Valuing Environmental Impacts: Practical Guidelines for the Use of Value Transfer in Policy and Project Appraisal

Technical Report

Submitted to

Department for Environment, Food and Rural Affairs

February 2010

eftec
73-75 Mortimer Street
London W1W 7SQ
tel: 44(0)2075805383
fax: 44(0)2075805385
eftec@eftec.co.uk
www.eftec.co.uk

Report prepared for the Department for Environment, Food and Rural Affairs

by:

Economics for the Environment Consultancy (eftec)
73 – 75 Mortimer St, London, W1W 7SQ
Tel: 020 7580 5383
Fax: 020 7580 5385
www.eftec.co.uk

Main contributors (in alphabetical order):

Prof. Ian Bateman (University of East Anglia)
Dr. Roy Brouwer (Institute for Environmental Studies, VU University, Amsterdam)
Matthew Cranford (eftec)
Dr. Stephanie Hime (eftec)
Ece Ozdemiroglu (eftec)
Allan Provins (eftec)

Acknowledgements:

The study team would like to thank Prof. Ken Willis (University of Newcastle upon Tyne) and Prof. Stale Navrud (Norwegian University of Life Sciences) for their input to the development of the guidelines and the members of the Steering Group for their comments on the previous version of the Technical Report.

eftec offsets its carbon emissions through a biodiversity-friendly voluntary offset purchased from the World Land Trust (http://www.carbonbalanced.org/), and only prints on 100% recycled paper.

CONTENTS

EXECUTIVE SUMMARY

ES1. Background

Value transfer (which is also known as 'benefits transfer') is a process of using secondary valuation evidence sourced from previously undertaken studies to apply to a new decision-making context. Its particular appeal lies in it being a quicker and lower cost approach compared to the alternative of specifically commissioning a primary valuation study. There is, however, a degree of uncertainty and potential for error inherent within the value transfer process. This arises from reliance on expert judgement in identifying and applying suitable valuation evidence in different contexts and also in some cases a lack of suitable studies from which to source valuation evidence.

This report and the accompanying *Value Transfer Guidelines* focus on the use of value transfer in appraisal and provide guidance for improving the quality and accuracy of valuing environmental impacts.

ES2. Objective

The objective of the Technical Report is to review and define best practice for value transfer. This includes reviewing key technical issues – for example the application of geographical information systems (GIS) in value transfer - and providing recommendations that inform the Value Transfer Guidelines.

In reviewing best practice, there is a need to reconcile the 'state of the art' with the practical use of value transfer by analysts tasked with appraising environmental impacts. The state of the art is typically driven by developments in academic research and as demonstrated in this report can involve sophisticated analysis. Where these developments improve the accuracy of value transfer there is a clear need for practical application to be based on such best practice principles. However, appraisal effort is governed by both time and resource constraints, meaning that the state of the art is not always feasible. There are also instances where greater uncertainty in evidence can be accommodated in decision-making, implying that less sophisticated, but still robustly implemented, analysis is sufficient.

ES3. Overview of requirements for value transfer

Value transfer is applicable to a wide range of both market priced and non-market goods. The transfer of values for market priced goods is typically straightforward and much attention instead focuses on the potential for undertaking transfers for non-market goods such as those provided by the environment. The typical application is one in which willingness to pay (WTP)[1] estimates from a previously undertaken study (the 'study good') are transferred to some policy context concerning a proposed change in provision of the good in question (the 'policy good'). The change in provision of the policy good could be a change in quality (e.g. water quality),

[1] Willingness to pay (WTP), either in terms of WTP for a gain in provision or WTP to prevent a loss in provision, is the most commonly estimated measure of the value of environmental goods and services.

quantity (e.g. the size of a protected area) or access (e.g. provision of visitor facilities at a woodland site).

Estimating the total value of some provision change for an environmental good requires that three essential questions are addressed:

i). *What is the change in provision of the good in question*: understanding the change in the quantity and/or quality of the good, as determined by relevant scientific and technical assessments, is an essential precursor to value transfer analysis.

ii). *How valid and robust is the available economic valuation evidence*: assessment of the quality of source valuation evidence should be informed by three broad categories of validity test:
- *Scope sensitivity:* are valuations responsive to the scale (or 'scope') of the provision change under assessment; i.e. WTP should not fall as the scope of a good increases (put simply 'more' is better than 'less' subject to satiation).
- *Tests of theoretically derived expectations:* economic theory establishes prior expectations which can be tested for; for example it is expected that WTP will increase with an individual's income and fall as the availability of substitutes increases.
- *Procedural invariance:* economic theory suggests that WTP should not vary due to 'irrelevant factors' related to the methodology used to estimate it. For example, in stated preference studies, tests of 'procedural invariance' can be useful indicators of whether respondents hold well-formed preferences characteristic of valid economic values or are simply 'constructing' those preferences with respect to the ad-hoc heuristics they see in the questionnaire design.

iii). *How do changes in the provision of a good and the presence of substitutes alter its marginal value*: in many cases WTP should be expected to be 'non-constant' implying that substantial care is required in estimating aggregate (total) benefits and costs of policies and project. This is dependent on the marginal benefit individuals obtain from a unit increase in the provision of a good (so-called 'marginal utility') and/or available alternatives (so-called 'substitute' goods).

Ensuring that these three questions are addressed appropriately is integral to robust and defensible value transfer analysis.

ES4. Approaches to value transfer –unit value or value function transfer?

Much academic effort has focused on the development and testing of value transfer approaches, which can be broadly categorised as: (i) unit value transfer; and (ii) value function transfer.

Unit value transfer is the simplest approach and is frequently used in the appraisal of environmental impacts. The validity of the approach is dependent upon the correspondence between the context of the study good valuation and the context for the policy good. At some level the two contexts will always be dissimilar; for instance the distinct ecosystem habitats and the sites that study and policy goods are found in are all essentially unique. The key issue however is the degree to which this dissimilarity affects values, which in turn will determine the appropriateness of unit value transfer.

Value function transfer relies on the application of statistical models that describe how the value of the change in provision of the study good changes with various explanatory factors (the value function variables). The value function is used to 'predict' the value for the policy good context by applying the values of the explanatory variables at the policy good site.

A common expectation is that because value function transfers allow greater control for differences between the policy good and study good context, they should be preferred on the grounds of the likely higher accuracy of estimated values. There are however two important qualifications to this expectation:

(i) *A value function should be specified to focus on factors which are generic across the study and policy good contexts*

This precludes the inclusion of context-specific and ad-hoc variables that significantly assist in improving the estimation of models to explain the study site data but have no relevance to the policy site. Instead the value function to be transferred should focus on general relationships that economic theory suggests should hold across the study and policy good contexts. In particular:

- The extent (or 'scope') of the change in provision under consideration;
- The costs of using the good - for a physically located good this mainly relates to the proximity of the site to an individual's home and travel and time costs;
- The availability of substitutes; and
- The individual's income constraints.

(ii) *Unit value transfer may generate as accurate a result as value function transfers if the policy and study good contexts are* <u>*very similar*</u>

In practice the value function transfer approach 'comes into its own' when applied in relation to dissimilar sites (but not necessarily dissimilar contexts) and these differences relate primarily to expectations based principles set out above. Where close correspondence between the scope of the change, costs of use, availability of substitutes and income constraints of the study and policy good can be demonstrated, then unit value transfer is likely to be sufficient. In cases where the correspondence for one or more of these relationships is questionable, value function transfer should be preferred.

Overall the recommendation for best practice is that when transferring across similar goods and sites, unit value approach is likely to be sufficient. When transferring across similar goods, but dissimilar sites, value function transfer is more appropriate and the specification of those functions should be restricted to include only generic variables for which there are prior economic expectations.

The principles for the 'choice' between unit value and value function transfer of course give rise to the question of how to assess if policy and study goods and sites are sufficiently similar.

Lack of scrutiny for this question has been a failure of value transfers to date, and, moreover source studies often report a relatively haphazard set of statistics from which such assessments could be made. While studies often provide data characterising certain aspects of the sample this does not always extend to the underlying population and information regarding the physical characteristics of valued goods or sites is rarely systematically presented.

1 INTRODUCTION

- *This section introduces the background to the Technical Report, overviews its contents and places it in the wider context of the value transfer guidelines project.*
- *The primary focus of the report is in defining best practice for value transfer.*

1.1 Background

The Ecosystems Approach Action Plan (Defra, 2007) identifies a specific action that Defra, in partnership with the Environment Agency, Natural England and the Forestry Commission, develop a value transfer strategy for use in valuing ecosystem services. This is aimed at addressing a core principle of the Action Plan to ensure that the value of ecosystem services is fully reflected in policy and project decision-making. The focus of these guidelines is the ability to use value transfer in appraisal as well as improving the quality and accuracy of the practice of valuing environmental and ecosystem service impacts in general[2].

1.2 Objective

Value transfer is a process by which economic valuation evidence that has been generated in one context, i.e. from a previously undertaken primary valuation study, is applied in another context for which valuation evidence is required. From the perspective of the appraisal of environmental and ecosystem service impacts, value transfer is a particularly appealing approach to economic valuation, given its potential expediency and (typically) lower cost compared to primary valuation. However, there is a degree of uncertainty and potential for error inherent within the value transfer process. This stems from the reliance on expert judgement in identifying and applying suitable valuation evidence in different contexts and paucity of suitable studies. Thus a greater level of subjectivity is implied with the application of value transfer analyses in comparison with a primary valuation approach.

The overall objective of this study is to produce a set of user-friendly Guidelines for the use of value transfer in policy and project appraisal. The target audience for the Guidelines is primarily economists and policy analysts in Central Government departments and executive agencies. The guidelines will also raise the profile of value transfer with all policy and research specialists involved in evidence-based policy decision-making. By providing recommendations for best practice, the Guidelines (and this accompanying Technical Report) are intended to ease the concerns about consistency and dependency upon expert judgement.

The specification for this study outlines a detailed set of objectives for the value transfer guidelines:

[2] Note on terminology: While the use of the term 'benefits transfer' is commonplace, the approaches described here and in the Guidelines can equally well be applied to the transferral of cost estimates. Given this, we use the more generic term 'value transfer' to embrace both the benefits and cost side.

I. *Review best practice for value transfer, account for key technical issues that arise and make recommendations for how these issues can be addressed in the guidelines*. This is to include the role of geographical information systems (GIS) both in immediate timescale and longer term.

II. *Review the policy and project appraisal contexts for use of value transfer ensuring that these are taken into account in the design of the guidelines*. In particular, different contexts may affect the degree to which value transfer is feasible and determine when primary valuation may be required.

III. *Provide clear steps for the practical use of value transfer in the Guidelines*. Illustrative examples to demonstrate how analysis would be undertaken in practice are to be provided in the value transfer guidelines and key resources for assisting the analysis should be highlighted.

IV. *Identify existing valuation databases as part of the 'key resources'*. These are important sources of data for value transfer.

V. *Provide a protocol for primary valuation studies with respect to subsequent use of results in value transfer*. Here the intention is to facilitate the future use of value transfer by ensuring that primary valuation studies, especially those that are publicly funded, report sufficient detail and data to enable subsequent use of results via value transfer.

VI. *Provide a range of case studies to demonstrate the practical use of the guidelines*. These are intended both to provide practical examples and to address particular issues of importance. As such they demonstrate the breadth of potential value transfer applications in terms of techniques, appraisal contexts, environmental or ecosystem impacts and the limitations of value transfer.

VII. *Development of a dissemination plan for the Guidelines*. In particular this is to feature a web-based 'interactive resource package' which makes use of the case studies in demonstrating the practical use of value transfer and a training workshop for the guidelines target users.

The above objectives recognise both the 'state of the art' and practical context for the value transfer. Largely the former is driven by developments in academic research and can entail sophisticated analysis. To the extent that these developments improve the accuracy of value transfer then striving to establish 'state of the art' practice in value transfer is a desirable goal. That said, policy analysis is itself subject to time and resource constraints; in practice the guidelines need to reflect the capabilities and requirements of the decision-making process. Balancing these two objectives is key challenge for this study in order to ensure resulting guidance is rigorous and realistic in the expectations for its use and for the application of value transfer.

1.3 Purpose of the Technical Report

This report focuses primarily on Objective I with respect to defining best practice for value transfer. Within this, aspects of Objectives III, V and VI are also addressed. The other objectives are covered in other study outputs.

Section 2 provides an overview in terms of the fundamental requirements for value transfer.

The Guidelines highlight two fundamental approaches to value transfer: (i) unit value transfer (sometimes adjusted for the characteristics of the case study under consideration); and (ii) value function transfer transferring the relationship between economic value and the factors which determine it. **Section 3** provides a detailed technical account of both approaches, discussing their implementation and validation. Key principles set out in Section 3 are reflected in the Guidelines.

Sections 4 to **6** provide case studies of the use of function transfer that adhere to the principles set out in Sections 2 and 3. Specifically they examine certain issues that arise when conducting value transfers for spatially fixed resources. These issues are addressed with respect to a common good[3], namely, the non-market benefits of water quality improvements. These case studies consider (i) transfers across diverse European countries and the choice of unit value of value function approach; (ii) transfers across UK regions using value function transfer and GIS; and (iii) the use of transfer techniques for estimating aggregating values. These various examples illustrate ways in which economic theory provides a guide for building transferable value functions and how the natural variation of the real world can be incorporated and allowed for within value transfer studies.

Finally **Annex 1** provides a brief summary of value transfer literature, while **Annex 2** contains technical content to support the case study example in Section 4.

[3] We use the term 'good' as per standard economics, to indicate any item, act or resource which generates utility (either positive or negative) and hence affects welfare. Changes in the provision of goods (or their substitutes or complements) can therefore result in welfare gains or losses. It is the value of those gains and losses which is the objective of value transfer.

2 FUNDAMENTAL QUESTIONS WHEN CONSIDERING VALUE TRANSFER

This section addresses three fundamental questions which have to be considered prior to undertaking any value transfer exercise:

i). What is the change in provision of the good in question (i.e. the change in the quantity and/or quality of the good)?
ii). How valid and robust are the available estimates of marginal economic value?
iii). How do changes in the provision of a good and the presence of substitutes alter its marginal value?

2.1 Overview of value transfer

The transfer of study findings from one situation to another is an attractive alternative to commissioning expensive and time consuming original research to inform decision-making. As we will discuss in this section, such undertaking may involve both: (i) the transfer of findings regarding the quantity (and/or quality) of a good to be provided by a policy change[4]; and (ii) the transfer of unit ('marginal'[5]) values. Only by understanding both the physical changes in provision and their marginal value can we predict the total value of that change. The term value transfer (or 'value transfer') most commonly refers to that combined exercise of transferring or predicting physical provision changes and transferring marginal values

Value transfer is applicable to a wide range of goods including both market priced and non-market goods. The transfer of market priced goods is typically more straightforward although the prediction of physical provision changes may be challenging and prices may require adjustment if the original data are from an earlier period or if the transfers are being made between dissimilar contexts. However, for some considerable time, interest has focussed upon the potential for undertaking transfers for non-market goods such as those provided by the environment[6]. Here there are typically no externally available values and so these are taken from primary valuation studies. These may be conducted using a variety of methods including revealed and stated preference techniques (Bateman et al., 2002a; Champ et al., 2003). We refer to these as 'study' values from which the objective of transferral is to estimate 'policy' values for the decision-making context under consideration. The most generic application is one in which study derived values are transferred to some policy context concerning a proposed

[4] While we use the example of a policy change here, we might be equally interested in changes arising from alterations in market conditions, or in environmental change (or some combination of policy, market and environment drivers) or the impact of some proposed project.

[5] In the terminology of economics, the term 'marginal' refers to an additional unit of a good. Therefore a marginal value is the extra value generated by an additional unit of that good.

[6] Useful examples of value transfers for environmental goods include Barton, 2002; Bateman and Brouwer, 2006; Bateman and Jones, 2003; Bergland et al., 1995; Brouwer, 2000; Brouwer and Spaninks, 1999; Desvousges et al., 1992; Downing and Ozuna, 1996; Johnston et al., 2005, 2006; Leon-Gonzalez and Scarpa, 2008; Lindhjem and Navrud, 2008; Moeltner et al., 2007; Muthke and Holm-Mueller, 2004; Navrud and Ready, 2007a,b; Ready et al., 2004; Santos, 2007; Scarpa et al., 2007; Zandersen et al., 2007. Note that there is also a growing literature concerning the transferral of non-environmental non-market goods such as transport alternatives and health goods.

change in provision of the good in question. Note that we frequently refer to study and policy 'goods' or 'sites'. However, the techniques described are not restricted to locationally defined areas and can be applied to a range of goods including marketed goods and non-market, non-located goods (e.g. health; see for example, Brouwer and Bateman, 2005a).

The calculation of the total value of some provision change for a good or service requires three vital pieces of information:

a) A robust estimate of the marginal (i.e. per unit) value (typically measured as the marginal willingness to pay; WTP)[7];
b) Clear understanding of the change in provision of the good under consideration; and
c) Knowledge of how (a) might alter as (b) changes.

These requirements can be formulated within the following questions, which must be answered before undertaking a value transfer for the provision of any good as a result of some change in its provision (whether induced by a policy, market or environmental forces or any other driver of change):

iv). What is the physical change in provision of the good in question (i.e. the change in the quantity and/or quality of the good)?
v). How valid and robust are the available estimates of marginal value?
vi). How do changes in the provision of a good and the presence of substitutes alter its marginal value?

In this section we consider methodological approaches for answering the above questions and, where appropriate, provide illustrative case study examples.

2.2 How much of the good under evaluation is provided both with and without any change (calculating the physical provision change)?

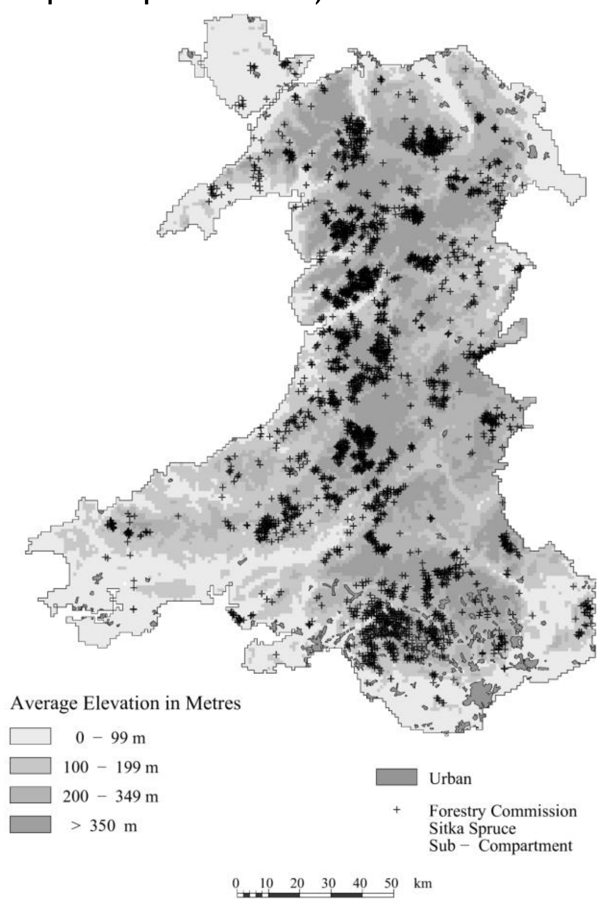
One of the major requirements of practical value transfer is to ensure, from the outset, that the change in provision is understood and quantified. It is clearly unreasonable to expect either primary valuation studies or value transfer to derive robust values for a good when the quantity change in provision (and/or the quality change) is unknown.

In some cases the provision change will be known a-priori. Indeed the impetus for the value transfer study may be the creation of a specified provision of a good (e.g. the creation of a 100

[7] The common usage of the term 'willingness to pay' is the WTP for a gain in provision. This is more properly (but much less frequently) called the 'compensation surplus' (or for market priced goods, the 'compensating variation'), literally the variation (here reduction) in an individual's income which just offsets (compensates for) an increase in the provision of the good under consideration. While this is the most common measure of the economic value of a provision change it is actually only one of four such measures (Hicks, 1943). The others being the amount an individual is willing to accept (WTA) in compensation to just offset a loss in the provision of a good (correctly termed the compensating loss); the amount of income an individual that is willing to pay to avoid a loss of provision (the equivalent loss); and the amount of income compensation which is just equivalent to foregoing a gain of the good (the equivalent gain). Bateman, et al., (2000a) provide stated preference estimates of all four 'Hicksian' welfare measures within a common public goods case study.

acre wood[8]). Provided that the final provision change is known with certainty then this bypasses the need for further consideration of this issue. However, this is not always the case. The provision change consequences of, say, a given land use policy may not be known from the outset and have to be assessed before value transfer valuation can proceed. In the case of environmental goods, prediction of the quantity change in provision typically requires a prior basis of natural science. As an example of this we consider a case study concerning a good for which the marginal value can be derived from the market (for simplicity) but the provision change is not a-priori known. This concerns the case of estimating the quantity (and subsequently the value) of timber arising from a potential policy of widespread afforestation of a large area of the UK (e.g. the entirety of Wales). This example is drawn from a wider cost-benefit analysis in which a range of competing land use options are assessed (Bateman, et al., 2003). However, for the analysis of potential timber values the initial task is to estimate the potential change in provision from afforestation. To answer this, a large database of tree growth observations was accessed. This detailed information on timber yield from locations spread across the study area as shown in **Figure 2.1**.

Figure 2.1: Location of Forestry Commission sub-compartments for one species (Sitka spruce) in Wales (superimposed upon elevation)



Source: Bateman et al. (2003)

[8] With due acknowledgement to A.A. Milne and apologies to the metric system.

Examination of Figure 2.1 shows that we have observations on tree growth scattered over the study area, but that there are also many locations where we do not have such records[9]. Therefore an initial task is to take what information we have and use it to predict how trees would grow in all areas (only then will we eventually be able to compare the value they generate with the value of other land uses). We could use a variety of methods to estimate tree growth in all areas. We could just assign every area with the growth rate of the nearest observation. This might be acceptable for a rough approximation. However, if we want a more accurate answer then we need to build a model of tree growth that can be applied in all areas.

While we do not have tree growth observations in all areas, there is a variety of data concerning the physical environment characteristics of all areas of the UK. These data include variables such as elevation, rainfall, daily temperature, soil type and a wide range of agro-climatic variables, land use, etc. These variables determine not only the growth of trees but the production (and hence policy response) of many ecosystem goods and services. Their effect can be quantified through modelling exercises. For example, **Table 2.1** details a regression model predicting the yield class (i.e. growth rate) of a particular tree species as measured in m^3/ha/year.

Table 2.1: Best fitting regression model predicting Sitka spruce yield (m^3/ha/year)

Predictor	*Coefficient*	*t-ratio*
Constant	16.710	47.92
Elevation	-0.009	-22.31
Rainfall	-0.002	-15.65
Shelter	0.024	3.20
Good soil	0.805	10.00
Poor soil	-4.883	-5.05
Area of forest	0.004	10.43
Year planted	0.050	10.31
1st Rotation	-1.928	-17.64
Mixed crop	-0.308	-4.02
Parkland	0.948	10.10
Ancient forest	0.927	3.00
Uncleared forest	2.641	11.61
Low inputs	-0.085	-10.49
Lower inputs	-0.434	-4.59
Lowest inputs	-5.142	-6.73
Observations (n) = 4307		
R-sq(adj.) = 42.8%		

Models such as those given in Table 2.1 tell us a lot about how the physical characteristics of the environment determine the change in provision arising from some policy or other driver. The first column set outs the determinants of timber yield. The value in the middle column describes the relationship between each variable and the timber yield. The 'Constant' tells us what the timber yield would be if all of the other variables had a value of zero (here it says that yield would be

[9] Such problems are common. For example, the UK has a good network of river water monitoring stations some of which are used to provide estimates of faecal pollution. However, the majority of river stretches are not monitored; therefore a scheme to clean up rivers needs to take the information available and model it to estimate the responsiveness of unmonitored areas to clean-up policies. Such modelling is required to predict policy impacts on the physical provision of almost any environmental good.

16.71m^3/ha/year). Of course it is implausible that all the other variables would have a value of zero and their coefficients tell us how yield changes as each one of those variables increases by one unit. The last column tells us how significant each of those variables are in determining yield. These are t-ratios and any value outside the range from +1.96 to -1.96 is considered to be statistically significant (and the further outside that range the more significant a factor is). All of the variables listed here are very highly significant in determining timber yield.

So, for example, the 'Elevation' variable shows us that if the woodland was one unit (here 1m above sea level) higher then yield would decline by 0.009m^3/ha/year. This is a tiny change but this result helps answer more useful questions such as the impact of planting a forest 100m higher above sea level, which is now a non-trivial loss of 0.9m^3/ha/year. Similarly the 'Rainfall' variable shows that as the amount of rain in an area increases by one unit (here 1mm) so timber yield falls by 0.002m^3/ha/year; a substantial effect when one considers that rainfall across Wales varies from 1000mm to 3000 mm annually (Met Office, 2009) thus inducing a variation in yield of 4m^3/ha/year from this factor alone. The remainder of the model is similarly readily interpretable, identifying those factors that either increase or decrease timber yield and in each case quantifying that relationship.

Notice that the model is composed of two types of variables, those linked to the physical environment and those which describe the effects of management on trees. As an example of the latter the Mixed Crop variable shows us that any given species tends to grow at a slightly lower rate when mixed with other species than when grown in a monoculture; illustrating a familiar landscape diversity versus timber output dilemma for the forest planner. We can set these latter management factors in line with any desired policy (e.g. preferring mixed species stands over monocultures). Turning to the physical environment variables, the UK enjoys a wealth of data with information on all these factors being held at a very high degree of resolution for the entire country. We can now use this information to predict timber yield for every location across the country and so estimate the physical provision change which would be induced by an afforestation project at any given location. Results from such an analysis are given in **Figure 2.2**, which indicates the physical provision change of an afforestation project and shows where the highest timber returns would be achieved.

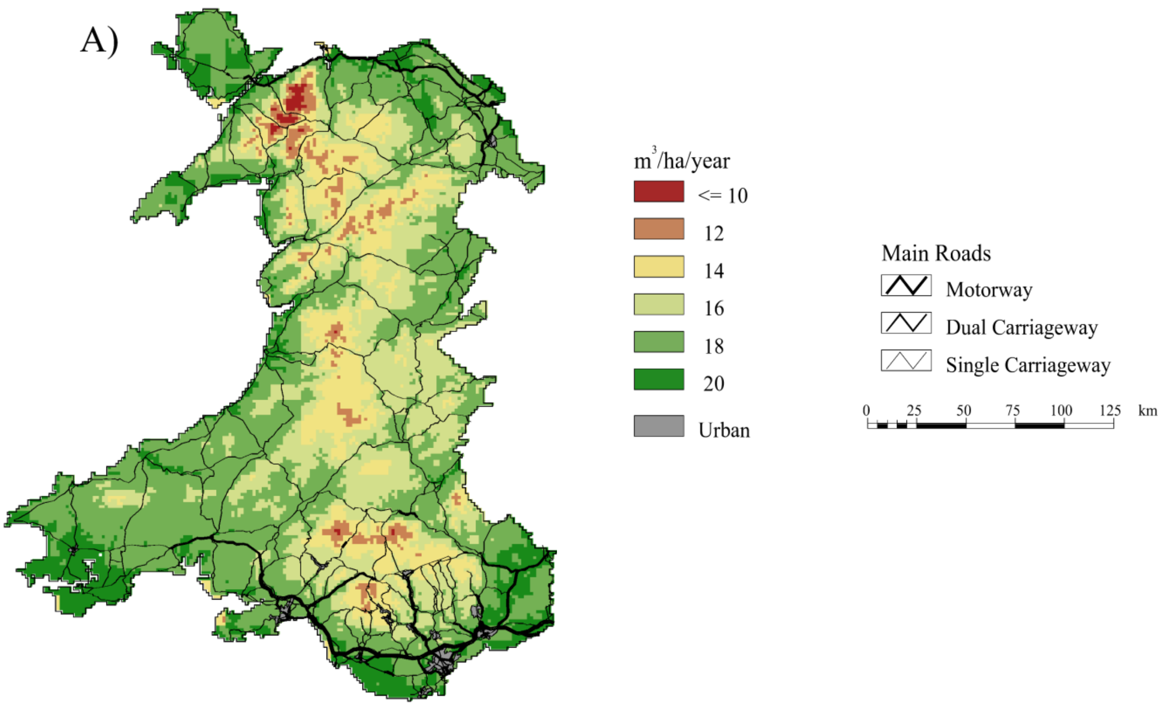
The calculation of the physical effects of a given policy is only the first part of the value transfer task. We now need to convert these physical units into monetary values. To do this we need WTP estimates. When considering a market priced good such as timber these can of course be inferred from market data[10]. However, when undertaking value transfers for non-market goods we often have to resort to a variety of valuation approaches[11] (a variety of methodological reviews are now

[10] Although note that market prices may need to be adjusted to their shadow price equivalents in order to be suitable for use within cost-benefit analyses. Such adjustment takes account of market distortions, subsidies, etc. For a highly accessible introduction to cost-benefit analysis, see Pearce (1986).

[11] An understanding of changes in physical provision is not only vital to the accurate valuation of the total quantity change under consideration, but can also be important in conveying that change to individuals where surveys are being used to estimate values. In such cases the units of physical change in provision of a good may need to be translated into terms which (a) have economic meaning and (b) can be understood by those members of the public expressing values. For example, an ecosystem service such as river water purity is not in itself an economic benefit, but the increase in fish numbers it generates does result in a benefit to anglers. Analysts therefore need to translate from the ecosystem service indicator (e.g. the level of nitrate in a water body) to final economic services (e.g. fish population) and then on to measures which can be readily understood by survey respondents (e.g. hourly expected fish catch rates).

available, see, for example, Bateman et al., 2002a; Champ et al., 2003). This brings us to our next question; ensuring whether the source valuations studies are sufficiently valid, robust and appropriate for policy use.

Figure 2.2: Example of a physical provision change transfer: Predicting the timber yield produced by planting trees in different locations across Wales



2.3 How valid and robust are the source valuations?

Later in this report we discuss at some length the issue of how to validate a value transfer analysis. That procedure is concerned primarily with taking results from existing primary valuation studies and testing whether some transfer procedure can accurately replicate the result that would have been obtained had a primary valuation study been undertaken for some policy relevant issue or at some previously unsurveyed location. This process implicitly assumes that the primary studies themselves are valid. Clearly this is a crucial assumption and constitutes a potential weakness in analysis. This needs to be addressed by undertaking quality assessments regarding the validity of source studies. Therefore the principles which we subsequently discuss at length in respect of value transfer validation also need to be applied to validation of the source studies.

We subsequently discuss three broad categories of validity test as follows:

- *Scope sensitivity:* Whether valuations are responsive to the scale (or 'scope') of the provision change under assessment. For reasons discussed subsequently, scope sensitivity is a rather weak test. Put simply, because of the possibility that individuals may be satiated by some lower level of provision, the only clear expectation is that total WTP should not fall as the scope of a good increases.

- *Tests of theoretically derived expectations:* Economic theory provides a number of prior expectations which can be tested for. For example we might expect that WTP might increase with the individual's income and fall as the availability of substitutes increases. Similarly, we would expect average household WTP to decline as the distance from a spatially confined resource increases as the availability of substitutes increase and the cost of access to the resource increases.

- *Procedural invariance:* Economic theory suggests that WTP should not vary due to "irrelevant factors" such as whether it is elicited using a payment ladder with a lot of values on it or one with smaller number of values. Similarly the choice between two options should not change when a third option, worse than either of the others is introduced[12]. Tests of such 'procedural invariance' can be useful indicators of whether valuation survey respondents hold the well formed preferences characteristic of valid economic values or are simply 'constructing' those preference responses with respect to the ad-hoc heuristics they see in the questionnaire design. There is however a caveat here. While high levels of procedural invariance provide warnings of problems, even purchases of market goods are subject to some procedural influences; indeed this is the premise of effective marketing[13].

While we feel that the above points provide guidelines for assessment of the validity of source valuations, a serious problem is that very few studies undertake or report the exhaustive tests suggested above. Instead the majority of studies at best only report some tests of scope and a few expectations based results (e.g. the impact of income on WTP) with very few reporting procedural invariance tests. Further discussion on the above points can be found in Section 3 below and also in Annex 2 of the Guidelines.

[12] For an example of such tests see Bateman, et al., (2008b).

[13] Standard economic theory labels all such observations as 'anomalies' and would cast doubt upon the validity of such responses for cost-benefit analysis. However, 'behavioural economics' recognises that individuals are swayed by factors such as the decision context, the framing of questions, etc., and that these effects occur in both market and non-market valuations (although there is evidence to suggest that they are more prevalent and extreme amongst non-market valuations; see Horowitz and McConnell, 2002). The challenge is that, once we move away from the somewhat unhelpful simplicity of standard economic theory (where all anomalies are inadmissible; a test which most supermarket purchases would fail!) there is no clear guidance as to what degree of anomaly is acceptable. There is a wider research need to find a validity standard for non-market valuations which is similar to the level of anomaly observed in purchases of marketed goods.

2.4 How do changes in the provision of a good and the presence of substitutes alter its marginal utility?

Assuming that we have successfully undertaken the transfer of physical provision changes in which we are interested (Question (i) in the introduction to this section) and have validated the source studies upon which our valuations are based (Question (ii)), we can now undertake value transfer exercise such as those illustrated in **Figure 2.3**. Here we examine the various value changes which are generated by a proposed change of land use, in this case from agricultural use to multiple purpose woodland. Such changes will generate timber values, carbon storage flux and recreation values (as illustrated in the first three maps of Figure 2.3). After summing these gains we need to subtract the value of agricultural production which would be lost through any conversion to woodland (valued in a manner similar to that discussed for timber above although, as discussed by Bateman et al., 1999a, the farming model is necessarily more complex to allow for the various sectors of the agricultural economy). Given that all of these values are now expressed in the common unit of money we can consider them all simultaneously and identify the net benefits of converting land from agriculture to woodland. This is shown in the final map of Figure 2.3 where the green shading highlights areas where such land use conversion would generate substantial social net benefits.

The analytic methodology underpinning Figure 2.3 offers substantial help to decision making. In particular its use of spatially sensitive information allows the ready identification of optimal areas for the implementation of certain policies, making the most of limited resources. For example, it is clear that woodland conversions generate the greater net benefits in urban fringe areas but would actually constitute a social loss if implemented in some other areas. Significantly, this example highlights an important issue which has to be addressed in all non-market valuation exercise whether primary or via value transfer. This concerns diminishing marginal utility.

The marginal utility of a good describes the benefit which individuals obtain from a unit increase in provision of a good. For goods such as timber and carbon storage the levels of production which could be generated by afforestation of even the entirety of a large area such as Wales would have a negligible impact upon the marginal utility of further production. To prove this consider the volumes of carbon which could be fixed by such an undertaking relative to the many, many times greater magnitude of carbon storage which is required to even begin to offset the problems of present and more particularly future climate change. Analyses such as those by Sedjo, et al. (1997) show that even the full afforestation of the entire USA would be far from sufficient to generate the sequestration required to address global warming. Therefore, the marginal benefit of further sequestration would stay roughly constant before and after any plausible level of afforestation. Hence the task for the decision maker is to obtain a valid estimate for the sequestration of one tonne of carbon and apply this to the estimated change in physical provision of such services. No such simplicity applies to the recreation benefits which such afforestation would generate.

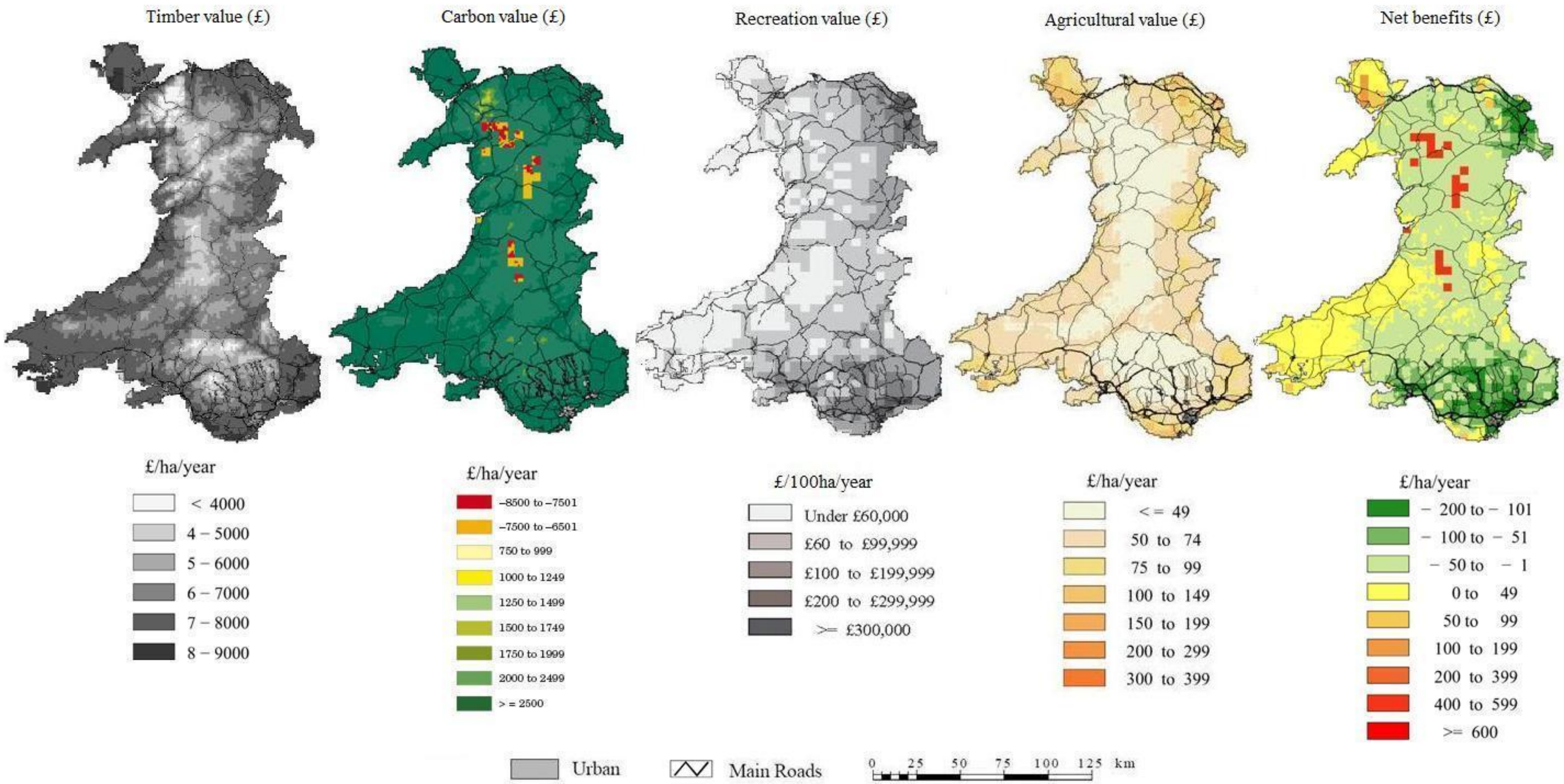
Recreation benefits, like many goods, exhibit significantly diminishing, spatially confined, marginal utility; the benefit of creating one new recreational forest in an area significantly reduces the benefit of creating a second nearby recreational forest. Although this effect is of potentially major significance, there is a paucity of good valuation evidence regarding diminishing marginal utility (and hence diminishing WTP) for most non-market environmental goods. Nevertheless, those studies that have examined this issue find it to be highly significant (e.g. Egan et al., 2009; Lanz et al., 2009). In effect, the greater the provision of a good the more satiated we become with it and the lower the marginal utility (and hence marginal WTP) of that good.

The above phenomena does not just apply to changes in provision of the good under evaluation, marginal utility can also decline because of the presence of substitute goods. For example, while the provision of woodlands can generate benefits by providing recreational walks, these will be at least partially substituted for in areas where there are riverside walks. As Jones et al., (2002) show, this substitution effect can arise from quite diverse resources including man-made recreation sites and urban attractions[14].

Given the importance of diminishing marginal utility and substitution effects we make these a key feature of the empirical case study examples we use to illustrate principles and practice of value transfer, to which we now turn.

[14] Note that complementary relations can also arise. For example forests with lakes may attract additional visitors.

Figure 2.3: Multiple value transfers feeding into an analysis of the net benefits of land use change



Notes: Taken together the maps illustrate the spatially explicit cost-benefit analysis of a potential land use change calculating the value of retaining agriculture as opposed to conversion to multi-purpose woodland. From left to right the individual maps show (i) The annualised equivalent of the social NPV of timber arising from a perpetually replanted woodland; (ii) The annualised equivalent of the social NPV of net carbon flux (live wood, products, waste and soils) under woodland; (iii) Estimated annual value of predicted visits to woodland sites (calculated on a per site basis); (iv) Predicted subsidy adjusted (shadow) value of sheep farming; (v) Cost benefit analysis of retaining sheep farming as opposed to conversion to woodland (defined as timber, carbon storage and recreation). Here negative values (shown in dark to light greens) indicate areas where woodland values exceed agricultural values, while positive values (shaded from yellow to red) indicate the opposite result highlighting areas which should stay in agriculture. Further details are given in Bateman et al., (2003) from which the above figures are adapted.

3 VALUE TRANSFER METHODS

- *The section opens by setting down principles for the design or evaluation of valuation studies for the purposes of value transfer.*
- *Principles for undertaking the major approaches to value transfer are presented; these approaches being (i) unit value transfer and (ii) value function transfer.*
- *Principles for choosing whether to use unit value or value function transfer methods are set out.*

3.1 Introduction

Decision making is an essential yet costly undertaking and resource constraints inevitably mean that that the decision process itself has to pass cost-benefit tests. Analysts have for many years sought methods which will reduce decision costs, and the extrapolation of assessments from one case to another is clearly attractive. Given the significant costs of valuing preferences for non-market goods, it is not surprising that this area has now generated a considerable literature concerning the transfer of value estimates, most particularly in the area of environmental valuation (Brouwer, 2000). Such transfer exercises typically involve estimating the value of a given change in provision of a good at some target 'policy site' from analyses undertaken previously at one or more 'study site'. The most fundamental problem for value transfers is in assessing whether a given transfer is correct or not when the 'true' value of the policy site is a-priori unknown.

The literature – see Annex 1 for an overview - has placed great emphasis upon the development and testing of value transfer methods (e.g., Desvousges et al., 1992; Bergland et al., 1995; Downing and Ozuna, 1996; Brouwer and Spaninks, 1999; Ready et al., 2004; Zandersen et al 2007; Johnston and Duke 2009)[15]. These methods can be broadly categorised into two types (Navrud and Ready, 2007a). The simplest approach is to transfer unit (mean) values from study to policy sites (e.g. Muthke and Holm-Mueller, 2004). Such transfers are frequently used in practical decision making but are crucially dependent upon the pertinence of differences between transfer sites. Clearly at some level all sites are dissimilar (e.g. the unique ecosystem habitats or the spatial pattern of substitutes around a site are unique); it is the degree to which this dissimilarity affects values which will determine the appropriateness of such 'univariate transfers'.

The principal alternative is to use statistical techniques to estimate models (value functions) relating the values given in source studies to the characteristics of the good under assessment. These are then used to predict new values for policy situations. This multivariate 'value function transfer' approach assumes that the underlying utility relationship embodied in the parameters of the estimated model applies not only to individuals at the study sites but also to

[15] For ease of exposition we omit discussion of parallel approaches such as meta-analysis (e.g. Bateman and Jones, 2003; Lindhjem and Navrud, 2008) and Bayesian approaches to modelling value functions (e.g. Moeltner et al., 2007, Leon-Gonzalez and Scarpa, 2008).

those at policy sites. While parameters are kept constant, the values of the explanatory variables to which they apply are allowed to vary in line with the conditions at the policy site. Both methods are discussed in greater detail below and Section 4 of this report presents examples of both approaches in practice. However, a vital precursor to both approaches is that there are sufficient good quality source studies. Recent years have seen the publication of a number of guidance manuals for valuation studies (see, for example, Garrod and Willis, 1999; Haab and McConnell, 2002; Bateman et al., 2002a; Champ et al., 2003; Bockstael and McConnell, 2006; Kanninen, 2006; Navrud and Ready, 2007b; Haab et al., forthcoming) and several of these provide useful guidelines regarding the quality assessment of valuation studies (Bateman et al., 2002a is particularly relevant for the UK context). Potential sources of valuation studies should be carefully scrutinised against these guidelines. However, a common failing is that studies are not designed with subsequent transfers in mind. Therefore we now consider a set of principles for the design of valuation studies to facilitate subsequent value transfers.

3.2 Value transfer design principles for valuation studies

Good practice within individual case studies, while necessary, may not be sufficient within the value transfer context. One of the basic requirements which has been stressed from the early days of value transfer, is the need for a common design format for valuation studies (Desvousges et al., 1992). We can identify a series of principles for designing studies for subsequent value transfer and their subsequent analysis, as follows:

- Study design should be developed from economic theoretic principles and employ a theoretically consistent utility specification;
- The design should permit robust validity testing[16];
- Survey questionnaires must be appropriate and common to all study and foreseeable policy sites with questions being consistently phrased with common response options and levels and common data coding;
- Studies should ideally report information on the location of the good and the location of respondents so that any reduction in values as distance from the site increases (distance decay) can be assessed;
- Substitute availability should either be assessed;
- GIS analyses provide ready quantification of off-site locational issues such as distance from respondent's home to site (proxying use in a readily transferable manner and allowing the estimation of distance decay in values[17]) and the availability of substitutes (again via

[16] This testing should include both transfer error analyses and tests based on theoretically derived expectations. As discussed elsewhere in this report, we consider the common 'scope sensitivity' test (Arrow et al., 1993) to be necessary but insufficient for this purpose.

[17] As discussed in Bateman et al. (2006a), quantification of distance decay is vital to the accurate aggregation of WTP values, avoiding the overestimates which arise when mean values derived from samples which are often collected near to sites are applied to wider populations. Indeed distance decay

distance measures). GIS also facilitates the ready transfer of functions containing such variables to other policy sites;

- Information on the do-nothing and post provision change scenarios must be readily reported;

- The framing of valuation questions and the method through which valuation responses are elicited should be reported[18];

- Where studies produce value estimates for multiple sites, unit value transfer and value function transfer tests should be carried out;

- Where studies produce value estimates for dissimilar sites then value function transfer tests should be carried out;

- Empirical specifications of value functions should be theory driven, being developed from the specified structure of a theoretically consistent utility model rather than including ad-hoc, context specific variables chosen primarily for their contribution to statistical fit at any given site;

- Similarly analysts should avoid the lure of over-reliance upon readily available, context specific, survey derived variables such as those concerning the minutiae of respondent characteristics and instead focus inclusion of those theoretically derived variables (e.g. income, use, substitute availability) which are liable to be common across sites;

- Findings should be provided in full; and

- Data, questionnaires, coding and original analysis should conventionally be made available to encourage reanalysis and transfers.

Once a set of suitable quality source valuation studies is identified, value transfer can proceed. As mentioned, two methods are available for this transfer; unit value and value function transfer. We discuss the implementation of these methods in the following subsections.

functions also allow the analyst to define the 'economic jurisdiction'; that spatial area beyond which the value of improvements falls to zero. Again explicit recognition of this area avoids overestimation of total WTP values.

[18] In theory once a 'correct' framing and elicitation method are identified then there is no need to test for the effects of varying these design features. However, as the literature regarding elicitation options in the CV method illustrates, consensus can be elusive and debate prolonged (e.g. Bateman et al., 1995; Carson and Groves, 2007).

3.3 Undertaking a unit value transfer

Unit value transfers are relatively straightforward being undertaken by pooling data from all relevant primary studies and applying the estimates to the policy situation.

Despite the apparent simplicity of this approach a number of practical decisions need to be made. Where the source studies rely on very different numbers of observations this may be taken into account through a weighting approach. Similarly one should adjust for the number of valuation estimates produced by an individual study. Such considerations are often taken into account when analysts produce meta-analyses of previous studies (see for example, Brouwer et al., 1999; Bateman and Jones, 2003).

The only validation available for such approaches is to ensure that:

- The source valuation studies are of sufficient quality (as discussed previously);

- The good valued in the source studies is identical or highly similar to the policy good under consideration. This includes the nature of the good and its provision change in both quantity and quality terms. For example, in the case of a policy to increase water quality in rivers (the case study discussed in Section 4), ideally the status quo and post-change levels of water quality and the length of river affected at the policy site should match that of the study sites from which valuations are taken; and

- The context is identical or highly similar. Continuing the previous water quality example, the accessibility of rivers, distance to populations, the characteristics of that population, the availability of substitutes and their quality should be constant across source study and policy cases.

Clearly few cases will exactly satisfy the conditions set out above. Analysis of the similarity of source studies and the policy case is a vital part of determining the suitability of unit value transfers and is considered in detail in Section 4. Although statistical tests of the similarity of the above characteristics can be undertaken their power is likely to be weak unless there are a lot of studies available. To some extent then it may be a matter of judgement as to whether the conditions between the policy and site sites are sufficiently similar. Given this, value function transfer approaches have been developed and are reviewed below. Note that some commentators discuss 'adjusted unit value transfers'. Here unit values are modified using information on the relationship of values to the characteristics of the good or valuing population.

A further issue is the nature of unit values themselves. Analysts should realise that valuation data could be skewed with a few respondents holding very high values. This results in mean values (which are the simple average of all the observed valuation responses) often being considerably higher than median values (that value which half of the sample would agree to pay for the good in question). This is no different for transferred or original study results. However the latter are typically qualified with information concerning the nature of the value distribution which allows the decision maker to inspect the stability of values. Transfer analyses (both unit value and function transfers) should seek to preserve this information and provide it to the decision maker.

3.4 Undertaking a value function transfer

Here statistical techniques are used to estimate value functions from study site data. These are then used to estimate values for policy sites. This is achieved by assuming that the underlying utility relationship embodied in the parameters (coefficients) of the estimated model applies not only to individuals at the study sites but also to those at policy sites. While parameters are kept constant, the value of the explanatory variables to which they apply are allowed to vary in line with the policy site.

Put simply suppose that analyses at some study site(s) j lead us to estimate the willingness to pay (WTP) of individual i for an improvement at those sites to be a function of a set of variables X then we can write that:

$$WTP_{ij} = \beta \mathbf{X} \qquad (3.1)$$

Note that X is a matrix of different variables, some being characteristics of the site (X_j, e.g. the length of walks available, etc.), some being characteristics of the individual (X_i, e.g. their household income, etc.) and some being characteristics of the relationship of the individual and the site (X_{ij}, e.g. the distance from the individual's home to the site, etc.). When transferring this value function to policy sites we allow for the fact that the values of the X variables may change (there may be longer walks, the individuals concerned may have different incomes and the distance between their home and the policy site may be different). However, the key assumption of value function transfer is that the relationship between those X variables and the individual's WTP stays the same, i.e. the ß values stay constant. For example, we assume that the impact of changes in income have a constant effect upon WTP. Section 4 of this report provides a worked example of the value function transfer approach.

The value function transfer method affords the practitioner with somewhat more in the way of validation checks than is the case with simple mean transfers. These quality assessment rules include the following:

- Ensure that the source valuation studies are of sufficient quality (as discussed previously);
- Ensure that the factors determining values in the source studies are relevant to the policy case and are incorporated within the transfer function (the case study given in Section 4 of this report discusses the issue of function specification in considerable detail);
- Ensure that the context is comparable. For example, continuing the previous water quality example, the accessibility of rivers, distance to populations, the characteristics of that population, the availability of substitutes and their quality should not be too dissimilar across source study and policy cases;
- As part of the above, ensure that the values of the explanatory variables included in the transfer function (the X_j and X_i from the above model) cover the range likely to occur in the policy case and avoid the problem of extrapolating relationships out of the range of the data upon which they were based; and

- Ensure that the relationships embodied within transfer function emphasise only those factors highlighted by economic theory as likely to be generic to all situations. We discuss these factors and theoretical expectations below.

3.5 Expectations based principles for developing transferable value functions

In specifying value functions for transfer purposes we have to ensure that the predictors used are of generic relevance to both study and policy sites. This rules out the inclusion of context specific, ad-hoc variables. While such predictors may significantly assist in optimising the statistical fit of models to study site data, the danger is that they will be of different (or even no) relevance to policy sites. In effect the assumption that parameter estimates from study sites will hold for policy sites begins to fail. Due to the multiplicative role of coefficients, such failure can result in major transfer errors. Statistically-driven, best-fit approaches to specifying value functions may not be ideal when the purpose of that function is for transferring values. Instead the focus might best be restricted to generic factors. Economic theory provides us with a number of expected relationships which should hold across contexts and it is these factors which should guide the specification of transferable value functions.

So what general relationships does economic theory suggest should hold across contexts? While theory leaves much to the discretion of individuals, basic microeconomic principles identify a number of factors which should and should not influence preferences and WTP. Within the context of a spatially fixed water quality public good considered in our case study, these factors include:

- The extent (or 'scope') of the change in provision under consideration;

- The costs which an individual faces for using the good (for a physically located good this mainly relates to the proximity of the respondent's home and is exhibited as the distance decay effect);

- The availability of substitutes (again a spatial relationship); and

- The individual's income constraints.

Theory also identifies some factors which should not influence WTP. For example, responses should not vary with the procedure used to elicit WTP providing that the questions being asked are objectively identical (i.e. procedural invariance as mentioned above).

These factors provide a rich source of expectations and a set of theoretically consistent variables for inclusion in our theory driven value functions. We consider each of these factors in turn below. Furthermore, their inclusion within our value functions means that we have a number of clear expectations against which we can validate our findings.

Scope

The US NOAA Blue-Ribbon Panel (Arrow et al., 1993) highlighted the responsiveness of WTP responses to changes in provision as the principal form of validity assessment for CV studies. This is based upon the expectation that:

> "[u]sually, though not always, it is reasonable to suppose that more of something regarded as good is better so long as an individual is not satiated. This is in general translated into a willingness to pay somewhat more for more of a good, as judged by the individual" (Arrow et al., 1993; p. 4604).

This 'scope sensitivity' assessment has come to be "regarded by many as an acid test" (Carson et al., 1996, p. 3) of survey-derived values. However, as Banerjee and Murphy (2005) point out, statistically significant sensitivity to scope is of itself an insufficient test of preference consistency. There are very few non-market goods for which we have prior expectations regarding the degree of increase in WTP might be reasonable. Indeed, given that individuals may become satiated with environmental goods (e.g., it might be reasonable for a respondent to think that once they had access to one nearby clean river they were not willing to pay anything for a second), then the only definite expectation that economic theory provides for us is that marginal WTP should not be negative for an increase in provision of a good. We consider a finding of significant scope sensitivity as a necessary but insufficient test of study validity[19].

Distance decay

Arguably some of the clearest expectations arise from the spatially fixed nature of environmental public goods (Zandersen et al., 2007) such as open-access water quality. Theory suggests that usage and net benefits will be related to the travel costs faced by households and indeed WTP values tend to decline markedly as the distance to the site in question increases. This trend in average WTP reflects a fall in the proportion of users to non-users as distance increases (Bateman et al., 2006). As users typically hold higher values than non-users a 'distance decay' (ibid.) in values is to be expected. This can readily be tested for by recording the home address of survey respondents and using GIS or similar software to assess distance and/or travel times. Analysis of empirical distance decay trends in values can then be undertaken.

Substitutes

Such accessibility measures can also be used to test a further clear expectation regarding location; that WTP for an improvement at a given site should decline as the availability of suitable substitutes rises. An operational issue here concerns the definition of substitutes. Reliance upon self-reported substitutes involves some challenging questions for survey

[19] Findings of significant scope sensitivity are more convincing for cross-respondent valuations obtained with split sample contingent valuation studies (where different groups of respondents are presented with different scopes of the good in question) than from within-respondent values (where the same respondent reacts to different scopes of the good). This is because the latter can be criticised as potentially merely reflecting a respondent's desire for internal coherence in their responses (Ariely et al., 2003). This internal coherence argument weakens the importance of scope sensitivity tests for choice modelling studies where respondents choose between multiple options, each defined by varying scope of the goods in question.

respondents and generates variables which are not available to the decision maker wanting to estimate values for unsurveyed sites. Jones et al., (2002) use GIS to calculate distances to multiple potential substitutes. The latter study then allows the data to determine which are the significant substitutes which is an appealing strategy.

Income

While a variety of socio-economic and demographic variables may empirically influence stated values, theoretical expectations emphasise the role of income in terms of the budget constraints it may impose on WTP. We might expect, ceteris paribus, that those with higher incomes will have higher WTP. In a manner analogous to the scope test, this is a fairly weak expectation dependent upon the value of the good in question. Nevertheless, income effects do appear to have some microeconomic foundations and should be included in transferable value functions.

Procedural invariance

As well as the factors that should affect WTP, economic theory also indicates certain issues which should not affect WTP and this provides a last set of design concerns for our transferable value functions. In essence, economic theory posits that, prior to giving a valid WTP response, individuals should have well formed preferences, conforming to standard assumptions and robust against what theory would see as irrelevant issues, such as the way in which a given question was framed. Tests of such procedural invariance are therefore important ways of validating stated preference responses. However, findings from such tests are not uniformly supportive with some results suggesting that individuals may determine their assessments of certain goods not solely by reference to what might be recognised as their economic preferences, but also by inferring information from the manner in which a question is framed (Lichtenstein and Slovic, 1971; Tversky and Kahneman, 1974; Kahneman et al., 1982; Slovic, 1995; Hsee, 1996; Birnbaum, 1999; Doyle et al., 1999; Ariely et al., 2003; Kahneman, 2003; Bateman et al., 1997a,b, 2005a, 2007). The problem with such 'constructed' preferences is that they are highly malleable; changing with the frame within which a question is posed. Such phenomena have excited considerable interest within the valuation literature (Kahneman and Knetsch, 1992; Schkade and Payne, 1994; Bateman et al., 2008a) with commentators arguing that framing effects have to be addressed within the design of studies (DeShazo, 2002; Bateman et al., 2009).

Worked examples of testing both unit value and value function transfers are given in Section 4 of this report. These address all of the issues raised above. However, we now turn to consider principles for guiding the choice between these two value transfer approaches.

3.6 Principles for choosing between unit value and value function transfer approaches

Pearce et al. (1994) argue that, in principle, because value function transfers allow the analyst greater control over differences across sites, they should yield lower transfer errors than simple unit value transfers.

We add two important qualifiers to that argument. First, as discussed previously, any value function which is to be used for transfer purposes should be carefully specified to focus upon factors which are generic across the study and policy situations. Using variables which are of different relevance at the policy site than at the study sites is liable to lead to potentially gross transfer errors. As discussed, we emphasise the use of those variables highlighted by theory as being generic for such undertakings (this may be problematic if reported value functions do not conform to such specifications in which case either new functions need to be estimated from the source data or unit value transfers should be considered). Related to this it should be noted that, because of the multiplicative nature of value functions the potential for error arising from poor quality analyses is generally higher than for simple unit value transfers.

Our second caveat is that value function transfers, even when using well specified functions, may still not outperform simple unit value transfers if conducted for policy changes in provision and contexts which are very similar to those given in source studies. To some extent unit values smooth out the variation which inevitably arises when a sample survey and subsequent statistical analysis is undertaken. Value functions will, in comparative terms, give a greater reflection of the variability of a sample. Estimated coefficients can be influenced by small numbers of individuals in a sample, especially where they have relatively extreme characteristics or unusual values. Since samples from two (or more) sites are likely to contain different proportions of such (relatively) extreme observations (even when those sites and their surrounding populations are relatively similar), when estimated coefficients are used within a value function transfer they may yield considerably greater errors than simple unit value transfers. The value function approach assumes that the relationships reflected in estimated coefficients is constant across sites. Even if this is the underlying truth, the vagaries of any sampling exercise may yield variations in estimated coefficients which result in higher transfer errors when the value function approach is used than when simple unit value transfers are applied.

While the above argument seems likely to apply for similar sites, the opposite seems likely to apply for dissimilar sites. Here the differences between the values of the underlying population may be so gross that simple mean transfers yield substantial errors. This is where the value function approach comes into its own. Accepting that the sampling variation effects described above will still apply, nevertheless the value function method now has the ability to adjust for the gross differences between dissimilar sites; be they physical or socioeconomic/demographic in nature. Simple unit value transfers will be unable to make such adjustments here and in principle are liable to yield larger errors than value function approaches[20].

[20]The issue of similarity in determining the appropriate choice of transfer methodology to some extent makes sense of the conflicting evidence available in the value transfers literature, wherein some studies report higher errors from univariate transfers than value function approaches while others find the opposite result (see, for example, Bergland et al., 1995; Barton, 2002; Chattopadhyay, 2003; Ready et al., 2004; Brouwer and Bateman, 2005).

An immediate question which the above principles generate is how an analyst would determine (a-priori and without survey evidence from policy sites) whether sites where similar or not. Unfortunately the assessment of similarity (and its link to methodology) is relatively under-researched and, like many of the design principles noted previously, source studies often report a relatively haphazard set of statistics from such assessments could be made. While original surveys often provide data characterising certain aspects of the sample this does not always extend to the underlying population and information regarding the physical characteristics of valued goods or sites is rarely systematically presented. One of the most prevalent omissions is data on the geographic location of environmental goods and respondent populations. This is unfortunate as this can be a major driver of estimated and transferred value (see Chapter 6 of this report). Furthermore, any assessment of similarity has to focus upon variables which can be obtained from secondary sources such as those that can be accessed for previously unsurveyed policy goods or sites.

In Section 4 of this report we present an example of a similarity assessment for a case study embracing both similar and dissimilar sites. The analysis examines data regarding the characteristics of both sites and their surrounding populations, examining whether such information is sufficient to determine similarity and hence the appropriate choice of transfer method. Results show that where the policy site is similar to the study sites then simple unit value transfers yield a lower level of error than the function transfer method. However, the inclusion of a dissimilar site reverses this finding, indicating that value function transfers may be more appropriate for heterogeneous cases.

Section 4 also discusses the issue of how to use case studies to develop and test transfer methods. This requires full valuation studies of all the goods or sites under investigation; i.e. strictly speaking all sites are ‘study’ sites. Methodological development then proceeds by omitting information for one of the sites and treating it as a ‘policy’ site by estimating a transfer value for it by using information from the remaining ‘study’ sites. Formal validation of that transfer is then possible by comparing the transferred value for a site with the actual value observed when that site is surveyed. By repeating this process site by site we build up a large volume of information on the validity of the method under investigation. However, as such undertakings are essentially of research rather than decision making concern they are only considered within that case study setting.

4 CASE STUDY 1: CHOOSING BETWEEN UNIT VALUE AND VALUE FUNCTION TRANSFER METHODS IN THE CASE OF THE BENEFITS OF THE EU WATER FRAMEWORK DIRECTIVE

- *The principal objective of this case study is to illustrate how the analyst should choose between implementing unit value or value function transfer methods. The case study shows the vital role of 'similarity':*
  - *When transferring between similar contexts and goods then unit value transfer can yield lower transfer errors than function transfer.*
  - *When transferring between relatively dissimilar contexts the variation within study source and/or policy contexts means that the value function transfer can yield lower errors than unit value transfer.*
- *The case study also shows that when value function transfer is more appropriate, it is important to use a function specified for transfer purposes. The use of models that give the best statistical best fit to the study context may yield higher transfer errors than a model specified to only contain those variables which are likely to be of generic importance to both the study and policy contexts.*
- *The case study also illustrates a topical issue; the value of water quality improvements.*

4.1 Case study objective

The objective of this case study is to use a real world study to examine the choice between unit value and value function transfer. In order to do this we use a study that values all sites of concern, ignore the data on one of those (as if it was the 'policy site'), predict the value of the policy site from the rest of the data. The comparison of this predicted (transferred) value to the value estimated in the study for the 'policy site' shows us the scale of the 'transfer error'. This is a 'test' case study and not an example of what routine value transfer applications are expected to do.

4.2 Developing and implementing the case study design

The design of the study used here followed a set of valuation design principles set out in Bateman et al., (2002a) and Section 3 of this report. Initial concerns for study design were to identify a public good and the relevant locations to provide a rigorous yet policy relevant test of our methodology. Considering the latter locational issue, recall that the underlying objective of value transfer is simple: to take information on the value of provision changes at some surveyed study site(s) and with it estimate values for provision changes at some unsurveyed policy site(s). However, we first need to be sure that the transfer methods employed are valid and reliable. To achieve this requires survey data from at the very least two sites. Transfer then involves using data from, say, site A to predict values at site B. Validation then compares the value of site B as predicted by transfers from site A with the actual value obtained from the survey of site B (with the transfer error being expressed in terms of the percentage difference between the two WTP estimates; see, for example, Bergland et al., 1995). So, while the objective is to develop methods for transferring to unsurveyed sites, methodological development requires data at all sites.

As set out in Section 3 of this report, when considering whether to employ unit value or value function transfers, a central issue is the degree of similarity both within the source 'study' site valuation and between those and the policy site. With this in mind the case study embraces sites from both similar and dissimilar contexts, taking data from studies in five European countries: Belgium, Denmark, Lithuania, Norway and the UK. As detailed in subsequent results, together these include both similar and dissimilar countries capturing some of the economic extremes of Europe and providing a robust test for the methodology.

While the case study deliberately sought variation in the study site contexts, we need to value a common good in all cases. It was felt that this good should be typical of those assessed within non-market valuation studies; one which generates both use and non-use values, of relevance across all case study areas and of policy interest. As can be seen by the number of meta-analyses of surface water (e.g., Johnston et al., 2005, 2006; Moeltner et al., 2007), this is a common target for non-market valuation that has from the earliest of studies reported significant use and non-use values arising from water quality improvements (e.g. Desvousges et al., 1987). Furthermore, this literature supports the common sense notion that open-access water quality is of interest to almost all populations, allowing us to undertake studies in multiple countries and transfer between them. Finally, with the introduction and gradual ongoing implementation of the EU Water Framework Directive (WFD; European Parliament, 2000) this is a topic of great policy interest. The WFD represents a fundamental change in the management of water quality in Europe with a general requirement to improve all European waters to "good ecological status" by 2015. In the five northern European case study areas the main water quality problem is eutrophication[21]. Moreover, there is a common policy need for information to justify time derogations and the setting of less restrictive targets in cases of disproportionate costs as determined through economic assessment of costs and benefits (WATECO, 2004). These issues provide a common ground to the valuation scenario.

A vital early task in any stated preference study is the clear definition of the good concerned, its status quo conditions and the change(s) in provision which we will ask survey respondents to value. This in turn requires an understanding of the physical science determining these states. While there are numerous pollutants that affect open access waters, the WFD focuses upon those which affect their ecological status and in particular those nutrients that are delivered to waterways via routes such as diffuse pollution from agriculture (Davies and Neal, 2004; Hutchins et al., 2006). To some considerable degree the pathways linking pollution to ecological impact is still the subject of ongoing research (UKTAG, 2008). However, this does not prevent the analyst from valuing certain states of the world on the assumption that ongoing research will indicate how such states might subsequently be attained. Furthermore, individuals do not hold values for reducing pollution per se but rather for the effects that such reductions may induce in terms of recreation suitability, ecological quality, etc.

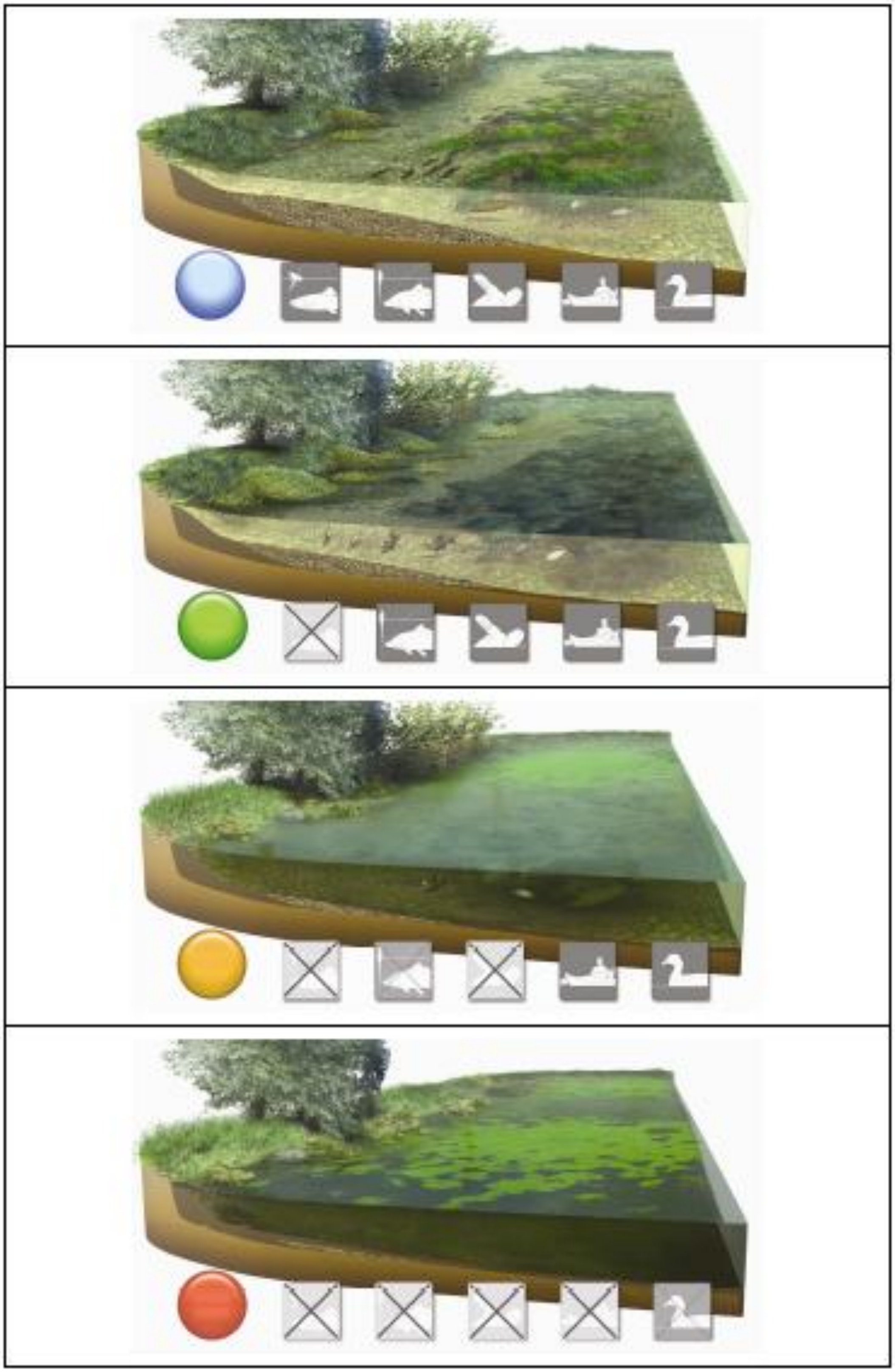
[21] This contrasts with southern Europe where an increasingly serious problem is water scarcity.

Clear and comprehensible information is essential for ensuring understanding of a good and its provision changes within a stated preference survey. The extensive literature on information provision stresses the advantages of visual as opposed to textual or numeric approaches (e.g., Peters et al., 2005a,b; Fagerlin et al., 2005; Zikmund-Fisher et al., 2005; Bateman et al., 2009) and this is reflected within the water quality valuation literature (e.g., Carson and Mitchell, 1993). With this in mind, a novel 'water quality ladder' was developed for the present application (full details being given in Hime et al., 2009 with a summary provided in Annex 2 to this report). This defined four levels of water quality based upon chemical, physical, flora and fauna characteristics as well as use characteristics. Following discussions across the various case study partners, a set of photographs of generic water quality characteristics was agreed for each quality level. These were then passed to a graphic artist to produce the generic water quality ladder shown in **Figure 4.1**. This ties together the ecological and use attributes of water bodies to be applicable to a wide range of lowland slow flowing rivers as well as lakeshores. The simple colour coding scheme shown in the figure allowed clear definition of quality levels in the survey interview. Qualitative face-to-face testing with a pilot sample confirmed that this form of information was clearly comprehended by respondents who were able to recall patterns in quality change following the interview process.

With the nature of the good clarified, the next task was to define the current level of provision and changes in that provision, which together determine scope. For rigorous scope sensitivity testing we require a clear definition of the status quo and at least two changes in provision of the good (a single provision change does not allow us to examine the shape of a value function or assess changes in the marginal value of a resource as its provision alters). These changes in provision need to be defined in terms of both quality and quantity. To enhance the consistency of our design, in each country the case study was applied to a water body whose status quo quality was best described by the yellow level of the water quality ladder while the quality of improved stretches was specified as attaining the blue level[22]. The two changes in provision were then distinguished by defining a waterbody improvement (which we will term the "Large Improvement") and halving this to produce a second scope of change (which we term the "Small Improvement"). The contrast between the values for these two quantities provides an insight into the rate at which marginal WTP diminishes as the scope of improvements increases. Such information is vital to prevent overestimation of values when considering more major improvements than those considered here. Each provision change was presented to respondents in map form with **Figure 4.2** illustrating maps from the UK case study.

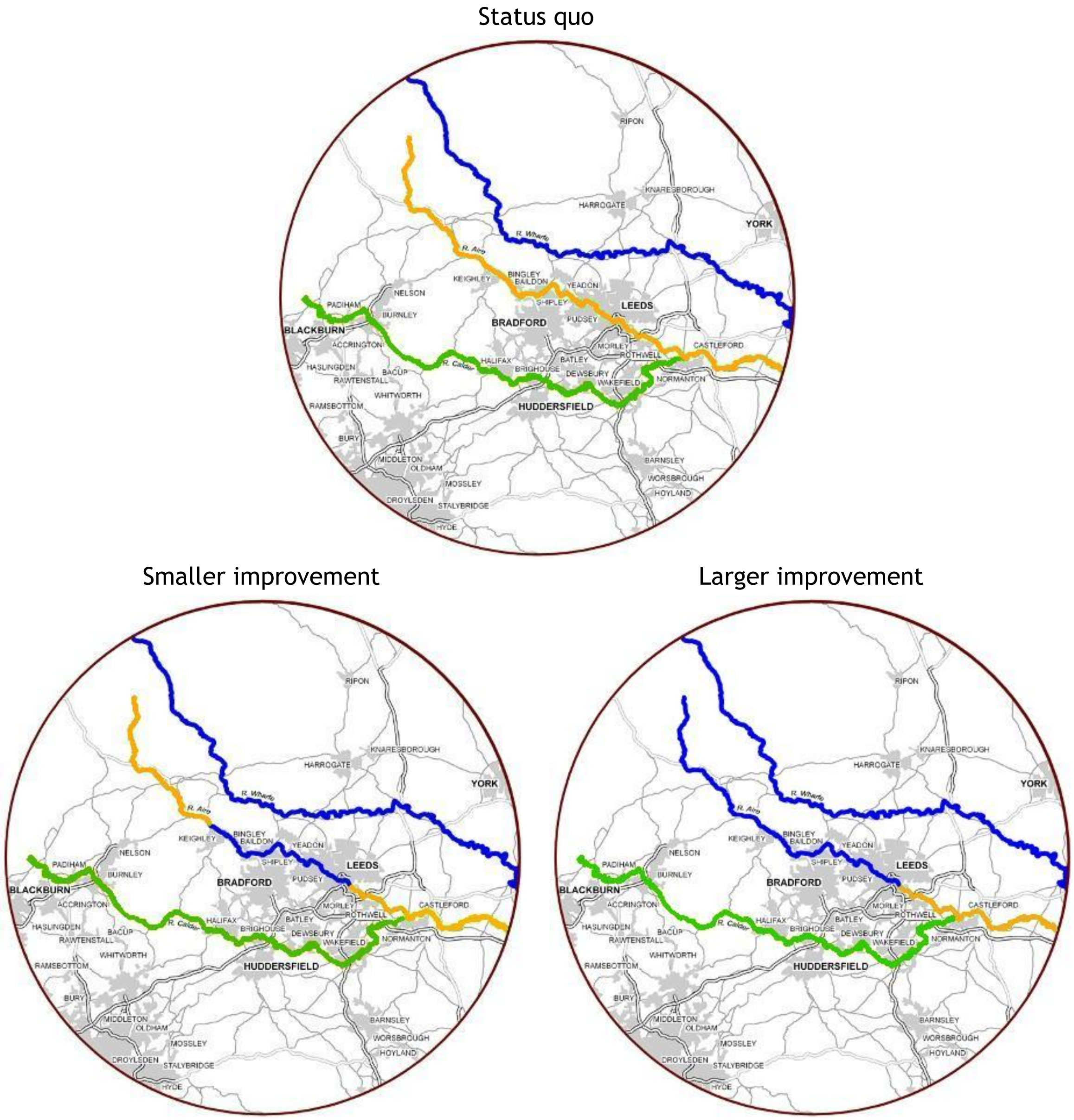
[22] Note that a further treatment in the Norwegian study also examined an improvement from 'red' to 'green' quality. However, for comparability, this dataset was not used within the present analysis.

Figure 4.1: The generic water quality ladder



Source: adapted from Hime et al., (2009): Copyright protected.

Figure 4.2: Maps from the UK case study depicting status quo provision of river water quality and smaller and larger improvements



Source: adapted from Hime et al., (2009)

Scope sensitivity and diminishing marginal WTP testing involves examination of valuation changes between the Small and Large improvements. There are a number of ways in which the data required for such an assessment can be gathered. One simple route is to ask each respondent to state their WTP for both the Small and Large Improvements. Such within-respondent scope tests are fairly weak and susceptible to criticism. Clearly an across-respondent test becomes feasible if a split sample approach is adopted where some individuals face an initial valuation question concerning the Small Improvement while others face the Large Improvement.

A split sample approach also readily facilitates a test of procedural invariance. Using an 'exclusive list' question format (Bateman et al., 2004), immediately after providing their initial WTP answer respondents are asked to imagine that the Small (or Large; depending on question ordering) Improvement they had just considered had not occurred and that they were still at their status quo point. Now a second valuation question is asked eliciting WTP for the Large (or Small) Improvement. By ensuring both valuations are made from the common status quo level we avoid the sequencing problems highlighted by Carson et al. (1998) which result in the non-comparability of values for a given good made from different baselines. Repeating such a procedure across the sample elicits four values: WTP responses to the first and second questions concerning either the Small or Large Improvement (yielding values denoted 'Small Improvement 1st'; 'Large Improvement 1st'; 'Small Improvement 2nd' and 'Large Improvement 2nd'). The procedural invariance expectation is therefore that values for the Small or Large Improvement should not change with the order in which they were elicited.

Given that the valuation literature provides clear evidence regarding the potential for changes in the WTP elicitation method having significant impacts upon responses (e.g. Bateman et al., 1995) a common approach was used in all countries. This consisted of a payment card presented in local currency units but which, when converted into Euros, included the same amounts for all countries. The payment card amounts were chosen after considering the differences in purchasing power parity (PPP) between countries and the impact upon the statistical efficiency of WTP estimates of different payment card levels[23]. The WTP question was prefixed by a standard budget constraint reminder

The common questionnaire also contained uniform questions such as household income and respondent age to be included in the value function. While theory is mute regarding the influence of such variables, they are commonly included in valuation functions and yield some empirical regularities. Arguably such variables, if they are genuine regularities reflecting preference relations which hold universally, could usefully be included within transferable value functions. However, our concern is that once we abandon the parsimonious guidance of economic theory there is no clear demarcation of which effects are likely to be common to all areas. In effect we stray toward the ad-hoc inclusion of context specific variables which our central argument repudiates. Therefore, while in our subsequent empirical analysis we examine the effect of including empirical regularities (providing a falsifiable test of the hypothesis that transferable value functions should be limited to variables for which we have economic expectations), the central thrust of our argument would be to exclude such variables from the specification of value function for transfer purposes.

To ensure that the data contained a sufficient level of variation in terms of the distance to the improvement site and to substitutes an efficient spatial sampling strategy was developed. In essence this strategy considered a regular grid of potential interview locations around the study site, assessing each location in terns of its distance (to site and substitutes), socioeconomic and

[23] A concern with payment cards is that they may be subject to range bias (Covey et al., 2007) where respondents infer that values in the centre of a range are somehow 'correct'. Following the findings of Rowe et al., (1996) we address this by using a payment card with values which ranged from zero to amounts that were clearly implausibly high and therefore not to be construed as having any information value. We also eschew the common habit of using a logarithmic style card with increasingly wide differences between values at the upper end of the range. Again this may be construed as suggesting such values are less plausible. Instead a card using evenly spaced amounts was used. For details see Annex 2 to this report.

population density characteristics. Survey locations were chosen such that a full range of data in each of these dimensions was captured. The home address of each survey respondent was recorded and GIS routines were subsequently employed to calculate individual specific distances and travel times to the improvement and all substitute sites.

Sample sizes were designed to support not only conventional parametric validity testing but also cross sub-sample analyses of the procedural invariance tests. In Belgium, Denmark and Norway, the surveys were conducted online. In Lithuania and the UK surveys were conducted using face-to-face interviews. Response rates ranged from 12% in the online Belgian survey up to 55% in the Lithuanian face-to-face survey. All surveys were undertaken during 2008.

4.3 Results

4.3.1 Descriptive statistics and the (dis)similarity of study sites

A total of 3,589 questionnaires were completed across our five study countries[24]. Response options and data coding were common across studies and monetary variables were PPP adjusted and data pooled into a single analysis. Table 4.1 presents descriptive statistics of each sample together with WTP sums disaggregated by size of provision change and ordering of question presentation.

The wider representativeness of the final sample is satisfactory with most sample descriptors differing by less than 5% from national statistics. The section of **Table 4.1** headed 'Respondent characteristics' shows those variables which could be obtained from secondary sources such as the census (for socioeconomic and demographic variables) and open source GIS data[25] (for the physical characteristic descriptors) to allow assessments of similarity for unsurveyed policy sites (although given the representativeness of our samples we use report their values for convenience). The first two rows of this section consider distance from respondent's home address to the improvement site and their nearest substitute site. Neither of these statistics suggests any clear dissimilarities across study sites with the distance to improvement site in the range households typically travel for recreation and all countries showing that on average respondents have a substitute site closer to them than the improvement site (both factors being patently important determinants of values yet typically ignored in transfer studies). However, the following row shows that there is one major source of dissimilarity between our study countries. Tests confirm that PPP-adjusted household income is substantially lower in Lithuania, being roughly one quarter to one third of the level in the other countries sampled. Our strong, theoretically derived, expectation is that this major dissimilarity would result in a significant difference in stated WTP between Lithuania and other countries. Following our central hypothesis we therefore expect that unit value transfers will generate higher errors than value function transfers when applied across the full set of sites. However, income differences are insignificant across the remaining four countries. Therefore, again following our hypothesis, if we were to omit Lithuania then we would expect that unit value transfers would

[24] The sample size proved sufficient to support not only the conventional parametric validity testing reported here but also cross sub-sample analyses of the procedural invariance tests reported in the Annex to this report.

[25] See Bateman et al. (2002b) for a review of such sources for use within valuation studies.

outperform value function transfers for the four remaining similar countries. Subsequently we formally test both of these hypotheses.

Table 4.1 Descriptive statistics and WTP by country

	Lithuania	Belgium	Denmark	Norway	UK	Total
Sample size						
Number of respondents	500	768	754	1133	434	3589
Respondent characteristics						
Mean distance to the improved site (km)	20	21	30	22	10	22
Mean distance to nearest substitute site (km)	1	3	24	7	5	9
Mean annual pre-tax household income tax; € PPP (s.d. in brackets)	9531 (7823)	40877 (19002)	34854 (17708)	24884 (11452)	26686 (16709)	28310 (17730)
Mean Age	48	45	50	45	50	47
Urban (% urban)	63%	45%	79%	41%	78%	58%
Gender (% women)	49%	36%	44%	48%	46%	45%
WTP values in € PPP (standard deviation in brackets)						
Protest bids (% of country sample)	8%	5%	2%	12%	2%	7%
Average WTP- Small Improvement	6 (23)	47 (66)	25 (38)	42 (82)	19 (29)	31 (61)
Average WTP- Large Improvement	8 (38)	48 (70)	36 (52)	47 (86)	26 (35)	37 (66)
Average WTP- Small Improvement 1st	6 (15)	50 (67)	29 (42)	45 (88)	22 (32)	34 (64)
Average WTP- Large Improvement 1st	10 (52)	49 (70)	31 (41)	45 (78)	25 (32)	36 (64)
Average WTP- Small Improvement 2nd	6 (29)	43 (66)	21 (33)	38 (76)	16 (25)	28 (57)
Average WTP- Large Improvement 2nd	7 (15)	47 (69)	40 (61)	48 (93)	26 (38)	37 (69)

Notes: Income and WTP recalculated based on Purchasing Power Parity indices (World Bank, 2008). Protest bids are excluded in the estimation of WTP descriptive statistics.

The remaining rows of this section of Table 4.1 detail various other sample characteristics which, although not highlighted by theory as determinants of WTP, have been used by analysts seeking ad-hoc variables to improve the statistical fit of study site value functions. None of these variables suggests any further major dissimilarities across countries. We incorporate such factors within the subsequent test of our hypothesis that models containing such ad-hoc variables, while providing a better fit to study site data, may yield higher transfer errors than functions specified solely from theoretically derived, generic predictors.

The final section of Table 4.1 overviews our WTP valuation results. Here the first row details protest rates identified using the guidelines in Bateman et al. (2002a). These are consistently within the bounds of acceptability suggested by the literature (Mitchell and Carson, 1989; Champ et al., 2003) and pure protestors were excluded from further analyses[26].

Following guidelines for international value transfers (Navrud and Ready, 2007b), WTP responses (and income data) were corrected for differences in purchasing power between countries using indices from the World Development Indicators (World Bank, 2008) and then converted into 2008 Euros. Even a cursory inspection of these results strongly suggests that our classification of Lithuania as dissimilar to the other countries is reflected in WTP values. For both the Small and Large improvements Lithuanian WTP is less than one-third that of the lowest mean given in any of the other countries. This proportion directly echoes the difference in incomes noted above. Recall that in a real world value transfer we would not a-priori have values for the policy sites. What is clear here is that the clear dissimilarities flagged up by variables which can be obtained from secondary sources (the 'respondent characteristics' discussed previously) do seem to provide relevant indicators of when simple unit value transfers can or cannot be relied upon. Subsequently we formally test these inferences.

A further point to note in Table 4.1 is clear evidence of diminishing marginal WTP. Recall that the Large Improvement provides double the length of highest quality river than the Small Improvement. It does appear that in general the former is accorded a higher value but it is clearly not double the latter. This is not of itself an anomalous result as it is perfectly feasible that respondents may have a rapidly diminishing marginal WTP for additional improvements once an initial length of high quality river has been provided. However, a final point to note in Table 4.1 is some evidence of a failure of procedural invariance in the form of an ordering effect in valuation responses. While there is no particular ordering pattern within the Large Improvement values, four countries yield Small Improvement 1st values which are higher than their respective Small Improvement 2nd WTP with the remaining country giving the same value for both. This finding echoes that of Bateman et al., (2004) who find that the value of small goods is elevated when elicited first in an ordering. Again formal parametric tests of these trends are given below (with nonparametric confirmation of these findings detailed in the Annex to this report).

[26] We retain the 4% of WTP responses that clash with prior expectations in that the smaller improvement is accorded a higher value than the larger good. While this is an issue to be highlighted (and is often not tested for in non-market valuation studies) the rate of apparent irrationality or misunderstanding of the scenario is consistent with findings in experimental economic tests. While some studies have omitted data from such respondents we argue that this may give a misleading indication of the consistency and validity of findings and so retain all responses within subsequent analyses. While some of these responses may reflect respondent's lack of comprehension of the different schemes, it is also reasonable to assume that some reflect a personal rationality that larger schemes may be less likely to proceed; a perception which has been linked with lower WTP (Powe and Bateman, 2004).

4.3.2 Specification and estimation of value functions

In order to compare simple unit value transfers with value function transfers we first need to estimate the latter functions. To test our assertions regarding the importance of correct specification of transferable value functions we develop a 'theory-driven' model from economic principles. This includes all those variables for which economic theory holds expectations and which should be generic to all sites: the change in provision; the costs of using the good (its distance from the valuing individual); the availability of substitutes (distance to substitutes) and budget constraints (the household income). Because income is the main dimension of dissimilarity between sites we also estimate a further model excluding this factor so that comparison with the theory-driven model can illuminate the impact of income in this analysis. We then contrast the theory-driven model by a 'statistically-driven' model which supplements the former with ad-hoc variables regarding which theory has no generic expectation but which empirical regularities observed in the literature suggest should improve the fit to the data at survey sites. In the next subsection we then transfer the various functions, calculate transfer errors and contrast these with those arising from simple unit value transfers.

Value functions were estimated by pooling data across the five countries. Both parametric and non-parametric analyses were conducted. As both identify common patterns in the data we focus upon the more readily interpretable parametric and report non-parametric results in the Annex to this report.

As our data contain both non-zero and valid zero bids, we have a WTP distribution which is censored. Given this, we specify a panel Tobit regression model which allows both this censoring and the fact that each respondent provides us with two WTP answers. The structural equation for such a random effects Tobit model is

$$y_{it}^* = X_{ikt}\beta_k + \mu_i + \varepsilon_{it} \tag{4.1}$$

where y^* is a latent variable observed for values greater than zero and censored otherwise, k indexes the number of independent variables included in the model, i is the individual index and t indexes our repeated responses. The random disturbances can be combined to form the composite error term of the model written as $w_{it} = \mu_i + \varepsilon_{it}$ which is assumed to be normally distributed. The latent variable y_{it}^* represents respondent i's unobserved willingness to pay to improve water quality at choice t, whereas the observable censored dependent variable y_{it} assumes value y_{it}^* when $y_{it}^* > 0$ and zero when $y_{it}^* < 0$.

The panel specification of equation (4.1) captures both inter- and intra-respondent variation in WTP as well as incorporating the effect of observable and unobservable variables. In the random effects model the unobservable or un-measurable factors that differentiate respondents are assumed to be characterized as randomly distributed variables. Observable variables are incorporated in the usual way. Therefore, the random effects model can be thought of as a regression model with a random constant term. We employ simulated maximum likelihood estimation procedures to obtained unbiased, consistent and efficient estimates of the parameters β_k[27]. **Table 4.2** reports the resulting models[28].

The p-values of the models in Table 4.2 show that they are globally highly significant. The parameter estimates on all variables for which we have prior expectations are statistically significant (at $\alpha = 5\%$ with most significant at $\alpha=1\%$) and conform to those theory derived priors. The move from the model excluding income to the full theory-driven model has relatively little impact on the variable parameters although it does change the intercept and, as we discuss subsequently, significantly reduces transfer errors.

Briefly reviewing the relationships reported in Table 4.2, both scope sensitivity to changes in the quantity of improvement provided and any ordering effect are inspected by assigning the Small Improvement 2nd responses as our base case WTP values. As can be seen the values accorded to the Large Improvement are very clearly larger than this base case suggesting clear scope sensitivity. Both parametric and non-parametric ordering tests[29] confirmed there was no significant difference in this effect between the Large Improvement 1st and Large Improvement 2nd values and so these have been pooled. However, values for the Small Improvement 1st, although below those for the Large Improvement, are significantly above those for the Small Improvement 2nd base case. This confirms the presence of the procedural invariance suspected from our inspection of Table 4.1. Very few value transfer studies conduct such tests and this finding suggests that such analyses may be worthwhile[30]. Given this result, in our subsequent function transfers we not only allow for the scope sensitivity difference between the Small and Large improvement but allow for ordering effects by conduct separate analyses for both the Small Improvement 1st and Small Improvement 2nd values.

[27] This model was estimated using the STATA 10 package. An alternative Heckman model within an initial hurdle identifying non-protest from protest respondents was also estimated but failed to yield significant improvements over the model reported here. Further, in order to test the stability of value transfer results we run a weighted panel Tobit model that takes into account the difference in sub-sample sizes. Results were relatively similar to those presented here which are preferred for their parsimony and ease of interpretation.

[28] Results for further specifications are reported in the Annex to this report.

[29] Non-parametric findings are reported the Annex to this report.

[30] It is worthwhile briefly considering why this may have arisen. One possibility is that this reflects a partial failure of the 'exclusive list' format resulting in a perceived change in the incentive compatibility of questions between the first and second response. It may also reflect the arguments of Carson and Groves (2007) regarding incentive properties of repeated valuation questions.

Table 4.2: Results from different models specification using random effects Tobit panel data model

	All theory driven variables except income	*Full theory-driven model*	*Statistically-driven model (including ad-hoc empirical regularities)*
Variable	Coefficient (s.e.)	Coefficient (s.e.)	Coefficient (s.e.)
Constant	12.58 (3.995)	-1.95 (5.076)	5.45 (7.934)
Large improvement	9.07 (1.055)	9.32 (1.073)	9.30 (1.073)
Small improvement 1st	5.47 (1.381)	5.15 (1.404)	5.09 (1.404)
Distance to the improvement site (km)	-0.19 (0.073)	-0.22 (0.074)	-0.19 (0.074)
Distance to nearest substitute site (km)	0.16 (0.037)	0.14 (0.037)	0.14 (0.038)
Income (net household income in € per year)	-	0.0008 (0.0002)	0.0008 (0.0002)
Age of respondent (in years)	-	-	-0.32 (0.099)
Urban (respondent lives in urban area=1; otherwise =0)	-	-	9.93 (3.232)
Norway	13.37 (4.294)	15.24 (4.359)	17.02 (4.518)
UK	-14.88 (5.228)	-17.36 (5.254)	-16.47 (5.254)
Belgium	27.85 (4.447)	22.76 (4.773)	24.87 (4.938)
Lithuania	-64.63 (5.053)	-50.43 (5.537)	-49.86 (5.569)
Sigma μ	73.47	75.20	72.79
Sigma ε	23.39	23.22	23.21
Rho	0.91	0.91	0.91
No. of observations	5790	5474	5466
Number of censored observations	1593	1457	1455
Log-Likelihood	-23186	-22142	-22098
Wald chi^2(*K*=restriction for overall significance)	573 (8)	525 (9)	552 (11)
p-value	0.000	0.000	0.000

Continuing with our inspection of Table 4.2, we find significant distance decay: as the distance between the respondent's home and the improvement site increases so WTP decays. Expectations are also borne out with respect to the substitution effect with WTP for the improvement site increasing as the distance to the nearest substitute rises. Income also has the expected positive impact on WTP.

The country dummy for Denmark is omitted making this the baseline from which country departures are to be interpreted. The good/order dummy Small Improvement 2nd is omitted making this the baseline from which the Large Improvement 1st, Large Improvement 2nd and Small Improvement 1st departures are to be interpreted. Rho = var_μ/(var_μ+var_ε) and represents the percent contribution to total variance of the panel-level variance component.

The statistically-driven model extends the former analysis by adding in two ad-hoc variables for which economic theory has no prior expectations but which have been incorporated within previous analyses. Both the respondent's age and whether they live in an urban area are found to be significant predictors of WTP and result in an increase in the degree to which this model fits the data. From a statistical perspective these are stronger models of the study site data, however our contention is that their inclusion of such ad-hoc variables may increase transfer error relative to the full theory-driven model which only contains generic variables.

The *sigmas* represent the variances of the two error terms $\mu_{i\,i}$ and ε_{it}. Their relationship is described by the variable *rho*, which informs us about the relevance of the panel data nature. If this variable is zero, the panel-level variance component is irrelevant, but as can be seen from the results in Table 4.2, the panel data structure of the WTP answers has to be taken into account and is retained for all subsequent value function transfers to which we now turn.

4.3.3 Value transfer and error analyses

We conduct both simple unit value transfers and value function transfers for both our full dataset[31], including the dissimilar country (Lithuania) and excluding this to focus solely upon the remaining more similar countries. Value function transfers are undertaken using both theory-driven and statically-driving models. Together this allows us to test all of our various hypotheses regarding the appropriate methodology for value transfers in different contexts.

Unit value transfers are relatively straightforward being undertaken by pooling data from all countries except that which we are transferring to; the former being our 'study' sites and the latter the 'policy' site. Transfer errors are then calculated as an absolute percentage by comparing the unit value from the study sites with the actual mean of the policy site. We repeat this for each scope/ordering combination.

[31] We also conducted a number of preliminary transfers of non-pooled, individual country, values for each scope and ordering of the good. Non-parametric tests of the Kristofersson and Navrud (2005) equivalence hypothesis (whether we can take the mean WTP value of a randomly chosen country and validly transfer that to any other country in the sample) are generally rejected at a 20% error rate (ibid.). Similarly, simple mean transfers (taking one country as the policy site and another as the study site) are generally rejected. The only exceptions are the values of the large and small improvement in Denmark and Norway and the large improvement value in Denmark and Belgium. These results do not change if we account for the question ordering. Further details are given in the Annex to this report.

Value function transfers again compare the observed value of the 'policy' site[32] with that predicted from the other 'study' sites. However, now this prediction is obtained for each country in turn by estimating a value function (such as those shown in Table 4.2) on data from the study sites (i.e. omitting data from the policy site), then applying the coefficients[33] to the values of the predictors at the policy site to yield a function transfer value for that policy site[34]. Defining this value as $\hat{WTP}_{k|s}$ for policy site *k* estimated from study sites *s* and the directly observed policy site mean value as $\overline{WTP}_k$ then we calculate the value function transfer error as $\left(\frac{\hat{WTP}_{k|s} - \overline{WTP}_k}{\overline{WTP}_k}\right)\%$.

Table 4.3 presents the transfer errors obtained from both mean and value function transfers when we consider our full dataset including the dissimilar (Lithuanian) site. The upper section of this table details results for the unit value transfers, disaggregated into values for the Large and Small improvement with the latter further disaggregated to allow for the significant framing (ordering) effect observed in these values. We can see that this mix of similar and dissimilar sites yields high levels of overall error (shown in the final column) when the unit value approach is applied with an overall raw error rate of 116%. This is likely to be unacceptably high for policy purposes and so function transfers seem worthy of investigation. However, as an aside, even a cursory inspection of the unit value transfer results for individual countries clearly bears out our expectation that it is the dissimilar Lithuanian site which is the principle cause of these high error rates. Subsequently we investigate the impact of restricting our analysis to just the similar countries.

[32] Some analysts compare value function estimates from study sites with the unit value from the policy site (e.g. Van den Berg et al. 2001, Brouwer and Bateman 2005). However, others compare the former value with that predicted by a function estimated from the policy site data (e.g. Barton 2002; Chattopadhyay 2003). In the present study we tested both approaches but found no difference in the pattern of results provided. Consequently we report the unit value comparisons here (directly comparable with the simple univariate transfers) and present the comparisons with predicted value in the Annex to this report.

[33] Given the Tobit specification, the parameters to be used for function transfer must be adjusted for censoring. Discussions of this adjustment can be found in Halstead et al. (1991), Haab and McConnell (2003) and Brouwer and Bateman (2005).

[34] An interesting issue is whether one should apply the estimated coefficients to the unit value of the policy site predictors or to the values that apply for each individual. Normally the former approach is the only one available and we follow that method here. However, all of the predictors in our theory-driven model can be assessed at the individual level using secondary source data (from GIS and census datasets). This raises the possibility that one could synthesise a dataset for an unsurveyed policy site consisting of the change in provision, distance to the improvement site, distance to substitutes and household income. One could then apply coefficients from a value function transfer at this individual response level thus obviating the need to rely upon mean predictor values in the transfer process. We suspect that this may further reduce value function transfer errors.

Table 4.3: Transfer errors (%) from unit value and function transfer methods: Dataset including the dissimilar (Lithuanian) site						
WTP measure[1]	Lithuania	Belgium	Denmark	Norway	UK	Average errors (weighted)[2]
UNIT VALUE TRANSFERS						
Small improvement 1st	508	49	6	41	48	130 (102)
Small improvement 2nd	392	53	23	43	69	116 (94)
Large improvement	391	39	10	37	34	102 (81)
Average error (weighted)	430 (420)	47 (45)	13 (12)	40 (40)	50 (46)	**116** (90)
VALUE FUNCTION TRANSFERS						
Reduced theory-driven model (Provision change, distance to site and distance to substitute)						
Small improvement 1st	69	100	98	95	89	90(84)
Small improvement 2nd	125	106	116	101	104	111(101)
Large improvement	48	94	96	91	84	82 (78)
Average error (weighted)	80 (72)	100 (98)	103 (102)	96 (94)	92 (90)	**94** (88)
Full theory-driven model (Provision change, distance to site, distance to substitute and income)						
Small improvement 1st	30	74	65	80	43	58(62)
Small improvement 2nd	81	76	71	83	40	70(71)
Large improvement	13	66	69	75	43	53 (58)
Average error (weighted)	41 (34)	72 (71)	68 (68)	79 (78)	42 (42)	**61** (64)
Statistically-driven model (Provision change, distance to site, distance to substitute, income, age and urban)						
Small improvement 1st	35	81	93	101	72	77 (81)
Small improvement 2nd	78	84	103	107	78	90 (92)
Large improvement	12	74	86	94	64	66(72)
Average error (weighted)	41(34)	80 (78)	94(92)	101(99)	71(70)	**77** (82)

Notes: Transfer errors calculated by comparison with the mean WTP values estimated at each site.
1. Small Improvement 1st = WTP for the smaller improvement elicited as the first valuation question asked; Small Improvement 2nd = WTP for the smaller improvement elicited as the second valuation question asked (significantly lower than Small improvement 1st); Large improvement 1st or 2nd= WTP for the larger improvement elicited as either the first or second valuation question asked (no significant difference between these responses).
2. Raw averages given outside parentheses. Figures in parentheses are average errors weighted by relative sample size

The remainder of Table 4.3 details results for our various function transfer analyses. This starts with the 'reduced theory-driven' model containing all of those variables suggested by economic theory except for the major source of dissimilarity between countries; income. As can be seen, even this model generates a substantial reduction in overall error relative to the unit value approach. This error reduction is further improved when we add in the income variable to specify our full theory-driven model. Given that this is the major source of dissimilarity across sites it is not surprising that this variable generates a larger improvement than any of the others in this model which nearly halves the rate of overall error generated by the unit value approach. We now test our hypothesis that the theory driven model, although not providing quite such a good fit to study site data as the statistically driven model (recall the results Table 4.2) will nonetheless provide a lower rate of transfer error than the latter. To test this we now calculate transfer errors for the statistically driven model. These are more than one quarter higher than the transfer errors associated with the theory driven model. The only change here is the addition of the ad-hoc variables, not prescribed by theory as being generic components of utility functions. This, we contend, provides strong support for the methodological principles proposed at the start of this report.

The results presented in Table 4.3 support the hypothesis that, when faced with a heterogeneous set of sites, analysts should prefer function transfer over unit value transfer (and within the former should restrict the specification of models to those variables regarding which economic theory has clear expectations). However, the individual country results show that unit value transfers can yield both high rates of error when transfers from similar countries are applied to a dissimilar country (Lithuania) but lower rates when (mainly) similar countries are used to predict for other similar countries. To investigate the potential for error reduction here we now exclude the dissimilar Lithuanian case study and repeat the previous analyses for the remaining similar countries, results being reported in **Table 4.4**.

Comparison of Tables 4.3 and 4.4 clearly shows that the exclusion of the dissimilar Lithuanian sites dramatically reduces the rate of error associated with unit value transfers such that they now fall well below those achieved by value function transfers. Including the dissimilar data reverses this relationship. This supports our central hypothesis that the choice of method depends crucially upon the degree of similarity of the sites under consideration; an issue which we now discuss further in our concluding remarks.

Table 4.4: Transfer errors (%) from unit value and function transfer methods: Dataset excluding the dissimilar (Lithuanian) site

WTP measure[1]	Belgium	Denmark	Norway	UK	Average errors (weighted)[2]
UNIT VALUE TRANSFERS					
Small improvement 1st	30	45	22	91	47 (40)
Small improvement 2nd	35	67	24	119	61 (51)
Large improvement	19	19	19	69	32 (27)
Average error (weighted)	28 (26)	44 (38)	22 (21)	93 (87)	**47** (40)
VALUE FUNCTION TRANSFERS					
Reduced theory-driven model (Provision change, distance to site and distance to substitute)					
Small improvement 1st	92	85	89	82	87 (88)
Small improvement 2nd	97	101	94	96	97 (97)
Large improvement	84	86	83	77	83 (83)
Average error (weighted)	91(90)	91(89)	89(87)	85(83)	**89** (89)
Full theory-driven model (Provision change, distance to site, distance to substitute and income)					
Small improvement 1st	79	65	81	53	69 (72)
Small improvement 2nd	82	72	84	54	73 (76)
Large improvement	70	68	75	50	66 (68)
Average error (weighted)	77(75)	69 (69)	80 (79)	52 (52)	**69** (72)
Statistically-driven model (Provision change, distance to site, distance to substitute, income, age and urban)					
Small improvement 1st	74	67	92	47	70 (75)
Small improvement 2nd	76	67	97	43	71 (76)
Large improvement	65	63	84	41	63 (68)
Average error (weighted)	72 (70)	66(65)	91(90)	44(43)	**68** (73)

Notes: Transfer errors calculated by comparison with the mean WTP values estimated at each site.
1. Small improvement 1st = WTP for the smaller improvement elicited as the first valuation question asked; Small improvement 2nd = WTP for the smaller improvement elicited as the second valuation question asked (significantly lower than Small improvement 1st); Large improvement 1st or 2nd= WTP for the larger improvement elicited as either the first or second valuation question asked (no significant difference between these responses).
2. Raw averages given outside parentheses. Figures in parentheses are average errors weighted by relative sample size.

4.4 Conclusions

The central objective of this study was to develop principles for choosing between the different methodologies available for conducting value transfers. Our analysis shows that the crucial issue concerns the degree of heterogeneity between the various sites across which transfers are to be undertaken. Results show that the pertinent dimensions of similarity or difference can be assessed using data (which is readily available from secondary sources) regarding the characteristics of those sites and their surrounding populations. Using such we find that when analysis is restricted to only include similar sites transfer errors are minimised when unit value transfer is used. However, when we also included dissimilar sites, value function (using generic variables) is to be preferred. The functions can better address the higher heterogeneity of such a group of sites and provided lower errors. These errors were minimised when we specified transfer functions to only include those generic variables which economic theory expects to be present in typical utility functions. Specifying value functions to include ad-hoc variables for which theory has no expectations resulted in an improvement in the statistical fit of those functions to study site data but led to higher error when those functions were transferred to predict value at other sites.

5 CASE STUDY 2: TRANSFERRING A VALUE FUNCTION WITHIN A UK REGION USING GIS

- *The principal objective of this case study is to illustrate how a value function can be used to estimate values for a given policy.*
- *The case study also illustrates the use of geographical information systems (GIS) as a means of mapping the spatial distribution of benefits arising from a project.*
- *GIS can be used to highlight the effects of varying the implementation of the policy, changing its extent or targeting it to a particular area.*
- *Such flexibility allows the analyst to see how, for a given fixed cost, the benefits of a scheme can alter according to its implementation.*

5.1 Incorporating spatial factors into value function transfers

In this second example we use the same study design as applied in Case Study 1 (in Section 4), but now we use a value function to transfer estimates of the benefits of improving water quality across a relatively small region of the UK. This is another ‘test’ case study and not an example of what routine value transfer applications are expected to do.

Ideally one would develop a transferable value function from first principles, starting with a structural model of utility. Ferrini et al., (2008) undertake such development for a spatially dispersed set of goods (again rivers) showing how both use and non-use values can be isolated from such a model. They argue that the value of a given improvement can be calculated by examining the utility generated by the improved site divided by the utility from all other sites and an ‘outside good’ of all other recreation options. Although Ferrini et al. report a simulation of their model it is both complex and highly non-linear. A rough approximation to that model[35] would lead to an empirical specification along the lines of equation (5.1):

$$WTP_{ij} = \left(\frac{\beta_1 \Delta Qual_j * \beta_2 \Delta Quant_j * \beta_3 LnD_{ij}}{\sum_{k=0}^{K} \beta_4 Qual_k * \beta_5 Quant_k * \beta_6 LnD_{ik}} \right) * \beta_7 S_i \tag{5.1}$$

Where:

WTP_{ij}	=the willingness to pay of individual i for a specified improvement in water quality at site j
$\Delta Qual_j$	= a measure of the nature of the water quality change at the improvement site j (e.g. from ‘yellow’ to ‘blue’ quality)

[35] As Ferrini et al. argue utility is actually a function of the site j utility pre- and post the improvement and the utility of all other options pre-and post the improvement.

$\Delta Quant_j$	= a measure of the quantity of water quality change at the improvement site *j* (e.g. X km of improved river stretch)
D_{ij}	= the distance (km) from individual *i*'s home to the nearest part of the improved site
$Qual_k$	= a measure of the water quality at substitute site *k*
$Quant_k$	= a measure of the quantity of that water quality at substitute site *k*
D_{ik}	= the distance (km) from individual *i*'s home to the nearest part of the substitute site
S_i	= a matrix of socioeconomic and demographic characteristics of the individual (e.g. income, etc.).

An intuitive explanation of equation (5.1) is that the numerator within the bracket details reasons why a respondent may pay for improvements at a given site if it was the only one available. This numerator captures the combined factors of the change in quality and the length of river affect, modified by the proximity of that river to the respondent. The denominator of the bracket literally divides the former effects by the influence of the *k* available substitute sites. Therefore the more substitute sites, the higher their quality and length of high quality and the closer they are so the greater the influence of this denominator in reducing WTP for the improvement site. The S_i matrix which multiplies the overall bracket indicates that, whatever the situation regarding the improvement site and its substitutes, WTP will also be modified by a respondent's socioeconomic circumstances (e.g. their ability to pay for improvements).

We generate values for the explanatory variables in equation (5.1) by considering a real world area: that around the River Aire in Yorkshire as illustrated in the various maps of Figure 4.2 above. This Figure specifies values for $\Delta Qual_j$ and $\Delta Quant_j$ for the baseline situation and for two alternative scenarios; an improvement from 'yellow' to 'blue' quality (as defined in the water quality ladder shown in Figure 4.1) for a shorter or longer stretch of the river. The quality of all other rivers stays constant throughout (thus defining $\Delta Qual_k$ and $\Delta Quant_k$). A regular grid (based on the Ordnance Survey grid) was used to define nearly 5000 possible outset locations (home addresses) across the area. A GIS was then used to calculate travel distances from each outset location to each access point both on the Aire and the potential substitute Rivers of the Wharfe and Calder. This gives values for the D_{ij} and D_{ik} variables in equation (5.1). Finally, for illustrative simplicity a single socioeconomic variable is considered, household income, with values for each grid square being taken from ONS Census ward data (thus defining S_i).

Value transfer exercises would normally obtain values for the various beta coefficients shown in equation (5.1) through reviews of the literature. However, there are very few studies which have explicitly incorporated issues to do with the spatial location of sites into their analyses and, to our knowledge, none which have considered the spatial configuration of substitutes[36]. This is an issue which policy makers will have to tackle if they are to be able to conduct robust value transfers as such a dearth of high quality, raw valuation results is arguably the biggest challenge to value transfer. Yet this situation can change rapidly; the promulgation of water valuation studies engendered by the Price Review 2009 process being a prime example.

[36] The ChREAM project (Bateman et al., 2006c) has recently completed a survey of over 2000 households in the study area considered here. This study includes all of the locational attributes of improvement and substitute sites. Results from this analysis will be available during 2009.

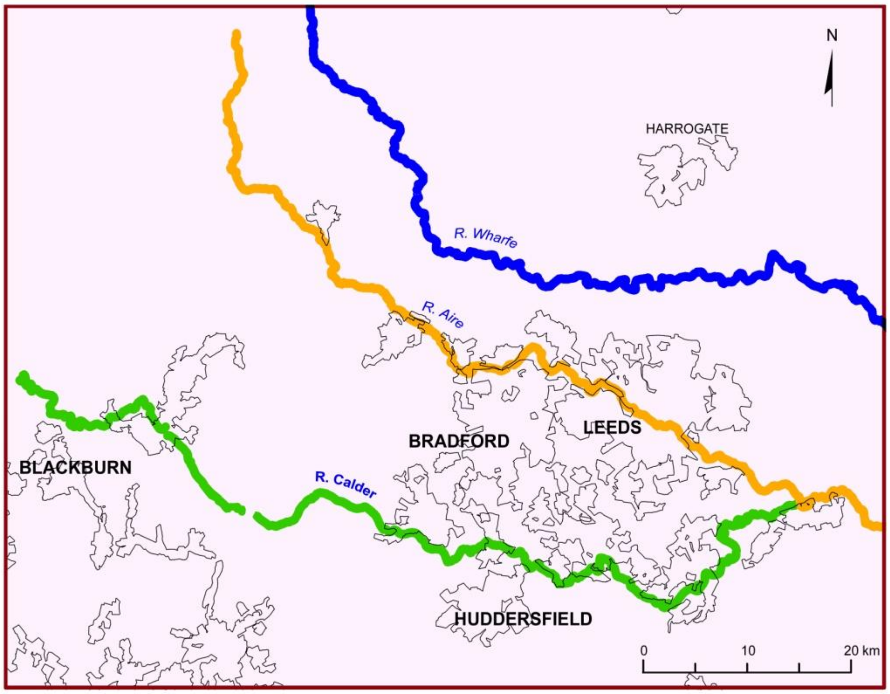
Therefore for illustrative purposes we adopt the approach of Ferrini et al., and use plausible beta values which are in line with the following, theory driven, expectations:

$\beta_1 > 0$	WTP rises as the quality improvement increases;
$1 > \beta_2 > 0$	WTP increases but at a declining rate as the quantity of improved river increases;
$\beta_3 < 0$	the further the improved site for the individual's home the lower will be their WTP for that improvement;
$\beta_4 < 0$	the higher the quality of substitutes the lower the WTP for improvement at site j;
$0 > \beta_5 > -1$	the greater the quantity of high quality substitute river stretches so the lower the WTP for the improved site;
$\beta_6 > 0$	the more distant the substitutes the higher the WTP for the improvement site, and
$\beta_7 > 0$	this positive sign applies for factors which increase WTP such as higher incomes. The sign will change for those factors which constrain WTP.

Applying such beta value to the explanatory variables defined previously we can now generate WTP values for each grid square across our study area.

To illustrate the potential for using such functions within a value transfer context, **Figure 5.1** illustrates a hypothetical status quo situation for the case study area. Here the River Aire (the central of the three rivers shown) is described as the yellow level of the water quality ladder, while the River Calder (to the south of the Aire) is designated as green quality and the River Wharfe (to the north) is described as the best (blue) quality.

Figure 5.1: Baseline water quality

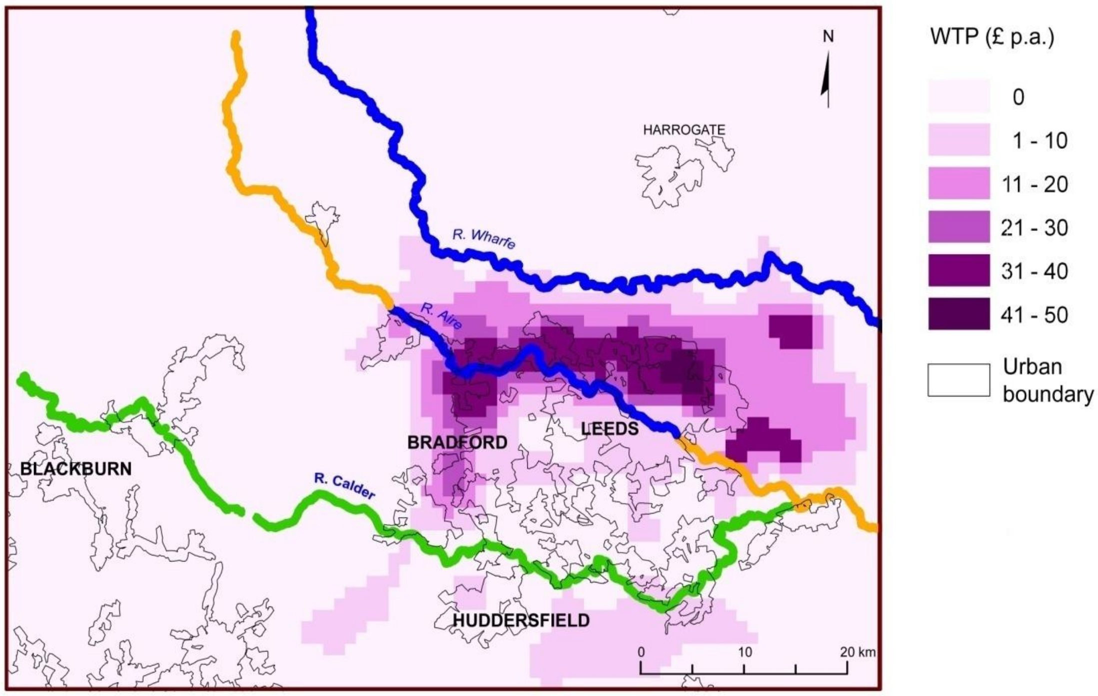


Source: Hime et al. (2009)

Suppose that a policy was being considered to improve a given stretch of the River Aire as it passes through central Bradford and Leeds (as shown by the blue section of the River Aire in Figure 5.2). The policy target provides us with the change in quality and quantity variables in equation (5.1). We can now use a GIS to calculate the distance variable for each grid square across the entire surface of the map shown in Figure 5.1 (in fact we need not be limited by this map or any arbitrary decision boundary; Section 6 of this report discusses the notion of an 'economic jurisdiction' for assessment being defined as the distance at which WTP values become negligible). Furthermore, a GIS can also bring in grid referenced socioeconomic data (from sources such as the UK Census) to quantify the S_i matrix of equation (5.1). Therefore in each gird square we have all the date needed to calculate equation 5.1 and assess the WTP for that square.

For illustrative purposes we have chosen plausible values for the β coefficients within equation (5.1) and calculated expected WTP per household per annum in each grid square for the proposed improvement in water quality. **Figure 5.2** illustrates these values. The pattern of WTP amounts reflects the nature of the provision change, its location (reflecting distance decay), substitute availability and ability to pay. As can be seen, the provision change has generated significant positive WTP but there is a clear distance decay effect with values falling as distance from the improvement increases. There is also a substitution effect as values fall rapidly near to the high quality River Wharfe and are zero for virtually all areas north of that river. In contrast the substitution effect of the lower quality River Calder is much weaker and positive values are recorded even some way south of this river. Finally there is the clear effect of socioeconomic drivers such as income with WTP being substantially higher in the wealthier areas to the north of Leeds and Bradford than in the poorer inner city areas of both cities[37].

Figure 5.2: Map of estimated mean willingness to pay (per household, per annum), for an improvement to the central Bradford and Leeds stretch of the River Aire

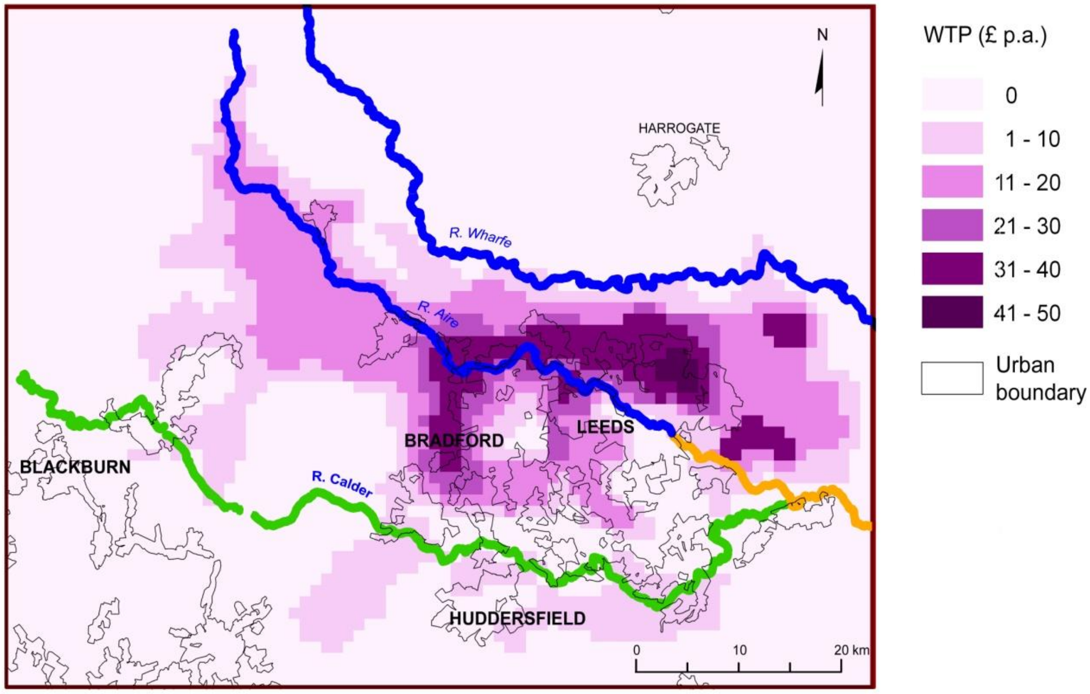


Source: Hime et al. (2009)

The methodology described above is highly flexible. For example, the decision maker can examine the values that would be generated by a more substantial improvement to the River Aire such as that illustrated by the blue section of the river in **Figure 5.3**. Again patterns reflect the factors captured in our value function at equation (5.1), the main difference from Figure 5.2 being the longer area of non-zero values now generated along the upper reaches of the Aire. Interestingly, the relatively lower incomes of this area mean that WTP per household in this additional area is not as high as some of the high values near to the initial stretch. This pattern would be strongly amplified if we then calculated the total WTP in each area by multiplying the WTP per household by the number of households in each grid square. This is an important final step and one which emphasises the higher values generated by locating improvements near to population centres. Such an additional analysis is illustrated in the case study presented in Section 6 of this report.

[37] As discussed in Bateman (2009), we could add distributional weights to these results to compensate for the income inequality driving this latter pattern.

Figure 5.3: Map of estimated mean willingness to pay (per household, per annum), for a larger (two stretch) improvement in water quality on the River Aire



Source: Hime et al., (2009)

5.2 Conclusions

Figures 5.2 and 5.3 illustrate the potential for value function transfers to spatially target improvements so as to make the most efficient use of limited resources. They also allow the decision maker to consider the redistributional aspects of environmental improvements (although it should be highlighted that simultaneous maximisation of environmental and distributional goals would only occur by chance and there is always a trade-off when considering multiple goals). For example, the above figures show that in this case *direct (non-consumptive) use value* benefits would be higher from creating improvements in high density urban areas than in remote rural areas[38]. To really assess the magnitude of this difference however we need to move from the per-household measures shown in the above figures, to aggregate values which allow for the varying distribution of population across areas. Our final case study (Section 6) addresses this issue and illustrates the vital need to consider spatial issues within aggregation procedures.

Note that the above case study examines WTP for the use value attributes of the good in question. Of course many goods, especially those non-market goods endowed by the environment, also generate non-use values (Pearce and Turner, 1990). As we note elsewhere (Bateman et al., 2006a) non-use values can have different location-related properties than use values and one should not presume that they will necessarily positively correlate with the location of use values.

[38] It should be recognised however that an account for indirect use value (e.g. reduced water treatment costs for domestic and commercial uses) and non-use value would also be required to determine the overall efficient allocation of resources in this example.

6 CASE STUDY 3: USING VALUE FUNCTION TRANSFERS TO CALCULATE THE AGGREGATE VALUE OF ENVIRONMENTAL IMPROVEMENTS

- *The principal objective of this case study is to extend the previous theme of applying spatially explicit value functions to estimate the distribution of the benefits across different areas.*
- *The case study focuses on the specific issue of how to calculate aggregate WTP using this approach, showing how this avoids the potential for major errors which may arise when applying more conventional aggregation methods.*
- *The example also introduces the concept of the administrative jurisdiction; that area which is designated by external factors (e.g. administrative boundaries) for decision making. This is contrasted with the economic jurisdiction, the area within which individual values for the good in question are non-trivial. The consequences of using either approach are investigated.*

6.1 Background

Cost-benefit analysis requires the aggregation of individuals' benefits in order to compare with the total costs of a project or policy. Because the methods for measuring non-market benefit values are based on analyses of individual behaviour, there is a problem in knowing how changes in a resource will affect aggregate values. This will depend on both the benefits per person and the population of beneficiaries. It is thus necessary to identify the extent of the market (or the affected population) and how this varies with changes in the good. The extent of the market may well be more important in determining aggregate values than any changes related to the precision of the estimates of per-person values.

In considering the extent of the market it is useful to distinguish between the *administrative jurisdiction*, concerning some administrative area, and the *economic jurisdiction* incorporating all those who hold non-zero economic values regarding a project.

The highest profile UK application of an administrative jurisdiction approach to aggregation occurred in 1998 when the Environment Agency for England and Wales (EA) refused a water company application to abstract water from the River Kennet in southern England (ENDS, 1998; Moran, 1999). The EA approach followed recommendations from FWR (1996) to use water company operations boundaries as the relevant area for aggregation. Total benefit value was then calculated by multiplying the population within this jurisdiction by a sample mean WTP estimate transferred from a contingent valuation (CV) study of the River Darrent (Willis and Garrod, 1995). Aside from concerns regarding differences in the river types, commentators criticised the use of a sample mean taken from one area and applied without adjustment to an entirely different area which not only had differing geographic size but also contained a much larger population with very different socioeconomic characteristics (Moran, 1999; Bateman et al., 2000b). These academic reservations were reflected in judicial review through which the

EA ruling was overturned after an appeal in which the plaintiff attacked both the valuation and aggregation procedure employed.

In principle the use of a sample mean WTP value need not necessarily lead to biased estimates of aggregate values. If a representative[39] sample is drawn from the entire economic jurisdiction (or indeed some larger sample area) then multiplying the sample mean WTP by the population of the sampled area should give an accurate estimate of aggregate values in that sampled area. Indeed even if the sample area is a subset of the economic jurisdiction then the sample mean approach to aggregation still gives unbiased estimates of total value within that subset area. However, problems may well arise where a unit value from some subset area is used as the basis for estimating aggregate values for the entire economic jurisdiction (or some non-coincident administrative jurisdiction, as in the River Kennet case). There are two sources of error here. First, as noted above, underlying values for changes to some spatially confined resource are likely to decay with increasing distance from that resource. Failure to account for this distance decay will lead to error if this unit value is used to aggregate values for a larger (or indeed smaller) area.

This problem is exacerbated by common practice where valuation studies indeed do fail to sample entire economic jurisdictions, but instead focus survey effort upon areas around the resource in question where values are highest. The application of consequent sample means in an unadjusted manner to some large (typically administrative) jurisdiction seems liable to generate over-estimates of aggregate values. A second problem arises when, as seems likely, the probability of responding to a valuation survey is positively related to underlying values. In such cases self-selection bias (Heckman, 1987) is likely to increase with distance from the resource. Thus, while underlying values should exhibit distance decay, sampled values will understate this as with relatively higher values self-select themselves into the sample at a greater rate than those with lower values. Aggregation procedures need to recognise and address both of these problems. A worked example of how to control for sample self-selection bias is provided by Bateman et al. (2006d).

Clearly given unlimited survey resources one would ideally sample from the entire feasible economic jurisdiction. But even here we face the problem that, a-priori, the extent of this area is unknown. In this report we argue that, given these challenges, a superior approach to aggregation is to use survey data to adjust for self selection bias, capture underlying distance decay and define the economic jurisdiction. This, we argue, is best achieved through the identification and estimation of a *spatially sensitive valuation function*. Such an approach explicitly addresses self-selection and incorporates distance decay relationships into defining the limits of the economic jurisdiction while allowing for variability in the socioeconomic characteristics of the encompassed population within the aggregation process. We use this approach to generate estimates of aggregate WTP and contrast this with measures based upon both the use of an administrative jurisdiction and the reliance upon sample means. The impact of using different aggregation approaches is then contrasted with the variability induced by uncertainty in the estimate of mean WTP.

[39] The concept of representativeness is worthy of comment. Typically this refers to issues such as the representation of population socioeconomic and demographic characteristics within a sample (Morrison, 2000). We do not discount the importance of such issues, but argue that the issue of spatial representativeness is also important, particularly when we are aggregating values for spatially confined resources.

6.2 Expectations and existing evidence regarding distance decay in underlying values

We can begin by considering the spatial distribution of values for some open access, public goods resource at its present level of quality. For the sake of simplicity we can identify two types of individual; users of a resource, and non-users. Users will hold use and option-use values and may well hold non-use values. Non-users may hold option and non-use values or have negligible WTP[40]. It seems reasonable to assume therefore that, *ceteris paribus*, users will typically hold higher WTP values than non-users.

Should we expect that the values held by those who are presently users of a resource will decline as distance from that resource increases? There are a number of issues here. It is clearly true that, as distance and travel time increase so the costs of accessing a site also rise. Also it may well be that the availability of substitutes increases with distance, lowering their opportunity cost. But these are cost side issues and do not dictate a reduction in benefits. The benefits held by someone who travels halfway round the globe to visit the Grand Canyon are likely to be higher than the *average* resident of nearby Phoenix[41,42]. Therefore, it is not *a-priori* clear that the benefits of those who are already users will necessarily exhibit strong distance decay (although rising costs mean that their *net* benefits do decay with increasing distance).

Turning to consider those who are non-users of a resource, again it is not clear that values should decay across space. After adjusting for income etc., why should non-use values for preserving the Amazon rainforest be higher in the USA than in more distant Europe? Empirically there is some (limited) evidence that there may be a cultural identity or 'ownership' dimension to non-use values with, those who live closer (or even in the same country) as a resource expressing relatively higher non-use values (Hanley et al., 2003; Bateman *et al.,* 2005c). However, the theoretical basis for this association is unclear.

What is evident from the above is that, as distance from a resource increases and opportunity costs rise, so both the number of users and the proportion of users to non-users will decline. Given that users will generally hold higher values than non-users these trends will result in distance decay in mean per household values. This will be compounded by any empirical reduction in non-user values across space.

All of the above holds for situations in which there is no potential change in resource quality. However, valuation studies invariably estimate welfare measures for situations in which there

[40] The possibility of negative WTP also arises although, as we argue subsequently, this does not seem credible for the resources considered in our case studies.

[41] This example does highlight a practical problem facing almost all aggregation exercises. As most benefit-cost studies extend at most to national borders, values beyond that extent are typically ignored. This problem occurs not only in unit value aggregations but also in the application of valuation functions as these omitted observations have no opportunity to influence the shape of the fitted curve. The extent of the consequent error is likely to directly depend upon the international prominence of the resource in question, with higher prominence resources (such as the Grand Canyon) suffering higher error than lower prominence good (such as the resource considered in our second case study). However, such cross-border values are typically ignored in national decisions.

[42] Now this lack of relation need not always hold. For example, Parsons' (1991) commentary upon the travel cost method notes that certain of those individuals with high use values for recreational resources will internalise some costs by choosing to live in locations near to those resources. However, this will be less of an issue for resources which have been degraded for long periods of time (as in the case of the River Tame discussed subsequently).

is some potential change in resource provision or quality. Two measures dominate the stated preference literature in this respect; equivalent loss and compensating surplus. Equivalent loss studies ask survey respondents to state their WTP to preserve some resource such that the future level of resource quality stays the same as its present quality. In such a situation, while we would still expect to observe distance decay in values for the overall sample (because users are clustered nearer to the site)[43], there seems no theoretical reason to expect such a trend to be seen within just the responses of those who are present non-users. If they were non-users prior to the project then, as quality has not changed, they should remain so after the project has completed[44]. This is not the case for compensating surplus studies where respondents are asked to state their WTP for some improvement in the resource such that the future level of resource quality exceeds the present quality. Here some present non-users may see themselves as likely to become users of the improved good. This trend will become more significant the larger is the proposed improvement in resource quality[45]. Importantly, this non-user to user conversion is more likely to occur for households nearer to the resource than those further away. Hence we would therefore expect to observe a distance decay effect not only across the overall sample of respondents, but also amongst those who are presently non-users.

We therefore have clear expectations of a difference in the nature of distance decay relationships according to the welfare measure under investigation. These expectations are borne out through inspection of those (relatively few) prior studies which have tested for distance decay in WTP responses. Equivalent loss studies examining values for the *preservation* of water quality (Sutherland and Walsh, 1985), conservation areas (Imber et al., 1991), endangered species (Loomis, 2000) and remote mountain lakes (Bateman et al., 2005c) reflect expectations of significant distance decay in overall WTP values but not in present non-users' values. Conversely, but again in line with expectations, compensating surplus studies valuing *increases* in wetland or bird life (Pate and Loomis, 1997) or *improvements* in forests (Mouranaka, 2004), river flows (Hanley et al., 2003) or river water quality (Bateman et al., 2006b) demonstrate distance decay not only in overall sample values but also in the values stated by present non-users.

Summarising the above discussions: we have argued that in a substantial number of cases the reliance upon sample means will fail to yield accurate measures of aggregate WTP. As a more robust and generally applicable alternative we propose an approach based upon the estimation of a spatially sensitive valuation function which takes into account issues of self-selection, the expected distance decay of values and the impact of variation in the socioeconomic circumstances of the relevant aggregation population. With reference to existing literature we have also argued that we should expect that the distance decay characteristics of any such valuation function will depend in part upon the choice of welfare measure, with compensating surplus but not equivalent loss measures exhibiting distance decay amongst the values of present non-users although both measures should give distance decay across the overall sample. Furthermore, aggregate values are likely to vary considerably according to whether

[43] Note that the WTP of those that were already users will not change due to such a project.

[44] While this seems to be the economic-theoretic expectation, the very act of being surveyed might convert some prior non-users into higher value future users if it raises awareness of the resource. Another possibility that a 'focussing illusion' (Schkade and Kahneman, 1998; Loewenstein and Frederick, 1997; Loewenstein and Schkade, 1999; Kahneman and Sugden, 2005) may lead to non-users overstating WTP amounts although whether this would have a distance decay dimension is unclear.

[45] And, of course, this improvement in site quality may also increase user WTP although it is not obvious that there should be a spatial dimension to this particular increase.

administrative or economic jurisdictions are used; a variability which has the potential to outstrip that caused by uncertainty in estimates of mean WTP measures. In the following case study we attempt to address each of the issues raised above. In so doing we develop a method for generating spatially sensitive value functions by harnessing the spatial analytic power of GIS.

6.3 Aggregating a compensating surplus measure: WTP to improve urban river water quality

This case study extends the GIS-based valuation function approach to aggregation developed in the prior analysis. Data is taken from a face-to-face contingent valuation survey of WTP for water quality improvements to the River Tame, i.e. a compensating surplus measure, for which a sample of 675 responses was gathered in interviews conducted at the respondents' home address. The river rises on the west of the Birmingham, flows eastward through the city and turns north till it meets with other elements of the Trent catchment. Its urban course means that it has suffered long term degradation and is now classed as one of Britain's most polluted rivers with fish stocks virtually non-existent and plant growth, insects, bird and animal life all severely limited. However, the river does have ecological and recreational potential and passes through high density residential areas, playing fields and a country park[46].

In order to assess the impact upon aggregate values of changes in the quality of a resource the study design presented respondents with three nested levels of river quality improvement (referred to here as the *Small*, *Medium* and *Large* improvements). The survey instrument described these changes in terms of consequent improvements to water quality, fishing, plants and wildlife, boating and swimming. Respondents were told that these three improvements were alternatives to each other, of which only one would be implemented, and therefore should be evaluated relative to a common baseline of the current situation, which was also described[47]. An open-ended elicitation format was applied and the multiple valuation design of the study meant that each respondent provided three values, one for each of the improvements. In order to allow for possible intra-respondent correlation of these values, responses were subsequently modelled using a generalised least squares random effects approach (Greene, 1990).

The GIS was used to calculate distances from each respondent's home address to the River Tame[48]. This was combined with survey data detailing individuals' knowledge of water quality in the River Tame, visitation patterns, attitudinal data, etc. However, two survey deficiencies should be acknowledged here. First, household income data were not consistently collected and had to be subsequently proxied from small area census data (a less than ideal situation but, we feel, an adequate response given our use of census data at the smallest 'output area'

[46] Full details of the case study area are given in Bateman et al., (2006b)

[47] Tests for ordering effects showed that these were insignificant, which we argue is a consequence of respondents' prior awareness of all schemes before valuations were elicited (Bateman et al., 2004).

[48] Bateman et al., (2005b) reports analyses for three alternative distance measures; a simple measure taken by hand from a paper map (to simulate analyses where a GIS is not available); distance to major access points along the river; and the Euclidian distance measure reported here. All three provide similar results (verified through Hausman tests) although the GIS measures were found to be somewhat superior to the simple paper map approach.

level)[49]. Second, the addresses of those who refused to be interviewed were not recorded[50]. Therefore, we cannot adequately incorporate self-selection bias into the present study and recognise that consequent aggregation estimates in both the sample mean and valuation function approaches are upwardly biased. While the relatively small economic jurisdiction of this good (discussed subsequently) reduces the absolute size of any error relative to our previous study, this is a deficiency which should be addressed in future applications. However, as this error affects both of our aggregation approaches it does (inadvertently) allow us to control for this factor when comparing the other treatments considered in this study.

Analysis indicated that the semi-logarithmic form of equation (6.1) provided a best fit specification of the spatially sensitive valuation function:

$$\ln(WTP_i + 1) = \beta_0 - \beta_1 Distance_i + \sum_{k=1}^{K} \beta_k x_{ki} + \varepsilon_i \qquad (6.1)$$

where *i* denotes individuals and:

Ln(WTP+1)	= Natural log of (willingness to pay + 1) (continuous);
Large	= Large environmental improvement (1 = yes; 0 = otherwise);
Medium	= Medium environmental improvement (1 = yes; 0 = otherwise);
Distance	= Euclidian distance (in metres) from respondent's home to the River Tame;
Env	= Respondent's interest in environmental issues (1 = not interested at all; 5 = very interested);
Know	= Did respondent already know the River Tame (1 = yes; 0 = otherwise);
Retire	= Is respondent retired (1 = yes; 0 = otherwise);
Age	= Respondent age (continuous), and
Poor	= Proxy low income variable (percentage of people aged 16-74 unemployed in the respondents' local census output area).

Note that the dependent variable is calculated by first adding a value of one to the stated WTP in order not to lose any zero bidders in the subsequent log transformation process[51]. This specification also means that the semi-logarithmic approach is not incompatible with identifying the limits of the economic jurisdiction (where fully untransformed WTP is predicted to fall to zero)[52]. **Table 6.1** estimates this model, reporting results for both the full sample of respondents and the sub-sample of present non-users.

[49] While point data are more informative than census tract or other polygon location data, Case (1991) and Swinton (2002) discuss conditions under which one can proxy for the other. The former paper also discusses spatial components of the error term, an issue which is not considered here. This may have reduced the efficiency of our estimates.

[50] This was contrary to instructions to the survey company. In the event only a non-response tally was collected, showing that some 761 households refused to be interviewed.

[51] The use of Tobit analysis is only applicable in those cases where the latent variable can, in principle, take negative values and the observed zero values are a consequence of censoring and non-observability. Actually, in line with Randall et al. (2001), we do not believe that it is reasonable to suppose that individuals may lose welfare from the projects being valued in our research. Therefore, while for sensitivity purposes we report maximum likelihood random effects Tobit specifications in Bateman et al., (2005b), we believe these downwardly bias predictions of WTP.

[52] The geostatistics and GIS literatures also provide an alternative approach in the form of Kriging for semivariograms with a nugget and a sill to identify areas beyond which there is no spatial effect (see, for example, Isaaks and Srivastava, 1989; Deutsch and Journel, 1992).

Considering the model for all responses we find, as expected, highly significant distance decay in values. The two scope dummies, *Large* and *Medium* act as intercept shifters upon this distance decay of values and indicate not only the expected higher values for larger improvements, but also (given the semi-logarithmic form of this function) a somewhat more rapid decline for the latter[53] although values for larger schemes always remain above those which they nest. Other covariate effects are as expected with values being positively correlated with environmental concern and prior knowledge of the study area and negatively related to the constraints of retirement. The expected negative correlation with higher deprivation was also observed.

We now use our estimated value function (based on responses from the whole sample) to calculate the level of WTP for different areas progressing away from the study site. This is achieved by holding the non-distance individual level variables (*Env*, *Know*, *Retire* and *Age*) at their mean values, using the GIS to integrate with output area census data to generate values for the socioeconomic deprivation proxy (*Poor*), and for each level of improvement (*Large*, *Medium* and the base case of a *Small* improvement), calculating the predicted household WTP for each output area. Given the significant distance decay relationship estimated in this model it is clear that at some point values will decline to zero. This defines the economic jurisdiction within which there are positive WTP values. The maximum limits of this area for each level of improvement are given in the last three rows of Table 6.1. As would be expected we see that the limit for the *Medium* improvement is wider than that for the *Small* improvement, while limits for the *Large* improvement are greater than either of the nested schemes. However, it is interesting to note that the limit for even the largest scheme, at less than 30 km.

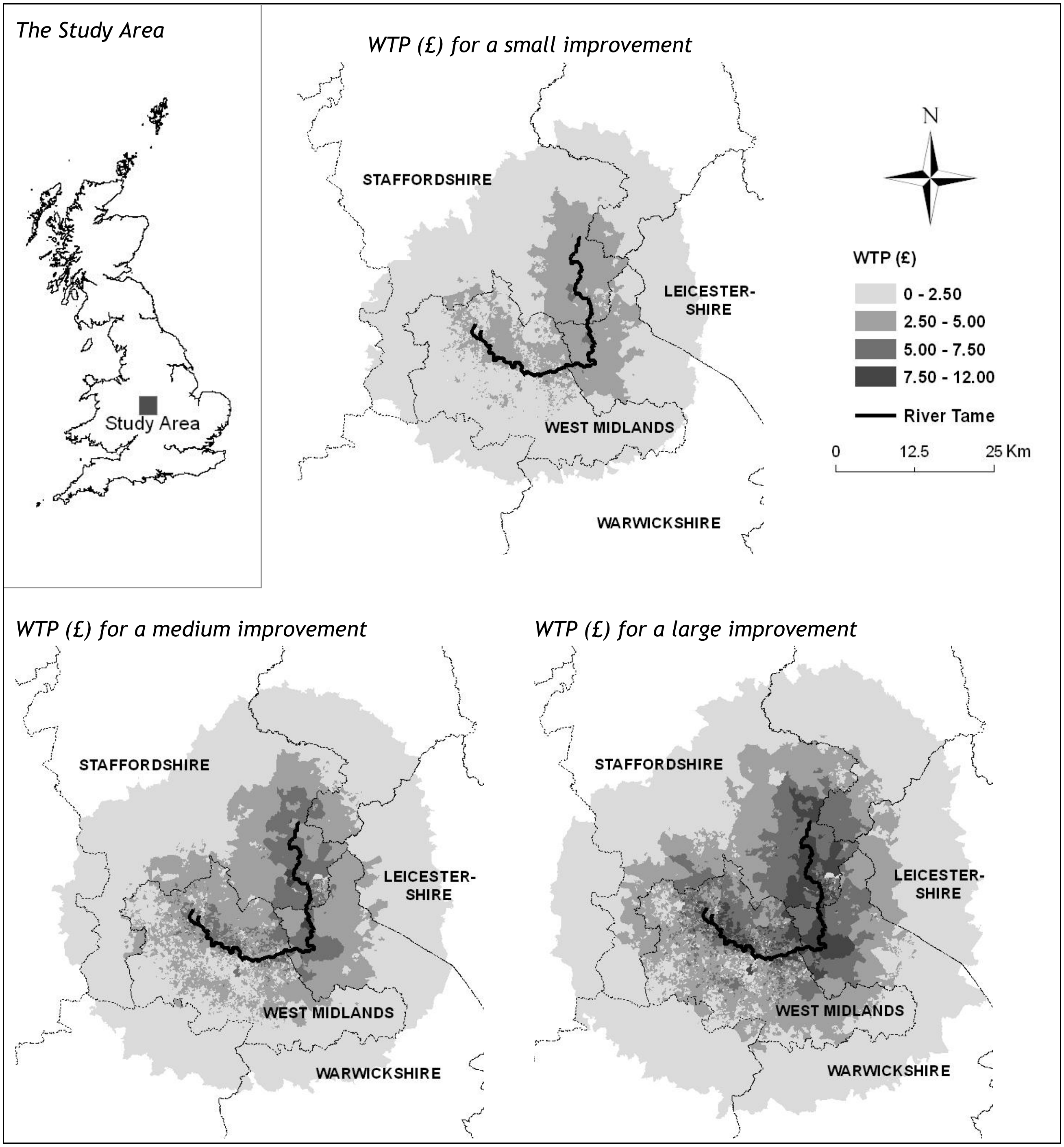
[53] More rapid rates of change were investigated by adding two terms for the interaction between distance and the two scope variables. However, in this case both were found to be clearly statistically insignificant and so are omitted from the final model reported above.

Table 6.1: Spatially sensitive valuation functions for both the full sample and for present non-users only

	All respondents	*Present non-users only*
Distance	-0.0000771 (2.97)**	-0.00008 (2.56)*
Large	0.623 (14.69)**	0.676 (11.42)**
Medium	0.330 (7.78)**	0.329 (5.56)**
ENV	0.219 (2.90)**	0.298 (3.00)**
Know	0.356 (2.36)*	0.503 (2.46)*
Retire	-0.601 (3.35)**	-0.613 (2.45)*
Poor	-0.100 (3.06)**	-0.100 (2.28)*
Constant	1.001 (2.79)**	0.583 (1.22)
Observations	1179	660
Number of groups	393	220
sigma_u	1.41	1.41
sigma_e	0.60	0.62
Rho	0.85	0.84
Wald chi2 prob>chi2	261.06 0.000	167.58 0.000
r-squared	0.12	0.16
Max Distance: *Large*	27.75km	
Max Distance: *Medium*	23.94km	
Max Distance: *Small*	19.66km	

One of the advantages of a GIS-based methodology is the ability of the software to readily generate graphic representations of findings. After removing those areas that were beyond the boundaries of the economic jurisdiction estimated for the specific distance decay function being considered, i.e. those with negative WTP values, maps were produced showing the spatial distribution of values for each of the improvements. These are illustrated in **Figure 6.1** and show how the size of the economic jurisdiction varies according to the improvement scenario considered, as well as how WTP varies in each output area census unit according to both its proximity to the River Tame and its socioeconomic characteristics. This latter facet of the results is amply illustrated by the contrast in values either side of the upper reaches of the Tame (its west to east reach). Here higher values are predicted for the more affluent north bank, with lower values estimated for the poorer central city area on the south bank.

Figure 6.1: Maps of estimated mean WTP (per household, per annum) of Census output areas for various water quality improvements



Turning to consider the model for present non-users values given in the final column of Table 6.1, a clear and highly significant distance decay effect is observed (p = 0.011). This result is characteristic of compensating surplus studies and seems to support the contention that differences will arise across welfare measures because of their differing impact upon the quality of sites. Equivalent loss studies present scenarios in which final quality is maintained at initial levels. Irrespective of their distance to the site, this does not induce present non-users to become higher value users. However, compensating surplus measures present scenarios in which site quality increases. This induces some present non-users to convert to users, a conversion which is greater for those nearer to the site[54]. This in turn results in a distance-decay in the values stated by those who are at present non-users of the resource, as observed here.

We can now use the above analysis to provide a spatially sensitive estimate of aggregate benefits for the economic jurisdiction and compare these with the standard approach to estimating aggregate benefits for an administratively defined jurisdiction (such as the one adopted by the Environment Agency in the River Kennet enquiry). Under the latter approach we can define the aggregation population as simply those households which live within the relevant local water company area. Again following the EA approach we can then calculate aggregate benefit estimates by simply multiplying this population by the sample mean WTP. Resulting estimates are reported as the first column of results within **Table 6.2**. So as to allow comparison between aggregation errors and errors due to uncertainty regarding mean WTP, we also report a confidence interval (CI) around these estimates based upon the 95% CI for the sample mean. The second column of results adopts the same approach to aggregation with the one refinement that the aggregation is now applied across the economic jurisdiction as defined by our spatially sensitive valuation function (as reported in Table 6.1). The population within the economic jurisdiction is substantially smaller than that of the administrative jurisdiction suggesting immediately that the latter is liable to lead to overestimation of aggregate benefits as it includes households for which WTP is at best zero (and arguably negative). Furthermore, as the scope of the good declines so the economic jurisdiction becomes even smaller. This will progressively lead to greater error arising from reliance upon the administrative jurisdiction approach. Indeed the administrative jurisdiction method leads to estimates which are just over double that for the economic jurisdiction for the large improvement and more than two and a half times too high for the small improvement. These errors dwarf those due to uncertainty in the estimate of mean WTP which range from 17% for the large improvement to 20% for the small improvement.

[54] This spatial trend can be observed in the present distribution of users to non-users in the sample. This itself exhibits a highly significant distance decay (p<0.01) such that the proportion of users in the sample falls from nearly 50% near to the site to almost zero at a distance of 9 km. Further analysis of this trend is given in Bateman et al (2005b).

Table 6.2: Aggregate benefits estimates based on sample mean and valuation function approaches

Quality change		***Aggregation using sample mean WTP***		***Aggregation using WTP estimated from distance decay function***
		Administrative Jurisdiction[1]	Economic Jurisdiction[2]	Economic Jurisdiction[2]
Large improvement[3]	Number of households	3,494,438	1,647,777	1,647,777
	Aggregate WTP	£82,049,404	£38,689,804	£5,040,526
	95% CI for aggregate WTP	(£68,001,763-£96,062,101)	(£32,065,740-£45,297,390)	
Medium improvement[4]	Number of households	3,494,438	1,486,415	1,486,415
	Aggregate WTP	£54,687,955	£23,262,395	£3,350,233
	95% CI for aggregate WTP	(£44,938,473-£64,437,437)	(£19,115,297-£27,409,493)	
Small improvement[5]	Number of households	3,494,438	1,336,736	1,336,736
	Aggregate WTP	£34,525,047	£13,206,952	£1,997,502
	95% CI for aggregate WTP	(£27,780,782-£41,269,313)	(£10,627,051-£15,786,852)	

Notes: Estimates are in £, 1999 values, per annum.

1. Local Water Utility Company area (Severn Trent and South Staffordshire Water Company Ltd.)
2. Area for which mean WTP > 0
3. Sample mean WTP for *Large* improvement = £23.48 (95% CI = £19.46: £27.49)
4. Sample mean WTP for *Medium* improvement = £15.65 (95% CI = £12.86: £18.44)
5. Sample mean WTP for *Small* improvement = £9.88 (95% CI = £7.95: £11.81)

While the change from administrative to economic jurisdiction substantially alters aggregation estimates, the final column of results shows that, at least in this case, an even greater source of error arises from reliance upon sample means within the aggregation process. As shown in Table 6.1, WTP values decline significantly across space such that unless samples are fully representative of the underlying population mean values can be poor indicators of value for that population. Given that, ahead of any valuation survey, we are unlikely to know the extent of the economic jurisdiction, ensuring sample representativeness of this *a-priori* uncertain area can be a difficult if not impossible matter to assess. Application of the valuation function allows us to estimate how household WTP varies across the economic jurisdiction, here calculating values for each census output area, taking into account its distance from the site and those area characteristics included within the model (other variables being held at their mean values). Resulting values, shown in the final column of Table 6.2, are superior to those given elsewhere in the table as they are both based upon the economic jurisdiction and best capture the variability of values across that area. Comparison with other estimates is revealing. In particular when compared with the administrative jurisdiction and sample mean aggregation approach we see that the latter are more than 16 times too high.

6.4 Conclusions

This case study has considered some of the factors which can influence the calculation of aggregate WTP estimates. Our analysis confirms the findings of Smith (1993) and Loomis (2000) that the choice of whether to aggregate across an administratively defined or economic jurisdiction can have a very substantial impact upon estimates of aggregate value. Similarly we have shown that the use of simple approaches such as aggregation via sample means can severely bias such estimates, and is very likely to occur given that the survey analyst is very unlikely to have prior knowledge of the correct area over which to aggregate. As an alternative to such over-simplified approaches we have argued for the use of a spatially sensitive valuation function, explicitly incorporating sample self-selection and expected distance decay in values to both define the limits of the economic jurisdiction and investigate how values vary within that area.

7 REFERENCES

Ariely, D., Loewenstein, G. and Prelec, D. (2003) 'Coherent arbitrariness': stable demand curves without stable preferences, *Quarterly Journal of Economics*. 118 (1): 73–105.

Arrow, K., R. Solow, P.R. Portney, E.E. Leamer, R. Radner, H. Schuman, (1993) Report of the NOAA panel on contingent valuation, *Fed. Reg*. 58, 4601–4614.

Banerjee, S. and Murphy, J.H. (2005) The scope test revisited, *Applied Economics Letters*, 12(10): 613-617.

Barton, D.N. (2002) The transferability of benefit transfer: contingent valuation of water quality improvements in Costa Rica. *Ecological Economics*, 42: 147-164.

Barton, D. N. (1999) The Quick, the Cheap and the Dirty. Benefit Transfer Approaches to the Non-market Valuation of Coastal Water Quality in Costa Rica. Ph.D. Thesis, Department of Economics and Social Sciences, Agricultural University of Norway.

Bateman, I.J. (2009) *Bringing the real world into economic analyses of land use value: Incorporating spatial complexity*, Report to the UK Foresight Land Use Futures Project, Government Office for Science, Department for Business, Innovation and Skills.

Bateman, I.J., Day, B.H., Dupont, D. and Georgiou, S. (forthcoming) Procedural invariance testing of the one-and-one-half-bound dichotomous choice elicitation method, *Review of Economics and Statistics*, in press.

Bateman, I.J., Day, B.H., Jones, A. P. and Jude, S. (2009) Reducing gains/loss asymmetry: A virtual reality choice experiment (VRCE) valuing land use change, *Journal of Environmental Economics and Management*, 58(1): 106-118.

Bateman, I.J., Burgess, D., Hutchinson, W.G. and Matthews, D.I., (2008a) Contrasting NOAA guidelines with Learning Design Contingent Valuation (LDCV): Preference learning versus coherent arbitrariness, *Journal of Environmental Economics and Management*, 55: 127–141. http://dx.doi.org/10.1016/j.jeem.2007.08.003

Bateman, I.J., Munro, A. and Poe, G.L. (2008b) Asymmetric Dominance Effects in Choice Experiments and Contingent Valuation, *Land Economics*, 84: 115 - 127.

Bateman, I.J., Dent, S., Peters, E., Slovic, P. and Starmer, C., (2007a) The Affect Heuristic and the Attractiveness of Simple Gambles, *Journal of Behavioral Decision Making*, 20(4): 365-380.

Bateman, I.J., Day, B.H., Loomes, G. and Sugden, R., (2007b) Can ranking techniques elicit robust values? *Journal of Risk and Uncertainty*, 34:49–66.

Bateman, I.J. and Brouwer, R. (2006) Consistency and construction in stated WTP for health risk reductions. A novel scope sensitivity test, *Resource and Energy Economics* 28: 199-214.

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

Bateman, I. J., Cole, M. A., Georgiou, S., and Hadley, D. J. (2006b) Comparing contingent valuation and contingent ranking: A case study considering the benefits of urban river water quality improvements, *Journal of Environmental Management*, 79: 221-231.

Bateman, I.J, Brouwer, R., Davies, H., Day, B.H., Deflandre, A., Di Falco, S., Georgiou, S., Hadley, D., Hutchins, M., Jones, A.P., Kay, D., Leeks, G., Lewis, M., Lovett, A.A., Neal, C., Posen, P., Rigby, D., and Turner, R.K. (2006c) Analysing the agricultural costs and non-market benefits of implementing the Water Framework Directive: The Catchment Hydrology, Resources, Economics And Management (ChREAM) project, Rural Economy and Land Use (RELU) programme special issue of the *Journal of Agricultural Economics*, Vol. 57, No. 2, 221–237.

Bateman, I.J., Day, B.H., Dupont, D., Georgiou, S., Matias, N.G.N. Morimoto, S. and Subramanian, L. (2006d) Cost-benefit analysis and the prevention of eutrophication, in Pearce, D.W. (ed.) *Environmental Valuation in Developed Countries: Case Studies*, Edward Elgar, Cheltenham, UK., pp. 317-342.

Bateman, I.J., Kahneman, D., Munro, A., Starmer, C. and Sugden, R. (2005a) Testing competing models of loss aversion: An adversarial collaboration, *Journal of Public Economics*, 89(8): 1561–1580, doi:10.1016/j.jpubeco.2004.06.013.

Bateman, I.J., Georgiou, S. and Lake, I. (2005b) The Aggregation of Environmental Benefit Values: A Spatially Sensitive Valuation Function Approach, *CSERGE Working Paper series,* Centre for Social and Economic Research on the Global Environment, University of East Anglia.

Bateman, I.J., Cooper, P., Georgiou, S., Navrud, S., Poe, G.L., Ready, R.C., Reira, P., Ryan, M. and Vossler, C.A. (2005c) Economic valuation of policies for managing acidity in remote mountain lakes: Examining validity through scope sensitivity testing, *Aquatic Sciences*, 67(3), 274-291.

Bateman, I.J. and Munro, A. (2005) An experiment on risky choice amongst households, *Economic Journal*, 115(502): 176-189.

Bateman, I.J., Cole, M., Cooper, P., Georgiou, S., Hadley, D. and Poe, G.L., (2004) On visible choice sets and scope sensitivity, *Journal of Environmental Economics and Management*, 47: 71-93.

Bateman, I.J. and Mawby, J. (2004) First impressions count: A study of the interaction of interviewer appearance and information effects in contingent valuation studies, *Ecological Economics*, 49(1): 47-55;

Bateman, I.J. and Jones, A.P., (2003) Contrasting conventional with multi-level modelling approaches to meta-analysis: An illustration using UK woodland recreation values, *Land Economics*, 79(2): 235-258.

Bateman, I.J., Lovett, A.A. and Brainard, J.S. (2003) *Applied Environmental Economics: a GIS Approach to Cost-Benefit Analysis,* Cambridge University Press, Cambridge.

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

Bateman, I.J., Jones, A.P., Lovett, A.A., Lake, I. and Day B.H. (2002b) Applying geographical information systems (GIS) to environmental and resource economics, *Environmental and Resource Economics*, 22(1-2): 219-269.

Bateman, I.J., Langford, I.H., Jones, A.P. and Kerr, G.N. (2001) Bound and path effects in multiple-bound dichotomous choice contingent valuation, *Resource and Energy Economics*, 23(3): 191-213;

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

Bateman, I.J., Langford, I.H. and Nishikawa, N. and Lake, I. (2000b) The Axford debate revisited: A case study illustrating different approaches to the aggregation of benefits data, *Journal of Environmental Planning and Management*, 43(2), 291-302.

Bateman, I.J., Ennew, C., Lovett, A.A. and Rayner, A.J. (1999a) Modelling and mapping agricultural output values using farm specific details and environmental databases, *Journal of Agricultural Economics,* 50(3) 488-511.

Bateman, I.J. and Langford, I.H. (1997) Budget constraint, temporal and ordering effects in contingent valuation studies, *Environment and Planning A*, 29(7): 1215-1228.

Bateman, I.J., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997a) A test of the theory of reference-dependent preferences, *Quarterly Journal of Economics,* 112(2): 479-505.

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

Bateman, I.J., Langford, I.H., Turner, R.K., Willis, K.G. and Garrod, G.D. (1995a) Elicitation and truncation effects in contingent valuation studies, *Ecological Economics*, 12(2):161-179.

Bateman, I.J., Brainard, J.S. and Lovett, A.A. (1995b) 'Modelling woodland recreation demand using geographical information systems: a benefits transfer study', Global Environmental Change Working Paper 95-06, Centre for Social and Economic Research on the Global Environment (CSERGE), University College London and University of East Anglia, Norwich.

Batker, D., Swedeen, P., Costanza, R., de la Torre, I., Boumans, R. and Bagstad, K. (2008) *A new view of the Puget Sound economy: the economic value of nature's services in the Puget Sound basin*. Seattle, WA: Earth Economics.

Bergland, O., Fried, B. and Adams, R. (1996) 'Temporal stability of values from the contingent valuation method', Paper presented at European Association of Environmental and Resource Economist (EAERE) Conference, Lisbon.

Bergland, O., Magnussen, K. and Navrud, S. (1995) 'Benefit Transfer: Testing for Accuracy and Reliability,' available in chapter 7: Florax, R.J.G.M., Nijkamp, P. and Willis, K.G. (eds.) (2002) *Comparative Environmental Economic Assessment,* Edward Elgar, UK.

Birnbaum, M. H. (1999). How to show that 9 > 221: Collect judgments in a between-subjects design. *Psychological Methods*, 4(3), 243–249.

Bockstael, N.E. and McConnell, K.E. (2006) *Environmental and Resource Valuation with Revealed Preferences: A Theoretical Guide to Empirical Models*, The Economics of Non-Market Goods and Services: Volume 7, Springer, Dordrecht.

Bockstael, N., Freeman, A.M. and Kopp, R., Portney, P.R. and Smith, V.K. (2000) 'On measuring the economic values for nature', *Environmental Science and Technology*, 34: 1384–89.

Boyle, K.J., Poe, G.L. and Bergstrom, J.C (1994) 'What do we know about groundwater values? Preliminary implications from a meta analysis of contingent valuation studies', *American Journal of Agricultural Economics* 76, 1055-1061.

Boyle, K.J. and Bergstrom, J.C. (1992) 'Benefit Transfer Studies: Myths, Pragmatism, and Idealism.' *Water Resources Research*, Vol. 28, No. 3: 657-663.

Brander, L.M., Ghermandi, A., Kuik, O., Markandya, A., Nunes, P.A.L.D., Schaafsma and M., Wagtendonk, A. (2008) 'Scaling up ecosystem services values: methodology, applicability and a case study. Final Report to the European Environment Agency.

Brander, L.M., Florax, R.J.G.M. and Vermaat, J.E. (2006) 'The empirics of wetland valuation: a comprehensive summary and a meta-analysis of the literature' *Environmental and Resource Economics*, 33, 223-250.

Brisson, I. and Pearce, D.W. (1995) 'Benefits transfer for disamenity from waste disposal', Global Environmental Change Working Paper WM 95-06, Centre for Social and Economic Research on the Global Environment (CSERGE), University College London and University of East Anglia, Norwich.

Brouwer, R. (2000) Environmental Value Transfer: State of the Art and Future Prospects. *Ecological Economics*, 32: 137-152.

Brouwer, R., Bateman, I.J., Barton, D., Georgiou, S., Martín-Ortega, J., Navrud, S., Pulido-Velazquez, M. and Schaafsma, M. (2009) Economic Valuation of Environmental and Resource Costs and Benefits in the Water Framework Directive: Technical Guidelines for Practitioners, VU University Amsterdam, the Netherlands.

Brouwer, R. and Bateman, I.J. (2005a) Benefits Transfer of Willingness to Pay Estimates and Functions for Health-Risk Reductions: A Cross-Country Study. *Journal of Health Economics* 24: 591-611.

Brouwer, R. and Bateman, I.J. (2005b) 'The Temporal Stability and Transferability of Models of Willingness to Pay for Flood Control and Wetland Conservation'. *Water Resources*

Research 41(3).

Brouwer, R., Langford, I.H., Bateman, I.J., and Turner, R.K., (1999) A meta-analysis of wetland contingent valuation studies, *Regional Environmental Change*, 1(1): 47-57.

Brouwer, R. and Spanninks, F.A. (1999) The validity of environmental benefits transfer: Further empirical testing, *Environmental and Resource Economics,* 14(1): 95-117.

Carson, R.T., Flores, N.E. and Hanemann, W.M. (1998) Sequencing and valuing public goods, *Journal of Environmental Economics and Management*, 36; 314-323.

Carson, R.T. and Groves, T. (2007) Incentive and informational properties of preference questions, *Environmental & Resource Economics*, vol. 37(1), pages 181-210.

Carson, R., Hanemann, M., Kopp, R., Krosnick, J., Mitchell, R., Presser, S., Rudd, P. and Smith, V.K. with Conaway, M. and Martin, K. (1996) Was the NOAA panel correct about contingent valuation?, *Resources for the Future Discussion Paper 96-20*, Washington D.C.

Carson, R. T. and Mitchell, R. C. (1993) The value of clean water: The public's willingness to pay for boatable, fishable and swimmable quality water, *Water Resources Research*, 29(7): 2445–2454.

Case, A. C. (1991) Spatial patterns in household demand, *Econometrica*, 59:953-965.

Champ, P. A., Boyle, K. and Brown, T.C. (eds.) (2003) *A Primer on Non-market Valuation*, The Economics of Non-Market Goods and Services: Volume 3, Kluwer Academic Press, Dordrecht.

Chattopadhyay S. (2003) A Repeated Sampling Technique in Assessing the Validity of Benefit Transfer in Valuing Non-Market Goods, *Land Economics*, 79 (4): 576–596.

Costanza, R., d'Arge, R., deGroot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., and van den Belt, M. (1997) 'The value of the world's ecosystem services and natural capital', *Nature*, 387, 253–260.

Covey, J., Loomes, G. and Bateman, I.J., (2007) Valuing risk reductions: Testing for range biases in payment card and random card sorting methods, *Journal of Environmental Planning and Management*, 50(4): 467-482.

Davies, H. and Neal, C. 2004. GIS based methodologies for assessing nitrate, nitrite and ammonium distributions across a major UK basin, the Humber, *Hydrol. Earth Sys. Sci.*, 8:823-833.

Defra (2007) *Securing a healthy natural environment: An action plan for embedding an ecosystems approach*, December.

DeShazo, J.R. (2002) Designing transactions without framing effects in iterative question formats, *J. Environ. Econ. Manage*. 43: 360–385.

Desvousges, W.H, Johnson, F.R. and Banzhaf, H.S. (1998) *Environmental Policy Analysis with Limited Information: Principles and Applications of the Transfer Method*, New Horizons in Environmental Economics, Edward Elgar, Cheltenham, UK and Northampton, MA, USA.

Desvousges, W.H., Naughton, M.C. and Parsons, G.R. (1992) Benefit transfer: conceptual problems in estimating water quality benefits using existing studies. *Water Resources Research*, 28(3): 675-683.

Desvousges, W.H., Smith, V.K. and Fisher, A. (1987) Option price estimates for water quality improvements: a contingent valuation study of the Monongahela River, *Journal of Environmental Economics and Management*, 14: 248-267.

Downing, M. and Ozuna, T. (1996) Testing the reliability of the benefit function transfer approach, *Journal of Environmental Economics and Management* 30: 316-322.

Doyle, J. R., O'Connor, D. J., Reynolds, G. M., & Bottomley, P. A. (1999) The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases, *Psychology and Marketing*, 16(3), 225-243.

EA (2007a) General quality assessment (GQA): Water chemistry guide, Environment Agency, available from http://www.environment-agency.gov.uk/ (accessed November 2007).

EA (2007b) Freshwater Fish Directive, Environment Agency, available from http://www.environment-agency.gov.uk/ (accessed November 2007).

EA (2007c) *Coarse fish biology and management*, Environment Agency, available from http://www.environment-agency.gov.uk/ (accessed November 2007).

EA (2007d) *General Quality Assessment of rivers: chemistry*, Environment Agency, available from http://www.environment-agency.gov.uk/ (accessed November 2007).

EA (2007e) *Fisheries Habitat Improvement*, Environment Agency, available from http://www.environment-agency.gov.uk/ (accessed November 2007).

EA (2007f) Environments for Fish, Environment Agency, available from http://www.environment-agency.gov.uk/ (accessed November 2007).

EA (2007g) River water quality standards, Environment Agency, available from http://www.environment-agency.gov.uk/ (accessed November 2007).

Egan, K.J., Herriges, J.A., Kling, C.L. and Downing, J.A. (2009) Recreation Demand Using Physical Measures of Water Quality, *American Journal of Agricultural Economics*, Vol. 91, 1, pp. 106-123.

ENDS (1998) Water abstraction decision deals savage blow to cost-befit analysis, *ENDS*, 278:16-18.

Eshet, T., Baron, M.G. and Shechter, M. (2007) 'Exploring Benefit Transfer: Disamenities of Waste Transfer Stations', *Environmental and Resource Economics*, 37(3), 521-547.

European Parliament (2000) Directive 2000/60/EC (The EU Water Framework Directive), *Official Journal* (OJ L 327): 22 December 2000.

Fagerlin, A., Wang, C., Ubel, P. A. (2005) Reducing the Influence of Anecdotal Reasoning on People's Health Care Decisions: Is a Picture Worth a Thousand Statistics? *Medical Decision Making*, 25:398–405.

Ferrini, S., Fezzi, C., Day, B.H. and Bateman, I.J. (2008) Valuing Spatially Dispersed Environmental Goods: A joint revealed and stated preference model to consistently separate use and non-use values *CSERGE Working Paper* EDM 08-03, Centre for Social and Economic Research on the Global Environment, University of East Anglia.

Freeman, A.M. (2002) 'How much is Nature really worth? An economic perspective. In: Valuing nature: the Shipman Workshop papers. Brunswick, ME: Bowdoin College.

FWR (1996) *Benefits assessments manual for surface water quality improvements*, Foundation for Water Research, Marlow.

Garrod, G. And Willis, K.G. (1999) Economic Valuation of the Environment: Methods and Case Studies, Edward Elgar Publishing, Cheltenham.

Haab, T. C. and McConnell, K.E. (2002) Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation, Edward Elgar Publishing, Cheltenham.

Haab, T. C., Huang, J-C. and Whitehead, J. (forthcoming), Preference Data for Environmental Valuation: Combining Revealed and Stated Approaches, Routledge.

Halstead J.M., Lindsay B.E., Brown C.M., (1991) Use of the Tobit model in the contingent valuation: experimental evidence from the Pemigewasset wilderness area, *Journal of Environmental Management* 33: 79:89.

Hanley, N., Schläpfer, F. and Spurgeon, J. (2003) Aggregating the benefits of environmental improvements: Distance-decay functions for use and non-use values. *Journal of Environmental Management*, 68, 297-304.

Heckman J.J. (1987) Selection Bias and Self-Selection, in P. Newman, M. Milgate and J. Eatwell (eds.), *The New Palgrave - A Dictionary of Economics*, Macmillan.

Hicks, John R. (1943) The four consumer surpluses, *Review of Economic Studies*, 8: 108-116.

Hime, S., Bateman, I.J., Posen, P. and Hutchins, M. (2009) A transferable water quality ladder for conveying use and ecological information within public surveys, *CSERGE Working Paper EDM 09-01*, Centre for Social and Economic Research on the Global Environment, University of East Anglia.

Hoehn J. (2006) 'Methods to address selection effects in the meta-regression and transfer of ecosystem values', *Ecological Economics* 60: 389–98.

Horowitz, J. K., and McConnell, K. E. (2002) A Review of WTA / WTP Studies, *Journal of Environmental Economics and Management*, 44: 426-447.

Hsee, C. K. (1996) The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, 67, 247–257.

Hutchins, M. G.; Deflandre, A.; Boorman, D. B. (2006) Performance benchmarking linked diffuse pollution and in-stream water quality models. *Archiv fur Hydrobiologie Supplement*, 161 (Large Rivers Supplement 17), 133-154.

Imber, D. Stevenson, G. and Wilks, L (1991) *A contingent valuation survey of the Kakadu Conservation Zone*, Research Paper No.3, Resource Assessment Commission, Canberra.

Johnstone, R.J. and Duke, J.M. (2009) Willingness to Pay for Land Preservation across States and Jurisdictional Scale: Implications for Benefit Transfer, *Land Economics* 85 (2): 217-237.

Johnston, R.J., E.Y. Besedin, R. Iovanna, C.J. Miller, R.F. Wardwell, and M.H. Ranson, (2005) Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: A meta-analysis, *Canadian Journal of Agricultural Economics,* 53: 221-48.

Johnston, R.J., M.H. Ranson, E.Y. Besedin, and E.C. Helm, (2006) What determines willingness to pay per fish? A meta-analysis of recreational fishing values, *Marine Resource Economics,* 21: 1-32.

Jones, A.P., Bateman, I.J. and Wright, J., (2002) Estimating arrival numbers and values for informal recreational use of British woodlands, report to the Forestry Commission, Edinburgh, published at http://www.forestry.gov.uk.

Kahneman, D. (2003) A perspective on judgment and choice: Mapping bounded rationality,. *American Psychologist*, 58: 697–720.

Kahneman, D. and Knetsch, J.L. (1992) Valuing public goods: The purchase of moral satisfaction, *Journal of Environmental Economics and Management* 22(1): 57-70.

Kahneman, D., Slovic, P. and Tversky, A. (1982) *Judgement Under Uncertainty: Heuristics and Biases*, Cambridge University Press, New York.

Kahneman, D. and Sugden, R. (2005) Experienced utility as a standard of policy evaluation, *Environmental and Resource Economics*, 32 (1): 161-81.

Kanninen, B. (ed.) (2006) Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice, The Economics of Non-Market Goods and Services: Volume 8, Springer, Dordrecht. ISBN: 978-1-4020-4064-1

Kask, S.B. and Shogren, J.F. (1994) 'Benefit transfer protocol for long-term health risk valuation: a case of surface water contamination', *Water Resources Research*, 30 (10), 2813–2823.

Kay, D., M.D. Wyer, J. Crowther, C. Stapleton, M. Bradford, A.T. McDonald, J. Greaves, C. Francis, and J. Watkins, (2005) Predicting faecal indicator fluxes using digital land use data in the UK's sentinel Water Framework Directive catchment: the Ribble study, *Water Research* 39, 655-667.

Kirchoff, S., Colby. B.G. and LaFrance, J.T. (1997) 'Evaluating the Performance of Benefit Transfer: An Empirical Inquiry.' *Journal of Environmental Economics and Management*, 33: 75-93.

Kristofersson, D. and Navrud, S. (2007) 'Can Use and Non-Use Values be Transferred Across Countries?' in Navrud, S. and R. Ready (eds.) *Environmental Value Transfer: Issues and Methods*. Springer, Dordrect, The Netherlands.

Kristofersson, D. and Navrud, S. (2005) Validity Tests of Benefit Transfer – Are We Performing the Wrong Tests, *Environmental and Resource Economics*, 30 (3), 279-286.

Krutilla, J.V. and Fisher, A.C. (1975) *The Economics of Natural Environments Studies in the Valuation of Commodity and Amenity Resources*, John Hopkins Press, Baltimore, USA

Lanz, B., Provins, A.J., Bateman, I.J., Scarpa, R., Willis, K.G. and Ozdemiroglu, E. (2009) Investigating willingness to pay – willingness to accept asymmetry in choice experiments, presented at the inaugural *International Choice Modelling Conference*, Institute for Transport Studies, University of Leeds, 30th March to 1st April 2009.

Leon-Gonzalez, R. and Scarpa, R. (2008) Improving multi-site benefit functions via Bayesian model averaging: A new approach to benefit transfer, *Journal of Environmental Economics and Management* 56(1): 50-68.

Lichtenstein, S., & Slovic, P. (1971) Reversals of preference between bids and choices in gambling decisions, *Journal of Experimental Psychology*, 89, 46–55.

Lindhjem, H.L. and Navrud, S. (2008) How reliable are meta-analyses for international benefit transfers?, *Ecological Economics,* 66: 425-435.

List, J.A. (2002) Preference Reversals of a Different Kind: The "More Is Less" Phenomenon, *American Economic Review*, 92(5): 1636-1643.

Loewenstein, G. and Frederick, S. (1997) Predicting Reactions to Environmental Change. In M. Bazerman, D. Messick, A. Tenbrunsel & K. Wade-Benzoni (Eds.), *Psychological Perspectives on the Environment*, Russell Sage Foundation Press.

Loewenstein, G., & Schkade, D.A. (1999) Wouldn't it be nice? Predicting future feelings, in D. Kahneman, E. Diener & N. Schwarz (Eds.) *Well-being: Foundations of hedonic psychology* (pp. 85 - 105). New York: Russell-Sage.

Loomis, J. B. (2000) Vertically summing public good demand curves: an empirical comparison of economic versus political jurisdictions, *Land Economics*, 76(2), 312-321.

Loomis, J.B. (1992) 'The Evolution of a More Rigorous Approach to Benefit Transfer: Benefit Function Transfer.' *Water Resources Research*, Vol. 28, No. 3: 701-705.

Loomis, J.B., Le, H.T. and Gonzales-Caban, A. (2005) 'Testing transferability of willingness to pay for forest fire prevention among three states of California, Florida, and Montana', *Journal of Forest Economics*, vol. 11, pp. 125-140.

Loomis, J.B. and White, D.S. (1996) 'Economic benefits of rare and endangered species: summary and meta-analysis'. *Ecological Economics* 18; 197-206.

Luken, R.A., Johnson, F.R., Kibler, V., (1992) 'Benefits and costs of pulp and paper effluent controls under the Clean Water Act', *Water Resources Research*, 28 (3), 665–674.

Magnussen, K. (1993) Mini meta analysis of Norwegian water quality improvements valuation studies, Norwegian Institute for Water Research, Oslo. Unpublished paper.

Met Office (2009) *Wales: Climate*, http://www.metoffice.gov.uk/climate/uk/wl/print.html accessed February 2009.

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

Moeltner, K., K.J. Boyle, and R.W. Paterson, (2007) Meta-analysis and benefit transfer for resource valuation–Addressing classical challenges with Bayesian modeling, *Journal of Environmental Economics and Management*, 53: 250-69.

Moran, D. (1999) Benefits transfer and low flow alleviation: what lessons for environmental valuation in the UK? *Journal of Environmental Planning and Management*, 42(3), 425-436.

Morrison, M. (2000) Aggregation Biases in Stated Preference Studies, *Australian Economic Papers, 39(2), 215-230.*

Mouranaka, A. (2004) Spatial economic evaluation of artificial Japanese cedar forest management as a countermeasure for Japanese cedar pollinosis: An analysis using a model of multizonal contingent markets with data from cities, towns and villages in Yamaguchi Prefecture, Japan, *Geographical Review of Japan*, 77 (13), 903-923.

Muthke, T. and Holm-Mueller, K. (2004) National and international benefit transfer testing with a rigorous test procedure, *Environmental and Resource Economics*, 29: 323-336.

Navrud, S. (2007) *Practical tools for value transfer in Denmark – guidelines and an example*, Working Report No. 28, Environmental Protection Agency, Danish Ministry of the Environment.

Navrud, S. and Ready, R. (2007a) Review of methods for value transfer, in: Navrud, S. and Ready, R. (eds.), *Environmental value transfer: Issues and methods*. Springer, Dordrecht, The Netherlands.

Navrud, S. and Ready, R. (eds.) (2007b) *Environmental value transfer: Issues and methods*. Springer, Dordrecht, The Netherlands.

Neal, C. and Jarvie, H.P., (2005) Agriculture, community, river eutrophication and the Water Framework Directive, *Hydrol. Proc.*, 19, 1895-1901.

Neal, C., Jarvie, H.P., Neal, M., Love, A.J., Hill, L. and Wickham, H., (2005) Water quality of treated sewage effluent in a rural area of the upper Thames Basin, southern England, and the impacts of such effluents on riverine phosphorus concentrations, *Journ. Hydrol.*, 304(1-4), 103-117.

Parsons, G.R. (1991) A note on choice of residential location in travel cost demand models, *Land Economics*, 67: 360-364.

Pate, J. and Loomis, J.B. (1997) The effect of distance on willingness to pay values: A case study of wetlands and salmon in California, *Ecological Economics*, 20, 199-207.

Pearce, D.W. (1986) *Cost-Benefit Analysis*, 2nd ed., Macmillan, Basingstoke.

Pearce, D.W. (1998) 'Auditing the Earth', *Environment*, Vol. 40, No.2.

Pearce, D.W. and Turner, R.K. (1990) *The Economics of Natural Resources and the Environment*, Harvester Wheatsheaf, Hemel Hempstead.

Pearce, D.W., Whittington, D. and Georgiou, S. (1994) *Project and policy appraisal: Integrating economics and environment*. OECD, Paris.

Peters, E., Dieckmann, N., Västfjäll, D., Mertz, C. K. (2005b) *When five out of four people have trouble with fractions and other numbers: Numeracy and mood in decisions*, University of Oregon, Eugene, Oregon.

Peters, E., Slovic, P., Hibbard, J. (2005a) *Bringing Meaning to Numbers: Evaluability and Affect in Choice*, University of Oregon, Eugene, Oregon.

Plummer, M.L. (2009) 'Assessing benefit transfer for the valuation of ecosystem services', *Frontiers in Ecological and the Environment*, 7(1): 38–45.

Pompe, J.J. and Rinehart, J.R. (1995) 'Beach quality and the enhancement of recreational property-values', *Journal of Leisure Research*, 27: 143–54.

Powe, N.A. and Bateman, I.J., (2003) Ordering effects in nested 'top-down' and 'bottom-up' contingent valuation designs, *Ecological Economics*, 45: 255-270.

Powe, N.A. and Bateman, I.J. (2004) Investigating insensitivity to scope: A split-sample test of perceived scheme realism, *Land Economics*, 80(2): 258-271.

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

Ready, R., Navrud, S., Day, B., Dubourg, R., Machado, F., Mourato, S., Spaninks, F. and Vázquez Rodriquez, M.X. (2004) Benefit transfer in Europe: How reliable are transfers between countries?, *Environmental and Resource Economics*, 29: 67-82.

Rosenberger, R.S. and Phipps, T.T. (2007) 'Correspondence and convergence in benefit transfer accuracy: A meta-analytic review of the literature', in Navrud, S. and R. Ready (eds.) *Environmental Value Transfer: Issues and Methods*. Springer, Dordrect, The Netherlands.

Rosenberger, R.S., and Stanley, T.D. (2006) 'Measurement, generalization, and publication: Sources of error in benefit transfers and their management'. *Ecological Economics* 60 (2): 372–8.

Rosenberger, R.S. and Loomis, J.B. (2001) *Benefit transfer of outdoor recreation use values: A technical document supporting the Forest Service Strategic Plan* (2000 revision). Gen. Tech. Rep. RMRS-GTR-72. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Rowe, R.D., Schulze, W.D. and Breffle, W. (1996) A test for payment card biases, *Journal of Environmental Economics and Management*, 31: 178-185.

Santos, J.M.L. (2007) Transferring landscape values: how and how accurately? in Navrud, S. and Ready, R. (Eds.) *Environmental Value Transfer: Issues and Methods*, The Economics of Non-Market Goods and Services: Volume 9, Springer, Dordrecht.

Santos, J.M.L. (1998) *The Economic Valuation of Landscape Change: Theory and Policies for Land Use Conservation*. Edward Elgar, Cheltenham, UK.

Scarpa, R., Hutchinson, W.G., Chilton, S.M., and Buongiorno, J. (2007) Benefit value transfers conditional on site attributes: Some evidence of reliability from forest recreation in Ireland, in Navrud, S. and Ready, R. (Eds.) *Environmental Value Transfer: Issues and Methods,* The Economics of Non-Market Goods and Services: Volume 9, Springer, Dordrecht.

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

Schkade, D. A. and Payne, J. W. (1994) How people respond to contingent valuation questions: A verbal protocol analysis of willingness to pay for an environmental regulation, *Journal of Environmental Economics and Management* 26: 88–109.

Sedjo, R.A., Sampson, R.N., and Wisniewski, J. (eds.) (1997) *Economics of Carbon Sequestration in Forestry*, CRC Press, New York.

Shrestha, R.K., and J.B. Loomis (2001) 'Testing a meta-analysis model for benefit transfer in international outdoor recreation', *Ecological Economics* 39:67–83.

Silberman, J., Gerlowski, D.A. and Williams, N.A. (1992) 'Estimating existence value for users and nonusers of New Jersey beaches,' *Land Economics* 68: 225–36.

Slovic, P. (1995) The construction of preference. *American Psychologist*, 50, 364–371.

Smith, V.K. (1993) Nonmarket Valuation of Environmental Resources: An interpretive Appraisal, *Land Economics*, 69(1), 1-26.

Smith, V.K. and Osborne, L. (1996) 'Do Contingent Valuation Estimates Pass a "scope" Test? A Meta Analysis', *Journal of Environmental Economics and Management*, 31 (3), 287-301.

Smith, V.K. and Karou, Y. (1990) 'Signals or noise? Explaining the variation in recreation benefit estimates', *American Journal of Agricultural Economics*, 72 (2). 419-433.

Sturtevant, L.A., Johnson, F.R. and Desvousges, W.H. (1995) *A meta-analysis of recreational fishing*. Triangle Economic Research, Durham, North Carolina, USA.

Sutherland R.J. and Walsh, R.G. (1985) Effects of distance on the preservation value of water quality, *Land Economics*, 61(3), 281-291.

Swinton, S. M. (2002) Capturing household-level spatial influence in agricultural management using random effects regression, *Agricultural Economics*, 27:371-381.

Toman, M. (1998) 'Why not calculate the value of the world's ecosystem services and natural capital', *Ecological Economics*, 25: 57–60.

Tversky, A. and Kahneman, D. (1974) Judgement under uncertainty: heuristics and biases, *Science*, 185: 1124–1131.

UKTAG (2008) UK Environmental Standards and Conditions (Phase 1): Final Report - April 2008, The UK Technical Advisory Group on the Water Framework Directive (UKTAG) http://www.wfduk.org/.

Van den Berg, T.P., Poe, G.L., and Powell, J.R. (2001) Assessing the accuracy of benefits transfers: evidence from a multi-site contingent valuation study of ground water quality, in: Bergstrom J, Boyle K, and Poe G (Eds). *The Economic Value of Water Quality*. Northampton, MA: Edward Elgar.

Walsh, R.G., Johnson, D.M. and McKean, J.R. (1990) 'Nonmarket values from two decades of research on recreation demand', *Advances in Applied Microeconomics*, vol. 5, 167-193, JAI Press Inc.

Walsh, R.G., Johnson, D.M. and McKean, J.R. (1992) 'Benefit transfer of outdoor recreation demand studies 1968-1988', *Water Resources Research*, 28, (3) 707-714.

WATECO (2004) *Economics and the Environment - The Implementation Challenge of the Water Framework Directive,* Luxembourg: European Commission.

Willis, K.G. and Garrod, G.D. (1995) The benefits of alleviating low flow rivers, *Water Resources Development*, 11, 243-260.

Woodward, R.T. and Wui, Y. (2001) 'The economic value of wetland services: a meta-analysis', *Ecological Economics*, 37, 257-270

World Bank (2008) Global purchasing power parities and real expenditures. World Bank, Washington, D.C., U.S.A.

Zandersen M., Termansen M., and Jensen F.S. (2007) Testing Benefits Transfer of Forest Recreation Values over a Twenty-Year Time Horizon, *Land Economics*, 83 (3): 412–440.

Zikmund-Fisher, B J., Fagerlin, A., Ubel, P. A. (2005) What's Time Got to Do with It? Inattention to Duration in Interpretation of Survival Graphs, *Risk Analysis*, Vol. 25, No. 3, pp589-595.

ANNEX 1: BRIEF SURVEY OF VALUE TRANSFER LITERATURE

This Annex provides a brief survey of the value transfer literature, including:

- *Tests of the validity of value transfer;*
- *Reported transfer errors;*
- *Tests of the temporal stability of values;*
- *Meta-analysis studies;*
- *Valuation of ecosystem services via value transfer; and*
- *The practical use of value transfer*

A1.1 Introduction

The development of guidelines for value transfer in this study follows a rich history in the academic literature and elsewhere which has scrutinised the use, testing and overall validity of value transfer approaches to valuing changes in the provision of market and non-market goods and services. As Navrud (2007) reports, the practice of value transfer can be traced back over 30 years to the calculation of the loss of recreational value resulting from the Hell´s Canyon (USA) hydroelectric project as documented in Krutilla and Fisher (1975).

More recently the issue of validity was placed at the forefront of issues concerning value transfer in the 1992 special issue of Water Resources Research (WRR) (vol. 28, no. 3). This triggered much interest in value transfer and a steady increase in the literature addressing the variety of issues it engenders.

The following provides a brief overview of the literature. A more comprehensive and detailed account of developments in value transfer – including application to use and non-use values and validity tests - is available in Navrud and Ready (2007b). In addition, a comprehensive account of value transfer, with a focus on environmental and health impacts from air pollution, is provided by Desvousges et al. (1998).

A1.2 Validity of value transfer

Significant emphasis on the development and testing of value transfer methods dates from the special issue of WRR cited above. Notable papers include Loomis (1992), Boyle and Bergstrom (1992) and Desvousges et al. (1992) which essentially set the value transfer research agenda that subsequent articles have explored. In particular, the requirement for assessment of the convergent validity of estimated values at policy and study sites; for example Loomis (1992) tested the transferability of travel cost demand functions for freshwater fishing in Oregon and Idaho and sea fishing in Oregon and Washington State. The results of this study were mixed. Transferral of functions from Oregon to Washington and Oregon to Idaho was rejected by the statistical analysis. However, empirical support was found for transfer of travel cost demand functions within Oregon.

In response to the WRR agenda, Bergland et al. (1995) provided the first study that estimated and compared WTP functions derived from independent and simultaneous contingent valuation surveys, designed specifically for testing value transfer. Focusing on improvements in water quality for two Norwegian rivers, the study controlled for factors such as time lags, different questionnaire formats and different estimation techniques. However, the findings indicated a lack of transferability of mean WTP values and functions. The main conclusion from the study, mirrored by Downing and Ozuna (1996) and Kirchoff et al. (1997), was that site quality and socio-economic differences between sites can represent a key source of error in transfer of mean values and functions in value transfer testing studies.

Ensuing studies such as Brouwer and Spanninks (1999) and Brouwer and Bateman (2005a) sought to develop the statistical procedures for testing the reliability and validity of benefit transfer, in addition to continuing to build an evidence base on the validity of value transfers (Ready et al., 2004; Moeltner et al., 2007). For example Brouwer and Bateman (2005a) employ a stepwise testing procedure to determine how various contextual factors influence the performance of function transfer, assessing the importance of explanatory variables that describe individual's attitudes and perceptions of the environment; the conclusion being that socio-economic factors are a necessary but insufficient requirement of value functions for successful transfers. More recent studies have also highlighted the importance of similarity in the physical and biophysical characteristics of sites in determining the success of value transfers (Rosenberger and Stanley 2006; Rosenberger and Phipps 2007).

Contemporary studies in general have continued the trend of ambiguity in results of testing the validity of value transfer. For example Loomis et al. (2005) examine the equivalency of CV results for forest fire prevention from US studies in California, Florida and Montana. The study tests for equality of function coefficients across the three states and finds mixed evidence for function transfer. Brouwer and Bateman (2005b) investigate the transfer of CV WTP estimates for reducing health risks associated with solar UV exposure between four countries (England, Scotland, Portugal and New Zealand). Where contexts are similar (reported as transfers between England, Scotland, and New Zealand), mean unit value transfers are found to perform better than value function transfer. Where study and policy site contexts are different (e.g. transfers to Portugal), value function transfer are found to produce lower transfer errors.

Kristofersson and Navrud (2007) investigate the validity of value transfer between Norway, Sweden, and Iceland for use and non-use values for freshwater fish stocks, based on identical CV studies in the three countries. Both unit and function transfers are tested, with equivalency analysis applied to test the validity of value transfers (rather than 'standard' tests and null hypothesises of equivalence of values in the literature). Results indicate that the accuracy of value transfer relies heavily on the similarity of study sites.

Similarly Eshet et al. (2007) examine the accuracy of transferring values for the disamenity of housing locations close to waste transfer stations between four cities in Israel. Value functions derived from separate hedonic pricing studies are used to transfer values for each site. Inaccuracy of the transfers increases with the dissimilarity of sites although transfer errors are relatively low (see below). The study's findings also demonstrate that while tests of transferability of functions do not indicate transferability of coefficients, resulting transfer errors from predicted policy good values can still result in low transfer errors.

A1.2 Measuring transfer errors

In the value transfer testing literature, transfer errors are generally expressed as the mean absolute percentage error (MAPE). This is defined as the difference between observed value (the test's policy good value) and predicted value (the test's study good) divided by the observed value. Hence a transfer error of 50 percent indicates that the value from the study site used in the new policy context is 50 percent higher or lower than the 'true' value in the new policy context. **Table A1.1** provides an overview of results from a selection of studies that have tested the reliability of the transfer of WTP values, with the final two columns presenting the range of transfer errors found.

While the transfer errors results presented in Table A1.1 cover a large range (0-7028 percent), most studies find transfer errors in the range of 0-100; i.e. transfer value is one to two times more/less than the 'true' value. Brouwer et al. (2009) identify three principal sources of error in relation to the reliability of the value transfer:

- Errors associated with estimating the original measures of value at the study site(s) - measurement error in primary valuation studies and estimates can arise due to weak methodologies, unreliable data, analyst errors, and/or general biases and inaccuracies associated with valuation methods.

- Errors arising from the transfer of study good values to the policy good - so-called 'generalisation error' occurs in relation to the correspondence of characteristics of the study and policy goods and sites at which these are found. This may be in terms of population characteristics (income, culture, demographics, education etc.) or environmental/physical characteristics (quantity and/or quality of the good or service, availability of substitutes, accessibility, etc.). As such the degree of error is expected to be inversely related to the study and policy good correspondence. Generalisation error may also result from temporal transfer of values, where preferences and values for environmental goods may not remain constant over time (see also below).

- Publication selection bias – this can result in an unrepresentative stock of knowledge on values associated with environmental goods and services where the stock of existing evidence is skewed to certain types of results and that does not meet the information needs of analysts undertaking value transfer. Specifically, in the economics literature there is generally an editorial preference to publish statistically significant results and novel valuation applications rather than replications, which may result in publication bias.

<table>
<tr><th colspan="4">Table 1: Transfer errors found in value transfer literature (% MAPE)</th></tr>
<tr><th>Reference</th><th>Resource/activity</th><th>Unit transfer error</th><th>Function Transfer error</th></tr>
<tr><td>Loomis (1992)</td><td>Recreation</td><td>4-39</td><td>1-18</td></tr>
<tr><td>Parsons and Kealy (1994)</td><td>Water/recreation</td><td>4-34</td><td>1-75</td></tr>
<tr><td rowspan="2">Loomis et al. (1995)</td><td rowspan="2">Recreation</td><td>-</td><td>1-475</td></tr>
<tr><td>-</td><td>1-113</td></tr>
<tr><td>Bergland et al. (1995)</td><td>Water quality</td><td>25-45</td><td>18-41</td></tr>
<tr><td>Downing and Ozuna (1996)</td><td>Fishing</td><td>0-577</td><td>1</td></tr>
<tr><td rowspan="2">Kirchoff et al. (1997)</td><td>Whitewater rafting</td><td>36-56</td><td>87-210</td></tr>
<tr><td>Birdwatching</td><td>35-69</td><td>2-35</td></tr>
<tr><td rowspan="2">Bowker et al. (1997)</td><td rowspan="2">Whitewater rafting</td><td>-</td><td>14-160</td></tr>
<tr><td>-</td><td>16-57</td></tr>
<tr><td rowspan="2">Kirchoff (1998)</td><td rowspan="2">Recreation/habitat</td><td>-</td><td>2-475</td></tr>
<tr><td>-</td><td>3-7028</td></tr>
<tr><td>Brouwer and Spaninks (1999)</td><td>Biodiversity</td><td>27-36</td><td>22-40</td></tr>
<tr><td>Morrison and Bennett (2000)</td><td>Wetlands</td><td>1-191</td><td>-</td></tr>
<tr><td>Rosenberg and Loomis (2000)</td><td>Recreation</td><td>-</td><td>0-319</td></tr>
<tr><td rowspan="3">Piper and Martin (2001)</td><td rowspan="3">Rural water supply</td><td>-</td><td>6-20</td></tr>
<tr><td>-</td><td>89-149</td></tr>
<tr><td>-</td><td>3-23</td></tr>
<tr><td rowspan="4">Van den Berg et al. (2001)</td><td rowspan="4">Water quality</td><td>1-239</td><td>0-298</td></tr>
<tr><td>0-105</td><td>1-56</td></tr>
<tr><td>3-57</td><td>0-39</td></tr>
<tr><td>3-100</td><td>2-50</td></tr>
<tr><td>Shrestha and Loomis (2001)</td><td>International recreation</td><td>-</td><td>1-81</td></tr>
<tr><td rowspan="6">Chattopadhyay (2003)</td><td rowspan="6">Air quality</td><td>106-429</td><td>104-486</td></tr>
<tr><td>57-150</td><td>57-153</td></tr>
<tr><td>42-82</td><td>42-82</td></tr>
<tr><td>36-67</td><td>36-67</td></tr>
<tr><td>32-58</td><td>32-58</td></tr>
<tr><td>89-128</td><td>65-110</td></tr>
<tr><td>Ready et al. (2004)</td><td>International air and water quality</td><td>20-81</td><td>20-83</td></tr>
<tr><td rowspan="2">Jeong and Haab (2004)</td><td rowspan="2">Marine recreational fishing</td><td>-</td><td>4-230</td></tr>
<tr><td>-</td><td>2-457</td></tr>
<tr><td>Rozan (2004)</td><td>International air quality</td><td>-</td><td>19-44</td></tr>
<tr><td>Jiang et al. (2005)</td><td>Coastal land protection</td><td>-</td><td>53-85</td></tr>
<tr><td>Brouwer and Bateman (2005b)</td><td>Flood control benefits</td><td colspan="2">4-51</td></tr>
<tr><td>Eshet et al. (2007)</td><td>Environmental disamenity</td><td colspan="2">2-46</td></tr>
</table>

Source: Rosenberg and Stanley (2006) and Brouwer et al. (2009)

A1.3 Testing value transfer over time

A number of tests of the 'temporal' value transfer and its validity have been undertaken; for example Bergland et al. (1996), Brouwer and Bateman (2005b) and Zandersen et al. (2007). The longest time-span is investigated by Zandersen et al. (2007), for a 20-year temporal transfer validity tests for forest recreation in Denmark. A key finding is that updating the transfer model - with data from a targeted survey of socio-economic and use characteristics – reduces the transfer error by a factor of 4 to 25%. This highlights the significance of updating the transfer model where there have been significant changes in the determinants of WTP over time; in the Zanderson et al. case this was a significant decrease in car-based visits to forest sites.

Brouwer and Bateman (2005b) assessed the temporal stability of WTP for flood protection and conservation (from two separate CV surveys) over a 5-year period. Significant differences in WTP were found over time, although tests of function transfers indicated that simple models – specified on the basis of expectations from economic theory – were transferable over the 5-year period. This provides some evidence of stability in key determinants of WTP over relatively short time periods. Brouwer and Bateman also review previous studies (typically considering shorter periods of 1-2 years) and also find that these generally show no significant difference in real WTP values over time.

The main conclusion, as detailed by Navrud (2007) is that temporal validity tests have indicated that the real value of environmental goods can be considered to be stable if there have been no big 'shifts' in the determinants of WTP. Transfers from studies undertaken closer in time (say 5-10 years) are less likely to be subject to such shifts. Navrud also notes that applying findings from more recent studies also implies a greater likelihood that the results satisfy current best practice criteria for economic valuation methods[55]. In addition the standard practice of adjusting transferred values via a consumer price index (or GDP deflator) can be viewed as conservative in cases where the good of interest is increasing in scarcity over time and/or household's relative income increases and the income elasticity of the environmental good is high.

A1.4 Meta-analysis studies

Meta-analyses focus on relatively homogeneous environmental goods for which multiple value estimates and functions are available from a number of primary valuation studies. Essentially they provide a quantitative synthesis of existing valuation evidence, enabling investigation of the range of value estimates, producing summary statistics such as mean value, median value, confidence intervals etc., as well as identifying the key factors that influence estimated economic values via a meta-analysis function. In general, however the purpose of the meta-analysis studies is to examine empirical questions pertaining to methodological issues, rather than seeking to provide 'generally applicable' unit value estimates or valuation functions that can be used in practical value transfer exercises (see for example Hoehn, 2006).

[55] The example given by Navrud (2007) is contingent valuation studies performed before the NOAA Panel guidelines (Arrow et al. 1993), which would not meet the best practice established at that time.

Indeed, as Navrud (2007) notes, many meta-analysis studies are not particularly useful for value transfer, even within the US where most of analyses have been focused. In particular, the inclusion of methodological variables in meta-analysis functions such as 'valuation method', 'model specification', 'payment vehicle' and 'elicitation format' are not useful for predicting economic values for changes in the provision of environmental goods and services. Moreover, it is often the case that published studies provide insufficient detail on the characteristics of the study good and site, the change in quantity or quality of the good, and also socio-economic characteristics of the sample population to specify fully meta-analysis functions that could be practically applied. Navrud notes also that socio-economic characteristics are particular relevant in the case of between-country value transfer, particularly if there is cross-country heterogeneity in preferences for environmental goods.

Applications of meta-analysis are various but include examination of CV studies of both use and non-use values of water quality improvements (Magnussen 1993), CV studies of groundwater protection (Boyle et al 1994), travel cost studies of freshwater fishing (Sturtevant et al. (1995), CV studies of visibility changes at national parks (Smith and Osborne 1996), CV studies of endangered species (Loomis and White 1996), CV studies of wetlands (Brouwer et al. 1997) as well as broader valuation of wetlands (Woodward and Wui 2001; Brander et al., 2006; and Brander et al. 2008), CV studies of landscape changes (Santos 1998), CV studies of WTP for waste water treatment in coastal areas (Barton 1999), and outdoor recreation (Shrestha and Loomis 2001).

A1.5 Valuation of ecosystem services and value transfer

Use of value transfer in an ecosystem services valuation context has developed from an inauspicious start, which is that of the much rebuked application by Costanza et al. (1997). This documents an attempt to value the global contribution of 17 ecosystem services for 16 biomes, using value transfers from around 80 studies. Criticisms are numerous (see for example Toman 1998; Pearce; 1998; Bockstael *et al*. 2000, and Freeman 2002 and more generally a 1998 special issue of Ecological Economics vol. 25, no. 1), but the main theme of responses were along the lines that the exercise represented a 'serious underestimate of infinity' (Toman, 1998) in seeking to estimate the 'total' (or absolute) value of all ecosystem services; extrapolating economic value evidence out of its context of valuing marginal changes in the provision of goods and services provided by ecosystems.

As Plummer (2009) points out however, the episode demonstrates an extreme example of inadequate correspondence that can be a significant source of error in mapping ecosystem services using a single characteristic of study and policy goods and sites. A further instance cited by Plummer is that of Batker et al. (2008) which used value transfer and an ecosystem services mapping exercise to value flood and storm protection, aesthetic and recreational services, and cultural and spiritual services of beaches (among other land cover/land use types) in Puget Sound, Washington (US). Valuation evidence was applied from studies of beaches in South Carolina (Silberman et al., 1992) and New Jersey (Pompe and Rinehart, 1995). While these studies 'match' the Puget Sound context on a superficial basis – that of correspondence between land cover/land use, the exercise ignores the influence of the determinants of the value of ecosystem services provided by the areas of interest. For example the importance of beaches in South Carolina in protecting against hurricanes, tropical storms and other severe weather events (implying significant protection from damages and loss of life) and the warmer air and sea temperatures of New Jersey beaches (implying significant recreation value). Neither characteristics are evident at beaches in Puget Sound to the same degree and there are serious questions about the correspondence between sites. Plummer concludes that 'landscape scale' (e.g. a watershed) generalisation errors can be significant where there is no attempt to understand the prime determinants of economic values derived from ecosystem service mapping approaches to value transfer.

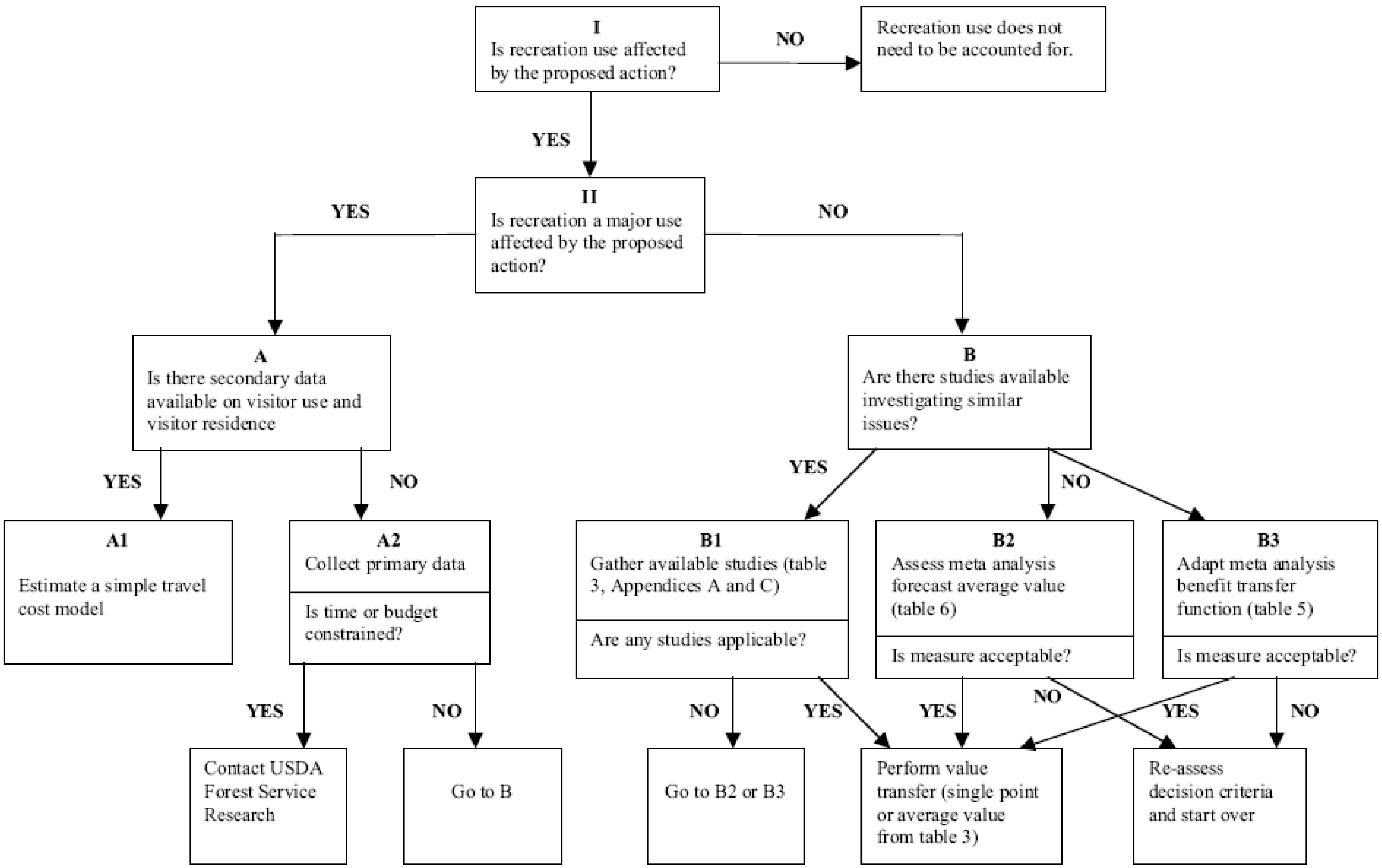
A1.6 Practical use of value transfer

Concurrent with the progressive development of methods, value transfer has been practiced and applied extensively in various contexts such as water quality management (Luken et al., 1992), associated health risks (Kask and Shogren, 1994), waste (Brisson and Pearce, 1995) and forest management (Bateman et al., 1995b). Numerous examples of value transfer in policy and project appraisal highlights that it is considered as 'routine' practice in a number of areas. A prime example is in relation to estimating recreational benefits associated with multi-use reservoir and forest sites in the US. For example Walsh et al. (1992) documented unit values for days spent participating in various recreational activities from almost 300 travel cost and contingent valuation studies available at the time[56]. This has subsequently been updated by Rosenberger and Loomis (2001) for the US Department of Agriculture (USDA) Forest Service to provide an annotated bibliography of studies from 1967 to 1998. This provides a basis of 760 benefit estimates from 163 individual studies along with practical guidelines, in the form of a decision tree, to determine how to value recreational activities (**Figure A1.1**).

More recently Navrud (2007) provides guidance for the application of value transfer in Denmark, as well as a practical example (the Skjern river restoration project), in order to establish a more consistent and reliable treatment of use and non-use values associated with environmental goods and services in economic analyses (in particular CBA). A seven step approach is recommended for unit value transfer (**Box A1.1**) with this approach viewed as the simplest and most transparent available[57]. While it is recognised that there is a limited number of primary Danish valuation studies available (although these are augmented by wider European studies and particularly those from Nordic countries), the guidance covers a range of environmental impacts, including: surface water quality, groundwater quality, marine and coastal areas, soil quality, landscape (aesthetics, cultural heritage and recreation aspects of e.g. forests and moorland), ecosystem functions and biodiversity.

[56] Walsh et al (1992) and the earlier Walsh et al. (1990) and Smith and Kaoru (1990) also represent early examples of meta-analyses of travel cost and contingent valuation studies.
[57] The steps recommended by Navrud (2007) map well to the principles examined in the Technical Report and the practical steps for value transfer (Steps 1-8) set out in the accompanying Value Transfer Guidelines document.

Figure A1.1: Framework for estimating outdoor recreation use value benefits (Rosenberger and Loomis, 2001)



Source: Figure 9: Framework for obtaining benefit measures for recreation (Rosenberger and Loomis, 2001)

Box A1.1: Practical tools for value transfer in Denmark – seven steps of the guidelines

Navrud (2007) recommends seven steps for undertaking value transfer:

1) Identify the change in the environmental good to be valued at the policy site
(i) Type of environmental good
(ii) Describe baseline, magnitude and direction of change in environmental quality

2) Identify the affected population at the policy site

3) Conduct a literature review to identify relevant primary studies

4) Assess the relevance and quality of study site values for transfer
(i) Scientific soundness; the transfer estimates are only as good as the methodology and assumptions employed in the original studies
(ii) Relevance; primary studies should be similar and applicable to the policy good context
(iii) Richness in detail; primary studies should provide a detailed dataset and accompanying information

5) Select and summarise the data available from the study site(s)

6) Transfer value estimate from study site(s) to policy site
(i) Determine the transfer unit
(ii) Determine the transfer method for spatial transfer
(iii) Determine the transfer method for temporal transfer

7) Calculate total benefits or costs
The use of these guidelines is illustrated by applying them to the Skjern River nature restoration project.

ANNEX 2: PARAMETRIC VALIDITY TESTING VIA A POOLED VALUE FUNCTION

This Annex provides the various additional materials referred to in Section 4 of this report in the following sub-sections:

- *Developing the water quality ladder;*
- *Valuation questions;*
- *Payment card;*
- *Further procedural invariance testing;*
- *Further specifications of the value transfer function, and*
- *Comparison of value transfer estimates with values predicted from policy site data*

A2.1 Developing the water quality ladder

A vital element of any stated preference valuation study is the definition of the good concerned, its status quo conditions and the change in provision which will we will ask survey respondents to value. This in turn requires an understanding of the physical science determining these states. While there are numerous pollutants that affect open access waters, the WFD focuses upon those which affect their ecological status and in particular those nutrients that are delivered to waterways via routes such as diffuse pollution from agriculture (Davies and Neal, 2004; Hutchins et al., 2006; Neal and Jarvie, 2005; Neal et al., 2005)[58]. However, individuals do not hold values for reducing pollution per se but rather for the effects that such reductions may induce in terms of recreation suitability and ecological quality. To some considerable degree the pathways linking pollution to ecological impact is still the subject of ongoing research (UKTAG, 2008). However, this does not prevent the analyst from valuing certain states of the world on the assumption that ongoing research will indicate how such states might subsequently be attained. With this in mind Hime et al., (2009) define four levels of river water quality based upon chemical, physical, flora and fauna characteristics. **Table A2.1** provides details of this characterisation exercise.

[58] In other research under the ChREAM program (Bateman et al., 2006c) we seek to extend the work of Kay et al., (2005) in examining the link between land use change and water borne faecal matter.

Table A2.1: Characterization of river water quality

Highest quality BLUE	GREEN	YELLOW	Lowest quality RED
Chemistry			
BOD Limit < 4mgl^{-1} Cat: A & B	BOD Limit >= 4mgl^{-1} and < 6mgl^{-1} Cat: C	BOD Limit >= 6 and < 8mgl^{-1} Cat: D	BOD > 8mgl^{-1} Cat: E & F
Freshwater fish directive limit game BOD Limit = 3 mgl^{-1}		Freshwater fish directive limit BOD Limit = 6	
Ammonia < 0.6 mgNl^{-1}	Ammonia < 1.3 mgNl^{-1}	Ammonia < 2.5 mgNl^{-1}	Ammonia > 2.5mgNl^{-1}
Assumed physical state			
Patches of faster flow	Lower flow rate; no fast patches	Low flow rate	Very low flow rate
Gravel / pebble substrate; No algae on rocks	Small gravel and sand substrate; little algae on rocks	Mud; algae on rocks	Mud; algae on rocks
Aquatic plants			
No algae; Water plants (described below); Good clarity	Greater amount of aquatic plants taking up more of the open space; Slight increase in water turbidity	Less aquatic plants with increases in algae; Further increase in turbidity and green hue to the water, Small number of algal mats	Large degree of siltation; Turbid water with a brown hue; Algal mat covering the substrate
Vegetation cover= 50%	Vegetation cover= 60%	Vegetation cover= 70%	Vegetation cover= 85%
Rhynchostegium riparoides (20); *Myriophyllum alterniflorum*[1] (20); *Leptodictyum (Amblystegium) fluviatile* (10); *Fontinalis antipyretica* (10) *Ran. penicillatus ssp. Pseduofluitans*[1] (4); *Pellia endiviifolia* (2); *Apium nodiflorum* (3); *Cal. hamulata*[1] (10); *Leptodictyum (Amblystegium) riparium* (3);*Rorippa nasturtium-aquaticum* (3); *Callitriche platycarpa*[1] (5); *Callitriche stagnalis*[1] (2); *Potamogeton crispus* (2); *Potamogeton natans*[2] (6)	*Apium nodiflorum* (20); *Leptodictyum (Amblystegium) riparium* (20); *Potamogeton crispus* (10) *Rhynchostegium riparoides* (15); *Myriophyllum alterniflorum*[1] (10); *Leptodictyum (Amblystegium) fluviatile* (5); *Fontinalis antipyretica* (5) *Callitriche hamulata*[1] (2); *Callitriche stagnalis*[1] (8); *Potamogeton crispus* (5)	*Apium nodiflorum* (5); *Leptodictyum (Amblystegium) riparium* (50); *Potamogeton crispus* (5) Algae *Cladopora etc. (40)*	algae *Cladopora etc.(100)*

Table A2.1: Characterization of river water quality (continued)			
Highest quality BLUE	GREEN	YELLOW	Lowest quality RED
Fish - general assessment			
Game and coarse	Same or higher coarse numbers, few game fish	Lower coarse fish, no game fish.	Very few fish
Fish - species breakdown			
Brown trout (mid) central area fastest flow	-	-	-
Minnow (high)	-	-	-
Vendace (mid)	-	-	-
Barbel (mid)	-	-	-
Chub (mid)	-	-	-
-	Bream	Bream	-
-	Common Carp (mid) mid-water	Common Carp (low) Whole area - not edges (silt)	-
-	Perch (less) mid-water	-	-
-	Roach (mid) mid-water	Roach (high) Whole area - not edges (silt)	-
-	Rudd (mid) mid-water	Rudd (low) Whole area - not edges (silt)	-
Pike (v. low)	Pike (v. low) mid-water	Pike (v. low) Whole area - not edges (silt)	-
-	-	Stickle Back (mid) edges as small fish, not where there is too much silt	-
Uses			
Game fishing	-	-	-
Coarse fishing	Coarse fishing	Restricted coarse fishing	-
Swimming	Swimming	-	-
Canoeing & boating	Canoeing & boating	Canoeing & boating	-
Bird watching	Bird watching	Bird watching	Restricted bird watching

Source: adapted from Hime et al., (2009)
Biological oxygen demand (BOD) and ammonia levels from UKTAG (2008) and EA (2007a,b). Aquatic plant frequency and species from Holmes et al., (1999) and JNCC (2005)
1 = Aquatic plant species which occur at up to 0.5m depth (EA, 2007c,d,e);
2 = Aquatic plant species which occur at 0.5 – 1.5m depth (EA, 2007c,d,e).
Numbers in parentheses to the left of plant community composition show the percentage breakdown of the total vegetation cover. Physical assessments and fish species information from EA (2007f,g).

The complexity of information given in Table A2.1 is of course far too high to reasonably allow its unadjusted use as survey information. Therefore, following discussions across the various case study partners, a set of photographs of generic water quality characteristics was agreed for each quality level. These were then passed to a graphic artist to produce the generic water quality ladder[59] shown in Section 4 (Figure 4.1). Qualitative face-to-face testing with a pilot sample confirmed that this form of information was clearly comprehended by respondents who were able to recall patterns in quality change following the interview process.

A2.2 Valuation questions

All countries adopted a common questionnaire. The UK questionnaire was computerized with a touch screen helping to identify the location of river stretches and recreational visits. The valuation section from that questionnaire is reported verbatim below.

<u>Extract from UK survey questionnaire</u>

I now want to ask a different type of question. Please look at this map; you live here [POINT TO MAP HOUSE ON MAP].

We have used the colours from the pictures to show the current water quality of rivers in this area. This is based on information from the Environment Agency, which is the official body that monitors river quality in the UK.

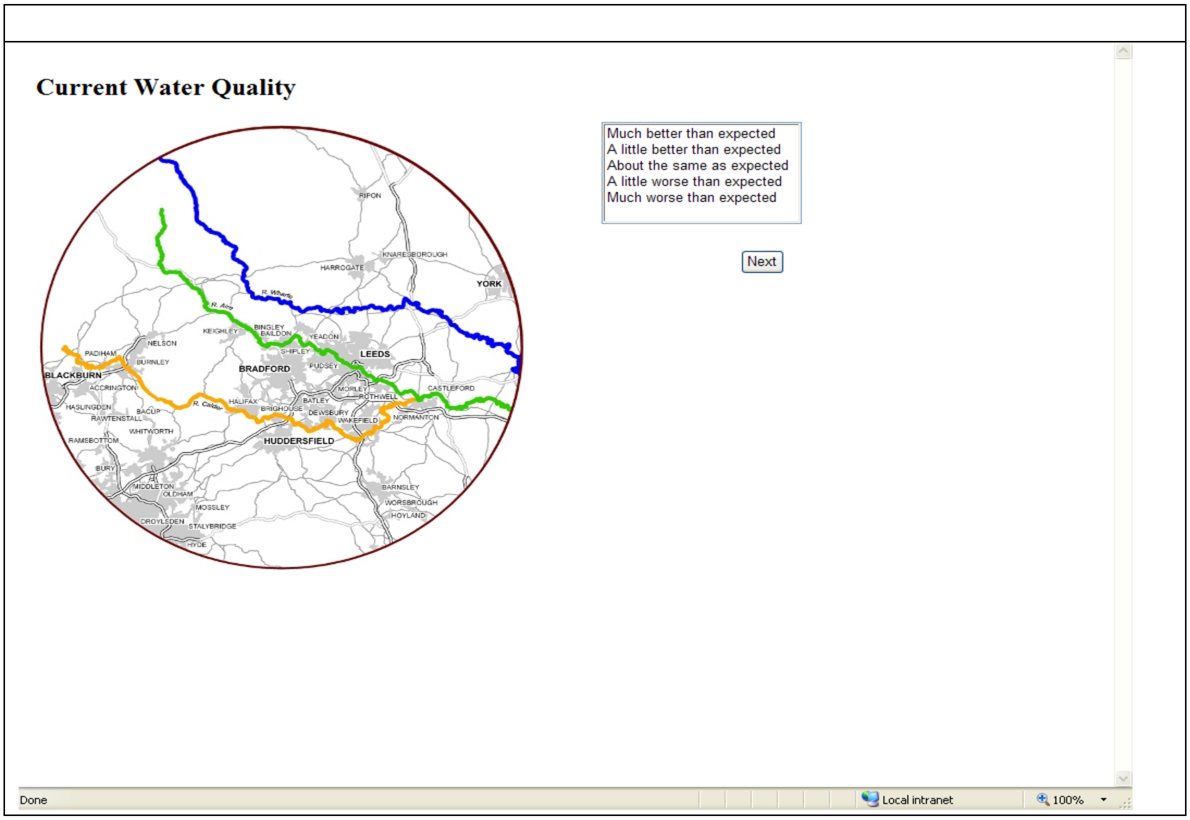
As you can see, at present the river closest to you [INDICATE RIVER CLOSEST TO RESPONDENTS HOME], is coloured [SAY COLOUR] which means that on average its water quality is like this [point to picture corresponding to colour]. This river [INDICATE 2ND CLOSEST RIVER TO RESPONDENTS HOME] is the next closest and is [STATE COLOUR OF THAT RIVER] on average its water quality is like this [POINT TO PICTURE CORRESPONDING TO COLOUR ON THE WATER QUALITY LADDER].

Finally, the furthest River [INDICATE THE RIVER 3RD CLOSEST TO RESPONDENTS HOME] is coloured [STATE COLOUR OF THAT RIVER]. So, on average, its water quality is like this [point to picture corresponding to colour].

E1. Looking at these categories [*SHOWCARD RIVERS QUALITY*], which phrase best, describes your reaction to the information concerning the general current water quality of rivers in the area?

[59] Note that this version of the water quality ladder ties together the ecological and use quality of rivers. This need not be the case as the drivers of ecological quality (mainly nutrients etc.) are, within reasonable limits, not those which determine suitability for use (E Coli levels and other faecal matter; see Kay et al., 2005). Furthermore, these various drivers need not be correlated (although in practise they frequently are). Therefore, in ongoing valuation work we break the deterministic link between ecological quality and use suitability shown in Figure 4.1. Nevertheless, our expectation is that, in terms of preferences, use and ecological utilities may well be empirically correlated (e.g. individuals dislike direct contact with water which has high algae levels even if they have low faecal matter and low health risk). Given this it may be that our estimated values are not contingent upon the ecology / use link specified in Figure 4.1.

Figure A2.1: Screen shot - Current water quality levels



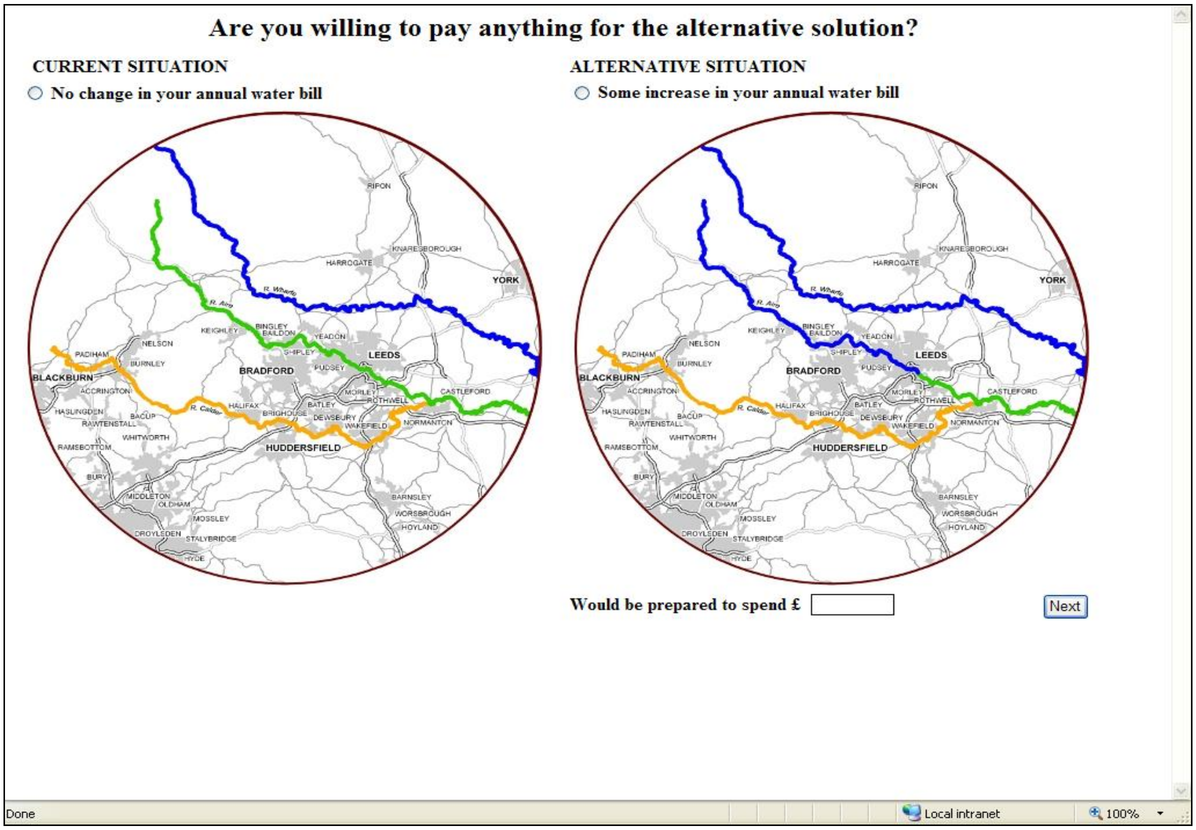
I now want to show you a second map [POINT TO “ALTERNATIVE” MAP]. This shows an alternative situation, where river water treatment works are undertaken to improve the stretch of river shown here [INDICATE CHANGED STRETCH]. Comparing the two maps you can see that in this stretch the river water quality has improved from YELLOW to BLUE.

We can see that’s a move from here [INDICATE INITIAL QUALITY] to here [INDICATE FINAL QUALITY]. All other parts of all the rivers stay as they currently are.

In a moment I will ask you a question about how much if anything your household might pay in increased water bills for this improvement. But before that please consider that any money you spend on improving river water quality obviously would not be available for spending on any other purchases. Please think about the location of the improvement, how close it is to your home, and whether you would benefit from it.

To help you work out how much, if anything, this scheme is worth to your household please consider this card. [GIVE RESPONDENT *PAYMENT CARD*]. For each amount please ask yourself whether or not your household would be prepared to pay this amount each year to get the improvement shown. Then tell me the amount which is the most your household would be prepared to pay on top of your normal yearly water bill in order to get this improvement.

Figure A2.2: Screen shot – Valuation response question using payment card elicitation method



IF PAYMENT CARD VALUE = 0

REASON A

E3. Looking at this list, please tell me the two most important reasons for your answer?

IF PAYMENT CARD VALUE > 0

REASON B

E3. Looking at this list, please tell me the two most important reasons for your answer?

Now I would like you to consider a second alternative.

Again this concerns an improvement from [INDICATE] YELLOW to BLUE quality but now for this stretch of the river.

As before tell me the amount which is the most your household would be prepared to pay on top of your normal yearly water bill in order to get this improvement

A2.3 Payment card

Given that there is a clear literature showing that changes in the elicitation method used to pose WTP questions have significant impacts upon responses (Bateman et al., 1995a; Bateman and Jones, 2003), this was standardised across all case studies using the common payment card[60] illustrated in **Figure A2.3** which was prefixed by a standard budget constraint reminder. Although presented in local currency units, when converted into Euros the payment card included the same amounts for all countries. The payment card amounts were chosen after considering the differences in purchasing power between countries and the impact upon the statistical efficiency of WTP estimates of different payment card levels.

Figure A2.3: Common payment card (converted to Euro equivalents)

☐ €0	☐ €30	☐ €65	☐ €100	☐ €135	☐ €190	☐ €350	☐ €700	☐ €1050
☐ €3	☐ €35	☐ €70	☐ €105	☐ €140	☐ €200	☐ €400	☐ €750	☐ €1100
☐ €5	☐ €40	☐ €75	☐ €110	☐ €145	☐ €225	☐ €450	☐ €800	☐ €1150
☐ €10	☐ €45	☐ €80	☐ €115	☐ €150	☐ €250	☐ €500	☐ €850	☐ €1200
☐ €15	☐ €50	☐ €85	☐ €120	☐ €160	☐ €275	☐ €550	☐ €900	☐ > €1200(specify)
☐ €20	☐ €55	☐ €90	☐ €125	☐ €170	☐ €300	☐ €600	☐ €950	☐ Other: € …..
☐ €25	☐ €60	☐ €95	☐ €130	☐ €180	☐ €325	☐ €650	☐ €1000	☐ Don't know

A follow up question sought respondents' motivations for their WTP response which also allowed assessment of any protest responses, rejecting the valuation scenario (Bateman et al., 2002a).

The UK case study additionally employed a dichotomous choice WTP elicitation approach with a separate sample of respondents.

A2.4 Further procedural invariance testing

As pointed out in Section 4, analyses of scope sensitivity are a necessary but often insufficient test for study validity. We emphasise the guidance obtained from experimental economics for such tests. Here the focus is upon whether findings pass tests of procedural invariance or exhibit anomalies. Our previous work provides a number of examples of such tests[61]. We can now formalise our procedural invariance test as follows defining:

[60] Note that the UK study also included a separate sub-sample for whom values were elicited using a single bound dichotomous choice elicitation method as recommended by Carson and Groves (2007). Results for this exercise are not presented within the present report as elicitation effects preclude ready comparison with data from the other studies considered here.

[61] See, for example: Bateman et al. (forthcoming); Bateman et al. (2009); Bateman et al. (2008a); Covey et al. (2007); Bateman et al. (2007b); Bateman and Munro (2005); Bateman et al. (2004); Bateman and Mawby (2004); Powe and Bateman (2003); Bateman and Langford (2001); Bateman et al. (2001); Bateman et al. (1997a,b); Bateman et al. (1995).

- A1 = Small quantity improvement, for which willingness-to- pay is denoted WTP (A1)

And;

- A2 = Large quantity improvement, for which willingness-to-pay is WTP (A2)

To ensure that our test is not undermined by quality differences, A^2 is defined so that it contains all of A^1 plus an additional quantity of the good (i.e. A^1 is 'nested' within A^2). Therefore in both quantity and quality terms $A^2 > A^1$. By varying the order of presentation randomly across respondents and denoting the 1st and 2nd question by subscripts, we therefore define the following four improvements over the status quo:

$$A_1^1,\ A_{2,}^1,\ A_1^2,\ A_2^2$$

and their corresponding WTP measures:

$$WTP(A_1^1),\ WTP(A_{2,}^1), WTP(\ A_1^2),\ WTP(A_2^2\ )$$

We can now define a series of both scope sensitivity and procedural invariance tests. Because of the diminishing marginal utility associated with many environmental goods the utility of A^2 is not necessarily greater than A^1 utility[62]. Therefore combining a weak scope sensitivity expectation with our procedural invariance expectation that (within an exclusive list format) WTP for a given good should not vary by order of presentation:

$$WTP(A_1^1) = WTP(A_2^1) \leq WTP(\ A_1^2) = WTP(A_2^2\ )$$

while procedural invariance with strong scope sensitivity implies that:

$$WTP(A_1^1) = WTP(A_2^1) < WTP(\ A_1^2) = WTP(A_2^2\ )$$

Furthermore we can also test for the consistency of scope sensitivity across question orders by defining:

$$A_2^2 - A_1^1 = \Delta BU$$

and

$$A_2^1 - A_1^2 = \Delta TD$$

[62] Note that Hsee (1996) shows that a further anomaly can arise when a high quality good is given a higher preference rating than the same good plus an additional inferior (but still in its right utility enhancing) good. List (2002) shows that the same result is replicated within an incentive compatible, real payment framework.

where BU denotes the bottom-up ordering (smaller good (A_1^1) valued before the larger good (A_2^2)) and TD denotes the top-down ordering (larger good (A_1^2) valued before the smaller good (A_2^1)). These provide two estimates of the magnitude of scope sensitivity and preference consistency would lead us to expect that:

$$\Delta BU = \Delta TD$$

Results: Non-parametric scope sensitivity and procedural invariance testing

Table A2.2 provides an initial inspection of the scope sensitivity and procedural invariance results by apportioning each sample to a set of mutually exclusive response types. This analysis is revealing, showing that only 26% of the sample exhibit strong scope sensitivity while 45% accord equal values to each improvement. Of course the latter result is perfectly in accord with prior expectations, suggesting satiation at the smaller improvement level (which seems plausible given the nature of the good). Equally reasonable is that a further 25% of the sample accord no value to either improvement. Given the closer proximity of substitute goods mentioned above this again seems highly plausible. This leaves some 4% of responses that strictly clash with prior expectations in that the smaller improvement is accorded a higher value than the larger good. While this is an issue to be highlighted (and is often not tested for in non-market valuation studies) the rate of apparent irrationality or misunderstanding of the scenario is consistent with findings in experimental economic tests. While some studies have omitted data from such respondents we argue that this may give a misleading indication of the consistency and validity of findings and so retain all responses within subsequent analyses.

Table A2.2: Classification of WTP response behaviour (sample percentages

	Belgium	Denmark	Lithuania	Norway	UK	*Total*
WTP (A_1^1) < WTP (A_2^2)	4	23	5	9	14	*11*
WTP (A_1^2) > WTP (A_2^1)	10	25	6	15	20	*15*
WTP (A_1^1)> WTP (A_2^2) or WTP (A_2^1)< WTP (A_1^2)	6	2	2	7	1	*4*
WTP (A^1) = WTP (A^2)	72	33	27	44	39	*45*
WTP (A^1) = WTP (A^2) = 0	8	17	60	25	25	*25*
	100	*100*	*100*	*100*	*100*	*100*

Resultant WTP levels disaggregated by both the size of improvement and the order in which valuations were sought are illustrated in **Figure A2.4**. A visual inspection suggests that, while each ordering treatment appears to yield scope sensitive results (with the larger improvement being accorded higher WTP) there appear to be considerable differences across the two treatments with greater scope sensitivity in the top-down treatment and a substantially higher WTP for the smaller improvement when presented as the first good encountered by respondents[63].

[63] This pattern accords with the previous findings of Bateman et al., (2004).

Figure A2.4: Scope sensitivity across ordering treatments

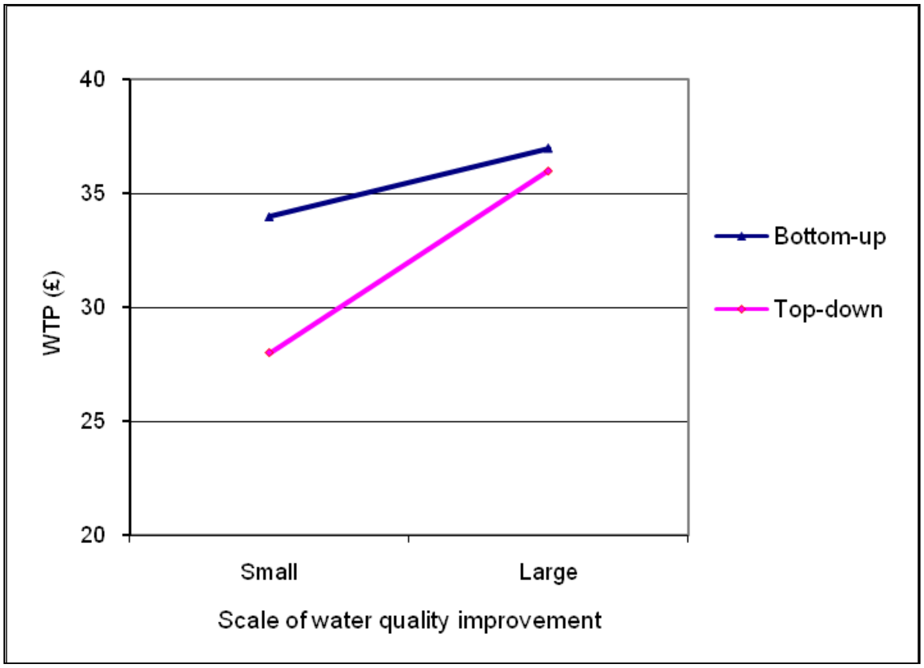


Figure A2.4 suggests that our WTP responses may reflect scope sensitivity but lack procedural invariance. To assess this, the expectations discussed in the main text were formulated into a series of testable hypotheses as set out in **Table A2.3**, which also reports results from nonparametric Kruskal-Wallis tests.

Table A2.3: Results from tests of scope effects and procedural invariance

Country	Scope effects				Procedural invariance (ordering effects)	
	All responses H_0^1: Small = Large	First WTP H_0^2: Small 1 = Large 1	Second WTP H_0^3: Small 2 = Large 2	Consistency H_0^4: $\Delta BU = \Delta TD$	Small good H_0^5: Small 1 = Small 2	Large good H_0^6: Large 1 = Large 2
Belgium	NS	NS	NS	S	S	NS
Lithuania	NS	NS	S	NS	NS	NS
Denmark	S	NS	S	NS	S	NS
Norway	S	NS	S	S	S	NS
UK	S	NS	S	S	S	NS

Notes: S = significant difference (α = 5%); NS = non-significant difference

Examining Table A2.3, H_0^1 (Small=Large) pools all WTP responses for the small improvement (irrespective of the order in which they were elicited) and compares these with pooled valuations of the large improvement. As argued previously, the only unambiguous expectation here is that WTP should not decline as the scale of the improvement rises. In the event all but two countries show significant increases in WTP as scope rises. As suggested, this is a relatively weak test but there are no anomalous reversals in scope sensitivity. This underlying weakness affects H_0^2 and H_0^3 which again report no anomalous scope reversals. However, it is interesting to note that when the first responses are tested (H_0^2: Small 1 = Large 1) none of the country studies report significant scope sensitivity[64] whereas, by contrast, when the second responses are tested (H_0^3: Small 2 = Large 2) all bar one yield significant scope effects. This indicates a worrying lack of scope consistency which is confirmed by H_0^4 (ΔBU=ΔTD) which reveals significant differences in the degree of scope across orderings in three out of five countries.

These results suggest the presence of framing effects, which are further assessed within our procedural invariance tests H_0^5 and H_0^6. Hypothesis H_0^5 tests whether values for the small improvement are robust against whether responses were elicited from either the first or second valuation question. Results show that in only one case do we fail to reject this hypothesis; clearly procedural invariance fails for these values. However, when repeating this test for valuations of the large improvement we cannot reject the hypothesis (H_0^6) of procedural invariance in any country. This pattern prompts a number of speculations regarding response behaviour, one being that respondents may be overvaluing the small good when it is the first they encounter and they are unaware that alternative goods are also available. However, other interpretations are also plausible (some of which we develop further in Bateman et al., 2004) and we merely conclude that, as per the few other studies that have carried out such tests, we have found evidence of framing effects. Accordingly we account for the presence of ordering throughout our mean and value function transfer analyses.

<u>Further specifications of the value transfer function</u>

In the main text of this report we present three value function specifications for transfer purposes:

- The reduced theory-driven model: WTP = f(provision change, distance to site and distance to substitute)
- The full theory-driven model: WTP = f(provision change, distance to site and distance to substitute and income)
- Statistically-driven model: WTP = f(provision change, distance to site, distance to substitute, income, age and urban)

[64] Note that this runs contrary to an incentive compatibility argument that first responses should give an unbiased and robust estimate of true WTP, devoid of strategic behaviour. Note however that this does not imply rejection of superficially similar argument of Carson and Groves (2007) as the latter only applies to responses to single referendum elicitation formats.

Table A2.4: Results from further value function specifications estimated using random effects Tobit panel data methods

Variable	Best single variable theory-driven model	Best fit model
	Coefficient (s.e.)	Coefficient (s.e.)
Constant	-4.63 (4.528)	-12.05 (10.289)
Large improvement	10.94 (0.989)	9.46 (1.105)
Small improvement 1st	6.050 (1.371)	4.87 (1.442)
Income (net household income in € per year)	0.0007 (0.0002)	0.0007 (0.0001)
Distance to the improvement site (in km)	-	-0.20 (0.078)
Substitute distance (distance to the nearest comparable site in km)	-	0.13 (0.038)
Age of respondent (in years)	-	-0.32 (0.102)
Urban (respondent lives in urban area=1; otherwise =0)	-	9.89 (3.346)
Users-non users (respondents visited one or more rivers or lakes in the last 12 months=1; otherwise=0)	-	15.65 (6.175)
Number of river and lakes recreation trips	-	0.037 (0.016)
Norway	11.51 (4.162)	18.69 (4.626)
UK	-15.70 (5.050)	-12.35 (5.47)
Belgium	22.18 (4.695)	31.94 (5.332)
Lithuania	-53.36 (5.447)	-45.79 (5.853)
Sigma μ	73.62	73.68
Sigma ε	23.31	23.31
Rho	0.91	0.91
No. of observations	5769	5268
Number of censored observations	1561	1430
Log-Likelihood	-23236	-21186
Wald chi^2(*K*=restriction for overall significance)	498 (7)	578 (13)
p-value	0.0000	0.000

Notes: (1) The country dummy for Denmark is omitted making this the baseline from which country departures are to be interpreted; (2) The good/order dummy Small improvement 2nd is omitted making this the baseline from which the Large improvement 1st, Large improvement 2nd and Small improvement 1st departures are to be interpreted; (3) Rho = var_μ/(var_μ+var_ε) and represents the percent contribution to total variance of the panel-level variance component.

For completeness, **Table A2.4** presents two further models:

- Best single variable theory-driven model: WTP = f(income). This is of interest because income is the strongest determinant of dissimilarity amongst our set of study countries. Further testing confirms that this provides the lowest function transfer error rate of any single variable model.
- The best-fit model: WTP = f(provision change, distance to site, distance to substitute, income, age, urban, user and visit frequency). Although this yields the highest level of explanation of study site data it was not selected for discussion in the main report as the variables 'user' and 'visit frequency' are not available from secondary sources but would have to be elicited through a survey of the policy site, thus defeating the objective of value transfers which is to avoid such additional survey work.

Comparison of value transfer estimates with values predicted from policy site data

In the main text we follow the approach of Van den Berg et al., (2001) and Brouwer and Bateman (2005a) by comparing the value estimated by function transfer with the simple mean WTP of the policy site. However, some other analysts compare the transferred value with one estimated on data from the policy site alone (e.g. Barton 2002; Chattopadhyay 2003). Here we assess the transfer error by first defining the transferred value as $\hat{WTP}_{k|s}$ for policy site *k* estimated from study sites *s* and the directly predicted policy site value as $\hat{WTP}_{k}$. The transfer error is then calculated as $\left(\frac{\hat{WTP}_{UK|s} - \hat{WTP}_{UK}}{\hat{WTP}_{UK}}\right)\%$.

Table A2.5 contrasts the simple mean WTP values with the directly predicted policy site values $\hat{WTP}_{k}$.

Table A52.: Policy site sample mean and directly predicted WTP values					
	Lithuania	Belgium	Denmark	Norway	UK
	Mean WTP				
Small 1st	6	43	21	38	16
Small 2nd	6	50	29	45	22
Large	8	48	36	47	26
	Predicted WTP				
	Reduced theory-driven model (Provision change, distance to site and distance to substitute)				
Small 1st	9	52	30	51	20
Small 2nd	10	56	30	55	23
Large	11	56	39	56	26
	Full theory-driven model (Provision change, distance to site and distance to substitute and income)				
Small 1st	11	57	33	54	24
Small 2nd	13	62	34	58	27
Large	14	62	42	59	30
	Statistically-driven model (Provision change, distance to site, distance to substitute, income, age and urban)				
Small 1st	8	53	28	51	20
Small 2nd	10	58	29	54	23
Large	11	58	37	56	26

Table A2.6 details for the full dataset (including the dissimilar Lithuanian site) function transfer errors calculated by comparing values estimated from study sites with the value predicted from the same functional form applied to the policy site data. Tobit adjustments for censoring are incorporated within these estimates. Comparison with the value function transfer errors compared to simple mean values (shown in the main text) shows that a similar pattern of errors across different functional forms.

Table A6: Transfer errors for value functions compared with policy site predicted values - data including dissimilar (Lithuanian) site

	Lithuania	Belgium	Denmark	Norway	UK	Average
Reduced theory-driven model (Provision change, distance to site and distance to substitute)						
Small 1st	57	93	66	93	84	79 (74)
Small 2nd	80	98	77	97	96	90 (84)
Big	32	88	69	88	77	71 (68)
Average (Weighted)	56 (50)	93 (92)	71 (70)	92 (91)	86 (84)	80 (75)
Full theory-driven model (Provision change, distance to site and distance to substitute and income)						
Small 1st	68	79	70	84	59	71 (73)
Small 2nd	89	82	81	88	54	80 (81)
Big	49	74	73	80	51	66 (68)
Average (Weighted)	69 (64)	78 (77)	75 (74)	84 (83)	55 (54)	72 (74)
Statistically-driven model (Provision change, distance to site, distance to substitute, income, age and urban)						
Small 1st	61	84	93	100	73	82 (85)
Small 2nd	83	87	102	105	82	92 (94)
Big	34	79	86	95	64	72 (76)
Average (Weighted)	60 (53)	83 (82)	94 (92)	100 (99)	73 (71)	82 (85)

Table A2.7 details for the reduced dataset of similar countries (excluding the dissimilar Lithuanian site) function transfer errors calculated by comparing values estimated from study sites with the value predicted from the same functional form applied to the policy site data. Tobit adjustments for censoring are incorporated within these estimates. Again comparison with the value function transfer errors compared to simple mean values (shown in the main text) shows that a similar pattern of errors across different functional forms.

Table A7: Transfer errors for value functions compared with policy site predicted values - data excluding dissimilar (Lithuanian) site

	Belgium	Denmark	Norway	UK	Average
Reduced theory-driven model (Provision change, distance to site and distance to substitute)					
Small [1st]	93	86	91	83	88 (89)
Small[2nd]	98	101	95	97	98 (98)
Big	87	87	86	77	84 (85)
Average (Weighted)	93 (91)	91 (90)	91 (90)	86 (84)	90 (90)
Full theory-driven model (Provision change, distance to site and distance to substitute and income)					
Small [1st]	83	70	85	62	75 (77)
Small[2nd]	86	82	89	69	82 (83)
Big	77	73	80	57	72 (74)
Average (Weighted)	82(81)	75(74)	85(83)	63(61)	76 (78)
Statistically-driven model (Provision change, distance to site, distance to substitute, income, age and urban)					
Small [1st]	78	67	93	54	72 (81)
Small[2nd]	81	75	98	49	77 (76)
Big	71	64	87	41	66 (70)
Average (Weighted)	77 (75)	69 (68)	93 (91)	48 (46)	72 (76)