Appendix 5g:

Understanding Customer Values_Revealed preference River Quality Report

AECOM

PR19 Understanding Customer Values: Work Package 3 – Revealed Preference Visitor Survey

Prepared for Yorkshire Water

1 Acknowledgements

AECOM and University of Exeter Consulting Ltd. would like to thank Professor Mike Christie (Aberyswyth University) for providing peer review of the work undertaken for this work package.

2 Quality information

Document name	Prepared for	Prepared by	Date	Approved by
Work Package 3	Yorkshire Water	Brett Day	01/11/17	Chris White
Work Package 3	Yorkshire Water	Brett Day	30/11/17	Chris White

3 Contents

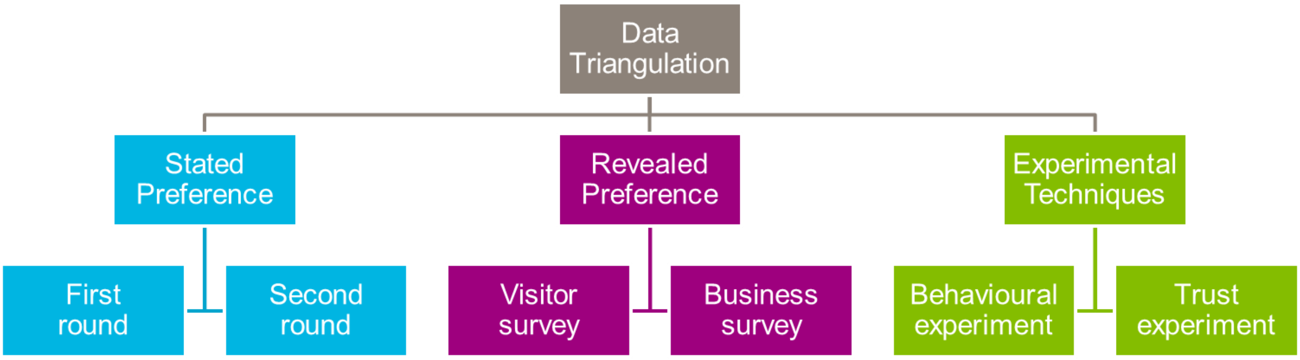
4 Work Package 3 – Visitor Survey

5 Context

The aim of this project is to undertake primary research to ascertain the values that Yorkshire Water (YWS) customers place on changes in service measures such as supply interruptions or drinking water failures. These values will then be used to populate the Decision Making Framework (DMF) in order to inform the investment planning process and support the wider Outcome Delivery Incentives (ODI) work stream.

In light of Ofwat's recommendations for improving the approach to understanding customer's values in PR19, the project includes six work packages (see Figure 1) which draw on a range of data to allow methodological triangulation; whereby data of different types are used to cumulatively refine and validate research outputs. The work summarised here pertains to Work Package 3 (the 3rd purple box in the final row of Figure 1); which covers the valuation of river water quality improvements using data collected from a visitor survey.

Figure 1. Overview of the six work packages



6 Aims

The aim of this work package is to develop estimates of the welfare values of river water quality improvements in the Yorkshire region. The work package focuses particularly on deriving a monetary value for the benefits of river water quality improvements that accrue directly to households living in the region; that is to say, YWS' customers.

There are at least two ways in which individuals might directly benefit from improved rivers. First, improvements might enhance the experience of recreational trips that involve spending time by, on, or in those rivers. Second, individuals may value the improvements not because they use those rivers for recreation, but simply because they prefer a world in which the region's rivers are of high water quality. Economists define this distinction as being one between 'use' values (in this case, arising from recreation) and 'non-use' values.

The aim of this work package is to bring together revealed and stated preference data to inform estimates of the value of water quality improvements in the Yorkshire region. The core of that data derives from a large-scale survey undertaken in the region in 2008 as part of the RELU-funded

Catchment Hydrology, Resources, Economics and Management (ChREAM) project.

While the data have been subject to some previous analysis, only recently have the research team developed the analytical methods that allow for the simultaneous analysis of the revealed and stated preference data. Indeed, one of the major achievements of the research presented in this work package is the estimation of models from the combined data that coherently inform on both the use and non-use values that are derived from water quality improvements.

While the model developed in this project is capable of making nuanced predictions regarding use and non-use values for any manner of river water quality changes across the Yorkshire region, the central objective of this work packge is more focused. Specifically, the goal is to develop a small set of generic values for water quality improvements that are suitable for use by YWS in a range of different project appraisal settings.

7 Method

This work package involved two approaches to quantifing and valuing the benefits YWS customers' receive from changes in river water quality. The first was a travel cost modelling approach based on rational choice behaviour, in which households choices of which recreational river site to visit are assumed to reflect the trade-offs they are prepared to make between the quality of that recreational site and the costs of travelling to that site. Using a dataset recording the observed recreational choices of 1,805 YWS customers, it was possible to use this approach to build a model of recreational choice behaviour. That model informs estimates of how households value environmental qualities and can be used to predict how their welfare might change if, for example, the river water quality at one or more recreational sites were to change.

The second approach was to ask the same 1,805 YWS customers to express their values for water quality improvements in a stated preference exercise. While the concept of stated preference surveys is relatively straightforward, their effective application to the valuation of river water quality is actually rather complex. One particular problem concerns how to convey to respondents the complex spatial reality of the landscape within which hypothetical changes in environmental quality occur. The approach adopted in this work package was that of a Visual Spatial Choice Experiment (VSCE); where hypothetical scenarios are presented to respondents in the form of colour-coded and annotated maps. In a VSCE, respondents are presented with a selection of such maps each illustrating a different spatial pattern of quality change and each associated with some particular cost. Of those on display, respondents are asked to identify which costly pattern of quality change is their most preferred.

Map-based presentations like the VSCE allow stated preference studies to elicit preferences for complex landscape-wide patterns of environmental change, explicitly presenting respondents with information on the quality and location of substitutes upon which they may condition their responses. How those responses should be modelled to properly reflect respondents' decision processes, however, remains an open question in the stated preference literature. One of the key methodological innovations of this work package is that it has developed a statistical approach to the analysis of VSCE data derived from a coherent model of preferences for landscape-wide environmental quality change. The starting point for that model of preferences is the same rational choice behaviour that underpins the discrete-choice travel cost model. Details of the data and methods applied in this workpackage can be found in the Appendix.

One important consequence of having a model of choices in the stated preference exercise that is based on the same assumptions as that describing choices in the revealed preference data is that it allows for those two datasets to be used simultaneously in the statistical analysis. By combining data sets this analysis takes advantage of the strengths of both. The travel cost data allows the approach to tie down elements of the preference function describing value from the use of rivers that are only poorly identified by the VCSE data. The VCSE data allows the approach to establish elements of the preference function relating to non-use value of rivers upon which the travel cost data provides no information.

Once estimated, the model of preferences for river qualities provides a spatially explicit value function that distinguishes value derived from use from that derived from non-use. More specifically, that model allows the approach to predict the use and non-use values that might be enjoyed by some particular households as a result of some particular set of changes in water quality in some specific set of river locations. Of course, that level of specificity is of little use to YWS. Rather, for the purposes of informing a wide range of different investment decisions what is required are more generic measures that are indicative of the value of water quality change to YWS customers in general.

This work package, therefore, developed a method that estimated those generic values by aggregating up values for all YWS customers and then finding the average of those population values across quality changes occuring at a variety of different locations across the region. More specifically, the approach began by considering some particular water quality change that might be of interest to YWS, for example, improving 10 km of river that currently has poor water quality up one level to the good category (where quality categories refer to the WFD ecological status classification). To find a generic value for such an improvement, the approach identified all river locations in the landscape that are currently poor quality, then randomly selected 10 km worth of those locations.

The value function estimated in the statistical analysis was then used to calculate how much value those particular set of quality changes bestow on each household in the Yorkshire region, a set of values that can then be summed to provide an aggregatge value for YWS customers. By repeating those calculations many times for different randomly selected sets of 10 km worth of river locations, the approach built up a distribution of such aggregate values. The figures taken as a generic aggregate value for improvements in 10 km of river from poor to good quality are taken to be the average of that range of values based on calculations from 100 random samples.

8 Results

Using the WFD ecological status classification, **Table 1** to **Table 3** present the results of the average aggregate welfare change associated with a one category improvement in water quality status. The results are shown for three scenarios:

- Average aggregate values for improvements from ‘bad’ to ‘poor’ quality
- Average aggregate values for improvements from ‘poor’ to ‘moderate’ quality
- Average aggregate values for improvements from ‘moderate’ to ‘good/high’ quality

For each scenario the value was estimated for three different sizes of programme; interventions that resulted in 1 km, 10 km and 50 km of river enjoying improvement. For each size of programme the aggregate value is decomposed into the benefits that households in the Yorkshire region enjoy from the improvement as a result of recreational use of the improved site (use value) and as a result of increased non-use value. The final column of each table records the total welfare improvement as the sum of use and non-use value.

Table 1. Average aggregate values for improvements from 'bad' to 'poor' quality

Length of River Improved	Average Aggregate Value		
	Use	Non-Use	Total
1 km	£8,566	£70,857	£79,423
10 km	£53,461	£689,348	£742,809
50 km	£288,030	£2,845,077	£3,133,107

Note: Values are expressed in 2008 £s per annum for households in Yorkshire region

Table 2. Average aggregate values for improvements from 'poor' to 'moderate' quality

Length of River Improved	Average Aggregate Value		
	Use	Non-Use	Total
1 km	£79,890	£123,444	£203,335
10 km	£353,938	£1,215,162	£1,569,100
50 km	£1,309,573	£4,963,025	£6,272,598

Note: Values are expressed in 2008 £s per annum for households in Yorkshire region

Table 3. Average aggregate values for improvements from 'moderate' to 'good/high quality

Length of River Improved	Average Aggregate Value		
	Use	Non-Use	Total
1 km	£3,157	£75,206	£78,363
10 km	£31,962	£762,092	£794,054
50 km	£152,849	£3,726,284	£3,879,133

Note: Values are expressed in 2008 £s per annum for households in Yorkshire region

While the values provided in **Table 1** to **Table 3** concern improvements in river cells of a particular current water quality level, the key objective of this work package is to provide a value that can be used in the DMF. **Table 4** provides those figures. Here rather than selecting only from river locations of one particular quality, river locations were selected at random and for each randomly chosen location the value of a one category increase in river water quality was calculated.

The procedure can be justified as being a best *a priori* guess at the value that would be generated by a future scheme delivering one category water quality improvements at some as yet to be specified location in the region; assuming that that the policy could be implemented in any location in the region with equal likelihood.

Table 4. Average aggregate values for general one category improvement in river water quality

Length of River Improved	Average Aggregate Value		
	Use	Non-Use	Total
1 km	£6,323	£77,228	£83,551
10 km	£42,331	£800,439	£842,771
50 km	£241,358	£3,911,587	£4,152,945

Note: Values are expressed in 2008 £s per annum for households in Yorkshire region

It is important to note that the model can be used to value any set of water quality changes across the landscape. Indeed, the model could be used to identify locations in which improvements deliver the most value. To illustrate this, Table 5 shows the values that could be achieved if improvements delivering a two-category increase in water quality were targeted at locations offering the most value to YWS customers. Notice how the targeting of improvements leads to at least a doubling in the value realised from those investments with some of the biggest gains being seen in use values.

Table 5. Aggregate values for spatially targeted two-category improvement in river water quality

Length of River Improved	Average Aggregate Value		
	Use	Non-Use	Total
1 km	£61,471	£315,685	£377,156
10 km	£191,882	£1,630,766	£1,822,648
50 km	£3,480,032	£5,427,393	£8,907,425

Note: Values are expressed in 2008 £s per annum for households in Yorkshire region

9 Implications

This work package provides a state of the art analysis of a data set recording stated and revealed preferences for water quality improvements in the Yorkshire region. The revealed preference data came in the form of records of recreational trips to rivers of different water qualities across the region. The stated preference data came in the form of choices over different hypothetical future patterns of water quality across the region.

- The research developed an original and state-of-the-art estimation method that allowed for the simultaneous analysis of the revealed and stated preference data. That method allowed for the coherent separation of welfare estimates into a use and non-use values.

- The model estimated from the data returns parameter estimates that broadly conform to prior expectations both in sign and magnitude. Most importantly, the models report significant sensitivity to river water quality both in observed choices over recreational sites and in stated choices over hypothetical future patterns of river water quality.

- The estimated model can be used to calculate welfare values at a fine resolution; that is to say, the value to residents of the Yorkshire region of some particular change in river water quality at some particular location in the region. The key objective of the research project, however, has been to develop generic welfare estimates that might be used across a range of appraisal scenarios. To that end, a procedure for estimating average aggregate welfare values was developed. The final output of that undertaking is reported in Table 4.

- Those generic values indicate that a one category increase in river water quality over a 1 km extent of river has an annual welfare value of around £85,000 to households in the Yorkshire Water region. Of that total value, around 10% is derived from recreational use of rivers the remainder coming in the form of non-use value.

- The generic welfare values scale roughly in proportion to the extent of river improved such that one category improvements to 10 km of river deliver roughly £850,000 of value per year and 50 km of improved river delivers roughly £4 million of value per year.

- It is important to note that the careful spatial targeting of interventions in order to deliver water quality improvements in areas that deliver the greatest benefit to YWS customers can realise returns than are at least double those suggested by the generic values.

Appendix 1: Methodology

10 Introduction

The purpose of the research recorded in this report is to develop estimates of the welfare values of river water quality improvements in the Yorkshire Water region, where water quality is defined according to the WFD ecological status classification. The report focuses particularly on deriving a monetary value for the benefits of river water quality improvements that accrue directly to households living in the region; that is to say, Yorkshire Water's customers. So, while improving river quality may ultimately benefit those customers through, say, reducing water bills or boosting the local economy as a result of increased tourism, the focus here is just on directly enjoyed benefits.

There are at least two ways in which individuals might directly benefit from improved rivers. First improvements might enhance the experience of recreational trips that involve spending time by, on or in those rivers. Second, individuals may value the improvements not because they use those rivers for recreation, but simply because they prefer a world in which the region's rivers are of high quality. Economists define this distinction as being one between *use values* (in this case, arising from recreation) and *non-use values*.

Deriving estimates of the value of water quality improvements requires analysts to observe how much money individuals are prepared to trade off in return for enjoying an increased level of quality. One situation in which such trade-offs might be observed is where individuals might incur real costs so as to enjoy the benefits of water quality. In this project, for example, we are going to use the expense incurred in travelling to a river for the purposes of recreation as an insight into the trade-off between money and water quality. Alternatively, individuals can be presented with carefully crafted survey instruments that ask them to make decisions that involve making some hypothetical payment in return for an improvement in river water quality. In the economics literature, observations of real trade-offs are termed *revealed preference* data and observations of hypothetical trade-offs are termed *stated preference* data.

While the debate continues, there are clearly advantages to both types of data. Obviously, revealed preference data is preferred on the grounds that it involves actual rather than hypothetical trade-offs. That fact is often used to support the validity of value estimates derived from revealed preference data. On the other hand, the real world might not present opportunities for gathering revealed preference data that can be used to estimate values. As an example, it is impossible to use revealed preference data to establish the value of a programme that intends to take river water quality beyond the range presently observable in the real world. Accordingly, stated preference methods might be useful in developing richer datasets that inform on trade-offs that might be impossible to observe in actual behaviour. Along the same lines, stated preference methods provide the only tool capable of capturing non-use values. By definition such values leave no trace in people's actual behaviour.

As described in this report, we bring together both revealed and stated preference data to inform on the value of water quality improvements in the Yorkshire Water region. The core of that data derives from a large-scale survey undertaken in the region in 2008 as part of the RELU-funded Catchment Hydrology, Resources, Economics and Management (ChREAM) project. While the data have been subject to some previous analysis (see Bateman et al., 2016), only recently have the research team developed the analytical methods that allow for the simultaneous analysis of the revealed and stated preference data. Indeed, one of the major achievements of the research presented in this report is the estimation of models from the combined data that coherently inform on both the use and non-use values that are derived from water quality improvements.

While the model developed in this project is capable of making nuanced predictions regarding use and non-use values for any manner of river water quality changes across the Yorkshire Water region, the central objective of this document is more focused. Specifically, the goal is to develop a small set of generic values for water quality improvements that are suitable for use by Yorkshire Water in a range of different project appraisal settings. As shown in the final chapter of this report we have succeeded in developing a procedure that delivers to that objective, returning generic value estimates of the use and non-use values that might be enjoyed by Yorkshire Water customers as a result of programmes designed to improve water quality.

11 Methods

11.1 Revealed Preferences: The Travel cost method (TCM)

Travel cost modelling is a revealed preference technique primarily used to establish the value of environmental quality enjoyed through recreation. At the core of the method is the observation that individuals spend time and money in order to visit recreational sites. Roughly speaking, the method identifies values by observing how much extra individuals are prepared to spend in travel costs in order to visit a site with higher environmental quality.

If there were only two sites, one with high quality and the other with low quality then observing how visitation and the incurred travel costs of visitors differed from the high to the low quality site would reveal the information needed to understand the value of the extra quality. The real world, unfortunately, is far more complex. Rather than two sites, real world landscapes consist of an array of recreational sites each offering different levels of environmental quality. That description certainly typifies the landscape of the Yorkshire Water region where a diversity of recreation sites both in the form of paths and parks provide users access to a variety of quality-differentiated river environments.

To deal with that complexity we adopt a travel cost modelling approach based on rational choice behaviour. In particular, households are assumed to first consider the qualities of the set of river sites, then assess the costs of visiting each before choosing to visit the site offering the best quality to cost trade-off. Given a sufficiently large dataset on observed recreational choices, it is possible to use this approach to estimate households' valuation function, that is to say, how much benefit they get from visiting a site with additional quality and how much disbenefit they endure from incurring additional costs in traveling to sites. In economics, this is often referred to as a households' *preference function*.

Armed with estimates of the preference function, and assuming choice behaviour in which households choose to visit their most preferred site, analysts can construct a model of recreational choice behaviour. That model provides the basis for valuation work. It can be used to predict how a household's welfare might change if, for example, the river water quality at one or more recreational sites were to change.

The discrete choice modelling approach captures many important behavioural realities. For a start, it allows for *distance decay*, that is to say, that the chances of a household visiting a site decline with distance (and travel costs) and hence the value of quality changes at more distance sites is likely to be less than at more proximate ones. Secondly, it neatly captures *substitution effects*. In particular, a household is likely to visit some particular site if they have other sites offering similar qualities in their vicinity. As a corollary, the value of quality improvements at a site are likely to be less the greater the abundance of sites with similar quality levels.

The difficulty with discrete choice travel cost modelling is that it is relatively data intensive. Analysts first need to collect together information on the location and qualities of the recreation sites visited by households in the sample. Moreover, so as to properly reflect the set of options from which those visited sites were (the choice set) it is also necessary to collect information on the location and qualities of all other possible substitute sites. Finally, the analyst needs to calculate the travel costs that each household might incur in travelling from their home to each of the site options in the choice set. In Section 3 we will discuss the construction of just such a dataset developed from observations of recreational trips to rivers recorded in the ChREAM survey.

11.2 Stated Preference Methods: Choice Experiments

Stated preference techniques for valuing environmental quality rely on carefully constructed questionnaires to directly ask individuals how much they are willing to pay for changes in that quality. Importantly, these methods allow individuals to express not only the value they might get from the environment as a result of using it for recreation, but also the value they might get from simply knowing that the environment of a location is in good condition; that is to say, non-use value.

While the concept of stated preference surveys is relatively straightforward their effective application to the valuation of river water quality is actually rather complex. One particular problem concerns how to convey to respondents the complex spatial reality of the landscape within which hypothetical changes in environmental quality occur. How can the analyst effectively convey the range of river locations in that landscape, their current quality, and the quality to which they will change in the hypothetical situation being valued?

The approach we adopt in this research is that of a *Visual Spatial Choice Experiment* (VSCE) where hypothetical scenarios are presented to respondents in the form of colour-coded and annotated maps (see Figure 5). In a VSCE respondents are presented with a selection of such maps each illustrating a different spatial pattern of quality change and each associated with some particular cost. Of those on display, respondents are asked to identify which costly pattern of quality change is their most preferred.

Map-based presentations like the VSCE allow stated preference studies to elicit preferences for complex landscape-wide patterns of environmental change. A VSCE explicitly presents respondents with information on both the locations of the rivers that change quality and those that do not, ensuring respondents choices over different hypothetical scenarios take account of the full spatial reality of the landscape. How those choices should be analysed so as to properly reflect respondents' decision processes, however, remains an open question in the stated preference literature.

In this report, we adopt a method coming out of our latest research. That method develops a statistical model that enables us to estimate the household preference function from choices in a VSCE. Importantly the starting point for that model of preferences is the same rational choice behaviour that underpins the discrete-choice travel cost model. In other words, analysis of the VSCE data and analysis of the travel cost data make identical and coherent assumptions regarding the household preference function and respondents' choice behaviour. We describe that model in some detail in Section 4, a section that can be skipped by the non-expert reader.

11.3 Combining Revealed and Stated Preference Data

While the revealed preference travel cost models inform us as to the value of water quality improvements from recreational use of river sites, they tell us nothing about value gained in non-use. Likewise, it is rarely possible to distinguish what part of the values revealed in a stated preference survey arise from recreational use and what part should be attributed to non-use. Right at the cutting-edge of academic work in the field of non-market valuation are methods that bring together complex travel cost and stated preference data in a single analysis.

One of the distinguishing features of the research reported in this document is that brings together the revealed and stated preference data in a single analysis. Combining data sets in this way is made possible by the fact that our new model of VSCE responses is based on the same rational choice behaviour that underpins the discrete travel cost model. Put simply, both models make the same assumptions regarding how households derive use and non-use values from the set of quality-differentiated river sites in the region. Accordingly, both datasets can be used together to tie down the exact structure of the preferences that give rise to those values; the travel cost model informing on use elements, the VSCE on both use and non-use.

Once estimated, the model provides a spatially explicit value function that distinguishes value derived from use from that derived from non-use. More specifically, that model allows us to estimate the change in use and non-use resulting from some particular set of changes in river water quality in some specific set of locations.

12 Data

12.1 River Survey

The core dataset for this research is provided by a survey of some 1,805 respondents from across the Yorkshire region that was collected as part of the RELU-funded Catchment Hydrology, Resources, Economics and Management (ChREAM) project in 2008. As shown Figure 2, the survey focused on a 70km square region centred on the River Aire and the Bradford-Leeds conurbation and incorporating other major rivers including the Wharfe and Calder.

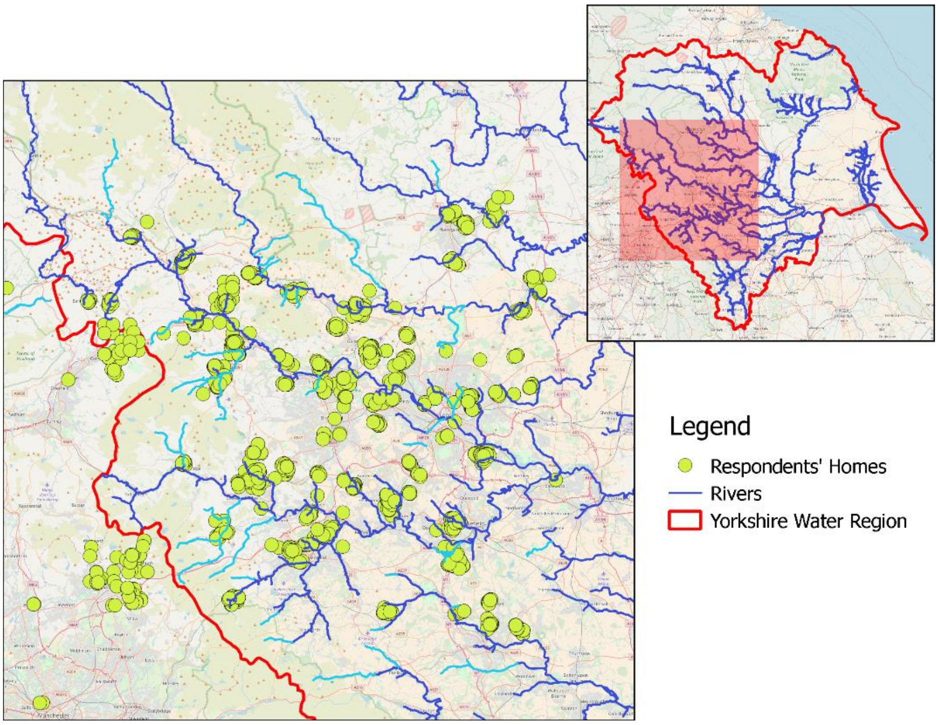


Figure 2: Respondent Sampling Locations

Randomly sampling respondents from the region would have resulted in a dataset that was dominated by households from the Bradford-Leeds conurbation. A fundamental requirement for our research, however, was to ensure a sample that exhibited a wide diversity in spatial location relative to rivers of different qualities. Accordingly, the survey adopted a spatial sampling approach in which surveying locations were chosen that evenly sampled the spatial extent of the study area (see Figure 2). Between 40 and 100 interviews were conducted in each sampling location giving a total sample size of 1,805.

Of course, using a non-random sampling approach makes it more difficult to use data from the sample to draw inference regarding the population as a whole. To overcome that problem, in our analyses we use weighting to adjust the sample for population inference (see Section 3.2).

The survey was explicitly designed to capture large quantities of spatially explicit data from respondents through a highly accessible custom-built computer aided personal interview (CAPI) system. During the interview, respondents were shown an interactive map on a computer screen indicating the respondent's home location and all of the surrounding rivers within an area the same size as the full survey area.

The first section of the survey collected data for application of the travel cost method. First, respondents indicated on the map the river locations that they had visited for recreation over the course of the last 12 months and indicated the frequency of their visits to each of those sites. Details of the total number of all outdoor recreation trips taken in the last 12 months was also collected.

Amongst respondents in the sample 18% made no recreational trips to a river site in the previous year, 27% made 1 to 5 trips to a river site, 12% made 6 to 10 trips, with the remaining 33% making more than 10 trips a year. The distribution of those trips across the study area is shown in Figure 3.

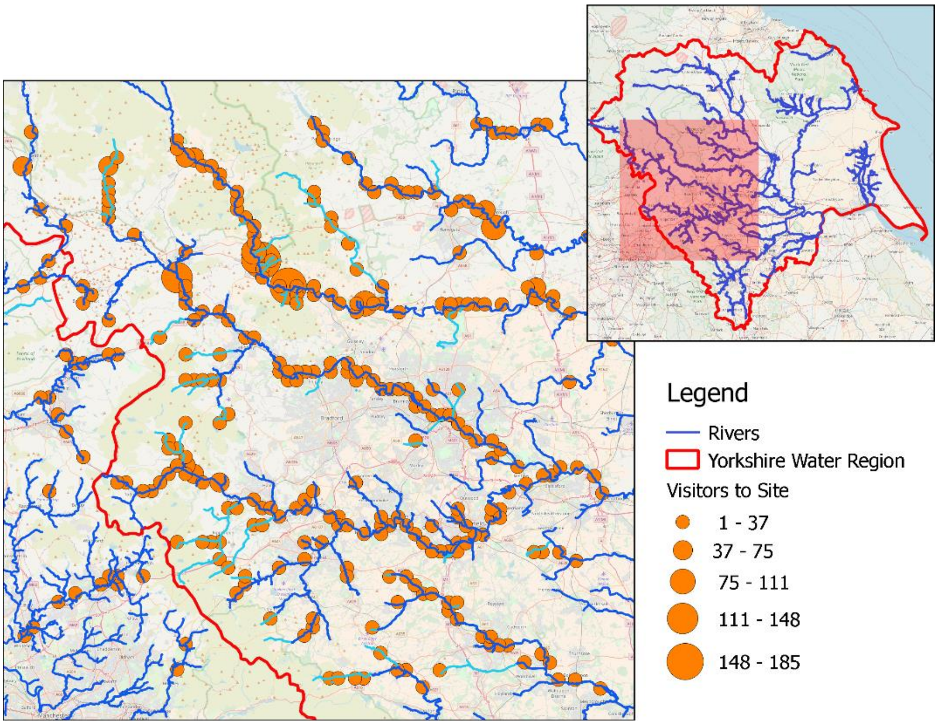


Figure 3: River Sites Visited by Sample

Respondents to the survey were asked to consider the level of river water quality that they experienced at recreational sites and introduced to the categorisation of water quality illustrated in Figure 4. The categorisation focused on the ecological status of rivers and was developed with hydrological and ecological experts following the procedure described in Hime et al. (2009).

Figure 4: Water Quality Definitions used in the Survey

The categorisation identified four levels of quality that moved from bad to poor to good to excellent with each level being associated with a quality colour; red, yellow, green and blue respectively. The picture for each level illustrates the appearance of the water, the river banks and bed, and the plant and animal species that might typically be associated with each ecological quality level. Those pictures also indicated the sorts of recreational activity that might be associated with each quality level, including wildlife watching, boating, swimming and coarse and game fishing.

The water quality categorisation in Figure 4 was developed before the formal adoption of the categories used by the Environment Agency in implementing the Water Framework Directive (WFD). The WFD ecological status categorisation identifies five levels of water quality; bad, poor, moderate, good, high. For the purposes of this research we assume that the bad, poor and moderate classes map to our bad, poor and good categories while the WFD good and high classes map to our excellent category.

In the second part of the survey, respondents participated in a VSCE. The VSCE concentrated on the main rivers in the study area; the Aire, the Wharfe and the Calder. To construct the choice experiment, those rivers were divided into nine river lengths of equal extent. To construct a scenario to describe a future possible state of the world, each river length was ascribed a particular water quality and that quality illustrated on a map by colouring river lengths with their ascribed quality colour. Each scenario was associated with a cost motivated as an annual increase in the household water bills payable by each household in the region. Finally, a choice task was constructed by pairing two scenarios, as illustrated in Figure 5. Using a fractional factorial design, 60 choice tasks were constructed and divided into five blocks of 12 tasks. In the VSCE each respondent was presented with a particular block of VSCE question and, therefore, provided answers to 12 choice tasks

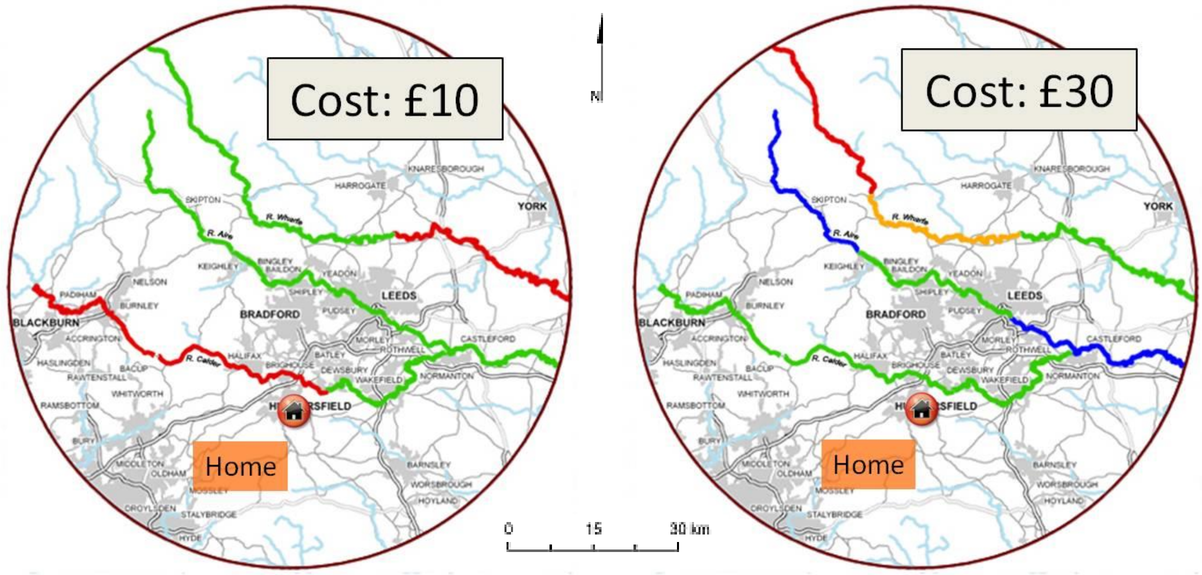


Figure 5: Typical Task from the Visual Spatial Choice Experiment

12.2 Sample Characteristics and Weighting

Of the 1,805 households in the sample, some observations were incomplete or lacked crucial information such that the final dataset consisted of the recreational activity for 1,794 households with 1,708 of those also providing a complete set of responses to the VSCE.

Our survey was carried out at the household level and the respondent from each household was chosen as being the adult with responsibility for paying the household water bill. We take that individual to be the same as the individual defined as the Household Representative Person (HRP) as captured in UK census data.

Using the boundaries of the Yorkshire Water operational area, we collated data from the 2011 census returns describing the socioeconomic characteristics of households in that region. Table 5 provides a comparison of the distribution of characteristics of households in the survey sample and compares those to the characteristics of households in that wider region.

Table 5: Comparison of Sample and Population Characteristics

Characteristics	**Yorkshire Water Region**	**Sample**	**Pearson χ^2 stat**	***p*-value**
Age of HRP				
<35	18.4%	23.4%		
35 to 54	38.4%	32.1%		
55 to 64	17.0%	17.5%		
>65	26.3%	27.0%	43.91	<0.001***
Household Size				
Small (1 or 2)	65.7%	57.7%		
Medium (3 to 5)	31.9%	38.5%		
Big (>5)	2.4%	3.8%	56.81	<0.001***
Children				
Yes	28.7%	35.5%		
No	71.3%	64.5%	40.98	<0.001***
Employment of HRP				
Part time	9.3%	13.5%		
Full time	41.7%	25.6%		
Self employed	10.1%	7.6%		
Unemployed	3.5%	3.6%		
Student	2.0%	3.7%		
Retired	26.2%	33.2%		
Looking after Home	1.8%	8.5%		
Other	1.5%	0.6%		
Sick	3.9%	3.6%	693.59	<0.001***
Residence				
Metropolitan	54.1%	58.5%		
Town	29.7%	22.0%		
Suburb	10.1%	16.6%		
Rural	6.1%	3.0%	144.63	<0.001***
Total Households	2,186,513	1,794		

Notes: Statistics report the probability that the sample could have been drawn at random from the Yorkshire Water Region population.

Given the sampling strategy, it is not altogether surprising that the distribution of household characteristics from the sample does not closely match that of the wider population. Perhaps the most evident differences are that our sample oversampled from suburban areas, contained more households with a retired HRP, fewer whose HRP was full time employed and relatively less small households. The final columns of Table 5 provide a statistical comparison of the sample and population and provide conclusive evidence that the sample does not closely resemble the population. To correct for the lack of representativeness of the sample, we calculated a set of sampling weights using the 'Anesrake' package for the R statistical software. In effect, these weights scale up or down the importance of the different sample observations such that households in our sample that are over-represented are given small weights and those that are under-represented are given large weights. Analysis of the weighted data should provide valid insights into the preferences of the population of the Yorkshire Water Region.

12.3 River Sites Giving Use Value: Recreation River Cells

To implement the travel cost method, it is not only necessary to know the river sites that households visited for the purposes of recreation, but also those locations that they might have visited instead. In other words, we require a comprehensive description of locations along rivers in the study region that might have formed the focus of a recreation trip. To construct such a dataset, we used the ORVal greenspace map. ORVal, the Outdoor Recreation Valuation Tool, is an on-going research endeavor undertaken by the research team on behalf of DEFRA that seeks to investigate the value of access to greenspaces across England and Wales. The ORVal greenspace map was developed as part of that project and brings together data from an array of spatial datasets to create a single map identifying every outdoor recreation site in the country (Day and Smith, 2016). The ORVal greenspace map identifies numerous features of those recreation areas including whether the site is a park or a path, the landcovers and habitats that make up the greenspace of that site (e.g. grass, moors, woods, agriculture etc.) and the bluespace that is accessible from that site (e.g. rivers, lakes, canals etc.).

As a first step, we took the OS map of main rivers for the study region and identified all the parks and paths in the ORVal greenspace map that gave access to those rivers. We then constructed a 1km square grid over the region and selected out all of the cells of that grid through which the region's rivers traversed. Finally, of those 1km cells along the rivers we selected out all of those that contained recreation sites. For the purposes of this research, these 1km cells were taken as our definitions of recreation sites. As shown in Figure 6 that procedure resulted in the identification of 579 recreation sites in the study area of which (comparing back to Figure 3) 280 had been visited by respondents in the sample.

The cells illustrated in Figure 6, therefore, represent each sample household's choice set and we assume that on each choice occasion they make the decision as to which of these recreation sites they will visit. Of course, other options are open to the household at each choice occasion. They could, for example, take a recreation trip to some other non-river recreation site. Likewise, they could decide not to take a recreation trip at all. As well as the frequency of trips to river sites in the study area, our survey collected information on the frequency over the previous year with which each household chose those other options. Accordingly, our specification of the recreation demand model (see Section 13) assumes that on every day of the year a household has a choice between 581 options; visiting one of the 579 recreation sites in the region, visiting some other outdoor recreation site and not visiting an outdoor recreation site at all.

With regards to choosing between the recreation sites a key consideration for households will be the costs of travelling from their home to that site. Developing estimates of those travel costs for the purposes of our modelling was not an inconsiderable task. Using the detailed road network for

England provided by the Ordnance Survey's Integrated Transport Network (ITN) dataset it is reasonably easy to use a GIS to calculate the fastest travel route from a particular household's home to a particular site. The problem we faced is that we have 1,794 respondents and 579 sites resulting in the need to compute 1,038,726 such optimal routing queries. The scale of that challenge was far in excess of the performance provided by standard GIS software. Accordingly, we turned to RoutingKIt, a highly efficient library of routing algorithms provided as C++ source code, that implement the Contraction Hierarchy method of routing analysis (https://github.com/RoutingKit/RoutingKit). Having written the wrapper code to link RoutingKit to our data, we achieved speed ups of three orders of magnitude over that provided by ArcGIS's Network Analyst.

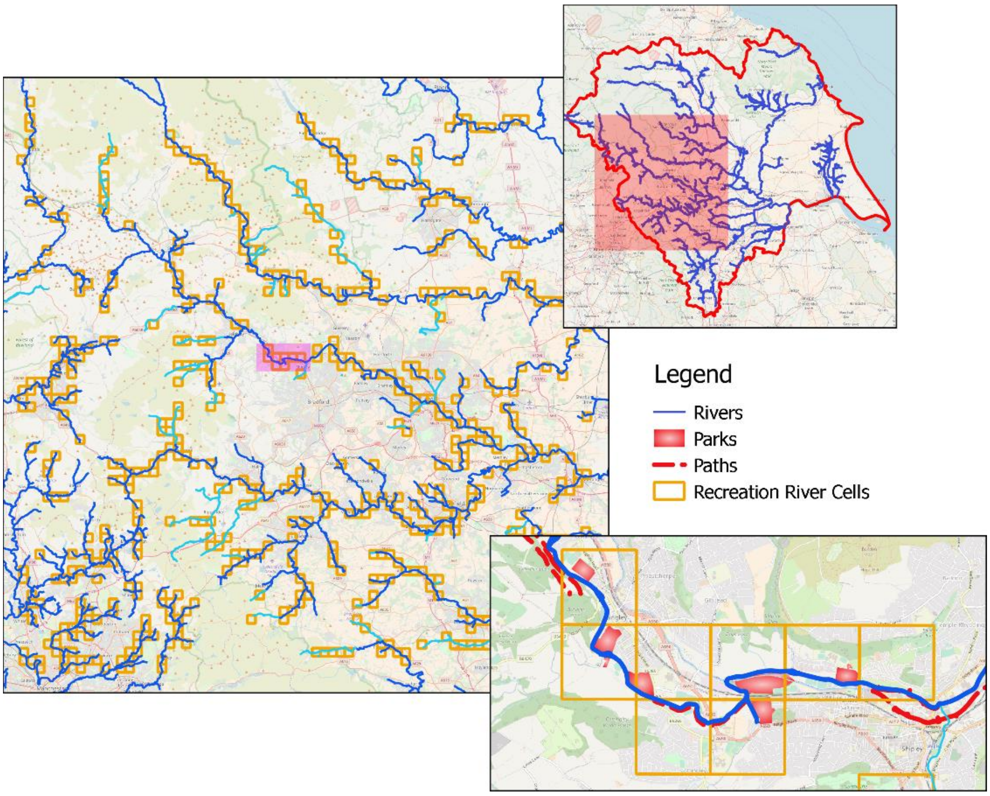


Figure 6: Recreation Cells defining Locations Providing Use Utility

In running the routing queries, we initialised the network with data on driving speeds along different categories of road and used RoutingKit to identify the fastest route from the home location to a site. For walking options, we assumed a standard walking speed of 5kph. Using formulae provided by DFT (Department for Transport 2014), we took the length of time driving at different speeds along the route to calculate a fuel consumption for an average family car. Subsequently, we calculated a fuel cost by multiplying fuel consumption by the price of fuel at the time the trip was taken (taken as an average of diesel and unleaded prices from AA fuel price reports).

We converted driving times into a monetary cost using results provided in recent research for DfT on the value of travel time (Department for Transport 2015). Those values were £2.30 per hour for trips under 8km, £3.47 per hour for trips between 8km and 32km, £6.14 per hour for trips between 32km and 160km and £9.25 per hour for trips greater than 160km (see Table 7.18 of DfT report). A total monetary cost for driving to a site was taken by adding the time costs to the fuel costs for the return journey.

Of course, the cost of travel is not the only variable that determines a household's choice as to which site to visit. The ORVal greenspace map also provided some details of the nature of the space available at the riverside parks and paths in each recreation cell. We used that data to add up the total area of park and path in each cell as well as the extent of different landcovers accessed through those recreation sites; particularly areas of grassland, moorland, woodland and agricultural landcover (the latter only being relevant for path-type recreation sites). In addition, recognising that a big park may only provide access to a small stretch of river, we also calculated the extent of river accessible through recreation sites in each cell.

Another decision-relevant factor concerns the context within which the recreation sites are situated. We might imagine, for example, that recreation sites in urban settings are considered differently from those in rural settings. To capture that factor, we used a GIS dataset of urban extents in the UK to calculate the proportion of each recreation cell that was in an urban area. Likewise, our study area contained two unique habitat areas in the form of the Yorkshire Dales and the Southern Pennines. We picked out three National Character Areas, the Yorkshire Dales (NCA 21), the Southern Pennines (NCA 36) and the Pennine Dales Fringe (NCA 22) and created variables that indicated whether a river cell was in each of those particular landscapes. We also created variables identifying whether recreation cells were in the Nidderdale AONB or the Yorkshire Dales National Park though the definition of those variables coincided strongly with those of the National Character Areas.

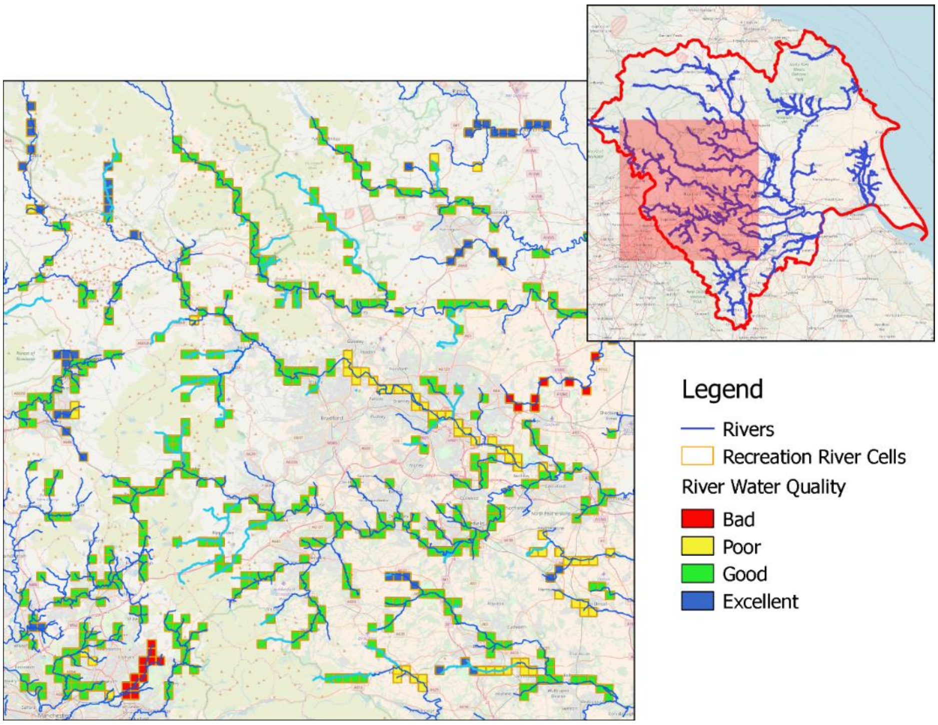


Figure 7: Water Qualities of Recreation River Cells

The final variable needed for our analysis of site characteristics was that describing the quality of river water accessible at recreation river cells. To approximate that variable, we used the Environment Agency's fine-scale WFD classification of water basins into different ecological status categories. While the survey data pertained to 2008 the first entries in the Environment Agency's WFD database were for 2009 and data for that year provided only partial coverage of the study area. In addition, we felt that a single year's data might not capture the general status of each water basin, with poor or good results in any one year being the result of features of that particular year (for example, poor weather or particular farming activities). Accordingly, we took as our measure of water quality the most frequent WFD ecological status category seen in each water basin across the years from 2009 to 2015, selecting the worst category where more than one category was equally most frequent. The water qualities available at each river recreation cell are shown in Figure 7.

12.4 River Sites Giving Use Value: River Cells

As we explain in more detail in the next section our model of how households derive value from rivers makes the self-evident assumption that use value can only be gained from river cells that provide recreational access. Such a restriction does not apply to non-use value. Indeed, we assume that households may gain non-use value from each and every stretch of river, though we hypothesize that that non-use value may decline with increasing distance of a river stretch from a household's home. To identify river locations offering possible non-use value we followed a similar procedure as used to identify river locations offering potential use value. We constructed the same 1km square grid over the study area and identified locations offering non-use values as all cells that were traversed by rivers.

For the purposes of the modelling exercise, it turned out that we could simplify further. Our information on non-use value comes from the VSCE exercise in which users get to state their preferences over different patterns of water quality change in the Aire, Wharfe and Calder (see Figure 5). In this exercise, respondents were explicitly asked to assume that the quality of all other river sites in the region remained unchanged. Since the non-use value provided by those other river locations was the same for all options in the VSCE we can assume that the quality at those sites had no bearing on respondent's decisions. Accordingly, in analysing the data all we needed to concern ourselves with were river cells along the Aire, Wharfe and Calder. Those cells are illustrated in Figure 8.

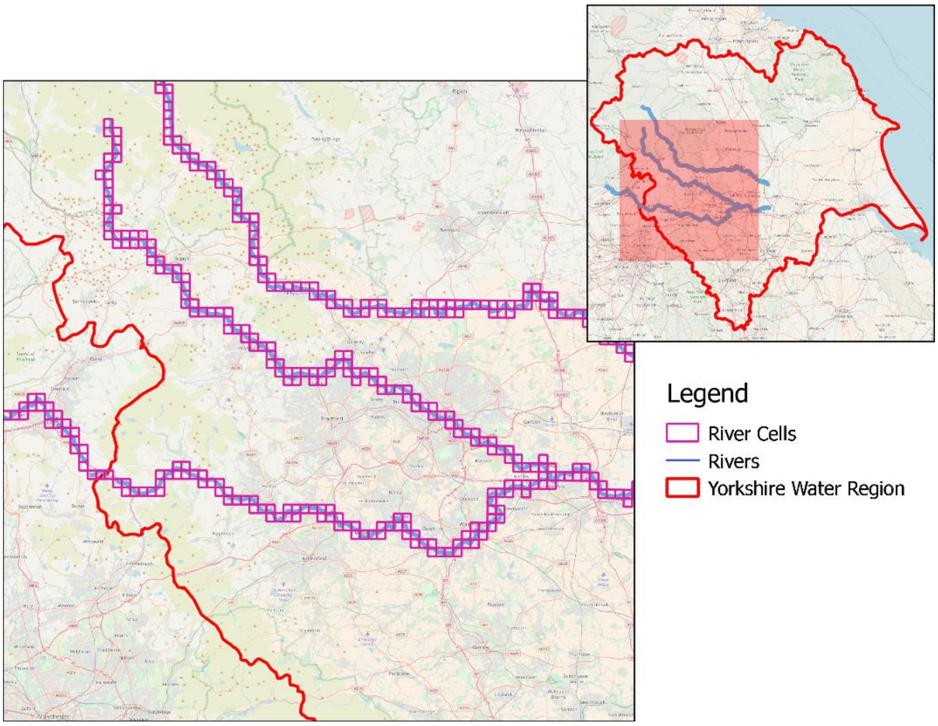


Figure 8: River Cells defining Locations Providing Non-Use Utility

We might imagine that non-use value could be affected by numerous characteristics of river cells not only their water quality. As we shall show more formally in the next section, we use the argument that these other characteristics of river sites remain the same across all options in the VSCE. It follows that our data informs only on how non-use values respond to changes in water quality but provides no direct information on how non-use value is impacted by changes in other river site characteristics.

Finally, to test our hypothesis that non-use value declines with distance, we needed to calculate a measure of proximity from each river cell to each household's home. While it would have been possible to calculate travel distances through the road network, that measure of distance did not seem to accord well with the idea that this value was not being obtained by actually travelling to the site. While numerous alternative measures are possible, we opted to use straight line distance as our measure of proximity.

13 Econometric Analysis

13.1 Structural Model

The data collected in our survey reports on the preferences of a sample of households indexed $i = 1,2,\dots,N$, living in a region endowed with an array of river cells, indexed $j = 1,2,\dots,J$. The welfare that a household realises from those river cells during time period t arises as a result of the qualities that those locations exhibit; our particular interest being their river water quality. That quality differs across river cells and may differ across time periods, though are assumed to remain constant for the duration of any one period. The qualities of river cells can also differ across possible states of the world, $s = 0,1,\dots,S$. The reality of the current state of the world is indicated $s = 0$ and the S alternative states of the world are those constructed to describe the qualities of river cells for the purposes of our VSCE valuation exercise. The qualities of river cell j, in period t, under scenario s is given by the vector $\boldsymbol{q}_{j,t,s}$.

River cells can be used for recreation, though to enjoy the recreational experience offered by river cell j, a household must travel to that location. We indicate the consumption levels of those trips by the vector $\boldsymbol{x}_{i,t,s} = \left(x_{1,i,t,s}, x_{2,i,t,s}, \dots, x_{J,i,t,s}\right)$ and take those to be goods whose purchase prices (comprising the costs of travel and the opportunity cost of travel time) are identified by the price vector $\boldsymbol{p}_i = \left(p_{i,1}, p_{i,2}, \dots, p_{i,J}\right)$.

Households might also gain utility from river cells without having to purchase any complementary market goods; perhaps through the pleasure they derive simply from knowing that such river cells exist or from the knowledge that others may benefit from their existence. Again, we assume that utility derived in this way arises as a consequence of the qualities of a river cell.[1] Following evidence from the stated preference literature (Bateman et al., 2006; Schaafsma et al., 2012) we allow for the possibility that the non-use value derived from a river cell with particular qualities may differ as distance from a household's home increases. Those distances are identified by the vector $\boldsymbol{d}_i = \left(d_{i,1}, d_{i,2}, \dots, d_{i,M}\right)$.

---

[1] The qualities which deliver value in non-use could potentially differ from those offering value in use. Our notation assumes, therefore, that the vector $\boldsymbol{q}_{j,t,s}$ is a comprehensive list of utility-relevant quality attributes, but that the contribution which a particular quality element makes to value in use or non-use may be zero.

Our assumptions lead us to the direct utility function;

$$u\left(u^{use}\left(\boldsymbol{x}_{i,t,s}, \boldsymbol{q}_{1,t,s}, \dots, \boldsymbol{q}_{J,t,s}\right), u^{non-use}\left(\boldsymbol{q}_{1,t,s}, \dots, \boldsymbol{q}_{J,t,s}, \boldsymbol{d}_i\right), z_{i,t,s}\right) \qquad (\forall\, i,t,s) \tag{1}$$

where z is a numeraire good with unit price. Observe that the utility function is separable into value derived from use of river cells, $u^{use}(\cdot)$, value derived from non-use, $u^{non-use}(\cdot)$, and value derived from consumption of other goods, $z_{i,t,s}$, and that utility is increasing in all three of those arguments. Moreover, we assume that $u^{use}(\cdot)$ exhibits weak complementarity such that the utility from use derived from the qualities of a river cell falls to zero when consumption of trips to that site is zero; that is to say, $\partial u^{use}/\partial \boldsymbol{q}_{j,t,s} = \boldsymbol{0}$ when $x_{j,t,s} = 0$. We also assume that preferences are strongly separable over time.

If the choice period is reduced to a length of time such that in each period, t, a household can make at most one recreational trip, then a household's consumption decision amounts to solving the discrete choice problem given by (Phaneuf and von Haefen, 2009);

$$\begin{aligned} \max_{x_{i,t,s}, z_{i,t,s}} \quad & u\left(u^{use}\left(\boldsymbol{x}_{i,t,s}, \boldsymbol{q}_{1,t,s}, \dots, \boldsymbol{q}_{J,t,s}\right), u^{non-use}\left(\boldsymbol{q}_{1,t,s}, \dots, \boldsymbol{q}_{J,t,s}, \boldsymbol{d}_i\right), z_{i,t,s}\right) \\ s.t. \qquad & y_{i,t} = \boldsymbol{p}_i' \boldsymbol{x}_{i,t,s} + z \\ & x_{j,t,s} \in \{0,1\} \qquad\qquad (\forall\, i,t,s) \\ & x_{j,t,s} x_{k,t,s} = 0 \; (\forall\, k \neq j) \end{aligned} \tag{2}$$

The conditional indirect utility function that arises from (2) takes the form;

$$u_{i,t,s|j} = u\left(u^{use}\left(\boldsymbol{q}_{j,t,s}\right), u^{non-use}\left(\boldsymbol{q}_{1,t,s}, \dots, \boldsymbol{q}_{J,t,s}, \boldsymbol{d}_i\right), y_{i,t} - p_{i,j}\right) \qquad (\forall\, i,t,s) \tag{3}$$

Observe that our assumptions regarding weak complementarity, imply that a household only derives use utility from the qualities of the river cell that they choose to visit in choice period t. In contrast, during that period households derive non-use utility from the qualities of all river cells. Our model of recreational behaviour is completed through the rational choice rule;

$$\text{choose to visit } j \text{ in period } t \text{ if:} \quad u_{i,t,s|j} > \left\{u_{i,t,s|k}\right\}_{\forall\, k \neq j} \qquad (\forall\, i,t,s) \tag{4}$$

Over the course of a year we assume that households face $t = 1,2,\dots,T$ recreational choice periods of equal length and that in each period households follow (4) in determining their recreational choice behaviour. Accordingly our model follows the tradition of repeated discrete choice models as per Morey et al. (1993).

In our VSCE exercise respondents are asked to consider alternative states of the world in which the qualities of the river cells differ from those experienced in the current state of the world ($s = 0$). The quality changes described in each alternative state of the world ($s = 1, 2, \dots, S$) cannot be achieved without cost, a cost to which households must contribute through a hypothetical coercive annual charge C_s. Since the year is divided into T equally-sized choice periods indexed $t = 1, 2, \dots, T$ we assume that the annual payment can be equivalently expressed as a series of per period payments; $c_s = C_s/T$. Preferences for these different states of the world have the same fundamental structure, though the conditional indirect utility function (3) must be modified to include the hypothetical payment;

$$u_{i,t,s|j} = u\left(u^{use}(\boldsymbol{q}_{j,t,s}), u^{non-use}(\boldsymbol{q}_{1,t,s}, \dots, \boldsymbol{q}_{M,t,s}, \boldsymbol{d}_i), y_{i,t} - p_{i,j} - c_s\right) \qquad (\forall\, i, t, s) \tag{5}$$

which reduces to (3) in the current state of the world since $c_0 = 0$;

In a typical hypothetical choice task, households are presented with a set of scenarios, $\mathbb{s}$, drawn from the S scenarios constructed for the VSCE. Respondents are asked to indicate which scenario is their most preferred. According to our model, to make that choice, respondents must first solve the site visitation problem (4) for each time period such that their declared preference over hypothetical scenarios should be made according to the choice rule;

$$\text{choose } s \text{ if:} \quad \sum_{t=1}^{T} \max_j(u_{i,t,s|j}) > \left\{\sum_{t=1}^{T} \max_j(u_{i,t,r|j})\right\}_{\forall\, r \neq s} \qquad (s, r \in \mathbb{s}) \tag{6}$$

where the summation over the T time periods in a year follows from our assumption of inter-temporal additive separability of the utility function.

13.2 Econometric Model

We develop our econometric model by first specifying a functional form for the conditional indirect utility function (4). Our separability assumptions are compatible with the additive form;

$$\begin{aligned} u_{i,t,s|j} &= v_{i,t,s|j}^{use} + v_{i,t,s}^{non-use} + v_{i,t,s|j}^{other} + \varepsilon_{i,j,t,s} \\ &= v_{i,t,s|j} + \varepsilon_{i,j,t,s} \qquad (j = 1, 2, \dots, J+1 \text{ and } \forall i, t, s) \end{aligned} \tag{7}$$

where $\varepsilon_{i,j,t,s}$ is an econometric error term introduced to capture the divergence between our model of use utility $(v_{i,t,s|j}^{use})$, non-use utility $(v_{i,t,s}^{non-use})$ and utility from expenditure on other goods $(v_{i,t,s|j}^{other})$, and the household's experienced utility $(u_{i,t,s|j})$.

Moreover we specify;

$$v_{i,t,s|j}^{use} = \alpha_{j,i,t} + \boldsymbol{q}_{j,t,s}\boldsymbol{\beta}_i \qquad (j = 1, 2, \dots, J \text{ and } \forall\, i, t, s) \tag{8}$$

where $\alpha_{j,i,t}$ is a site-specific utility element and $\boldsymbol{\beta}_i$ is the vector of coefficients describing the marginal use utilities of site qualities. Of course, in any choice period a household may choose not to make a recreational trip to a river cell. We give that option the index $J + 1$, and specify the use utility from choosing that option as;

$$v_{i,t|J+1}^{use} = \alpha_{J+1,i,t} \qquad (\forall i, t) \tag{9}$$

Observe that since this option does not involve visiting one of the J river cells, the use utility associated with choosing this option does not change across scenarios. We gather the parameters of the use element of household i's utility into the vector $\boldsymbol{\theta}_i^{use} = [\alpha_{1,i,t} \dots \alpha_{J+1,i,t}\ \boldsymbol{\beta}_i]$;

Our model of the non-use utility element of the preference function is given by;

$$v_{i,t,s}^{non-use} = \sum_{j=1}^{J} d_{i,j}{}^{\lambda_i} \left(a_{j,i,t} + \boldsymbol{q}_{j,t,s}\boldsymbol{b_i}\right) \qquad (\forall\, i, t, s) \tag{10}$$

where $d_{i,j}$ is the distance from household i's home to area j, $a_{j,i,t}$ is an area-specific element contributing to non-use utility, $\boldsymbol{b}_i$ is the vector of coefficients on site qualities and λ_i is a parameter that establishes the rate of distance decay in non-use utility. Notice from (10) that non-use utility is specified as a distance-weighted sum across the non-use utility provided by each river cell. The use of summation imposes the assumption that no substitution or complementarity relationships exist between sites in delivering non-use value. The power function used to describe that distance weighting, nests a number of plausible specifications: for example, $\lambda_i = 0$ suggests that non-use utility does not decline with distance, while $\lambda_i = -1$ suggests that the non-use utility declines inversely with distance. Again we use the notation $\boldsymbol{\theta_i^{non-use}} = \left[a_{1,i,t} \dots a_{J,i,t}\, \boldsymbol{b}_i\, \lambda_i\right]$ to denote parameters of the non-use element of utility.

Finally we assume a simple linear form for utility from other consumption, such that conditional on travelling to j;

$$v_{i,t,s|j}^{other} = \gamma_i\left(y_{t,i} - p_{j,i} - c_s\right) \qquad (j = 1, 2, \dots, J+1 \text{ and } \forall\, i, t, s) \tag{11}$$

Respondents to our survey provided both revealed preference and stated preference data. The revealed preference data details the visits each respondent made to the different river cells over the course of the last year. The stated preference data is collected from a VSCE in which respondents choose between quality-differentiated states of the world. Our objective is to build an econometric model that is derived from the coherent behavioural model described in equations (4) and (6) such that the parameters of the structural equations in equations (8), (9), (10) and (11) can be estimated simultaneously from both revealed and stated preference data.

Our econometric model proceeds through building a likelihood function describing the probability the analyst attributes to observing the choices seen to be made by respondents. As a result of the error term $\varepsilon_{i,j,t,s}$, probabilistic behavioural equations replace the deterministic choices envisaged by (4) and (6). As such, our econometric model of the probability of observing household i choosing to visit site j in period t can be written as;

$$\begin{aligned} P_{i,j,t,0}(\boldsymbol{\theta_i^{use}}, \gamma_i) &= Prob\left[u_{i,t,0|j} > u_{i,t,0|k} \quad \forall\, j \neq k\,\right] \\ &= Prob\left[v_{i,t,0|j}^{use} + v_{i,t,0}^{non-use} + v_{i,t,0|j}^{other} + \varepsilon_{i,j,t,0}\right. \\ &\left. > v_{i,t,0|k}^{use} + v_{i,t,0}^{non-use} + v_{i,t,0|k}^{other} + \varepsilon_{i,k,t,0} \;\; \forall\, j \neq k\,\right] \\ &= Prob\left[v_{i,t,0|k}^{use} + v_{i,t,0|j}^{other} - v_{i,t,0|j}^{use} - v_{i,t,0|j}^{other} > \varepsilon_{i,j,t,0} - \varepsilon_{i,k,t,0} \;\; \forall\, j \neq k\,\right] \end{aligned} \tag{12}$$

Since households derive the same level of non-use value independent of their choice of which area to visit, the non-use element of utility nets out of line 3 of equation (12). It follows, that the parameters determining values through non-use cannot be estimated from discrete-choice data on recreational behaviour.

In a similar vein, in a VSCE with M exercises indexed $m = 1,2,\dots,M$, the probability that household i chooses option s from the choice set $\mathbb{s}_m$ amounts to;

$$\begin{aligned} P_{i,s,m}(\boldsymbol{\theta_i^{use}}, \boldsymbol{\theta_i^{non-use}}, \gamma_i) &= Prob\left[\sum_{t=1}^{T} \max_j\left(u_{i,t,s|j}\right) > \left\{\sum_{t=1}^{T} \max_j\left(u_{i,t,r|j}\right)\right\}_{\forall\, r\neq s}\right] \\ &= Prob\left[\sum_{t=1}^{T} \max_j\left(v_{i,t,s|j} + \varepsilon_{i,j,t,0}\right) > \left\{\sum_{t=1}^{T} \max_j\left(v_{i,t,s|j} + \varepsilon_{i,j,t,0}\right)\right\}_{\forall\, r\neq s}\right] \qquad (13) \\ & \qquad (s, r \in \mathbb{s}_m) \end{aligned}$$

The nature of the probabilities in (12) and (13) are determined by the assumptions the analyst makes regarding the distribution of the error terms; for example, a frequent assumption is that the error terms are IID across choice options and choice occasions. Under that assumption, the probability of the revealed preference choices (12), defines a standard discrete choice travel cost model in which probabilities can be expressed in terms of differences in IID errors. A much more difficult econometric challenge is posed by the probability of stated preference choices (13) which involve sums of maxima across random variables. One of the central contributions of this research is to show how making the assumption that the error terms are drawn from the family of generalised extreme value (GEV) distributions (McFadden, 1978) results in closed form solutions to (12) and (13) that facilitate joint estimation of the parameters of use and non-use utility.

For the sake of notational simplicity, we illustrate that result by assuming that the error terms are independent draws from a Type I Extreme Value distribution with location parameter zero and scale parameter σ^2 $\left(IID\ EV(0,\sigma^2)\right)$. Under that assumption, (12) can be solved to give an expression for the probability of observing a particular recreational choice that takes the familiar multinomial logit (MNL) form;

$$P_{i,j,t,0}(\boldsymbol{\theta_i^{use}}, \gamma_i) = \frac{e^{v_{i,t,0|j}^{use}/\sigma_{RP}}}{\sum_{k=0}^{J+1} e^{v_{i,t,0|k}^{use}/\sigma_{RP}}} \qquad (\forall i, j, t) \qquad (14)$$

where the scale of the error terms, σ_{RP}, is subscripted RP to allow for the fact that that scale may differ between actual recreational decisions observed in revealed preference data and those made in response to the hypothetical scenarios presented in the VSCE.

To derive an equivalent expression for the probabilities of choice in the VSCE exercise data (13) is a more complex challenge. First we need to deal with the expression $\max_j\left(v_{i,t,s|j} + \varepsilon_{i,j,t,s}\right)$, which indicates the use utility a respondent expects to derive by solving the site-visitation problem and choosing which river cell to visit in time period t under state of the world s . Note that, in a VSCE respondents choose between states of the world but do not provide details of that anticipated recreational behaviour. Accordingly, information is not available with which to replace the maximisation expression with the utility of the particular site solving that maximisation problem. One way to proceed, follows from the observation that the set of arguments to the visitation problem are, by assumption, independent Type I Extreme Value variates with equal variance. It follows from properties of the Type I Extreme Value distribution that a household's maximum utility in scenario s must also be an extreme value variate[2];

$$u_{i,t,s} = \max_{j \in 1,..,J+1} u_{i,t,s|j} \sim EV\left(\sigma_s \ln \sum_{j=1}^{J+1} e^{\left(v_{i,t,s|j}^{use} + v_{i,t,s}^{non-use} + v_{i,t,s|j}^{other}\right)/\sigma_s}, \sigma_s \right) \qquad (\forall i,t,s) \tag{15}$$

where σ_s is the scale of the error terms relating to the utilities evaluated in response to scenario s. Since we have no reason to suspect that the error scales differ across scenarios, we impose the normalisation $\sigma_s = \sigma_{SP} = 1$ for all $s = 1, 2, \dots, S$. Accordingly, our specification allows us to write the utility enjoyed by household i in period t under scenario s as;

$$\begin{aligned} u_{i,t,s} &= \ln \sum_{j=1}^{J+1} e^{v_{i,t,s|j}^{use} + v_{i,t,s}^{non-use} + v_{i,t,s|j}^{other}} + \varepsilon_{i.t.s} \\ &= \ln \sum_{j=1}^{J+1} e^{v_{i,t,s|j}^{use} + v_{i,t,s|j}^{other}} + v_{i,t,s}^{non-use} + \varepsilon_{i.t.s} \qquad (\forall i,t,s) \end{aligned} \tag{16}$$

where $\varepsilon_{i.t.s}$ is a standard Type I Extreme Value variate.

Of course, the VSCE scenarios are framed as choices made over the duration of one year such that the final step in deriving the econometric specification for the utility derived from a particular choice experiment scenario is to sum over all periods;

$$\begin{aligned} u_{i,s} &= \sum_{t=1}^{T} \ln \sum_{j=1}^{J+1} e^{v_{i,t,s|j}^{use} + v_{i,t,s|j}^{other}} + \sum_{t=1}^{T} v_{i,t,s}^{non-use} + \sum_{t=1}^{T} \varepsilon_{i.t.s} \\ &= v_{i,s} + \sum_{t=1}^{T} \varepsilon_{i,t,s} \qquad (\forall i,s) \end{aligned} \tag{17}$$

In our VSCE households are presented with a series of tasks, $m = 1, 2, \dots, M$ each of which asks them to state a preference over two particular scenarios, s and r, such that the choice set $\mathbb{s}_m$ has only two members. Accordingly, replacing (17) into (13) reveals the probability of observing household i choosing option s in choice task m, to be;

---

[2] Other GEV distributions such as those resulting in the nested and cross-nested logit models have similar properties. To avoid burdening the reader with the notation needed for the expressions describing the maximum of iid variates under those alternative GEV distributions, we do not present those here.

$$
\begin{aligned}
P_{i,s,m}(\boldsymbol{\theta}_{\boldsymbol{i}}^{\boldsymbol{use}}, \boldsymbol{\theta}_{\boldsymbol{i}}^{\boldsymbol{non-use}}, \gamma_i) &= Prob\left[u_{i,s} > u_{i,r}\right] \\
&= Prob\left[v_{i,s} + \sum_{t=1}^{T} \varepsilon_{i,t,s} > v_{i,r} + \sum_{t=1}^{T} \varepsilon_{i,t,r}\right] \\
&= Prob\left[v_{i,s} - v_{i,r} > \sum_{t=1}^{T} \left(\varepsilon_{i,t,r} - \varepsilon_{i,t,s}\right)\right] \\
&= Prob\left[v_{i,s} - v_{i,r} > \sum_{t=1}^{T} e_{i,t,m}\right] \qquad (\,\forall i, m \text{ and } s, r \in \mathbb{s}_m)
\end{aligned} \tag{18}
$$

where, from a property of the Type I Extreme Value distribution, $e_{i,t} \sim Logistic(0,1)$. Observe that in differencing the utilities across the two scenarios any additive elements that are constant across scenarios are removed. For that reason, the data provides no means of identifying the area-specific non-use utility elements $a_{m,i,t}$.

To evaluate the probability in (18) we use a result from George and Mudholkar (1983) that shows how, as a convolution of standard logistic variates, the distribution of $\sum_{t=1}^{T} e_{i,t,m}$ can be very closely approximated by Student's *t* distribution. In particular;

$$
Prob\left[z > \sum_{t=1}^{T} e_{i,t,m}\right] \sim t_{5T+4}\left(0, \pi\left(\frac{15T+12}{5T^2+2T}\right)^{-\frac{1}{2}}\right) \qquad (\forall i, m) \tag{19}
$$

where $t_{5T+4}(\cdot)$ is the cumulative density function of Student's *t* distribution with $5T + 4$ degrees of freedom.

To complete our econometric specification, we note that our independence assumptions allow us to write the likelihood of observing household i's recreational visit and VSCE choices as;

$$
L_i(\boldsymbol{\theta}_i) = \prod_t \prod_j P_{i,j,t}(\boldsymbol{\theta}_i^{use}, \gamma_i)^{Y_{i,j,t}} \prod_m \prod_{s \in \mathbb{s}_m} P_{i,s,m}(\boldsymbol{\theta}_i^{use}, \boldsymbol{\theta}_i^{non-use}, \gamma_i)^{Y_{i,s,m}} \tag{20}
$$

Where $\boldsymbol{\theta}_i = [\boldsymbol{\theta}_i^{use} \quad \boldsymbol{\theta}_i^{non-use} \quad \gamma_i]$ is a vector gathering together all the parameters of the behavioural model. $Y_{i,j,t}$ records visit choices such that $Y_{i,j,t} = 1$ if household i chose to visit site j in choice period t and $Y_{i,j,t} = 0$ otherwise. And, $Y_{i,s,m}$ records SP choices where $Y_{i,s,m} = 1$ if household i chose s from the set of scenarios presented to them in choice task m and $Y_{i,s,m} = 0$ otherwise.

Since our data are not sufficiently rich to allow estimation of a parameter vector $(\boldsymbol{\theta}_i)$ for each respondent, in our application we assume that these parameters are held constant across all households. In effect our parameter estimates represent the population average of those parameters. The log likelihood for estimation is given by;

$$
\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln L_i(\boldsymbol{\theta}) \tag{21}
$$

Optimising (21) over the parameters of the model provides maximum likelihood estimates of both use and non-use parameters of the preference function.

Appendix 2: Results

14 Results

The model described in the previous section was estimated using code written in the Gauss programming language. While the parameters of the model were estimated simultaneously using the combined stated and revealed preference data, for ease of exposition, we report the results in a series of separate tables. Each table lists the parameter estimate with the robust standard error of that estimate presented in brackets beneath. The statistical significance of each estimate is presented in the final column of the tables in the form of a *p*-value which indicates the likelihood that the estimate might actually take a value of zero.

Table 6 reports our estimate of the cost parameter. Recall that costs are present in both the revealed and stated preference data taking the form of travel expenses in the former and hypothetical increases in water bills in the latter. As per expectations the parameter estimated on costs is negative and significant with greater than 99% confidence.

Table 6: Parameter Estimates: Costs

Parameter	**Estimate (std. err.)**	***p*-value**
Cost	-0.6696 (-0.0487)	<0.001***

Notes: Statistics report the coefficient estimate with the robust standard error below in brackets. Coefficients significant at the 90% level are highlighted with *, those at the 95% level with ** and those at 99% at ***.

Table 7 presents the first set of parameters relating to the recreation trip choice decision, in this case, a series of constants that indicate the baseline utility associated with broad categories of choice. Those broad categories comprise: not taking an outdoor trip; taking a trip to an outdoor location other than a river site; and taking a trip to a river site. Notice that the River Trip option is the baseline in this set of dummy variables, such that the parameters on the other options are taken to be comparisons to the utility offered by the choice of visiting a river site.

Notice that the parameters for the No Trip option and the Other Outdoor Trip are large, positive and significant. In other words, on any particular day households are more likely to do something other than make a trip to a river site. Finally consider the parameter estimate for the hypothetical bias variable included in the model to capture differences in attitudes to recreation trips when expressed as part of a choice experiment survey as opposed to being revealed through actual behaviour. Our expectations were that households may express a greater likelihood to take recreational trips to rivers in a choice experiment. The negative parameter estimate suggests the opposite may be true but the parameter is not estimated with sufficient certainty to register statistical significance.

Table 7: Parameter Estimates: Trip Choice Constants

Parameter	Estimate (std. err.)	*p*-value
No Trip	12.6749 (1.0151)	<0.001***
Other Outdoor Trip	9.1007 (0.8763)	<0.001***
River Trip	0	Baseline
River Trip (Hypothetical)	2.6998 (0.8433)	0.001***

Notes: Statistics report the coefficient estimate with the robust standard error below in brackets. Coefficients significant at the 90% level are highlighted with *, those at the 95% level with ** and those at 99% at ***.

Table 8 reports parameter estimates on variables examining the relationship between household characteristics and the likelihood of taking a recreational trip. Significant relationships are observed for the income and age variables. While those with higher disposable income are more likely to take trips to the outdoors, the relationship between age and participation in outdoor recreation exhibits a quadratic form. Participation increases with rising age of the Household Representative Person to reach a maximum at around 53 years of age before falling off amongst higher age groups.

Table 8: Parameter Estimates: Participation

Parameter	Estimate (std. err.)	*p*-value
Age	-0.0449 (0.021)	0.033**
Age Squared	0.4238 (0.2218)	0.056*
Employed	0.1359 (0.1779)	0.445
Retired	-0.0611 (0.2386)	0.798
Household Size	0.0065 (0.0102)	0.523
Children	0.0778 (0.0614)	0.205
Income	-0.216 (0.0561)	<0.001***
Urban	0.1219 (0.1407)	0.386

Notes: Statistics report the coefficient estimate with the robust standard error below in brackets. Coefficients significant at the 90% level are highlighted with *, those at the 95% level with ** and those at 99% at ***.

Table 9 lists parameter estimates on variables describing various features of recreation sites. Recall that each recreation site is defined as a 1km cell containing open greenspace allowing access to a river. Not surprisingly, we observe that households are more likely to take trips to cells that offer a greater area of greenspace and a larger extent of river. Compared to the baseline category of natural and semi-natural grassland households tend to disprefer sites with greater extents of moorland or woodland and, for recreational paths, prefer sites in agricultural landscapes.

Table 9: Parameter Estimates: Recreation Site Qualities in Use

Parameter	**Estimate (std. err.)**	***p*-value**
Site Area (ha)	0.0225 (0.0094)	0.016**
River Area (ha)	0.2064 (0.0251)	<0.001***
Urban Site	0.205 (0.2433)	0.400
Landcover: Woods	-0.0411 (0.0273)	0.132
Landcover: Moors & Heath	-0.147 (0.0536)	0.006***
Landcover: Agriculture	0.1425 (0.0524)	0.007***
National Park	-0.0194 (0.0192)	0.311
AONB	0.0037 (0.0187)	0.842
Yorkshire Dales	1.7391 (0.2004)	<0.001***
Southern Pennines	0.8235 (0.2124)	<0.001***
Pennine Dales Fringe	-0.007 (0.3618)	0.985
Water Quality: Bad	0	Baseline
Water Quality: Poor	2.8672 (0.8407)	<0.001***
Water Quality: Good	3.5451 (0.7863)	<0.001***
Water Quality: Excellent	3.6808 (0.7341)	<0.001***

Notes: Statistics report the coefficient estimate with the robust standard error below in brackets. Coefficients significant at the 90% level are highlighted with *, those at the 95% level with ** and those at 99% at ***.

We include dummy variables to capture the special landscapes offered by the Yorkshire Dale and the Pennines and, as expected these exhibit positive and significant coefficients; all else equal households prefer to visit locations in those landscapes than elsewhere. Those dummy variables isolate very similar areas to those identified by the boundaries of the National Park and AONB; so much so, that we observe no significant additional effect from sites with those designations.

Most importantly, for our purposes, observe the parameter estimates on the dummy variables identifying the ecological status of the river within each recreation cell. Here bad quality is the baseline (WFD category bad) and the parameters follow the expected pattern increasing in size from 2.867 for poor quality (WFD category poor) to 3.545 for good quality (WFD moderate category) up to 3.686 for excellent quality (WFD category good and high). One way to interpret these parameters is as indicating the additional utility provided by a site to a household per visit; that is to say the extra benefit a household receives from a recreational visit to a river site with water quality better than the bad ecological status category. The data suggest that water quality has a significant impact on recreational choice decisions.

Table 10 list parameters describing values derived from non-use. Recall that non-use value is assumed to be derived from all river cells and not just those providing accessible open greenspace for recreation.

Table 10: Parameter Estimates: Site Qualities in Non-Use

Parameter	**Estimate (std. err.)**	***p*-value**
Water Quality: Bad	0	<0.001***
Water Quality: Poor	0.0033 (0.0007)	<0.001***
Water Quality: Good	0.0071 (0.0014)	<0.001***
Water Quality: Excellent	0.0093 (0.0017)	<0.001***
Distance Decay in Non-Use Value	-0.984 (0.0502)	<0.001***
Relative Scale of CE	1.4543 (0.1169)	<0.001***

Notes: Statistics report the coefficient estimate with the robust standard error below in brackets. Coefficients significant at the 90% level are highlighted with *, those at the 95% level with ** and those at 99% at ***.

Consider the parameters on the ecological status of rivers for the non-use element of utility. Each of these parameters are statistically significant with a very high level of confidence and progress in a natural order with excellent status being preferred to good status being preferred to poor status being preferred to bad status. Again, these parameters have a natural interpretation as the additional non-use utility derived from a river cell each day as a result of it having water quality higher than the worse quality category.

Observe also that the distance-decay parameter on non-use utility is significantly different from zero indicating that the non-use utility a household enjoys from a river cell declines with the distance that river cell is from their home. Indeed, the parameter value of −0.98 suggests the rate of decline to be approximately equal to the inverse of distance.

Finally, the relative scale coefficient is a technical parameter relating to the variability of answers in the choice experiment compared to in observed recreational choice behaviour. The scale parameter of greater than one indicates that there is greater variability in the hypothetical choice experiment data.

15 Welfare Analysis

15.1 Marginal Welfare Values

Since the parameter estimates described in the last section inform on the structure of household preferences for water quality in use and non-use, they can be used to carry out a welfare analysis of the values derived from improving river water quality.

One relatively easy form of welfare analysis is to examine the 'marginal values' of water quality. To illustrate, consider the parameters estimated on water quality from use for recreation (see Table 9). As we have discussed, these parameters can be interpreted as the extra utility enjoyed by a household from a visit taken to a river site offering a higher level of water quality. We can convert that additional utility into a monetary welfare estimate by dividing through by the negative of the parameter estimated on the cost variable; in effect, the cost variable parameter captures the utility cost of expenditure and, as such, provides the conversion rate between utility and money.

Using that approach we discover that for each trip to a recreation site, households enjoy £4.28 of value from visiting a site with poor water quality, £5.28 of value from visiting a site with good water quality and £5.50 of value from visiting a site with excellent water quality. In all cases, those values are by way of comparison to the value realised in the base case of the site having bad water quality. Of course, households only get that value if they visit a site with improved water quality and that trip decision depends in complicated ways on a household's characteristics and the location of river recreation sites relative to their home. There is no simple way to aggregate these marginal values across households. Indeed, on their own, the marginal welfare values are not particularly helpful for directing policy decisions.

A similar analysis is possible for the non-use parameters from Table 10. Dividing through by the negative of the cost parameter indicates the additional value per day that a household enjoys from improving the water quality in a single river cell. If we imagine that that river cell is right next door to a household's home and multiply up to get a value per year we discover that improving from bad to poor quality results in the household enjoying £1.80 worth of extra non-use benefit per year, improving to good quality results in £3.87 of benefit per year and to excellent quality results in £5.07 of benefit per year. Of course, the distance decay parameter tells us that the non-use value offered by those improvements declines the further the site of the improvement is from a household's home. Indeed, while an improvement of river water quality to excellent might offer £5.07 in a next-door river cell, that value falls to £0.58 per year if the cell is 10km away and to £0.05 per year at a distance of

100km. Again, for policy purposes the value of a change is going to depend in complex ways on the distribution of households across the landscape relative to the location of water quality change.

15.2 Aggregate Welfare Values

To provide welfare values that might be more useful in informing policy choices, a different approach is required that attempts to deal with the complex spatial distribution of households, river cells and river qualities across the landscape. Our objective is to estimate a welfare figure that approximates the aggregate value that might be generated from improving water quality in a stretch of river at any non-specific location in the Yorkshire Water region.

Our solution to that challenge is based on calculating averages across a series of randomly generated water quality changes across the landscape. Roughly speaking, we begin by considering the type of change for which we are interested in estimating an average welfare value. So, for example, we may be interested in improving 10km of river that currently has poor water quality up one level to the good category. To do so we take all river lengths in the landscape that are currently poor quality and randomly select out 10kms worth of river locations. We then use the model parameters to perform the complex calculations needed to identify how much value those particular set of quality changes bestow on households in the Yorkshire Water region. By repeating that task many times for different randomly selected sets of 10kms worth of river locations, we can build up a distribution of the aggregate value of 10kms of river improvement from poor to good quality. The figures we report in this segment are the average of that range of values based on calculations from 100 random samples.

To simplify the process of sampling and valuation, we repeat the process of identifying 1km river cells carried out for the survey data area for the whole of the Yorkshire Water region. In particular, as shown in Figure 9 we begin by creating a grid of cells that overlay rivers identified from the Ordnance Survey's main rivers network dataset.

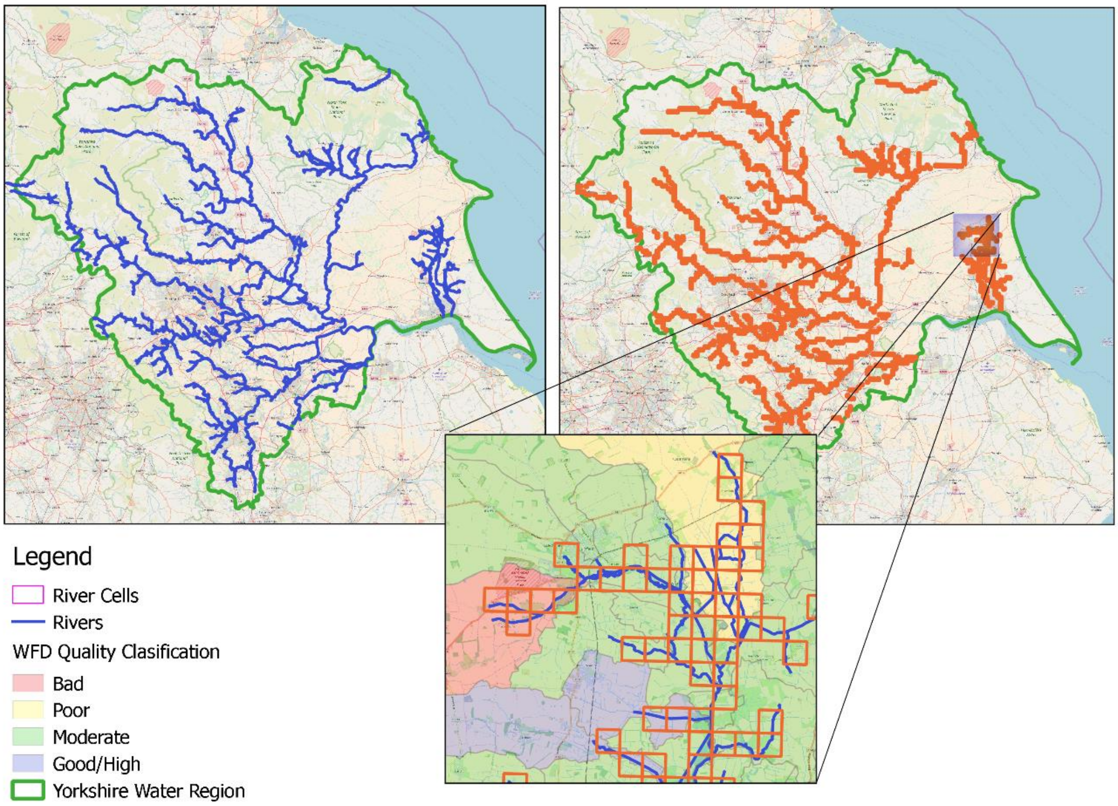


Figure 9: River Cells in the Yorkshire Water Region

Having reallocated cells containing only small fragments of rivers to the adjoining cells we end up with 1,624 river cells in the region. Each of those river cells, it is assumed, can convey non-use value to households in the region, though how much value each household derives from each river cell depends, of course, on their proximity to that cell. We assigned a water quality to each cell by overlaying the 2015 WFD ecological status data from the Environment Agency.

Next, using the ORVal greenspace map, we identified river cells that provided open greenspace for recreation and used a variety of datasets to establish the characteristics of the recreation sites in each of those cells needed for the model. Of the 1,624 river cells, 749 were identified as providing for recreation (see Figure 10).

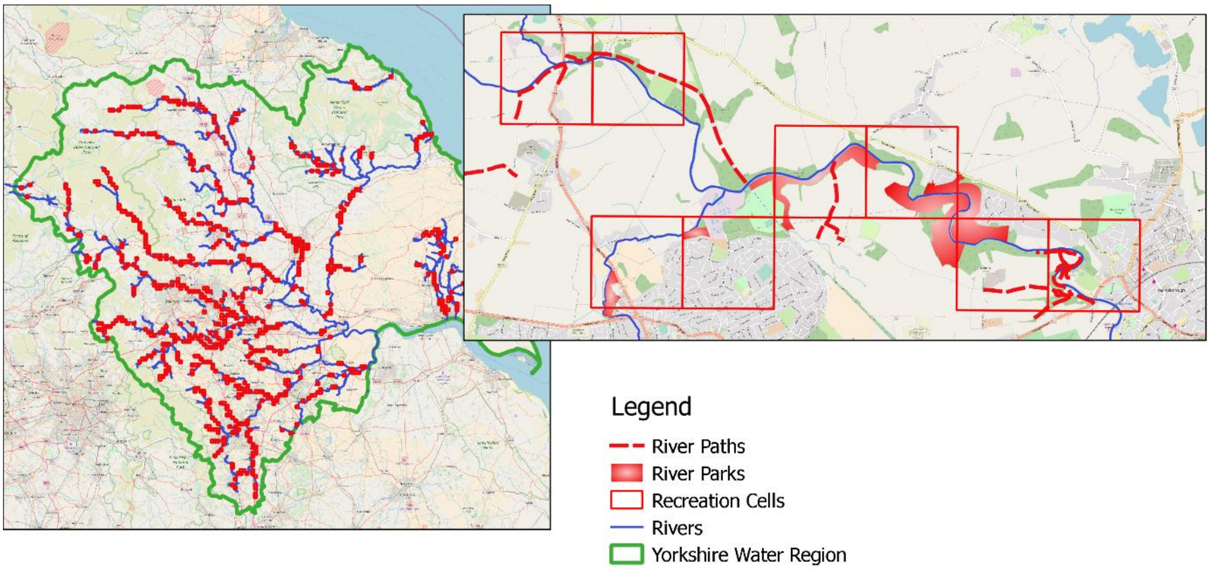


Figure 10: Recreation Cells in the Yorkshire Water Region

To understand the value of changes in water quality at the various cells we needed to establish the size and distribution of the population in the in the Yorkshire Water region. For that purpose, we collated population data at the scale of Lower Super Output Areas (LSOAs) of which 3,262 are located in the Yorkshire Region (see Figure 11).

We determined the characteristics of households living in each of those LSOAs from 2011 census data pulling out information on those characteristics used in estimating the model. In particular, we established the number of households in each LSOA, the age distribution of the Household Reference Person (HRP), the proportion households with children, the average household size, the proportion of HRPs that are employed and the proportion of HRP's that are retired. Information on the mean household income for each LSOA was sourced from Experian. Assuming each household in an LSOA lived at the population weighted centroid of the LSOA, we calculated straight line distances to each river cell and then travel costs to each recreation cell through the roads network again using the RoutingKit software (see Section 12.3).

With the data collated, our welfare averaging procedure was implemented by applying water quality improvements at some randomly selected set of rivers cells, assuming that each river cell represents approximately 1km of river length. The model was used to estimate the values of those changes for the households in each LSOA and then summed to give a welfare value of the changes for the 2,186,513 households living in region (figure from 2011 Census).

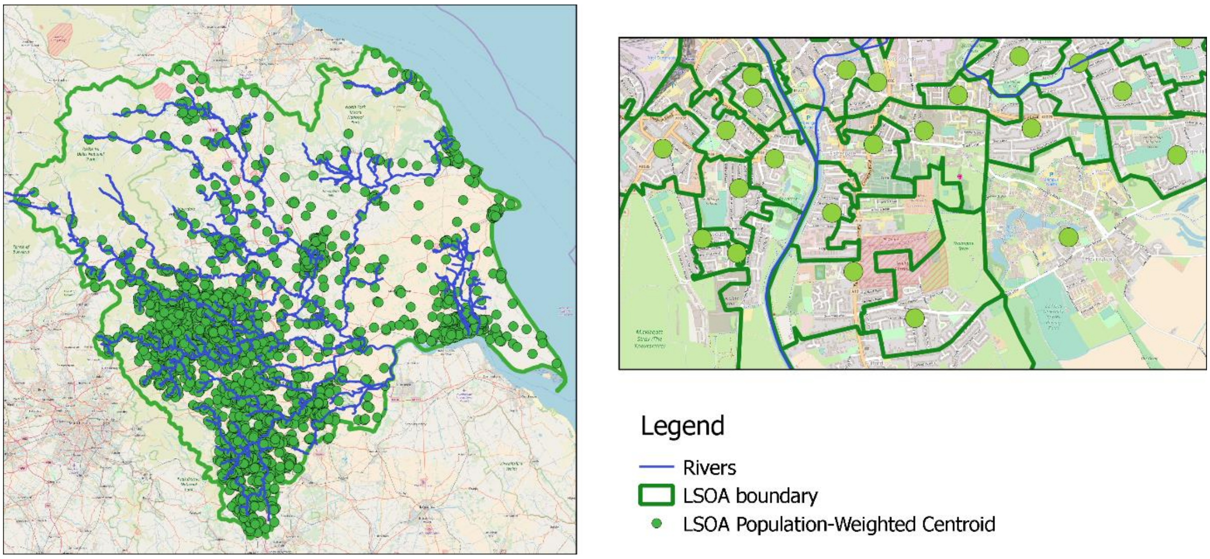


Figure 11: LSOA Population Weighted Centroids in the Yorkshire Water Region

15.3 Aggregate Welfare Estimates

Table 11 presents a first set of estimates from the welfare averaging procedure. The figures concern the valuation of a one category improvement in water quality status amongst river cells that are currently of bad quality. Notice the valuation was completed for three different extents. First a 1km extent in which just one poor quality river cell was improved at each iteration of the procedure. Second for a 10km extent in which 10 cells were improved at each iteration. And finally, for a 50km extent which entailed improving 50 randomly selected river cells at each iteration.

For each extent, the aggregate value is decomposed into the benefits that households in the Yorkshire Water region enjoy from the improvement as a result of recreational use of the improved site (use value) and as a result of increased non-use value. The final column of the table records the total welfare improvement as the sum of use and non-use value.

Table 11: Average aggregate values for improvements from Bad to Poor quality

Length of River Improved	Average Aggregate Value		
	Use	Non-Use	Total
1km	£ 8,566	£ 70,857	£ 79,423
10km	£ 53,461	£ 689,348	£ 742,809
50km	£ 288,030	£ 2,845,077	£ 3,133,107

Notes: Values are expressed in 2008 £s per annum for households in Yorkshire Water Region

With regards to a single cell (1km) improvement we observe an average aggregate use value of £8,566. In other words, of the 100 randomly selected bad quality river cells the average improvement in recreation values as a result of the one category improvement in quality was £8,566. Notice that for many of those random draws river cells will have been selected that do not contain greenspace access for recreational enjoyment of the river and hence will return a use value of zero – not all river cells are also recreational locations.

The value of those same improvements for non-use value are calculated as being £70,857, around 8 times larger than the use values. At first glance that might appear a little surprising. When we examined the marginal values for this improvement we noted a per visit use value of £4.28 per household compared to a paltry half a penny (£0.005) a day non-use value (£1.80 divided by 365 days). A number of factors are at work here. First, as just noted, only 46% of river cells also provide recreational access. More often than not, therefore, the random cell chosen for improvement will only provide non-use value. Second, while the benefit enjoyed onsite from a visit to an improved location is relatively high, households make limited number of visits to river locations, around once a quarter depending on household characteristics. Even when taking a trip, a household has at least another 748 recreation river cells that they might visit instead of the improved cell. Accordingly, while the visit value of an improvement is large, the likelihood of visits to a particular improved site is rather small. In contrast, while the daily non-use value from an improvement is very small, our model suggests that some small value is realised every day by every household in the region.

The final column records the sum of the use and non-use values suggesting that a programme to improve some 1km stretch of river from bad to poor river quality in the Yorkshire Water region would (on average) deliver £79, 423 per annum of welfare value to the residents of that region.

The remaining two rows of Table 11 record average aggregate values for 10km and 50km improvements of bad water quality respectively. Notice the values scale reasonably proportionately so that the use, non-use values increase roughly in line with the extent of the improvement.

Table 12 undertakes the same analysis but this time considering a one category improvement in river cells currently in poor water quality. Notice that the model suggests that the benefits of this improvement are regarded as giving greater value than the one category improvement from bad water quality. The exact reasons for that observation arise from a combination of the relative sizes of the preference coefficients estimated in the model as well as the specific spatial distribution of poor water quality river cells with respect to the location of recreation sites and to households' homes. This increase in value particularly impacts on use value such that the average aggregate use value are more similar in magnitude to the non-use values than was observed for the changes examined in Table 11.

Table 12: Average aggregate values for improvements from Poor to Moderate quality

Length of River Improved	Average Aggregate Values					
	Use		Non-Use		Total	
1km	£	79,890	£	123,444	£	203,335
10km	£	353,938	£	1,215,162	£	1,569,100
50km	£	1,309,573	£	4,963,025	£	6,272,598

Notes: Values are expressed in 2008 £s per annum for households in Yorkshire Water Region

Table 13 provides details of average aggregate values for improvements from moderate to good/excellent water quality. Here we observe relatively small use values, a fact that must reflect the similarity in the coefficients of the model for moderate and good water quality (see Table 9).

Table 13: Average aggregate values for improvements from Moderate to Good/Excellent quality

Length of River Improved	Green to Blue		
	Use	Non-Use	Total
1km	£ 3,157	£ 75,206	£ 78,363
10km	£ 31,962	£ 762,092	£ 794,054
50km	£ 152,849	£ 3,726,284	£ 3,879,133

Notes: Values are expressed in 2008 £s per annum for households in Yorkshire Water Region

While the values provided in Table 11 to Table 13 concern improvements in river cells of a particular current water quality level, the key objective of this project is to provide a value that can be used in generic project appraisal for Yorkshire Water. Table 14 provides those figures. Here rather than selecting only from river cells of one particular quality, we have selected any river cell at random and calculated the value of a one category increase in river water quality. So, if a randomly selected cell was currently poor water quality then the procedure calculated the value of a change to moderate water quality. Likewise, for a randomly chosen river cell that is currently moderate quality the procedure calculated the value of a change to good/excellent water quality. The procedure can be justified as being our best *a priori* guess at the value that would be generated by a future scheme delivering one category water quality improvements at some as yet to be specified location in the region; assuming that that the policy could be implemented in any location in the region with equal likelihood.

Table 14: Average aggregate values for general one category improvement in river water quality

Length of River Improved	One Category Improvement		
	Use	Non-Use	Total
1km	£ 6,323	£ 77,228	£ 83,551
10km	£ 42,331	£ 800,439	£ 842,771
50km	£ 241,358	£ 3,911,587	£ 4,152,945

Notes: Values are expressed in 2008 £s per annum for households in Yorkshire Water Region

Not surprisingly, the average aggregate values recorded in Table 14 are of a magnitude intermediate to the specific change valuations recorded in Table 11 to Table 13. They suggest that the value of a one category improvement over a 1km length of river is around £85,000 per year rising to around £4million pounds per year for a 50km improvement.

Of that total value only 10% to 20% is made up of use value from recreation, the remainder is non-use value. Of course, the value from use is derived from observations of actual recreation behaviour while the non-use value is identified solely from a hypothetical survey exercise. In applying these values in policy analysis, therefore, there is an argument to suggest that the use value element has greater validity than the non-use element. Indeed, we suggest that the use value and total value be taken as indicate reasonable bounds to the magnitude of welfare changes resulting from water quality improvements in rivers.

16 Summary and Recommendations

- This report provides a state of the art analysis of a data set recording stated and revealed preferences for water quality improvements in the Yorkshire Water Region. The revealed preference data came in the form of records of recreational trips to rivers of different water qualities across the region. The stated preference data came in the form of choices over different hypothetical future patterns of water quality across the region.
- The research developed an original and state-of-the-art estimation method that allowed for the simultaneous analysis of the revealed and stated preference data. That method allowed for the coherent separation of welfare estimates into a use and non-use values.
- The model estimated from the data returns parameter estimates that broadly conform to prior expectations both in sign and magnitude. Most importantly, the models report significant sensitivity to river water quality both in observed choices over recreational sites and in stated choices over hypothetical future patterns of river water quality.
- The estimated model can be used to calculate welfare values at a fine resolution; that is to say, the value to residents of the Yorkshire Water region of some particular change in river water quality at some particular location in the region.
- The key objective of the research project, however, has been to develop generic welfare estimates that might be used across a range of appraisal scenarios. To that end, a procedure for estimating average aggregate welfare values is developed in the report. The final output of that undertaking being reported in Table 14.
- Those generic values indicate that a one category increase in river water quality over a 1km extent of river has an annual welfare value of around £85,000 to households in the Yorkshire Water region. Of that total value, around 10% is derived from recreational use of rivers the remainder coming in the form of non-use value.
- The generic welfare values scale roughly in proportion to the extent of river improved such that one category improvements to 10km of river deliver roughly £850,000 of value per year and 50km of improved river delivers roughly £4million of value per year.
- Since the estimated of value from use are derived from observations of actual recreation behaviour while the non-use value estimates are identified solely from a hypothetical survey exercise, it is often argued that the former have greater validity. Indeed, within the context of the results recorded in this research, one way to proceed would be to take the use value and total value as indicating reasonable bounds to the magnitude of welfare changes resulting from water quality improvements in rivers.

17 References

Bateman, I.J., Day, B.H., Georgiou, S., Lake, I., 2006. The aggregation of environmental benefit values: Welfare measures, distance decay and total WTP. Ecological Economics 60, 450-460.

Bateman, I., M. Agarwala, A. Binner, E. Coombes, Day, B. , Ferrini, S., Fezzi, C., Hutchins, M., Lovett, A., and Posen, P., Spatially explicit integrated modelling and economic valuation of climate driven land use change and its indirect effects (2016), *Journal of Environmental Management*, 181, pp 172-84

Day, B., Bateman, I., Binner,A., Ferrini, S., and Fezzi, C., (2017). Structurally-consistent estimation of use and non-values for landscape-wide environmental change, *Journal of Environmental Economics and Management*, revise & resubmit Nov 2017.

Day, B. H. and G. Smith (2016). Outdoor Recreation Valuation Data Set Construction. Report for Defra, June 2016, available online at http://leep.exeter.ac.uk/orval/documents

Department for Transport (2014). Values of Time and Vehicle Operating Costs: TAG Unit 3.5.6, Department for Transport.

Department for Transport (2015). Provision of Market Research for Value of Travel Time Savings and Reliability, Phase 2 Report, Department for Transport.

George, E., Mudholkar, G., 1983. On the convolution of logistic random variables. *Metrika* 30, 1-13.

Hime, S., Bateman, I.J., Posen, P., Hutchins, M., 2009. A transferable water quality ladder for conveying use and ecological information within public surveys. CSERGE working paper EDM

McFadden, D., 1978. Modelling the choice of residential location, in: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (Eds.), Spatial Interaction Theory and Planning Models. Amsterdam, Amsterdam, pp. 75-96.

Morey, E.R., Rowe, R.D., Watson, M., 1993. A Repeated Nested-Logit Model of Atlantic Salmon Fishing. *American Journal of Agricultural Economics* 75, 578-592.

Phaneuf, D.J., von Haefen, R.H., 2009. Chapter 12 Estimating the Demand for Quality with Discrete Choice Models, Quantifying Consumer Preferences. Emerald Group Publishing Limited, pp. 317-347.

Schaafsma, M., Brouwer, R., 2013. Testing geographical framing and substitution effects in spatial choice experiments. *Journal of Choice Modelling* 8, 32-48.