaced the wtp amounts for all South West RBD respondents with half the England and Wales mean wtp. The findings from these sensitivity tests are contained in Appendix M to this report.

9.3.1 DCCV Analysis Approach

For the DCCV analysis, the first summary statistics we derive are the proportions of respondents accepting the offered DCCV improvement scenario by the cost amount presented in the scenario. We then derive lower bounds of mean DCCV wtp using the Turnbull non-parametric method. This method involves imposing a monotonicity restriction and then calculating a lower bound on wtp by assuming that the proportion willing to pay a given amount is the proportion of the sample that accepted that amount if/when it was offered to them in the DC scenario. The figure below illustrates the calculation of the Turnbull lower bound estimate.

Figure 9.3
Illustration of Turnbull Estimator for the Lower Bound on Mean DCCV WTP



In this figure, the demand for the environmental improvements presented to respondents in the CV scenario is represented as a monotonically decreasing function with respect to the price of the improvements, ie a downward sloping line. The area underneath this function is equal to the total willingness to pay of the population for the improvements. This function is unknown. Only the proportions of the sample that accept each of the six bill amounts included in the experimental design are known to the analyst. The Turnbull lower bound estimator of the mean wtp for the population proceeds by summing the shaded area in the figure. So, in this figure, the estimate is equal to:

$$(1/N) \times [n_4 t_4 + (n_3 - n_4) \times t_3 + (n_2 - n_3) \times t_2 + (n_1 - n_2) \times t_1]$$

The advantage of the Turnbull analysis is that it does not impose any restrictions on the distribution of preferences in the sample. However, it is relatively weak, in comparison with parametric regression methods, in identifying the effects of multiple explanatory variables on DCCV wtp.

To contribute to an assessment of the validity of DCCV responses, we first examine some descriptive cross-tabulations of Turnbull lower bound mean DCCV wtp. We then estimate numbers of logit models to investigate the determinants of DCCV choices. The logit analysis is complementary to the Turnbull analysis, in that it is efficient at estimating multivariate influences on DCCV wtp, but at the cost of imposing restrictions on the distribution of preferences in the sample, eg that the mean is equal to the median wtp.

The logit analysis of DCCV responses was conducted in a similar manner to the PCCV analysis. For each of eight variable groups, holding the functional form of all the seven other groups constant, we checked a range of plausible logit models to identify the most appropriate functional form.

A summary of the findings from this analysis, including a table showing our final DCCV model for aggregation to the population, is reported below. The table contains results for alternative samples *trim1* and *trim3* to serve as a sensitivity check on the main results from the *pref* sample.

9.3.2 CE Analysis Approach

The CE analysis has similar aims to that of the PCCV and DCCV analysis: to assess the validity of responses by examining the degree to which choices made imply a model of preferences that is consistent with expectations, and to derive a model which informs valuation. Specific to the CE exercise is an analysis of choice behaviour to attempt to understand whether there any lexicographic-type patterns observed in CE response behaviour, for example, examining whether there are numbers of people that always choose the most expensive, cheapest, or middle cost option in each card. Also specific to the CE work is more attempts to separately identify the timed and local relative values of improvements. There are multiple observations per person in the CE analysis, and a good degree of experimental variation, so the CE responses provide by far the richest information about relative preferences of any of the three methods.

Our analysis of the CE data has proceeded as follows:

- first, we produce descriptive statistics on the patterns of choices made by respondents; and
- then, we estimate econometric models using logit and mixed logit estimation methods, testing across variable groups to identify which variables should be included or excluded, and which functional form the included variables should take.

The majority of the analysis has used the *pref* sample, however we have also examined the sensitivity of our adopted model to the effect of alternative protest/outlier exclusion approaches.

10 WFD Valuation Results

10.1 Payment Card Contingent Valuation Results

This section reports the main results from our descriptive and econometric analyses of the PCCV wtp response data. Supplementary results are contained in Appendix L.

10.1.1 Descriptive PCCV Results

Table 10.1 presents the distribution of PCCV wtp amounts for the *pref* sample. The wtp amounts presented are the final values offered by respondents following revision by those who chose to do so, and includes those offered the 75% high quality scenario and those offered the 95% high quality scenario.

The mean PCCV wtp value for the sample is £50 per household per year; and the median PCCV wtp value for the sample is £30 per household per year. The lower quartile (25^{th} percentile) amount is £15 per household per year and the upper quartile amount (75^{th} percentile) is £60 per household per year.

Figure 10.1, beneath Table 10.1, plots the reverse cumulative distribution of PCCV wtp amounts. The plotted curve is interpreted as showing the proportion of respondents in the analysis sample, stating willing to pay amounts up to and including the amount on the horizontal axis, measured in pounds per household per year. For example, the table and the chart show that 86% of the sample are willing to pay up to £10 per household per year, and 20% of the sample are willing to pay up to £100 per household per year.

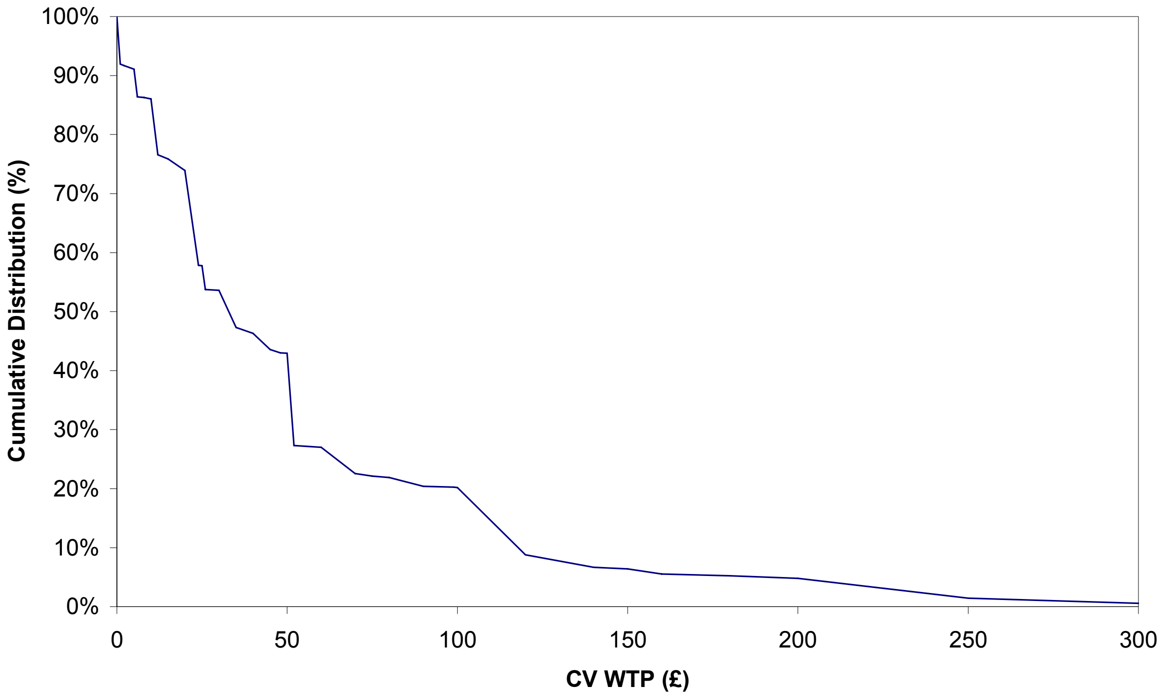
The reasons people gave for their wtp responses are reported verbatim in Appendix L. Notably, given that many areas of England and Wales had suffered severe flooding at the time of the survey fieldwork, only one respondent mentioned flooding as a (partial) reason for their wtp response.

Table 10.1
Payment Card CV WTP Distribution

Payment Card CV WTP (£/hh/yr)	Frequency	Cum. Dist (%)
0	112	100.0
1	3	91.9
2	9	91.7
5	65	91.1
6	1	86.4
8	4	86.3
10	131	86.0
12	10	76.6
15	27	75.9
20	223	73.9
24	1	57.9
25	56	57.8
26	2	53.8
29	1	53.6
30	87	53.6
35	14	47.3
40	38	46.3
45	8	43.6
48	1	43.0
50	217	42.9
52	4	27.3
60	62	27.0
70	6	22.5
75	3	22.1
80	21	21.9
90	2	20.4
99	1	20.2
100	158	20.2
120	29	8.8
140	4	6.7
150	12	6.4
160	4	5.5
180	6	5.3
200	47	4.8
250	12	1.4
300	8	0.6
No. respondents	1,389	
Median WTP (£/hh/yr)	30.0	
Mean WTP(£/hh/yr)	49.5	
Std dev	53.0	

Source: NERA

Figure 10.1
Reverse Cumulative Distribution of PCCV WTP



Source: NERA

Table 10.2 presents mean and median PCCV wtp amounts by income group.

Table 10.2
Payment Card CV WTP by Total Weekly Household Income

Total Weekly Household Income	Mean WTP (£/hh/yr)	Median WTP (£/hh/yr)	Sample Size (Freq)
Low (<£300)	38.6	20	419
Medium	58.6	50	463
High (>£1,000)	72.3	50	127
Missing Income Data	42.7	25.5	380

Source: NERA

As expected, PCCV wtp increases with total weekly household income.

Table 10.3
Testing Average WTP across Income Groups

H_0	H_1	t-stat	p-value
$WTP_{\_LOW} - WTP_{\_MEDIUM} = 0$	$WTP_{\_LOW} - WTP_{\_MEDIUM} < 0$	-5.57	0.0000
$WTP_{\_MEDIUM} - WTP_{\_HIGH} = 0$	$WTP_{\_MEDIUM} - WTP_{\_HIGH} < 0$	-2.45	0.0075

Source: NERA
Note: t-test performed with unequal variances

Table 10.4 presents mean and median PCCV wtp amounts for users and non-users of the water environment.

Table 10.4
Payment Card CV WTP by Household's Use of the Water Environment

Any Household use	Mean WTP (£/hh/yr)	Median WTP (£/hh/yr)	Sample Size (Freq)
Contact activities			
Non user	42.6	25	772
User	58.2	50	617
Fishing			
Non user	48.2	30	1,160
User	56.2	35	229
Other activities			
Non user	39.4	20	264
User	51.9	30	1,125
Any use (Contact, fishing or other)			
Non user	36.3	20	215
User	51.9	30	1,174

Source: NERA

As expected, users of the water environment, whether for contact activities, fishing, or other activities, all tend to have a higher PCCV wtp on average than non-users.

Table 10.5
Testing Average WTP across "Any Use"

H_0	H_1	t-stat	p-value
$WTP_{\_NON\ USER} - WTP_{\_USER}=0$	$WTP_{\_NON\ USER} - WTP_{\_USER} < 0$	-4.76	0.0000

Source: NERA
Note: t-test performed with unequal variances

Table 10.6 and Table 10.7 present descriptive results on PCCV wtp by measures of environmental attitudes.

Table 10.6
Payment Card CV WTP by Membership to an Environmental Organisation

Anyone in Household Member of an Environmental Organisation?	Mean WTP (£/hh/yr)	Median WTP (£/hh/yr)	Sample Size (Freq)
No	44.7	26	1015
Yes	62.5	50	374

Source: NERA

Table 10.7
Payment Card CV WTP by Opinion on National Spending on Water Pollution Control

	Mean WTP (£/hh/yr)	Median WTP (£/hh/yr)	Sample Size (Freq)
Protecting the environments so important that pollution control requirements and standards cannot be too strict, and continue improvements must be made regardless of cost	53.1	35	641
Protecting the environment is important and continue improvements should be funded, provided that they are not excessively cost	50.0	30	547
We are spending about the right amount on cleaning up the environment and don't need to increase or decrease this spending	33.4	20	95
We have made enough progress on cleaning up the environment and we should now concentrate on holding down costs rather than requiring stricter pollution controls	47.6	20	59
Pollution control requirements and standards have gone too far and they already cost more than they are worth	28.7	20	41
Don't Know	32.5	30	6

Source: NERA

As expected, those with pro-environment attitudes gave higher PCCV wtp amounts than others. Those that are members of environmental organisations, or consider that environmental improvements should continue regardless of cost or if the cost is not excessive, have higher wtp on average than other respondents.

Table 10.8 provides separate values for England and for Wales.

Table 10.8
Payment Card CV WTP by Country

Country	Mean WTP (£/hh/yr)	Median WTP (£/hh/yr)	Sample Size (Freq)
England	49.1	30	1,267
Wales	53.3	30	122

Source: NERA

Table 10.8 shows that there is little difference in the payment card CV wtp amounts given by English and Welsh residents, but that Welsh respondents provided slightly higher amounts..

Table 10.9 provides separate values for urban and rural households.

Table 10.9
Payment Card CV WTP by Rural / Urban Area

Rural / Urban?	Mean WTP (£/hh/yr)	Median WTP (£/hh/yr)	Sample Size (Freq)
Rural	50.5	30	302
Urban	49.2	30	1087

Source: NERA

Table 10.9 shows that there is very little difference between urban and rural areas in respect of mean and median PCCV wtp amounts.

Table 10.10 presents mean and median PCCV wtp amounts by respondent's sex.

Table 10.10
Payment Card CV WTP by Respondent's Sex

Sex	Mean WTP (£/hh/yr)	Median WTP (£/hh/yr)	Sample Size (Freq)
Male	56.3	40	591
Female	44.4	25	798

Source: NERA

Males have a significantly higher mean and median PCCV wtp than females. Given that the sample contains a larger number of females than males, this suggests that wtp estimates for the population will need to be weighted by sex.

Table 10.11 presents mean and median PCCV wtp amounts by respondent's level of education.

Table 10.11
Payment Card CV WTP by Respondent's Level of Education

Level of Education	Mean WTP (£/hh/yr)	Median WTP (£/hh/yr)	Sample Size (Freq)
Primary	30.5	20	234
O levels, GCSE or CSE (1+ passes), NVQ Level 1 or foundation level GNVQ	45.10	25	388
5+ O, CSE grade 1's or GCSE A-C; School certificate; 1+ A or As; NVQ level 2 or intermediate GNVQ	45.40	30	177
2+ A levels; 4+ As; Higher School Certificate; NVQ Level 3; or Advanced GNVQ	59.20	50	161
First Degree, Higher degree, NVQ Level 4/5; HNC; HND; Qualified teacher status; Qualified medical doctor, dentist; nurse; midwife; health visitor	66.50	50	376
Other / Don't know	28.50	20	38
Refused	29.00	20	15

Source: NERA

Respondent's level of education has a significant effect on PCCV wtp amounts, with the more educated giving larger values on average.

Supplementary results, included as Appendix M, provide further supporting descriptive evidence that willingness to pay amounts given are in line with expectation. For example, households receiving state financial support of some kind have significantly lower payment card CV wtp on average, in comparison with all other households; respondents that are working have a higher payment card CV wtp on average than those that are not working; and, the higher socio-economic classes tend to have higher payment card CV wtp amounts on average.

10.1.2 Econometric OLS PCCV Results

In this sub-section we report on our econometric analysis of payment card CV wtp amounts. Table 10.12 presents a summary of our findings from an exploratory econometric analysis of the variance of PCCV wtp amounts across respondents.

Table 10.12
Summary of Exploratory PCCV Econometric Results

Variable Group	Models Estimated[(1)]	Summary of Findings from PCCV Econometric Analysis[(2)]
Group 1 - Environmental change	Combinations of *delta_hl*, *delta_ll*, and *delta_hn*, plus the logarithms of these variables *ln_delta_hl*, *ln_delta_ll* and *ln_delta_hn*.	Models with logged variables fit the data better than those with variables measured as levels (ie not logged). Due to correlations between the variables, some combinations result in results counter to expectation. The adopted model for aggregation includes *ln_delta_hl* only from this group.
Group 2 - Affordability	Combinations of *ln_inc*, *ln_oecd_eqinc*, *ben*, *children*, *single_parent*, *q4a, housing_own housing_rent_social, emp_work emp_stud,* and *sec_1-sec_9*.. All models including income also include *income_miss* to capture values for those that did not report their income.	All measures of disposable income correlate positively with PCCV wtp. The variable *ln_inc* performed better than *ln_oecd_eqinc* in explaining variation in *ln_wtp* The variables *ben*, *single_parent* , *q4a, housing_own, housing_rent, social, emp_work, amp_stud and sec_1-sec_9* were all found to be statistically significant (at 10% level) in models including *ln_inc*. The adopted model for aggregation includes *ln_inc* and *children* only from this group.
Group 3- Use of the water environment	Combinations of *contact_use*, *fishing_use*, *otheract_use*, and *use*	All models showed all types of users of the water environment giving a higher wtp than non-users. The adopted model for aggregation includes *use* only from this group.
Group 4- Attitudes towards the environment	Combinations of *pol_control*, *club*, *water_pol*, w*priority _467; local_worse, national_worse,*	Results for *pol_control*, *club*, *water_pol*, and w*priority _467* move in line with expectation in all models: those with a pro-environment attitude tend to have a higher willingness to pay than others for the WFD benefits. The variables *local_worse, national_worse* were found not to be statistically significantly different from zero at the 10% level. The adopted model for aggregation includes *pol_control* only from this group.
Group 5 - Other demographics	Combinations of *sex, edu_12, edu_35, age_cont, age_contsq, age_18_29, age_65,*	Males were found to have a significantly higher PCCV wtp than females. High education is correlated with high PCCV wtp. Age did not have a significant effect on PCCV wtp. The adopted model for aggregation includes *sex* and *edu_12* and *edu_35* only from this group.
Group 6 - Geography	Combinations of *wales, urban, rbd1-rbd11*	The variables *wales* and *urban* were each found to be not statistically significant from 0 at the 10% level. Wales was included in the adopted model for aggregation however to provide the best aggregate estimate of WFD benefits values for Wales and England separately despite little statistical evidence of

		difference in the sample. The variables *rbd1-rbd11* were found to be jointly statistically significant at the 5% level, but were excluded from the adopted model for aggregation due to concerns that the samples were not representative at RBD level and so could provide misleading results. The adopted model for aggregation thus includes *wales* only from this group.
Group 7 - Survey instrument and interview features	Combinations of *cv_first; wb_1,-wb_4;* and *nochange_5 - nochange_20; int_sex and interactions of int_sex and sex*	The variable *cv_first* was found to be negative and statistically significantly different from zero at the 1% level in all models in which it entered. *wb1-wb4* were found to be jointly insignificant at the 10% level *nochange_5-nochange_20* were found to be jointly insignficant at the 10% level. *int_sex* enters significantly negatively, indicating that female interviewers tended to elicit lower wtp values on average. Interactions of *int_sex* and *sex* were insignificant, indicating that the effects of *sex* and *int_sex* are additive. The adopted model for aggregation includes *cv_first* and *int_sex* only from this group.
Group 8 - Comprehension of questionnaire	Combinations of *understood, effort2*, and *concentration*	All variables enter positively in all models. Each variable is statistically significant at the 5% level when entered into models individually. When entereed in pairs, or as a group of three variables, the variables all become insignificant at the 10% level. The adopted model for aggregation includes *understood* only from this group.

Source: NERA
(1) Variable definitions and summary statistics are provided in Appendix H. The dependent variable in all models summarised here was ln_wtp*; the sample used was* pref *in all cases, and the estimator used was OLS.*
(2) Statistical significance has been assessed using t-tests, based on robust standard errors.

Table 10.12 shows that PCCV results are generally in line with expectations. In the following, we summarise the results in respect of each of the groups considered.

10.1.2.1 Group 1 – Environmental change

In respect of environmental change, our analysis showed that, as expected, separate estimates could not be reliably estimated for local and national improvements, nor for the relative values of improvements from low to medium quality versus medium to high quality. This was expected from the PCCV elicitation method due to correlations between the variables and only a small degree of variation in the scenarios presented to respondents. The CE method is substantially better suited to deriving values for marginal changes in environmental attributes. Nonetheless, a functional form was identified from the PCCV responses, which uses the change in the log of high quality in the local area as an explanatory factor. This variable cannot be interpreted as measuring the value of change in high quality in the local area independently of the other environmental attributes due to the correlation across all environmental attributes. This correlation implies that the coefficient on *ln_delta_hl* must be interpreted as a measure of the value of change in all environmental attributes. Our interpretation of the coefficient on this variable is discussed following presentation of our adopted model for aggregation.

10.1.2.2 Group 2 – Affordability

Measures of affordability considered included: household income, equivalised household income, (using the OECD equivalisation, which adjusts income for household composition), a variable indicating whether the household receives any state benefits, a variable indicating whether or not there are any children in the household, a variable indicating whether the household is a single parent unit, a variable measuring the number of children in the household, and indicators of housing tenure status, employment status and socio-economic class. The results from our exploratory analysis showed that all measures of affordability correlate with PCCV wtp in the expected way, but that in combinations, many proved to be insignificant. For example, with income included as an explanatory factor the impact on wtp of receipt of benefits, and whether the respondent represented a single-parent household, became insignificant. Our approach to selecting variables from this group consisted of considering the statistical fit of the models alongside the parsimony of the functional form. This approach led us to a model including only household income and the variable indicating whether or not there are any children in the household as measures of affordability. The variable indicating whether or not there are any children in the household was actually estimated with a positive coefficient, indicating that its impact as a measure of affordability was less important than perhaps its effect on concern for the future of the environment.

10.1.2.3 Group 3- Use of the water environment

Several models were estimated which included different combinations and specifications of the variables measuring use of the water environment. Our specification search involved separating or grouping together different types of use, ie use for contact activities, use for fishing, and use for other activities, and different frequencies of use, ie never, rarely, sometimes, often. In all cases, the models showed that all types of users of the water environment gave a higher wtp than non-users. Considering the statistical fit of the models and the parsimony of the functional form, the adopted model for aggregation includes *use* only from this group, which indicates whether or not the household has used the water environment at any point over the last 12 months for any activity. The inclusion of this variable allows for a direct comparison of use versus non-use values for WFD improvements.

10.1.2.4 Group 4 – Attitudes towards the environment

The first three questions in the survey were designed primarily to elicit attitudes towards public spending on environmental improvements of different types. In addition, respondents were asked whether or not they were a member of an environmental organisation. Our analysis grouped the responses to all these questions in several ways to explore the impact of attitudes towards the environment on wtp. We also explored the impact on wtp of respondents' reactions to the local and national water environment maps. The hypothesis was raised by one of the peer reviewers on the project that respondents may offer higher valuations than they otherwise would have done if they perceived the maps to be depicting a situation that was worse than they had expected. Our analysis tested this hypothesis.

The variables included in the modelling included the {1,0} variable *pol_control*, which indicated whether or not the respondent answered (a) or (b) to Question 3, ie that (a) we should continue improvements regardless of cost, or (b) if not excessively costly. The alternative set of responses were: (c) we are already paying the right amount, (d) we should concentrate on holding down costs, and (e) we are already paying too much. The second variable *club*, was a {1,0} variable indicating whether or not the respondent is a member of an environmental organisation. The third variable, *water_pol*, indicated whether the respondent had listed improvements to the water environment as either first or second as a government spending priority, in relation to other spending categories such as education, health and transport. The fourth variable *wpriority_467* indicated whether or not the respondent listed items (4), (6) or (7) as answers to Question 2 as first or second priorities for water company spending. These items refer to (4) improving river water quality, (6) improving bathing water quality, and (7) protecting animal and plant life around waterways. The fifth and sixth variables, *local_worse*, and *national_worse*, were both {1,0} variables indicating whether or not, firstly the local area map and secondly the England and Wales map were thought to depict a water environment situation that was a great deal worse, or slightly worse than expected.

Results for *pol_control*, *club*, *water_pol* and *wpriority_467* all entered the models with signs in line with expectation: those with a pro-environment attitudes, members of environmental organisations, those who consider improvements to water environment to be a national spending priority, or those who consider improvements to water environment as a water company spending priority, tend to have a higher willingness to pay than others for the WFD benefits. When entered in combinations, the variable *pol_control* emerged as the best predictor of this group of variables. With this variable included, the variables *club*, *water_pol* and *wpriority_467* all became statistically insignificantly different from zero at the 10% level. The adopted model for aggregation thus includes *pol_control* only from this group.

In respect of respondents' reactions to the national and local area water status maps, our analysis found no statistically significant effect, ie wtp was not significantly higher for those who perceived the maps as representing a worse situation than they would have expected.

10.1.2.5 Group 5 – Other demographics

Demographic variables included in our exploratory analysis included respondent sex, age and education levels. Results from this analysis showed that males were found to have a significantly higher PCCV wtp than females. Also, those with high education were found to have higher PCCV wtp on average. Age did not have a significant effect on PCCV wtp. The adopted model for aggregation therefore includes sex and education only from this group. The functional form for education was tested by grouping the five levels in different ways. The best form included two dummy variables *edu_12* and *edu_35*. The first of these, *edu_12*, indicates whether or not the respondent achieved a maximum education of NVQ level 1 or NVQ level 2 or their equivalents, eg one or more GCSEs or O-levels (no qualifications). The second variable, *edu_35,* indicates whether or not the respondent achieved a minimum of two or more A-levels or the equivalent. The omitted category against which the coefficients on these variables are to be interpreted is the group with no qualifications.

10.1.2.6 Group 6 – Geography

We have estimated a number of models that explore the variation of PCCV values across geographical groups: for England and Wales separately, for urban versus rural, and for different RBDs. These geographical effects were explored by including indicator variables for the different groups within full (*pref*)-sample regressions. Given that the models estimated also include demographic and affordability measures, the coefficients on the geographic variables are interpreted as additional effects of the geography on wtp values, ie effects that are separate from the effects of demographic and socio-economic variation across areas.

The variables examined were *wales*, which indicates whether or not the respondent is from Wales, *urban,* which indicates whether or not the respondent lives in an urban area, as defined by the joint ONS/DEFRA/DCLG/CA/NAW classification of urban and rural areas, and dummy variables for each of the 11 RBD groups. Results from our analysis found that *wales* and *urban* were both not statistically significant from zero at the 10% level. The variable *urban* was therefore excluded to improve the parsimony of the specification, however Wales was retained in the adopted model for aggregation in order to provide the best aggregate estimate of WFD benefits values for Wales and England separately despite little statistical evidence of difference in the sample. The variables *rbd1-rbd11* were found to be jointly statistically significant at the 5% level, but were excluded from the adopted model for aggregation due to concerns that the samples were not representative at RBD level and so could provide misleading results. For indicative aggregation to RBD level, given the limited sample size, national econometric estimates are used and weighted to RBD population and environmental characteristics. The adopted model for aggregation thus includes *wales* only from this group.

10.1.2.7 Group 7 – Survey Instrument and Interview Features

As described in Section 3 above, the design of the survey instrument included a number of different versions allocated randomly to different respondents. These features included the type of water body shown, and described, to respondents as an example of the status level characterisation used in the survey, the order of the CV question in relation to the CE, and the amount respondents would told they must pay extra in future in order to maintain the current state of the water environment, allocated randomly from the range {£5,£10, £20}. We examined the effects of each of these features by including dummy variables, in various combinations, into regression models. In addition to these features, the interview experience varied across respondents according to which interviewer conducted the survey. A hypothesis raised by a peer reviewer during the course of the study was that male respondents might give higher wtp amounts if interviewed by a female than otherwise. We examined the effects of this variation by including a dummy variable indicating the sex of the interviewer, and also estimated models with this term interacted with, ie multiplied by, the variable indicating the respondent's sex.

The results of our analysis found that PCCV wtp amounts for those respondents asked the PCCV question before the CE were significantly lower than those where the order was reversed. The coefficient on the variable *cv_first* which captures this effect was statistically significantly different from zero at the 1% level in all models in which it entered. The variables, *wb1-wb4*, indicating the water body type example shown were found to be jointly insignificant at the 10% level in all models. This demonstrates that the choice of example had no systematic effect on wtp responses.

The variables *nochange_5-nochange_20*, measuring the amount respondents were told they would need to pay extra to maintain current water status levels, were also found to be jointly insignificant at the 10% level. This strongly suggests that wtp for WFD improvements is independent, within the range covered by the survey, of the increase in payments needed to avoid deterioration.

Finally, the variable *int_sex,* equal to one if the interviewer was female, and zero otherwise, was found to have a significantly negative effect on wtp, indicating that female interviewers tended to elicit lower wtp values than male interviewers on average. Interactions of *int_sex* and *sex* were insignificant, which indicates, contrary to the maintained hypothesis, that it is not the case that males are influenced by female interviewers. Rather, the results indicate that the effects of *sex* and *int_sex* are additive, ie female respondents give lower values than male respondents, independently of the interviewer's sex, and, female interviewers elicit lower values than male interviewers on average, also independently of the respondent's sex. We have no theory to explain this finding.

In line with our maintained approach to adopt a conservative approach with respect to survey instrument and interviewer effects where there is no clear theoretical guide to which is the more accurate, we include the variables *cv_first* and *int_sex* in our adopted PCCV model for aggregation, and set values for these variables to one in each case, which yields the lower results from the aggregated model for each effect.

10.1.2.8 Group 8 – Comprehension of questionnaire

The final group of variables we have examined relate to respondents' understanding of the questionnaire, level of effort committed by the respondent, and their degree of concentration throughout the survey. All of these variables are based on assessments by the interviewers at the end of each survey interview. The variables included in our analysis were *understood*, a {1,0} variable indicating whether or not the respondent was assessed as having understood the questionnaire either "completely", or "a great deal", *effort2*, a {1,0} variable indicating whether or not the respondent was assessed as having given the questions careful consideration, and *concentration*, a {1,0} variable indicating whether or not the respondent was assessed as having maintained concentration throughout the survey interview.

The results from our analysis showed that each of these variables had a positive and statistically significant effect on wtp at the 5% level when entered into models individually. When entered in pairs, or as a group of three variables, the variables all become insignificant at the 10% level. Due to its statistical performance as a predictor variable, *understood* was included as the only variable from this group in the adopted model for aggregation. Our approach to aggregation from this variable is to set it equal to one, thereby deriving estimates of wtp for populations that can be considered to have understood the questionnaire presented to them.

Table 10.13 presents our final adopted regression model for aggregation from the PCCV wtp responses. The table contains the adopted model estimated using our preferred sample, *pref*, and also using the *trim1, trim3* and *pref_cons* samples. This allows an examination of the sensitivity of our main PCCV results to the approach used to exclude protests and outliers.

Table 10.13
Adopted PCCV Regression Model for Aggregation

Variable	PCCV Regression Model			
	***pref* sample (coef | s.e.)**	***trim1* sample (coef | s.e.)**	***trim3* sample (coef | s.e.)**	***pref_cons* sample (coef | s.e.)**
ln_delta_hl(1)	0.671	0.517	0.249	0.916
	(0.487)*	(0.523)	(0.416)	(0.613)*
ln_inc(1)	0.258	0.246	0.246	0.310
	(0.045)***	(0.048)***	(0.041)***	(0.053)***
income_miss	1.411	1.304	1.392	1.660
	(0.280)***	(0.300)***	(0.256)***	(0.336)***
children	0.126	0.203	0.058	0.168
	(0.074)*	(0.080)**	(0.066)	(0.089)*
use(1)	0.266	0.346	0.158	0.237
	(0.103)***	(0.110)***	(0.092)*	(0.127)**
pol_control(1)	0.445	0.518	0.312	0.610
	(0.112)***	(0.115)***	(0.104)***	(0.136)***
sex	-0.181	-0.142	-0.213	-0.264
	(0.069)***	(0.074)*	(0.060)***	(0.081)***
int_sex	-0.218	-0.212	-0.136	-0.248
	(0.069)***	(0.075)***	(0.062)**	(0.080)***
edu_12	0.186	0.147	0.211	0.159
	(0.100)*	(0.107)	(0.089)**	(0.126)
edu_35	0.511	0.499	0.380	0.521
	(0.104)***	(0.112)***	(0.095)***	(0.127)***
wales	0.041	-0.038	0.104	0.085
	(0.130)	(0.140)	(0.115)	(0.152)
cv_first	-0.409	-0.466	-0.340	-0.492
	(0.065)***	(0.071)***	(0.058)***	(0.078)***
understood	0.258	0.215	0.272	0.348
	(0.119)**	(0.123)*	(0.110)**	(0.143)**
Constant	0.762	0.761	1.377	0.322
	(0.408)*	(0.430)*	(0.354)***	(0.507)
Observations	1389	1462	1319	1006
R2	0.32	0.26	0.20	0.60

Source: NERA
Dependent variable = ln_wtp; estimator used: OLS; robust standard errors in parentheses
t-test p-values (two-sided except where indicated): $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$; (1) Note:1-sided t-test used to calculate p-values.*

In the model presented above, all explanatory variables enter with coefficients that are statistically significantly different from zero at the 10% level except for *wales*. As discussed above, *wales* was retained in the adopted model for aggregation in order to provide the best aggregate estimate of WFD benefits values for Wales and England separately despite little statistical evidence of difference in the sample.

The results in Table 10.13 can be used to derive aggregate values for WFD environmental improvement scenarios by attaching values to each of the variables in the model and multiplying through by the respective coefficients.

Table 10.14
Input Assumptions for Aggregation of PCCV WTP Estimates

Variable	Coeff	Input assumptions for aggregation
ln_delta_hl	0.671	Derived for numbers of improvement scenarios as described in Section 11, all of which are based on improvements from the average proportion of high quality area in local (<30-mile radius) areas surrounding each Census ward.
ln_inc	0.258	Derived from average household income statistics for England, Wales, E&W and individual RBDs, sourced from ONS 2004/5 income data by Mid-Level Statistical Output Area.
income_miss	1.411	0
children	0.126	Derived from Family Resource Survey (2005/6) data, equal to 29.0% for England and Wales.
use	0.266	Full sample mean from present survey, equal to 83.9% for England and Wales.
pol_control	0.445	Full sample mean from present survey, equal to 85.4% for England and Wales.
sex	-0.181	50%
int_sex	-0.218	100%
edu_12	0.186	Derived from Family Resource Survey (2005/6) data for England, Wales, E&W, and individual RBDs, equal to 35.8% for E&W.
edu_35	0.511	Derived from Family Resource Survey (2005/6) data for England, Wales, E&W, and individual RBDs, equal to 34.3% for E&W.
wales	0.041	Derived from Census (2001) population data, equal to 5.6% for E&W
cv_first	-0.409	100%
understood	0.258	100%
Constant	0.762	1

Source: NERA

The sum of the coefficients multiplied by the values of the variables they correspond to yields the conditional mean of the log of wtp, as this was the dependent variable in the model. The exponential of this value gives the conditional median, rather than the mean wtp estimate, (see Goldberger, 1968). The conditional mean is derived by applying an adjustment equal to the mean of the exponential of the residuals from the model. Table 10.15 presents mean and median wtp results for the "95% scenario" by country based on this method applied to the OLS model and inputs described in the above table. Raw sample means and medians are presented alongside the derived results for comparison.

Table 10.15
PCCV OLS WTP Results for "95% Scenario", by Country

Elicitation Method / Model	England		Wales		England and Wales	
	Mean WTP £/hh/yr	Median WTP £/hh/yr	Mean WTP £/hh/yr	Median WTP £/hh/yr	Mean WTP £/hh/yr	Median WTP £/hh/yr
PCCV OLS Model	44.8	25.3	40.1	22.7	44.5	25.1
PCCV Sample Statistics	49.2	30.0	62.6	50.0	50.4	30.0

Source: NERA

Table 10.15 shows that the effect of the aggregation assumptions, plus the weighting from sample to population by local water environment quality levels, income, children, sex, and education, is to reduce the mean and median wtp. The table presents our best estimates of the

PCCV mean and median wtp for the 95% scenario. The model results can also be used to examine sensitivities to a wide range of factors. By varying the scope of environmental change, estimates can be developed for a range of WFD improvement scenarios. The sensitivity of wtp to these scenarios is discussed in Section 11. By varying the income amount, estimates can be developed for groups with affordability concerns versus other groups. This sensitivity is presented in Table 10.16 below. By varying the demographic variables, and the local environmental quality level, wtp values can be obtained for different geographical regions, such as RBDs. Indicative values for RBDs, based on this methodology, are presented in Table 10.17 for the 95% scenario. Further analysis would be able to derive RBD estimates for alternative WFD improvement scenarios. Results can also be subjected to sensitivity testing with respect to the assumptions we have made regarding survey instrument and interviewer effects. Such investigation will be possible using the spreadsheet model developed by NERA for the CRP steering group. The model, and accompanying guidance note, will be delivered following the present report. Finally, it will be possible to update wtp values in future by varying the *pol_control* value, eg by conducting a repeat of the Omnibus survey some time in the future when attitudes may have been expected to change. Results could thereby be updated to take account of any attitude changes.

Table 10.16 provides a breakdown of mean PCCV OLS wtp results by income group. The low income group is defined as a household with an income of £200 per week. The medium income group is defined as a household with an income of £500 per week. The high income group is defined as a household with an income of £1500 per week.

Table 10.16
PCCV OLS Mean wtp Results for "95% Scenario", by Country and Income Group

Country	Income Group		
	Low Income £/hh/yr	Med Income £/hh/yr	High Income £/hh/yr
England	33.6	42.6	56.5
Wales	31.7	40.1	53.2
England & Wales	33.5	42.4	56.3

Source: NERA

Table 10.16 shows, as expected, that higher income groups are willing to pay more than lower income groups. Given the specification of the PCCV model, which includes only income from the group of variables measuring affordability, the estimates for the low income group represent our best estimates of PCCV wtp for those with potential affordability concerns.

Table 10.17 provides indicative mean and median PCCV OLS wtp results by RBD for the 95% improvement scenario.

Table 10.17
PCCV OLS Mean and Median WTP Results for "95% Scenario", by River Basin District

River Basin District	Mean WTP (£/hh/yr)	Median WTP (£/hh/yr)
Anglian	44.5	25.1
Dee	42.6	24.1
Humber	42.3	23.9
North West	42.7	24.1
Northumbria	38.3	21.7
Severn	43.8	24.8
Solway Tweed	39.0	22.1
South East	46.3	26.1
South West	42.2	23.8
Thames	49.6	28.0
Western Wales	39.6	22.4

Source: NERA

Table 10.17 shows that households in the Thames, South East and Anglian RBDs are estimated to have the highest PCCV wtp, and households in Northumbria, Solway-Tweed and Western Wales RBDs are estimated to have the lowest PCCV wtp.

10.2 Dichotomous Choice Contingent Valuation Results

This section reports the main results from our descriptive and econometric analyses of the DCCV responses.

10.2.1 Descriptive DCCV Results

Table 10.18 shows the proportion of respondents accepting the DCCV improvement scenario by the cost amount presented in the scenario. Results are shown for each of the analysis samples.

Table 10.18
Proportion of Respondents Accepting DCCV Improvement Scenario, by DCCV Cost Amount and Analysis Sample

DCCV Cost Amount (£/hh/yr)	*pref* (%)	*trim1* (%)	*trim3* (%)	*pref_cons* (%)
5	85	83	90	88
10	92	88	93	94
20	82	82	85	80
30	79	76	82	69
50	68	68	72	59
100	61	60	63	38
200	41	41	41	7
Sample Size	1389	1462	1319	1006

Source: NERA

Table 10.18 shows that the proportion accepting the DCCV improvement scenario generally diminishes with the cost of the scenario, as expected. The single exception to monotonic diminishment is the fact that the DCCV acceptance proportion is higher at an offered cost of £5 than an offered cost of £10 for all samples. A likely explanation for this finding is that the price of £5 might be considered unrealistic.

It is notable that over 40% of respondents in *pref* that were offered the CV improvements at a cost of £200 accepted the package. Most of these, and a large number of those accepting the £100 DCCV package, are removed in the *pref_cons* sample because their PCCV values were lower.

In order to calculate wtp from the DC results, we first use the Turnbull non-parametric approach, following which we estimate logit models to parameterise the distribution and thereby allow for more power to test the effects of covariates in a multivariate context and to aggregate from the sample to the population.

10.2.2 Turnbull Estimates of DCCV WTP

Table 10.19 presents the calculation of the Turnbull estimate of the lower bound of mean wtp. In this table, the responses to the £10 case have been grouped with the responses to the £5 case in order to achieve monotonicity. The lower bound of £5 is used as the bill amount for this group in order to preserve the status of the estimator as a lower bound.

Table 10.19
Turnbull Estimate of Lower Bound Mean DCCV WTP

DCCV Cost Amount (£/hh/yr)	Accept (%)	N	Change in Reject Proportion (%)	WTP (£/hh/yr)
t			f*(t)	t* x f*(t)
5	88.3%	376	5.8%	0.29
20	82.5%	228	3.9%	0.78
30	78.5%	205	10.2%	3.07
50	68.3%	183	7.5%	3.76
100	60.8%	204	19.9%	19.85
200	40.9%	193	40.9%	81.87
Turnbull Lower Bound Estimate of Mean WTP (£/hh/yr)				109.62

Source: NERA

Table 10.19 shows the calculation of the Turnbull estimate of lower bound mean DCCV wtp. The column entitled "Change in Reject Proportion (%)" captures the width of the blocks ($n_i - n_{i+1}$) in Figure 9.3. These amounts are multiplied by the DCCV cost amount associated with the bin to calculate the area under that part of the demand curve. The sum of these values is the Turnbull estimate of lower bound mean DCCV wtp. The resulting value for this is £109.62 per household per year.

The Turnbull estimator is known to be distributed normally due to the central limit theorem. The standard error of the estimator is equal to:

$$\sqrt{\left(\sum_{j=1}^{6}\frac{F_j^*\left(1-F_j^*\right)}{T_j^*}\left(t_j-t_{j-1}\right)^2\right)},$$

where j indexes the bid amounts, F_j is the proportion rejecting the CV scenario at the j^{th} bid amount (t_j), and T_j is the total number of responses in the sample offered that bid amount. This calculation results in a value of £4.02 for the standard error of the Turnbull lower bound estimate.

The Turnbull estimate of £109.62 per household per year as a lower bound of the mean DCCV wtp is over twice the size of the mean PCCV wtp estimate. The two elicitation approaches are therefore yielding substantially different values for the benefits of WFD improvements. We return to this issue in Section 10.5.

Table 10.20 to Table 10.26 present some cross-tabulations of the Turnbull estimate by variables from the dataset. The standard errors around the estimates increase substantially as more categories are explored due to the fact that the estimator is relatively weak with regard to covariate analysis; t-tests are presented to show whether differences between means are statistically significant.

Table 10.20
Turnbull Lower Bound of Mean DCCV WTP by Total Weekly Household Income

Total Weekly Household Income	Turnbull Lower Bound of Mean WTP (£/hh/yr)	Std. Error	Sample Size (Freq)
Low (<£300)	81.0	6.4	419
Medium	119.8	7.2	463
High (>£1,000)	138.0	12.2	127

Source: NERA

Table 10.21
Testing Turnbull Lower Bound of Mean WTP across Income Groups

H_0	H_1	t-stat	p-value
$WTP_{\_LOW} - WTP_{\_MEDIUM} = 0$	$WTP_{\_LOW} - WTP_{\_MEDIUM} < 0$	-4.03	0.000
$WTP_{\_MEDIUM} - WTP_{\_HIGH} = 0$	$WTP_{\_MEDIUM} - WTP_{\_HIGH} < 0$	-1.28	0.100

Source: NERA
Note: t-test performed with unequal variances. p-value derived using normal approximation for the t distribution.

DCCV wtp increases with income, in line with expectation. The difference in wtp is only statistically significant at the 10% level between medium and high incomes. The t-test results show that the difference between DCCV wtp values at low and medium incomes is significant at the 1% level.

Table 10.22
Turnbull Lower Bound of Mean DCCV WTP by Respondent's Employment Status (Abbreviated Version)

Employment Status	Turnbull Lower Bound of Mean WTP (£/hh/yr)	Std. Error	Sample Size (Freq)
Working	116.5	5.7	707
Not working	99.1	5.8	654

Source: NERA

Table 10.23
Testing Turnbull Lower Bound of Mean WTP across Employment Status

H_0	H_1	t-stat	p-value
$WTP_{\_NOT\ WORKING} - WTP_{\_WORKING} = 0$	$WTP_{\_NOT\ WORKING} - WTP_{\_WORKING} < 0$	-2.14	0.016

Source: NERA
Note: t-test performed with unequal variances. p-value derived using normal approximation for the t distribution.

DCCV wtp is higher for those in work than those not working, in line with expectation. The difference is statistically significant at the 5% level.

Table 10.24
Turnbull Lower Bound of Mean DCCV WTP by Household's Use of the Water Environment

Any Household Use of the Water Environment	Turnbull Lower Bound of Mean WTP (£/hh/yr)	Std. Error	Sample Size (Freq)
Non user	73.9	13.9	215
User	111.2	4.3	1,174

Source: NERA

Table 10.25
Testing Turnbull Lower Bound of Mean WTP across Use of the Water Environment

H_0	H_1	t-stat	p-value
$WTP_{\_NON\ USERS} - WTP_{\_USERS} = 0$	$WTP_{\_NON\ USERS} - WTP_{\_USERS} < 0$	2.567	0.005

Source: NERA
Note: t-test performed with unequal variances. p-value derived using normal approximation for the t distribution.

DCCV wtp is higher for users than non-users, in line with expectation. The difference is statistically significant at the 1% level.

Table 10.26
Turnbull Lower Bound of Mean DCCV WTP by Opinion on National Spending on Water Pollution Control

Opinion on National Spending on Water Pollution Control	Turnbull Lower Bound of Mean WTP (£/hh/yr)	Std. Error	Sample Size (Freq)
Continue improvements regardless of cost OR Continue improvements if not excessive cost	113.4	4.4	201
Already paying the right amount OR Concentrate on holding down costs OR Already paying too much OR Don't know	75.1	12.7	1,188

Source: NERA

Table 10.27
Testing Turnbull Lower Bound of Mean WTP across Opinion on National Spending

H_0	H_1	t-stat	p-value
$WTP_{\_CONTINUE} - WTP_{\_NOT\ CONTINUE} = 0$	$WTP_{\_CONTINUE} - WTP_{\_NOT\ CONTINUE} > 0$	2.845	0.002

Source: NERA
Note: t-test performed with unequal variances. p-value derived using normal approximation for the t distribution.

DCCV is higher for those with pro-environment attitudes than for others, in line with expectation. The difference is statistically significant at the 1% level.

10.2.3 Econometric Logit DCCV Results

In this sub-section we report on our econometric logit analysis of DCCV responses.

Table 10.28
Summary of Exploratory DCCV Econometric Logit Results

Variable Group	Models Estimated[1]	Summary of Findings from DCCV Econometric Analysis[2]
Group 1 - Environmental change	Combinations of *delta_hl*, *delta_ll*, and *delta_hn*, plus the logarithms of these variables *ln_delta_hl*, *ln_delta_ll* and *ln_delta_hn*.	None of the variables entered the models with coefficients significantly different from zero at the 10% level. This finding suggests insensitivity to the scope of environmental improvements across treatments. The adopted model for aggregation excludes all environmental change variables, thus constraining wtp to be the same for the 75% and 95% scenarios.
Group 2 - Disposable income	Combinations of *income_cont*, *ln_inc*, *ln_oecd_eqinc*, *ben*, *children*, *single_parent*, *q4a*.. All models include *income_miss* to capture values for those that did not report their income.	Measures of income and oecd disposable income have a positive sign and are statistically significant at the 1% level. The variable *ln_inc* performs better than income_cont in the models and better than *ln_oecd_eqinc*. *ben, children, q4a* and *single_parent* are not statistically significantly different from zero at the 10% level in models that also include measures of income. The adopted model for aggregation includes *ln_inc* and *income_miss* only from this group.
Group 3- Use of the water environment	Combinations of *contact_use*, *fishing_use*, *otheract_use*, and *use*	None of the variables in this group were statistically significantly different from zero at the 10% level. The adopted model for aggregation does not include any variables from this group.
Group 4- Attitudes towards the environment	Combinations of *pol_control*, *club*, *water_pol*	*pol_control* and club are statistically significantly different from zero at the 1% level. *water_pol* is not statistically different from 0 at the 10% level The adopted model for aggregation includes *pol_control* and *club* only from this group.
Group 5 - Other demographics	Combinations of *sex, edu_12, edu_35, age_cont, age_contsq, age_18_29, age_65, housing_own housing_rent_social, emp_work emp_stud,* and *sec_1-sec_9*	*sex* negative and significant at the 5% level. *edu_12* positive and not statistically different from 0 at the 10%. *edu_35* positive and significant at the 5% level *housing* variables not statistically different from 0 at the 10% level *employment* variables not statistically different from 0 at the 10% level. *sec* variables not statistically different from 0 at the 10% level. The adopted model for aggregation includes *edu_35* only from this group
Group 6 - Geography	Combinations of *wales, urban, rbd1-rbd11*	No variables in this group are statistically significant at the 10% level, and none are included in the adopted model for aggregation.
Group 7 - Survey	Combinations of *cv_first;*	Water body example shown (*wb_1-wb_4*) are jointly

instrument and interview features	*dc_position; wb_1,- wb_4;* and *nochange_5 - nochange_20; int_sex* and interactions between *int_sex* and *sex*	insignificant at the 10% level. The cost of maintaining current conditions has no statistically significant effect at the 10% level. Those receiving the DCCV question before PC CV and any other CE questions have a significantly higher likelihood of accepting the scenario. *int_sex* enters significantly negatively, indicating that female interviewers tended to elicit lower wtp values on average. Interactions of *int_sex* and *sex* were insignificant, indicating that the effects of *sex* and *int_sex* are additive. The adopted model for aggregation includes *cvfirst_dcposition_0_1* and *int_sex* only from this group.
Group 8 - Comprehension of questionnaire	Combinations of *understood, effort2*, and *concentration*	All variables positive and statiistically significant at 1% or 5% level when entered individually. Statistical significance lowers when multiple variables are included. The best performing variable from a statistical perspective is *concentration* and this variable only is included in the adopted model for aggregation.

Source: NERA
(1) Variable definitions and summary statistics are provided in Appendix H. The dependent variable in all models summarised here was dc_choice; the sample used was pref *in all cases, and the estimator used was logit.*
(2) Statistical significance has been assessed using t-tests, based on robust standard errors.

Table 10.28 shows that most results from the DCCV analysis are in line with theoretical expectations. Variables related to the 75% or 95% treatment were not significant, which is not surprising given the lack of power of the test in DCCV data, and that all scenarios have 95% high quality eventually. The PCCV data and CE data allow much more powerful tests and do find a scope effect. The most informative data on preferences regarding the scope of environmental improvements is obtained from the CE. Results from this analysis are reported in Section 10.3.

Table 10.29 presents our final adopted model for aggregation from the DCCV responses. The table contains the adopted model estimated using our preferred sample, *pref*, and also using the *trim1, trim3* and *pref_cons* samples. This allows an examination of the sensitivity of our main DCCV results to the approach used to exclude protests, outliers and inconsistencies.

Table 10.29
Adopted DCCV Logit Model for Aggregation

Variable	DCCV Logit Model			
	pref **sample (coef | s.e.)**	***trim1*** **sample (coef | s.e.)**	***trim3*** **sample (coef | s.e.)**	***pref_cons*** **sample**
dc_bill(1)	-1.236	-1.143	-1.361	-2.831
	(0.102)***	(0.096)***	(0.105)***	(0.226)***
ln_inc(2)	0.395	0.337	0.433	0.670
	(0.094)***	(0.089)***	(0.101)***	(0.130)***
income_miss	2.066	1.699	2.276	3.465
	(0.556)***	(0.529)***	(0.601)***	(0.760)***
pol_control(2)	0.538	0.627	0.563	0.820
	(0.182)***	(0.172)***	(0.203)***	(0.231)***
club(2)	0.437	0.400	0.392	0.455
	(0.175)**	(0.166)**	(0.186)**	(0.239)*
edu_35	0.400	0.423	0.285	0.780
	(0.156)**	(0.147)***	(0.166)*	(0.214)***
cvfirst_dcposition_0_1	0.562	0.537	0.597	0.652
	(0.192)***	(0.179)***	(0.206)***	(0.262)**
concentration	0.562	0.518	0.535	0.404
	(0.185)***	(0.175)***	(0.200)***	(0.226)*
int_sex	-0.295	-0.291	-0.223	-0.359
	(0.161)*	(0.152)*	(0.171)	(0.206)*
sex	-0.284	-0.196	-0.365	-0.532
	(0.140)**	(0.132)	(0.151)**	(0.186)***
wales	0.268	0.252	0.520	-0.134
	(0.260)	(0.248)	(0.294)*	(0.312)
constant	-1.314	-1.208	-1.259	-2.597
	(0.568)**	(0.539)**	(0.610)**	(0.778)***
Observations	1389	1462	1319	1006
Pseudo R2	0.16	0.15	0.18	0.37

Source: NERA
Robust standard errors in parentheses; t-test p-values (two-sided except where indicated): $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$; (1) Note:1-sided t-test used to calculate p-values. (2)* dc_bill *was divided by 100 for inclusion in this model.*

The results in Table 10.29 can be used to derive aggregate values for WFD environmental improvement scenarios by attaching values to each of the variables in the model and multiplying through by the respective coefficients. The same methods were employed to derive these values as for the PCCV results described above. Table 10.30 presents mean and median wtp results for the "95% scenario" by country based on the DC CV logit model in Table 10.31, with sample averages for *pol_control* and *club*, and population values for *ln_inc*, *sex*, *edu_12*, *edu_35*. The variable *cvfirst_dcposition_0_1* is set equal to 0*, int_sex* is set equal to 1 and *concentration* is set equal to 1 to derive these results. PC CV OLS results and raw sample means and medians are presented alongside the derived results for comparison.

Table 10.30
DCCV Logit WTP Results for "95% Scenario", by Country

Elicitation Method / Model	England		Wales		England and Wales	
	Mean WTP £/hh/yr	Median WTP £/hh/yr	Mean WTP £/hh/yr	Median WTP £/hh/yr	Mean WTP £/hh/yr	Median WTP £/hh/yr
DCCV Logit Model	167.0	167.0	181.4	181.4	167.9	167.9
PC CV OLS Model	44.8	25.3	40.1	22.7	44.5	25.1
PCCV Sample Statistics	49.2	30.0	62.6	50.0	50.4	30.0

Source: NERA

The table above presents our best estimates of the DCCV mean wtp for the 95% scenario. The DCCV results produce a higher willingness to pay value than was found from the PCCV estimate of the same benefits. This reflects the respondents' different answers to the PCCV and DCCV format questions, readily observable in the high proportion of 'yes' responses to the £200 DCCV option, the largest bill presented in the DCCV questions. The unknown high tail of DCCV valuations presents a problem for reliance on models assuming continuity in this range.

As for the PCCV model, the DCCV results can also be used to examine sensitivities to the scope of environmental change, (discussed in Section 11), income, (presented in Table 10.31 below), and geographical area, (RBD estimates presented in Table 10.32 below). Sensitivity analysis may also be conducted with respect to our assumptions regarding survey instrument and interviewer effects.

Table 10.31 provides a breakdown of DCCV logit wtp results by income group. As above, the income groups are defined as low - £200 per week, medium - £500 per week, and high - £1500 per week.

Table 10.31
DCCV Logit WTP Results for "95% Scenario", by Country and Income Group

Country	Income Group		
	Low Income £/hh/yr	Med Income £/hh/yr	High Income £/hh/yr
England	131.5	160.8	195.9
Wales	152.0	181.3	216.4
England and Wales	132.6	161.9	197.0

Source: NERA; mean and median wtp results are the same by construction in the logit model.

Table 10.31 shows, as expected, that DCCV wtp results increase with income. Given the specification of the models, the estimates for the low income group represent our best estimates of the wtp for groups with potential affordability concerns.

Table 10.32 presents indicative RBD estimates of DCCV wtp, based on differences in current water environment quality, income and demographics.

Table 10.32
DCCV Logit WTP Results for "95% Scenario", by River Basin District

River Basin District	WTP £/hh/yr
Anglian	167.7
Dee	166.3
Humber	160.7
North West	161.7
Northumbria	158.3
Severn	167.7
Solway Tweed	161.1
South East	171.8
South West	165.4
Thames	175.5
Western Wales	177.9

Source: NERA

Table 10.32 shows that households in the Thames, South East and Western Wales RBDs are estimated to have the highest DCCV wtp for the WFD improvements. Households in the Northumbria, Humber and Solway-Tweed RBDs are estimated to have the lowest DCCV wtp.

10.3 Choice Experiment Results

This section reports the main results from our descriptive and econometric analyses of the CE responses.

10.3.1 Descriptive Results on CE Choice Behaviour

Table 10.33 shows the pattern of choices made by respondents with respect to the number of times they chose the cheapest (Option 1) or the medium cost option (varies over Options 2 and 3) or the most expensive option.

Table 10.33
Respondents by Number of Times Cheapest, Middle Cost and Highest Cost Options are Chosen

No. of Times Chosen	Cheapest (Option 1)	Medium cost	Most expensive
0	905	5	245
1	153	9	215
2	77	39	235
3	63	95	215
4	41	147	194
5	28	216	148
6	25	316	93
7	16	321	29
8	81	241	15
Total	1,389	1,389	1,389

Source: NERA

The above table shows a significant result: the majority of respondents (905/1389) never chose Option 1 (“no change”). This finding suggests that there is a high value associated with even the most limited improvement options offered to respondents in comparison with the status quo.

Only a small proportion (81/1389) always chose the no-change option, which further suggests that there is a widespread interest in seeing improvements occur, within the range of cost amounts offered in the improvement options.

The table also shows that the medium cost improvement option was chosen more frequently than the most expensive option. The vast majority of respondents chose the medium-cost option four times or more, and chose the most expensive option four times or fewer. This finding is in line with expectation, and indicates that most respondents paid attention to the cost as well as the improvements offered when making their choices.

The reasons people gave for the choices they made are reported verbatim in Appendix L.

Table 10.34 shows numbers of respondents by the maximum bill size that they chose during the full CE.

Table 10.34
Respondents by Maximum Bill Size Chosen

Maximum bill chosen by respondents in CE (£)	Freq.	Percent	Cum.
0	81	5.8	5.8
5	25	1.8	7.6
10	55	4.0	11.6
20	89	6.4	18.0
30	159	11.5	29.5
50	282	20.3	49.8
100	331	23.8	73.6
200	367	26.4	100.0
Total	1,389	100	

Source: NERA

The above table shows that a over a quarter of respondents chose a £200 option at least once during the CE. Half the sample chose an option costing £100 or more.

10.3.2 Fixed Coefficient Logit CE Results

A number of fixed coefficient CE models were estimated to explore the variance of choice responses by explanatory factors. Table 10.35 summarises the results from these models.

Table 10.35
Summary of Exploratory Fixed Coeficient Logit CE Results

Variable Group	Models Estimated[1]	Summary of Findings from Fixed Coeficient Logit CE Analysis[2]
Group 1 - Environmental change	*hl8*, *ll8*, *hn8* and *ln8*, and/or the logarithms of these variables, included in all specifications. Additional combinations of variables were selected from {*asc*, *h20_95*, *hl20*, *hn20*, *ll20*, *ln20*, *hl0*, *ll0*, *hn0*, *ln0*} and interactions of these terms with each other.	*hl8*, *ll8*, *hn8* and *ln8* all have the correct sign and are always significant at the 1% level in all models. Models with logs of the *hl8*, *ll8*, *hn8* and *ln8* variables perform slightly better than those with the variables entering linearly. Results show that marginal utility is diminshing with respect to environmental improvements. *asc* enters all models with a negative sign and is always statistically significant at the 1% level. *h20_95* is statistically significant at the 1% level when not entered into models in combination with other 20-year variables {*hl20*, *ll20*, *hn20*, *ln20*). The effects of these latter four variables cannot be separately, but each is individually significant at the 1% level with the expected sign. The current conditions variables {*hl0*, *ll0, hn0* and *ln0*} did not enter significantly (at the 10% level) in any models, individually, or interacted with the 8-year variables. The adopted model for aggregation includes *hl8, ll8, hn8, ln8, asc* and *h20_95* from this group.
Group 2 - Affordability	Combinations of *income_cont*, *inc_high*, *inc_med*, *inc_highmed, ln_inc*, *ln_oecd_eqinc*, *ben*, *children*, *single_parent*, *q4a*.. All variables interacted with *asc* and/or *bill*, and all models also included *income_miss* to capture values for those that did not report their income.	All measures of income enter the models with the expected sign but with a range of statistical significance levels. The adopted model for aggregation includes *ln_inc* from this group, along with *income_miss*, both interacted with *asc*.
Group 3- Use of the water environment	Combinations of *contact_use*, *fishing_use*, *otheract_use*, and *use.* All variables interacted with *asc* and/or *bill*.	All measures of use types enter the models with the expected sign but with a range of statistical significance levels. If use is included, the other types of use have no statistically significant additional impact at the 10% level. The adopted model for aggregation includes *use* from this group, interacted with *asc*.
Group 4- Attitudes towards the environment	*pol_control*, interacted with *asc* and/or *bill*.	*pol_control* enters the models with the expected sign and performs better when interacted with *asc* than when interacted with *bill*. The adopted model for aggregation includes *pol_control* from this group, interacted with *asc*.
Group 5 - Other demographics	Combinations of *sex, edu_12, edu_35, age_18_29, age_65.* All variables interacted with *asc* and/or *bill*.	*sex* is statistically significant at the 5% level when interacted with *bill*; with males displaying a lower sensitivity to the bill than females. *sex* is not significant at the 10% level when interacted with *asc*. There is generally no significant difference between low and medium education levels, but a significant difference between the highest education level and the other levels,

		at the 5% level when interacted with the bill, and at the 1% level when interacted with *asc.* Age has no significant effect in any models at the 10% level. The adopted model for aggregation includes *sex* interacted with *bill*, and *edu_35* interacted with both *bill* and *asc*
Group 6 - Geography	*wales,* interacted with *asc* and/or *bill*.	*wales* is significant at the 1% level when interacted with *bill*, but not significant at the 10% level when interacted with *asc*. The adopted model for aggregation includes *wales* interacted with *bill*
Group 7 - Survey instrument and interview features	Combinations of *cv_first; wb_1,- wb_4; seen_200, nochange_5 - nochange_20; and int_sex.* All variables interacted with *asc* and/or *bill*.	*cv_first* enters positively and statistically significantly at the 1% level when interacted with *asc*, showing that people are more likely to choose "no-change" in the CE if they have already stated a PCCV wtp amount. The variable is not significant when interacted with *bill*. The variables *wb_1 - wb_4* are jointly insignificant at the 10% level when interacted either with *asc* or *bill*, indicating that the water body type example shown had no systematic effect on choice behaviour. The variable *seen_200* enters negatively and significantly when interacted with *asc.* This indicates that respondents are less likely to choose the "no-change" option after having seen an option costing £200, than beforehand. The variable also enters positively and statistically significantly at the 1% level when interacted with bill, indicating that respondents are less sensitive to the bill after having seen a £200 option. This finding is partly an artefact of the modelling approach since it is not possible for respondents to choose options costing £200 if they have not been shown them. We therefore do not include this variable in the model for aggregation, despite its statistical significance. The cost amount for maintaining current quality levels does not have much effect on choice behaviour. *nochange_20* enters positively and statistically significantly at the 10% level, but not the 5% level, when interacted with bill, indicating weakly that respondents are less sensitive to the bill if they are told a higher amount for maintaining current conditions. No other variables or combinations are significant. The variable is not included in the model for aggregation due to its weak statistical significance. Interviewer sex was significant in the models when interacted with asc, and indicated that female interviewers elicited significantly more "no change" choices than male interviewers. The adopted model for aggregation includes *cv_first* and *int_sex* only from this group, both interacted with *asc* only.
Group 8 - Comprehension	*understood,* interacted with *asc* and/or *bill*.	The variable measuring respondents understanding of the questionnaire has no statistically significant effect in

of questionnaire	the choice models, either when interacted with *asc* or *bill*.

Source: NERA
(1) Variable definitions and summary statistics are provided in Appendix H. The dependent variable in all models summarised here was choice*; the sample used was* pref *in all cases, and the estimator used was logit.*
(2) Statistical significance has been assessed using t-tests, based on robust standard errors, which take account of the within-respondent correlation of choice behaviour.

Table 10.36 below shows our adopted fixed-coefficient logit CE models for aggregation, with results for the *trim1*, *trim3* and *pref_cons* presented alongside the *pref* sample results for comparison..

Table 10.36
Adopted Fixed Coefficient Logit CE Model

Variable[1]	Fixed Coefficient Logit CE Model			
	***pref* sample (coef | s.e.)**	***trim1* sample (coef | s.e.)**	***trim3* sample (coef | s.e.)**	***pref_cons* sample (coef | s.e.)**
hl8	0.872	0.795	0.944	0.882
	(0.092)***	(0.090)***	(0.095)***	(0.111)***
ll8	-0.613	-0.599	-0.685	-0.686
	(0.117)***	(0.115)***	(0.117)***	(0.133)***
hn8	1.026	0.983	1.078	1.036
	(0.100)***	(0.097)***	(0.103)***	(0.120)***
ln8	-0.837	-0.839	-0.904	-0.925
	(0.152)***	(0.150)***	(0.157)***	(0.183)***
h20_95	0.097	0.093	0.114	0.110
	(0.030)***	(0.029)***	(0.030)***	(0.035)***
bill	-1.406	-1.346	-1.477	-1.603
	(0.073)***	(0.070)***	(0.076)***	(0.091)***
male_bill	0.211	0.153	0.244	0.358
	(0.088)**	(0.085)*	(0.090)***	(0.106)***
edu35_bill	0.170	0.201	0.208	0.250
	(0.089)*	(0.085)**	(0.092)**	(0.109)**
wales_bill	0.583	0.554	0.642	0.656
	(0.117)***	(0.116)***	(0.118)***	(0.141)***
asc	2.253	1.965	2.720	2.659
	(0.551)***	(0.523)***	(0.515)***	(0.616)***
asc_ln_inc	-0.469	-0.401	-0.505	-0.550
	(0.090)***	(0.087)***	(0.092)***	(0.101)***
asc_incmiss	-2.272	-1.841	-2.440	-2.681
	(0.520)***	(0.505)***	(0.541)***	(0.596)***
asc_use	-0.579	-0.611	-0.477	-0.375
	(0.154)***	(0.146)***	(0.164)***	(0.188)**
asc_polcontrol	-0.752	-0.793	-0.707	-0.835
	(0.159)***	(0.151)***	(0.174)***	(0.192)***
asc_edu35	-0.502	-0.478	-0.254	-0.360
	(0.153)***	(0.145)***	(0.155)	(0.177)**
asc_cvfirst	0.417	0.457	0.330	0.491
	(0.126)***	(0.120)***	(0.131)**	(0.149)***
asc_intsex	0.447	0.475	0.261	0.303
	(0.143)***	(0.139)***	(0.146)*	(0.165)*
Observations	33336	35088	31656	24432
No. Respondents	1389	1462	1319	1006
Pseudo-R2	0.20	0.18	0.23	0.23
Log-Likelihood	-9730.24	-10511.18	-8909.49	-6872.50

Source: NERA; Robust standard errors in parentheses, which take account of the within-respondent correlation of choice behaviour.
** 2-sided* $p < 0.10$, *** 2-sided* $p < 0.05$, **** 2-sided* $p < 0.01$
(1) Variable definitions and summary statistics are provided in Appendix H. The dependent variable in all models summarised here was choice*; and the estimator used was logit.*

Table 10.36 shows that the adopted equation provides a reasonably good statistical fit for the data, with a pseudo-R^2 value of 0.20, indicating that 20% of the variance of choice behaviour is explained by the included variables. Furthermore, all variables enter into the model with the expected signs.

Table 10.37 shows marginal willingness to pay figures derived from the results in Table 10.36.

Table 10.37
Marginal Willingness to Pay Estimates from CE Model
(£ per Household per Year, in Perpetuity)

Variable	Definition	England £/hh/yr/%	Wales £/hh/yr/%	E&W £/hh/yr/%
hl8	Percent High Quality Locally in 8 years	0.71	1.38	0.73
ll8	Percent Low Quality Locally in 8 years	-0.50	-0.97	-0.51
hn8	Percent High Quality Nationally in 8 years	0.84	1.63	0.86
ln8	Percent Low Quality Nationally in 8 years	-0.68	-1.33	-0.70
h20_95	Dummy variable indicating "95% scenario"	0.08	0.15	0.08
asc (low inc)	Dummy variable indicating "No change", for the low income group	-0.56	-1.12	-0.57
asc (med inc)	Dummy variable indicating "No change", for the medium income group	-0.91	-1.80	-0.93
asc (high inc)	Dummy variable indicating "No change", for the high income group	-1.33	-2.62	-1.37
*asc (all incs)**	Dummy variable indicating "No change", for all income groups	-0.98	-1.80	-1.01

Source: NERA analysis, based on Table 10.36 pref *sample results.*
Calculated using average proportions in our sample for use *and* pol_control*, and population means for sex, income, education levels, and the population living in Wales.* cv_first *was set equal to 1 to derive the figures in this table.*
**alternative to use of separate income classes* asc *values.*

The results can be aggregated to the 75% and 95% CV scenarios by inserting the appropriate values for the variables in the table and summing. Table 10.38 presents the proportional improvement values for the CV scenarios, by country.

Table 10.38
Environmental Improvement Under the "75% Scenario" and "95% Scenario"

Variable	Change in Proportions, by Scenario and Country (∆%)					
	"75% Scenario"			"95% Scenario"		
	England	Wales	E&W	England	Wales	E&W
hl8	0.665	0.539	0.654	0.865	0.739	0.854
ll8	-0.601	-0.259	-0.571	-0.601	-0.259	-0.571
hn8	0.600	0.600	0.600	0.800	0.800	0.800
ln8	-0.440	-0.440	-0.440	-0.440	-0.440	-0.440
h20_95	0	0	0	1	1	1
asc	-1	-1	-1	-1	-1	-1

Source: NERA
Changes in local water quality conditions are derive using sample averages of current local conditions. Current conditions are 15% high quality, and 44% low quality in E&W; sample average local high quality is 8.5% for England, 21.1% for Wales and 9.6% for E%W; sample average local low quality is 60% for England 26% for Wales and 57% for E&W.

The wtp values for each scenario, by country, can then be calculated by multiplying the changes in variables in Table 10.38 by the wtp values in Table 10.37. Table 10.39 presents the results from this calculation.

Table 10.39
CE WTP Results for the "75% Scenario" and "95% Scenario" by Country and Income Class

Income Group	Change in Variable, by Scenario and Country					
	"75% Scenario"			"95% Scenario"		
	England (£/hh/yr)	Wales (£/hh/yr)	E&W (£/hh/yr)	England (£/hh/yr)	Wales (£/hh/yr)	E&W (£/hh/yr)
Low income	212.3	363.9	216.7	251.1	439.6	256.6
Medium income	247.3	432.1	252.6	286.2	507.8	292.5
High income	289.3	513.9	295.8	328.1	589.6	335.7
All incomes	254.8	432.4	260.0	293.7	508.0	299.9

Source: NERA; calculated by multiplying the changes in variables presented in Table 10.38 by the wtp values presented in Table 10.37

Under the assumption that the environmental improvements are as shown in Figure 9.2, we can derive an implied discount rate from the results in Table 10.39, and separate wtp values (per household per year) for contemporaneous percentage point changes in the amount of high quality and low quality in the local and national areas. These results can then be used to derive aggregate values for many different types of scenario that might be considered, including values for improvements (from low to medium quality and/or from medium to high quality) to any group of water bodies in England and Wales, over any time period. The values themselves can be scaled to the PCCV or the DCCV results by fixing the aggregate values at, eg the "95% scenario" wtp levels reported in the earlier parts of this section.

The derived implied discount rates and contemporaneous wtp values per percentage point change in each environmental status variable are presented below.

Table 10.40
Derived Contemporaneous Marginal WTP Results and Implied Discount Rates from Fixed Coefficient CE Model, By Country

Environmental Change Variable	Contemporaneous Marginal WTP		
	England (£/hh/yr/%)	Wales (£/hh/yr/%)	England and Wales (£/hh/yr/%)
High Quality Locally	0.17	0.31	0.17
Low Quality Locally	-0.12	-0.22	-0.12
High Quality Nationally	0.20	0.36	0.20
Low Quality Nationally	-0.16	-0.30	-0.17
Implied Discount Rate	10.7%	9.8%	10.7%

Source: NERA
WTP values are presented as perpetual annual payment, however the environmental change being valued is for a single year only.

The estimated discount rates implied by the CE modelling results are higher than the Green Book rate of 3.5%. In principle, there is no reason to expect that individual households should have the same discount rate for all public policies, nor that their discount rates should equal the social rate of time preference that is given by the Green Book rate. The mechanism for eliciting the discount rate used in the present study is limited, however, in its power to determine the true discount rate to be applied to WFD programmes as it is based on a single experiment, while the Green Book rate is based on numbers of studies covering a wide range of contexts. For consistency across public programmes, it would be acceptable to apply the

Green Book rate to derive benefits for alternative WFD improvement scenarios where the timing of the improvements varies. However, a note of caution is warranted. The Green Book discount rate is not consistent with the observed choice responses, and simply imposing this rate while holding the contemporaneous marginal wtp rates of Table 10.37 constant will lead to a substantial over-estimate of the value of potential improvement policies.

Results can also be obtained on contemporaneous marginal wtp for environmental changes given a discount rate constrained to any set value, such as the Green Book rate, by fixing on values estimated for the 8-year environmental attributes (*hl8*, *ll8*, *hn8* and *ln8*). This procedure yields the following values:

Table 10.41
Derived Contemporaneous Marginal WTP Results from Fixed Coefficient CE Model Under Green Book Discount Rate Assumption, By Country

Environmental Change Variable	Contemporaneous Marginal WTP		
	England (£/hh/yr/%)	Wales (£/hh/yr/%)	England and Wales (£/hh/yr/%)
High Quality Locally	0.10	0.19	0.10
Low Quality Locally	-0.07	-0.13	-0.07
High Quality Nationally	0.11	0.22	0.12
Low Quality Nationally	-0.09	-0.18	-0.10
Assumed Discount Rate	3.5%	3.5%	3.5%

Source: NERA
WTP values are presented as perpetual annual payment, however the environmental change being valued is for a single year only.

As Table 10.41 shows, the marginal wtp estimates are significantly lower under the constraint that the discount rate is equal to 3.5%. If this discount rate is to be used, then it is the values in Table 10.41 that should be used, rather than those in the Table 10.40.

10.4 Mixed Logit CE Results

10.4.1 Mixed vs Fixed Coefficient CE Models

The CE analysis presented above is based on standard fixed coefficient logit models, which are by far the most widely used procedure for examining choices, such as the choices made by respondents in the CE tasks. The popularity of the model arises from its simplicity and it robustness. One limitation of logit models is that the coefficients of variables that affect choices are the same for all respondents. However, in reality, willingness to pay varies over people, even over people with the same demographics.

Mixed logit is a generalization of standard logit that allows the coefficients of variables to vary randomly over respondents. Rather than estimating a coefficient for each explanatory variable, the models allow a distribution of the coefficients to be estimated. In particular, mixed logits provides estimates of the distribution of willingness to pay. These distributions can be important for policy analysis. For example, the mean willingness to pay for an improvement is useful for assessing whether the improvement brings more benefits, on average, than its costs; however, the median willingness to pay is important for determining whether a policy is politically acceptable, and whether a ballot proposition regarding the improvement would pass or fail.

Revelt and Train (1998) describe mixed logits for repeated choices, as for the CE tasks. We apply this model to the choices made by respondents in the CE tasks to estimate the distribution of respondents' willingness to pay for each attribute in the CEs. We estimate the models by Bayesian methods, as described by Train (2003, Ch. 12.) The Bayesian methods are faster than classical maximum likelihood (ML) for the types of parameter-rich models that we are estimating. They also avoid many of the numerical difficulties that often arise with ML when the likelihood function is not well-approximated by a quadratic. Importantly, the Bayesian estimator is asymptotically equivalent to ML, such that the estimates can be interpreted from a classical perspective the same as ML estimates. The asymptotic equivalence implies that the two estimators converge towards each other as sample size rises without bound. However, Huber and Train (2001) show that Bayesian and ML estimates of mixed logits provide very similar results even in samples that are far smaller than those used in the current study.

10.4.2 Mixed Logit CE Models

We first estimate a model under the specification that the coefficients of the water quality attributes are normally distributed. We allow the coefficients to be correlated, in order to reflect the possibility that respondents who, e.g., care greatly about one attribute might also care greatly about another. The estimation will reveal the direction and magnitude of such correlations. The coefficient of the bill is held fixed, following the suggestion by Ruud (1996). The willingness to pay for water quality improvements is therefore specified to be normally distributed, with mean and covariance equal to the mean and covariance of the water quality coefficients divided by the bill coefficient.

Table 10.42 presents the estimation results.

Table 10.42
Mixed Logit CE Model with Jointly Normal Coefficients

Variable Name	Definition	Parameter	Estimated Coefficient	Std Error
hl8	Percent High Quality Locally in 8 years	Mean	1.396	0.109
		Variance	3.707	0.652
ll8	Percent Low Quality Locally in 8 years	Mean	-1.186	0.147
		Variance	3.288	0.614
hn8	Percent High Quality Nationally in 8 years	Mean	1.830	0.137
		Variance	7.837	1.127
ln8	Percent Low Quality Nationally in 8 years	Mean	-1.402	0.242
		Variance	4.038	0.920
hn20	Percent High Quality Nationally in 20 years	Mean	2.370	0.121
		Variance	4.159	0.643
bill	Increase in water bill and other household payments	Fixed	-1.591	0.042

Source: NERA
Log-likelihood at convergence: -7273.8

All of the estimated means and variances are significant at the 95% confidence level. The significance of the variances indicates that the hypothesis that the coefficients are fixed, as in a standard logit, can be rejected. The log-likelihood at convergence is -7273.8, which is far higher than that obtained with the standard logit (which was around 10,000 for the specifications with fixed coefficients.) The scale of the coefficients are larger than those from the standard logit. This greater scale reflects the smaller error in the mixed logit compared to the standard logit. The ratios of coefficients, which are the economically meaningful concept, are about the same.

Table 10.43 gives the mean willingness to pay for a 1 percentage point improvement in each water quality attribute, calculated as the estimated means of the coefficients in Table 10.42 divided by the estimated bill coefficient.

Table 10.43
Estimated Mean Mixed Logit CE WTP

Attribute Name	Definition	Mixed Logit CE WTP (£/hh/yr/%)
hl8	Percent High Quality Locally in 8 years	0.88
ll8	Percent Low Quality Locally in 8 years	-0.75
hn8	Percent High Quality Nationally in 8 years	1.15
ln8	Percent Low Quality Nationally in 8 years	-0.87
hn20	Percent High Quality Nationally in 20 years	1.49

Source: NERA

The mean willingness to pay follows the same pattern as in the standard logit: for the eight-year level: a percentage point improvement in national quality is valued more than a percentage point improvement in the local area, and percentage point improvements in high quality are valued slightly more than reductions in low quality. However, unlike the standard logit model, the percent high quality in 20 years nationally is estimated to be valued more than the eight year levels. This perhaps reflects the concept that the water quality in 20 years represents what it will be thereafter, such that even with discounting, "forever" has a higher value than the upcoming 8 years. Overall, the mean willingness to pay for the attributes are somewhat (about 15%) higher than those estimated with the standard logit model. These mean values need to be considered, however, in light of the discussion to follow about the true distribution being highly skewed, which the normal distribution does not reflect.

Table 10.44 gives the estimated correlations among the willingness to pay for each quality attribute. To ease intreprepation, willingness to pay is expressed for an improvement: an increase in high quality, a decrease in low quality. (That is, we reversed the sign for the willingness to pay for percent low quality, such that they are also positive and represent the willingness to pay to *reduce* the percent low quality.)

Table 10.44
Estimated Correlations Between Willingness to Pay Amounts for Environmental Improvement Attributes

Attribute	*hl8*	*ll8*	*hn8*	*ln8*	*hn20*
hl8	1	0.461	0.044	0.450	0.465
ll8		1	0.342	0.487	0.515
hn8			1	0.375	0.425
ln8				1	0.771
hn20					1

Source: NERA

All of the correlations in the above table are positive. A positive correlation between any two attributes means that people who are willing to pay more than average for improvement of one water quality attribute are also willing to pay more than average for the other water quality attribute. The fact that all correlations are positive means that one person who is willing to pay more than another person for an improvement of one kind, also tends to be willing to pay more than that person for improvements of all kinds.

The highest correlation is between *ln8* and *hn20*: people who value decreases in low quality areas nationally in eight years also value increases in high quality areas in 20 years. This high correlation reflects a certain logic: people who are willing to pay more than average to get the worst areas cleaned-up throughout the nation within the next 8 years are also willing to pay more than average to assure high quality water throughout the nation in 20 years.

The estimated variances in Table 10.42 are quite large. This indicates that there are very large differences in how much people are willing to pay for water quality improvements. This fact is important in itself. However, it is important also for suggesting another issue. In particular, the normal distribution is continuous on both sides of zero, and the share on each side of zero depends on the mean and the variance of the distribution. The estimates imply that a sizeable share of the population is estimated to have a *negative* willingness to pay for improvements. In particular, the mixed logit CE model implies the following share of people have negative willingness to pay for improvements in the attribute:

hl8:	23%
ll8:	26%
hn8:	26%
ln8:	24%
hn20:	12%

These shares should not be taken literally as meaning that these people actually need to be paid to accept an improvement. The result is partly an artefact of the specification of a normal distribution: the normal distribution has support on both sides of zero and necessarily implies that some share of the population has negative willingness to pay for improvements. However, the large estimated shares below zero suggest that perhaps a large share of people are willing to pay very little for these quality improvements.

Further results from our limited mixed logit analysis are presented in Appendix L. Given the limited time available for the analysis of the survey results, we consider the mixed logit results obtained to date to be preliminary only, and that there is substantial potential for worthwhile further research, including analysis of alternative modelling and estimation approaches. We believe further research could significantly improve our understanding of the types of preferences held by the public for WFD improvements.

10.5 Brief Overview of Willingness to Pay Findings

Most households are willing to pay a notable sum for improvement scenarios achieving 95% high quality by 2027.

Some households have zero or very low (<£10 per year) willingness to pay for improvements.

Some households are willing to pay many times the median amount, so mean willingness to pay is higher than the median.

There is evidence that for a majority of people delivering improvements earlier than 2027 adds little to their willingness to pay, but some households value this highly, with the result that mean willingness to pay for earlier improvements is substantial.

Across the range of CE results, percentage improvements to the whole of England and Wales are valued only somewhat more than percentage improvements in the local area. The relative size of the areas is around 20-1, implying that 'per sq. mile' values would be much higher for local improvements than for distant ones. This finding is supported by descriptive statistics showing that the feeling of the majority of respondents that making improvements near where people live should be a priority.

The general pattern is for increases in good quality to be valued somewhat more than decreases in low quality, though not at all models exhibit this, and the qualitative work found the reverse.

There is evidence that the households likely to fall into groups where 'affordability' is a concern have lower willingness to pay. There is also a substantial degree of heterogeneity in the population's values that is not explained by measures observed in the data.

The large variation in the willingness to pay evident in the data in all elicitation formats leads estimates of mean willingness to pay to be sensitive to modelling assumptions about the high range in particular, and to the choices made in this range.

While estimates of the mean willingness to pay would generally be the appropriate main foundation for national-level cost-benefit analysis, relying on a hypothetical-compensation test, the distribution of willingness to pay should in this case be fully considered in policy-making.

The estimated willingness to pay values for wtp improvements are sensitive to the elicitation method used in expected ways, with choice-based methods eliciting substantially higher values than the PCCV method. We consider that the true values are very likely to lie between the PCCV estimate and the DCCV mean estimate. We recommend a conservative approach to determining the appropriate estimate of the benefits of WFD improvements, which leads to use of estimates from the lower end of this range.

Marginal values for environmental improvements are estimated much more precisely in the CE due to the substantial additional information available on preferences obtained by this method. Because of this, we recommend that where marginal estimates are required these be taken from the CE, consistent with the discount rate to be used, and scaled to the adopted total value estimate.

11 Aggregation and Sensitivity Analysis

11.1 Benefit Scenarios

Defra have given us the following scenarios for which to report aggregate benefits. We understand these scenarios are subject to change at a later stage.

Table 11.1
Benefit Scenarios

Scenario Name	Description
Maximum benefits	All benefits arise starting in 2015
Front loaded	50% of the benefits start in 2015, followed by 30% in 2021, and 20% in 2027
Even loaded	33% of benefits start in each of the periods 2015, 2021, 2027
Back loaded	20% start in 2015 followed by a further 30% in 2021 and 50% in 2027
Less stringent objectives	25% in each of the first 3 plans then no more after that so it assumes that we apply less stringent objectives equivalent to 25%
Nature assimilation lag	constraints in natural conditions, such as stocks of P in sediment (etc.) mean that 50% of benefits will not occur until 50+ years

Source: DEFRA

Table 11.2 to Table 11.7 translate these scenario descriptions into a form suitable for aggregation from our model based on the current national conditions and the average of current local environmental conditions in the sample.

Table 11.2
Water Environment Status by Year
(Scenario 1 - Maximum Benefits)

Attribute Name	Definition	Environmental Status, by Year				
		Now (%)	2015 (%)	2021 (%)	2027 (%)	2065 (%)
hl0(1)	Percent High Quality Locally	9.6	100	100	100	100
ll0(1)	Percent Low Quality Locally	56.6	0	0	0	0
hn0(2)	Percent High Quality Nationally	15.0	100	100	100	100
ln0(2)	Percent Low Quality Nationally	44.0	0	0	0	0

Source: NERA

Table 11.3
Water Environment Status by Year
(Scenario 2 - Front-Loaded)

Attribute Name	Definition	Environmental Status, by Year				
		Now (%)	2015 (%)	2021 (%)	2027 (%)	2065 (%)
hl0(1)	Percent High Quality Locally	9.6	54.8	81.9	100	100
ll0(1)	Percent Low Quality Locally	56.6	28.3	11.3	0	0
hn0(2)	Percent High Quality Nationally	15.0	57.5	83.0	100	100
ln0(2)	Percent Low Quality Nationally	44.0	22.0	8.8	0	0

Source: NERA

Table 11.4
Water Environment Status by Year
(Scenario 3 - Even-Loaded)

Attribute Name	Definition	Environmental Status, by Year				
		Now (%)	2015 (%)	2021 (%)	2027 (%)	2065 (%)
hl0(1)	Percent High Quality Locally	9.6	39.4	70.2	100	100
ll0(1)	Percent Low Quality Locally	56.6	37.9	18.7	0	0
hn0(2)	Percent High Quality Nationally	15	43.1	72.0	100	100
ln0(2)	Percent Low Quality Nationally	44	29.5	14.5	0	0

Source: NERA

Table 11.5
Water Environment Status by Year
(Scenario 4 - Back-Loaded)

Attribute Name	Definition	Environmental Status, by Year				
		Now (%)	2015 (%)	2021 (%)	2027 (%)	2065 (%)
hl0(1)	Percent High Quality Locally	9.6	27.7	54.8	100	100
ll0(1)	Percent Low Quality Locally	56.6	45.3	28.3	0	0
hn0(2)	Percent High Quality Nationally	15.0	32.0	57.5	100	100
ln0(2)	Percent Low Quality Nationally	44.0	35.2	22.0	0	0

Source: NERA

Table 11.6
Water Environment Status by Year
(Scenario 5 - Less Stringent Objectives)

Attribute Name	Definition	Environmental Status, by Year				
		Now (%)	2015 (%)	2021 (%)	2027 (%)	2065 (%)
hl0(1)	Percent High Quality Locally	9.6%	32.2%	54.8%	77.4%	77.4%
ll0(1)	Percent Low Quality Locally	56.6%	42.4%	28.3%	14.1%	14.1%
hn0(2)	Percent High Quality Nationally	15.0%	36.3%	57.5%	78.8%	78.8%
ln0(2)	Percent Low Quality Nationally	44.0%	33.0%	22.0%	11.0%	11.0%

Source: NERA

Table 11.7
Water Environment Status by Year
(Scenario 6 - Nature Assimilation Lag)

Attribute Name	Definition	Environmental Status, by Year				
		Now (%)	2015 (%)	2021 (%)	2027 (%)	2065 (%)
hl0(1)	Percent High Quality Locally	9.6%	54.8%	54.8%	54.8%	100.0%
ll0(1)	Percent Low Quality Locally	56.6%	28.3%	28.3%	28.3%	0.0%
hn0(2)	Percent High Quality Nationally	15.0%	57.5%	57.5%	57.5%	100.0%
ln0(2)	Percent Low Quality Nationally	44.0%	22.0%	22.0%	22.0%	0.0%

Source: NERA

11.2 Aggregate Estimates of WFD Benefits for England and Wales

Consistent with our view that a conservative approach will focus on benefit estimates from the lower end of the PCCV – DCCV range, we have derived aggregate values using the PCCV mean and PCCV median estimates of household wtp for each of the scenarios above, and for England, Wales and England and Wales. The tables from which these figures are derived are presented in Appendix C. The discount rate used in each case for the aggregation is the Green Book rate of 3.5%. Population numbers used for aggregation are shown in the tables.

CE results are presented for the "95% scenario" only. The numbers derived for the relative values of improvements from low to medium quality versus medium to high quality, and for local versus national-level improvements are potentially very useful for aggregating to bespoke scenarios that may be developed at a later date. Any policy mix that will improve any group of water bodies over any time period can potentially be valued separately. The aggregation model, to follow after this report, allows a range of such scenarios to be valued. For the scenarios described above, there are no variations in local and national improvements, or low-to-medium and medium-to-high improvements, and so there is little additional value in presenting the CE results for all of the scenarios.

Table 11.8 presents our main results on the aggregate values for WFD benefits in England, Wales, and in England and Wales by policy scenario.

Table 11.9 and Table 11.10 beneath this present aggregate value for the WFD benefits calculated from the median PCCV and mean DCCV respectively.

Table 11.8
Aggregate PCCV Estimates of WFD Benefits by Policy Scenario and Country

Policy Scenario	England			Wales			England and Wales		
	Mean WTP	Annual WTP (for 20.5m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 1.2m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 21.7m hh)	PV WTP (@3.5% discount rate)
	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million
Scenario 1 - Maximum	47.4	968.63	27,675.15	42.9	51.89	1,482.43	47.1	1,020.15	29,147.10
Scenario 2 - Front Loaded	40.9	836.10	23,888.55	37.0	44.79	1,279.60	40.7	880.57	25,159.10
Scenario 3 - Even Loaded	38.2	782.25	22,349.93	34.7	41.90	1,197.19	38.0	823.85	23,538.65
Scenario 4 - Back Loaded	35.7	731.06	20,887.37	32.4	39.16	1,118.84	35.5	769.94	21,998.30
Scenario 5 - Less Stringent Objectives	28.7	586.85	16,767.20	26.0	31.44	898.14	28.5	618.06	17,658.99
Scenario 6 - Nature Assimilation Lag	31.3	639.68	18,276.45	28.3	34.26	978.99	31.1	673.70	19,248.51

Source: NERA

Table 11.9
Median PCCV Estimates of WFD Benefits by Policy Scenario and Country

Policy Scenario	England			Wales			England and Wales		
	Median WTP	Annual WTP (for 20.5m hh)	PV WTP (@3.5% discount rate)	Median WTP	Annual WTP (for 1.2m hh)	PV WTP (@3.5% discount rate)	Median WTP	Annual WTP (for 21.7m hh)	PV WTP (@3.5% discount rate)
	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million
Scenario 1 - Maximum	26.8	547.47	15,641.87	24.3	29.33	837.87	26.6	576.58	16,473.81
Scenario 2 - Front Loaded	23.1	472.56	13,501.70	20.9	25.31	723.23	23.0	497.69	14,219.81
Scenario 3 - Even Loaded	21.6	442.12	12,632.08	19.6	23.68	676.64	21.5	465.64	13,303.94
Scenario 4 - Back Loaded	20.2	413.19	11,805.45	18.3	22.13	632.37	20.1	435.17	12,433.34
Scenario 5 - Less Stringent Objectives	16.2	331.69	9,476.75	14.7	17.77	507.63	16.1	349.33	9,980.78
Scenario 6 - Nature Assimilation Lag	18.9	385.97	11,027.62	17.1	20.67	590.70	18.8	406.49	11,614.14

Source: NERA

Table 11.10
Aggregate DCCV Estimates of WFD Benefits by Policy Scenario and Country

Policy Scenario	England			Wales			England and Wales		
	Mean WTP	Annual WTP (for 20.5m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 1.2m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 21.7m hh)	PV WTP (@3.5% discount rate)
	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million
Scenario 1 - Maximum	176.7	3,614.42	106,883.67	193.9	234.46	6,933.26	177.7	3,849.03	113,821.44
Scenario 2 - Front Loaded	152.6	3,119.89	92,259.50	167.4	202.38	5,984.63	153.4	3,322.40	98,248.03
Scenario 3 - Even Loaded	142.7	2,918.94	86,317.23	156.6	189.34	5,599.17	143.5	3,108.41	91,920.05
Scenario 4 - Back Loaded	133.4	2,727.93	80,668.70	146.4	176.95	5,232.76	134.1	2,905.00	85,904.87
Scenario 5 - Less Stringent Objectives	107.1	2,189.83	64,756.28	117.5	142.05	4,200.57	107.7	2,331.97	68,959.58
Scenario 6 - Nature Assimilation Lag	116.7	2,386.94	70,585.13	128.1	154.83	4,578.67	117.4	2,541.87	75,166.78

Source: NERA

11.3 Aggregate Sensitivity Analysis

11.3.1 Sensitivity to Elicitation Method and Model Selection

Table 11.11 presents a comparison of the results obtained from the different models estimated. Sample statistics are also included on the PCCV results for comparison. The tables from which these figures are derived are presented in Appendix C.

Table 11.11
WTP for Water Environment Improvement (95% by 2015) by Elicitation Method*

Elicitation Method / Model	England		Wales		England and Wales	
	Mean WTP £/hh/yr	Median WTP £/hh/yr	Mean WTP £/hh/yr	Median WTP £/hh/yr	Mean WTP £/hh/yr	Median WTP £/hh/yr
PCCV Sample Statistics	49.2	30.0	62.6	50.0	50.4	30.0
PCCV OLS Model	44.8	25.3	40.1	22.7	44.5	25.1
DCCV Turnbull Statistics*	101.6	100-200	129.9	>200	103.0	100-200
DCCV Logit Model	167.0	167.0	181.4	181.4	167.9	167.9
CE Logit Model	293.7	293.7	508.0	508.0	299.9	299.9

Source: NERA
** Note: DCCV Turnbull Statistics for Wales are calculated using choices made for both the 95% and 75% improvement scenarios. There were insufficient respondents in Wales that were shown the 95% scenario to calculate these statistics reliably using only these observations.*

11.3.2 Variations in Valuation within the Population

Table 11.12 shows how wtp varies by income group for the population, for each of the elicitation methods.

Table 11.12
Estimates of WTP for 95% Improvement Scenario by Elicitation Method and Income Group

Elicitation Method / Model	Low Income Group Mean WTP			Med Income Group Mean WTP			High Income Group Mean WTP		
	England	Wales	E&W	England	Wales	E&W	England	Wales	E&W
PCCV Sample Statistics	44	51.7	44.9	55.9	83.2	58.7	70.5	105.0	72.9
PCCV OLS Model	33.6	31.7	33.5	42.6	40.1	42.4	56.5	53.2	56.3
DCCV Logit Model	131.5	152.0	132.6	160.8	181.3	161.9	195.9	216.4	197.0
CE Logit Model	251.1	439.6	256.6	286.2	507.8	292.5	328.1	589.6	335.7

Source: NERA

There is also substantial heterogeneity in the population that is not explained by observed characteristics of respondents in the sample. This is shown in particular by the mixed logit results.

12 Conclusions

The research undertaken in CRP Projects 4b and 4c and presented in this report has designed and implemented a stated preference approach to valuing the benefits placed by households in England and Wales on water environment improvements brought about by the Water Framework Directive.

A representation of the water environment, to reflect the changes which are projected to be brought about by the WFD, was developed. This representation reduces the dimensions and levels which would be required for a full "scientific" depiction of the state of the water environment across the whole of England and Wales, projected for many years, to a much smaller set that household respondents are able to understand in the course of an interview. The representation also reflects the main WFD policy dimensions, being improvements of various extents by particular dates. The representation was developed by multiple iterations between the project team and a science group, and through a large number of test groups and pilot surveys with household respondents.

As expected, the stated WFD benefit values varied with the method employed to elicit them. In particular, for a set package of local and national WFD improvements achieved by set times, our payment card contingent value question produced lower stated willingness to pay answers than our dichotomous choice contingent value question, for a substantial number of respondents. Also, the sum of the marginal values across individual dimensions of improvement derived from our choice experiment answers implies a value for the package of improvements which is greater than the dichotomous choice contingent value mean and the payment card contingent value mean.

Table 12.1 presents a comparison of results from the differing elicitation methods we have used, for England, Wales and England and Wales, for the WFD improvement package leading to 95% high quality by 2015.

Table 12.1
WTP for Water Environment Improvement (95% by 2015) by Elicitation Method*

Elicitation Method / Model	England		Wales		England and Wales	
	Mean WTP £/hh/yr	Median WTP £/hh/yr	Mean WTP £/hh/yr	Median WTP £/hh/yr	Mean WTP £/hh/yr	Median WTP £/hh/yr
PCCV Sample Statistics	49.2	30.0	62.6	50.0	50.4	30.0
PCCV OLS Model	44.8	25.3	40.1	22.7	44.5	25.1
DCCV Turnbull Statistics*	101.6	100-200	129.9	>200	103.0	100-200
DCCV Logit Model	167.0	167.0	181.4	181.4	167.9	167.9
CE Logit Model	293.7	293.7	508.0	508.0	299.9	299.9

Source: NERA
** Note: DCCV Turnbull Statistics for Wales are calculated using choices made for both the 95% and 75% improvement scenarios. There were insufficient respondents in Wales that were shown the 95% scenario to calculate these statistics reliably using only these observations.*

The effects of elicitation methods on stated values are still a research area in the environmental valuation literature, but the observed effects are consistent with strategic behaviour where the payment card answers are biased downward: those whose willingness to pay is lower than what they think they will eventually be made to pay if the project goes ahead answer zero to the payment card willingness to pay question; those whose willingness to pay is above what they think the project will cost push their stated willingness to pay on the payment card down towards this cost level, hoping this will lessen the chance of being made to pay more if the project goes ahead. For different reasons, the payment card answers may be lower because of the greater focus of this method on the money sum, or in reaction to the wide range of value answers presented on the card, by comparison with the formats inherent in dichotomous choice questions which have all variables presented alongside each other and only one alternative money sum.

A single dichotomous choice question is generally regarded as incentive compatible in the sense that, if the respondent believes the offered alternative payment is what she will actually pay if the project proceeds, then a truthful answer is the best strategy. However in our case most respondents to the dichotomous choice question saw many possible payment figures first, an inevitable result of varying the order in which the questions were presented to check for ordering effects. There may then be an effect where the number of respondents stating they are willing to pay the lower figures falls (because they think the real figure they will be asked to pay will be more like an average dragged up by the predominantly higher figures seen earlier) and the number stating they are willing to pay the higher figures rises (because they think that all they will really be asked to pay is more like an average with the predominantly lower figures seen earlier). Because the willingness to pay values are all bounded below by zero this averaging belief effect will usually increase the estimated mean dichotomous choice about its true value.

Our results are consistent with these effects in that there are substantially more payment card zeros than is suggested by the dichotomous choice results, and substantially fewer respondents giving high payment card answers than is suggested by the dichotomous choice results. Of the respondents giving apparently inconsistent answers to the payment card and dichotomous choice questions, many more give “relatively too low” payment card answers than give “relatively too high” payment card answers. When respondents showing either type of inconsistency are dropped from the sample, the mean of the payment card stated values increases and the mean of the dichotomous choice stated values decreases towards a mid-point.

Given the current state of understanding of elicitation effects in general and as specifically present in our data we cannot conclude that the mid-point between the payment card and dichotomous choice mean willingness to pay values is the best estimate of the “true” mean willingness to pay for the WFD improvements as we have represented them, though we consider that that figure is very likely to lie in the range between the PCCV and DCCV means (£44.5 to £167.9 per household per year for the 95% package for England and Wales).

A further important caveat is that all the values we have elicited have been provided by respondents for their own understanding(s) of WFD improvement scenarios as we have represented them. The representation was developed with both care and substantial testing, the survey was carefully administered, and the respondents’ behaviour and answers satisfy a suite of validity tests. Nevertheless in a study this complex it is very difficult to be sure that respondents developed and valued “accurate” internal views of the WFD improvements as we have represented them, and there will always be some reasons to be cautious. Among the reasons for caution here are that the respondents tended to find the map representation worse than what they had expected (though they accepted it), and that their value responses were significantly lower when the interviewer was female.

Given the reservations expressed above we consider that a conservative approach to the valuation of WFD improvements is appropriate. Beyond being conservative when making choices during the modelling as we have done, and checking the sensitivity of the results to benefit values from the PCCV-DCCV range of mean values, a conservative approach implies that cost benefit tests of WFD programmes of measures using our evidence would rely largely on mean benefit values drawn from the lower end of the PCCV-DCCV mean range. Accordingly, we present aggregate values based on the mean PCCV figure, allowing for country differences in explanatory factors.

Table 12.2 presents our aggregate estimates of the values of six alternative WFD improvement scenarios for England, Wales and England and Wales, based on the PCCV mean. Across the range of scenarios, the total present value of WFD benefits based on the mean PCCV figures ranges from £18bn to £29bn, for England and Wales.

On a different front, the value responses in the survey for WFD improvements differ greatly across the population irrespective of the elicitation method. It is usual in cost benefit analysis to focus on mean or expected values, implicitly or explicitly invoking a hypothetical compensation test. However the great diversity of values and large potential cost sums involved in WFD implementation suggest that the implications of adopting other social welfare functions or decision rules should be explored. Applying a voting rule is consistent with a focus on median, rather than mean values. Using a social welfare function which gives more weight to respondents with low benefit valuations is likely to assist the group with "affordability" issues, though some respondents in this set did express substantial willingness to pay in our survey.

The relative values of improvements in different places or times are best approached with the relativities derived from our choice experiment work, chosen to be consistent with the adopted discount rate and scaled to match the adopted total benefit value figures where consistency is necessary.

The dataset of values obtained in our survey is very rich, and further work on a number of aspects of valuation may lead to interesting insights into environmental and WFD valuation in general. Further modelling investigations of different functional forms, and of mixed logit approaches in particular, may be fruitful. More broadly, more work on the dependence of values on features of the instrument, the interview setting, and the representation of national scenarios, would be useful for future researchers.

On a different front, the values we have estimated can be a basis for approaches to aggregation which allow more for the variations in local conditions. This would be a small step towards the ideal of benefit evaluations much more tailored to the many varied real and specific features of actual environmental situations which will potentially be affected by the implementation of the Water Framework Directive

Table 12.2
Aggregate PCCV Estimates of WFD Benefits by Scenario

Policy Scenario	England			Wales			England and Wales		
	Mean WTP	Annual WTP (for 20.5m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 1.2m hh)	PV WTP (@3.5% discount rate)	Mean WTP	Annual WTP (for 21.7m hh)	PV WTP (@3.5% discount rate)
	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million	£/hh/yr	£million / yr	£million
Scenario 1 - Maximum	47.4	968.63	27,675.15	42.9	51.89	1,482.43	47.1	1,020.15	29,147.10
Scenario 2 - Front Loaded	40.9	836.10	23,888.55	37.0	44.79	1,279.60	40.7	880.57	25,159.10
Scenario 3 - Even Loaded	38.2	782.25	22,349.93	34.7	41.90	1,197.19	38.0	823.85	23,538.65
Scenario 4 - Back Loaded	35.7	731.06	20,887.37	32.4	39.16	1,118.84	35.5	769.94	21,998.30
Scenario 5 - Less Stringent Objectives	28.7	586.85	16,767.20	26.0	31.44	898.14	28.5	618.06	17,658.99
Scenario 6 - Nature Assimilation Lag	31.3	639.68	18,276.45	28.3	34.26	978.99	31.1	673.70	19,248.51

References

Adamowicz W.K., G. D. Garrod and K. G. Willis (1995) *Estimating passive use benefits of Britain's inland waterways*, Report to British Waterways, UK

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

Azevedo, C.D., K.J. Egan, J.A. Herriges, and C.L. Kling (2003). *Iowa Lakes Valuation Project: Summary and Findings from Year One*. Final Report to the Iowa Department of Natural Resources

Bateman I.J, B. Day, D. Dupont, S. Georgiou, N. G. N. Matias, S. Morimoto and L. Subramanian (2006) Does phosphate treatment for prevention of eutrophication pass the cost-benefit test? *CSERGE Working Paper,* EDM 06-13.

Bateman, I. J., R. T. Carson, B. Day, W. M. Hanemann, N. Hanley, T. Hett, M. Jones-Lee, G. Loomes, S. Mourato, E. Özdemiroğlu, D. W. Pearce, R. Sugden and J. Swanson (2002) *Economic Valuation with Stated Preference Techniques: A Manual,* Edward Elgar Publishing, Cheltenham, UK

Bockstael, N. E., M. Hanemann and I. Strand (1989) *Measuring the benefits of water quality improvements using recreation demand models: Parts 1 and 2*, reports to the U.S. Environmental Protection Agency

Campbell, D., W. G. Hutchinson, and R. Scarpa (2006a) *Using mixed logit models to derive individual-specific WTP estimates for landscape improvements under agri-environmental schemes: evidence from the Rural Environment Protection Scheme in Ireland*, Rural Economy Research Centre, Working Paper 06-WP-RE-07

Campbell, D., W. G. Hutchinson, R. Scarpa, T. O'Leary, A. McCormack and B. Riordan (2006b) *Quantifying the landscape benefits arising from the Rural Environment Protection Scheme: results from a public survey*, Rural Economy Research Centre, Working Paper 06-WP-RE-08

Campbell, D., W. G. Hutchinson, and R. Scarpa (2006c) *Integrating landscape improvement indices and discrete choice experiments: evidence from the Rural Environment Protection Scheme in Ireland*, Rural Economy Research Centre, Working Paper 06-WP-RE-09

Campbell, D., W. G. Hutchinson, and R. Scarpa (2006d) *Lexicographic preferences for rural environmental landscape improvements: implications on individual-specific willingness to pay estimates*, Rural Economy Research Centre, Working Paper 06-WP-RE-10

Carson, R.T. and R. C. Mitchell (1993) The Value of Clean Water: The Public's Willingness to Pay for Boatable, Fishable, and Swimmable Quality Water, *Water Resources Research* **29**(7) 2445-2454.

Carson, R.T., N. E. Flores and N. F. Meade (2001) Contingent Valuation: Controversies and Evidence, *Journal of Environmental and Resource Economics*, **19**, 173-210

Champp, P.A., K. J. Boyle and T. C. Browne (eds) (2003) *A Primer on Non-Market Valuation*, Dordrecht, Kluwer

Chilton, S., J. Covey, M. Jones-Lee, G. Loomes and H. Metcalf (2004) *Valuation of Health Benefits Associated with Reductions in Air Pollution*. Report to DEFRA, UK.

Christie M., J. Warren, N. Hanley, K. Murphy, R. Wright, T. Hyde and N. Lyons (2004) *Developing measures for valuing changes in biodiversity,* Final Report to DEFRA, UK

CIS (2003) Overall Approach to the Classification of Ecological Status and Ecological Potential, Water Framework Directive Common Implementation Strategy

Desvouges, W., V.K. Smith, and M. McGivney (1983) *A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements*, Report to US EPA

Eftec (2002) *Valuation of Benefits to England and Wales of a Revised Bathing Water Quality Directive and Other Beach Characteristics Using the Choice Experiment Methodology,* A Report to DEFRA, UK

Eftec (2006) *Valuing our Natural Environment* A Report to DEFRA, UK

Environmental Economics Advisory Committee (1993) *Review of the contingent valuation method for the proposed RIA for RCRA Corrective Action Rule*, A report for EPA, US

Fisher, J., H. Sunman and N. Tambe (2002) *The environmental benefits of the environmental programme in the Periodic Review of the water industry (PR04). Final report.* Environment Agency Economics Unit Research Report No. 2002/1.

FWR (1996) *Assessing the Benefits of Surface Water Quality Improvements Manual*, FWR, Marlow, UK

Freeman, A. M., (1982) *Air and Water Pollution Control: A Benefit Cost Assessment*, John Wiley, New York.

Georgiou, S., I. J. Bateman, M. Cole, and D. Hadley (2000) Contingent ranking and valuation of river water quality improvements: testing for scope sensitivity, ordering and distance decay effects. *CSERGE Working Paper,* GEC 2000-18.

Goldberger, A. S. (1968) The interpretation and estimation of Cobb-Douglas functions, *Econometrica*, **35**, 464-472

Gramlich, F. W. (1977) The demand for clean water: The case of the Charles River, *Nat. Tax J.* **30**: 183-194.

Hoehn, J. P. and A. Randall (1989) Too many proposals pass the benefit cost test, *American Economic Review,* **79**(3), 544-551.

Huber, J. and K. Train (2001), On the similarity of classical and Bayesian estimates of individual mean partworths, *Marketing Letters* **12**, 259–269.

Huber, J. and W. K. Viscuzi (2006) *Economics of Environmental Improvement*, Report to EPA, USA

Ipsos-MORI (2006) *Valuing the Water Environment: A Survey of Scottish Public Attitudes*, Scottish Executive Social Research, UK

Jacobs (2006) *Report on Non-Use Benefits Valuation* A Report to the Environment Agency, UK

Kahneman, D. and J. L. Knetsch (1992) Valuing public goods: The purchase of moral satisfaction, *J. Environ. Econ. Manage*., **22**, 57-70.

Kaplowitz, M. F. Lupi and J. P. Hoehn (2004) Multiple methods for developing and evaluating a stated choice questionnaire to value wetlands, in Presser, S. et al (2004) *Methods for Testing and Evaluating Survey Questionnaires*, Wiley-Interscience

Li, C. J. Kuuluvainen, E. Pouta, M. Rekola and O. Tahvonen (2004), Using Choice Experiments to Value the Natura 2000 Nature Conservation Programs in Finland, *Environmental and Resource Economics*, **29**(3) 361-374

Magat, Wesley., Joel Huber and W. Kip Viscuzi (2000) An Iterative Choice Approach to Valuing Clean Lakes, Rivers and Streams, *Journal of Risk and Uncertainty*, **21**(1) 7-43

McClelland, G.H., W.D.Schulze, J.K.Lazo, D.M.Waldman, J.K.Doyle, S.R.Elliott, and J.R.Irwin (1992) *Methods for measuring non-use values: A contingent valuation study of groundwater cleanup*, A report for US EPA

McClelland, G.H., W.D.Schulze and J.K.Lazo (1993) *Additional explication of methods for measuring non-use values: A contingent valuation of groundwater clean-up*, A report for US EPA

Mitchell, R. C. and R. T. Carson (1981) *An experiment in determining willingness to pay for national water quality improvements*, report to the U.S. Environmental Protection Agency, Resources for the Future, Washington, D. C.

Mitchell, R. C. and R. T. Carson (1984) *A contingent valuation estimate of national freshwater benefits*, report to the U.S. Environmental Protection Agency, Resources for the Future, Washington, D. C.

Mitchell, R. C., and R. T. Carson (1986) *The use of contingent valuation data for benefit/cost analysis in water pollution control*, report to the U.S. Environmental Protection Agency, Resources for the Future, Washington, D. C.

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.

Mitchell, R.C. and R.T.Carson (1989) *Existence values for groundwater protection,* report to the U.S. Environmental Protection Agency

Mourato S., S. Georgiou, E. Özdemiroglu, J. Newcombe, and A. Howarth (2003) *Bathing water directive revisions: what are the benefits to England and Wales? A stated preference study.* CSERGE Working Paper ECM 03-12.

MVA (2003) *Joint National Research into Customers' Views on Water and Sewerage Services: Stage II*, research commissioned by Department for Environment, Food and Rural Affairs, Welsh Assembly Government, Office of Water Services, WaterVoice, Water UK, Environment Agency, Drinking Water Inspectorate, English Nature, Wildlife and Countryside Link

Pearce, D., G. Atkinson and S. Mourato (2006) *Cost-Benefit Analysis and the Environment*, Paris, OECD

Revelt, D. and K. Train (1998), Mixed logit with repeated choices, *Review of Economics and Statistics* **80**, 647–657.

Ruud, P. (1996), *Simulation of the multinomial probit model: An analysis of covariance matrix estimation*, Working Paper, Department of Economics, University of California, Berkeley.

Scarpa, R. (2003) *The Recreational Value of Woodlands*, Social and Environmental Benefits of Forestry Project, Phase 2, Report to the Forestry Commission, Edinburgh, UK

Smith, V. K., and W. H. Desvousges (1986) *Measuring Water Quality Benefits*, Kluwer, Academic, Hingham, Mass.

Train, K. E. (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press

United States Environmental Protection Agency (1994) *President Clinton's Clean Water Initiative: Analysis of Benefits and Costs*. EPA 800-R-94-002 (NTIS Document No. PB94-154101). Washington DC: Office of Water United States Environmental Protection Agency.

Willis K.G., R. Scarpa and M. Acutt (2005) Assessing water company customer preferences and willingness to pay for service improvements: a stated choice analysis. *Water Resources Research,* **41**(2) WO2019.1-WO2019.11.

WRc, Oxera and Institute of Hydrology (1999) *Potential Costs and Benefits of the implementing the Proposed Water Resources Framework Directive*, Final Report to DETR, UK

WRc (2007) Scoping the basic science input into the WFD benefits analysis: Technical support and production of maps of water body quality, A report to DEFRA, UK

NERA

Economic Consulting

NERA Economic Consulting
15 Stratford Place
London W1C 1BE
United Kingdom
Tel: +44 20 7659 8500
Fax: +44 20 7659 8501
www.nera.com

NERA UK Limited, registered in England and Wales, No 3974527
Registered Office: 15 Stratford Place, London W1C 1BE

MMC Marsh & McLennan Companies