Open access • Journal Article • DOI:10.1007/S10640-011-9476-8

Making Benefit Transfers Work: Deriving and Testing Principles for Value Transfers for Similar and Dissimilar Sites Using a Case Study of the Non-Market Benefits of Water Quality Improvements Across Europe — Source link

Ian J. Bateman, Roy Brouwer, Silvia Ferrini, Silvia Ferrini ...+10 more authors

Institutions: University of East Anglia, VU University Amsterdam, University of Siena, University of Copenhagen ...+3 more institutions

Published on: 03 May 2011 - Environmental and Resource Economics (Springer Netherlands)

Topics: Valuation (finance)

Making benefit transfers work: Deriving and testing principles for value transfers for similar and dissimilar sites using a case study of the non-market benefits of water quality improvements across Europe

Bateman, I.J.; Brouwer, R.; Ferrini, S.; Schaafsma, M.; Barton, D.N.; Dubgaard, A.; Hasler, B.; Hime, S.; Liekens, I.; Navrud, S.

published in
Environmental and Resource Economics
2011

DOI (link to publisher)
10.1007/s10640-011-9476-8

document version
Publisher's PDF, also known as Version of record

Environ Resource Econ (2011) 50:365–387
DOI 10.1007/s10640-011-9476-8

Making Benefit Transfers Work: Deriving and Testing Principles for Value Transfers for Similar and Dissimilar Sites Using a Case Study of the Non-Market Benefits of Water Quality Improvements Across Europe

I. J. Bateman · R. Brouwer · S. Ferrini · M. Schaafsma · D. N. Barton · A. Dubgaard · B. Hasler · S. Hime · I. Liekens · S. Navrud · L. De Nocker · R. Ščeponavičiūtė · D. Semėnienė

Accepted: 1 April 2011 / Published online: 3 May 2011

Abstract We implement a controlled, multi-site experiment to develop and test guidance principles for benefits transfers. These argue that when transferring across relatively similar sites, simple mean value transfers are to be preferred but that when sites are relatively dissimilar then value function transfers will yield lower errors. The paper also provides guidance on the appropriate specification of transferable value functions arguing that these should

I. J. Bateman (✉) · S. Ferrini · M. Schaafsma · S. Hime
Centre for Social and Economic Research on the Global Environment (CSERGE), School of Environmental Sciences, University of East Anglia, Norwich, UK
e-mail: i.bateman@uea.ac.uk

S. Ferrini
Dipartimento di Economia Politica, Università di Siena, Siena, Italy

R. Brouwer
Department of Environmental Economics, Institute for Environmental Studies, VU University, Amsterdam, The Netherlands

D. N. Barton
Norwegian Institute for Nature Research, Oslo, Norway

A. Dubgaard
Institute of Food and Resource Economics, University of Copenhagen, Copenhagen, Denmark

B. Hasler
National Environmental Research Institute (NERI), Aarhus University, Aarhus, Denmark

I. Liekens · L. De Nocker
Flemish Institute for Technological Research, Boeretang, Belgium

S. Navrud
Department of Economics and Resource Management, Norwegian University of Life Sciences, Aas, Norway

R. Ščeponavičiūtė · D. Semėnienė
Center for Environmental Policy, Vilnius, Lithuania

be developed from theoretical rather than ad-hoc statistical approaches. These principles are tested via a common format valuation study of water quality improvements across five countries. While this provides an idealised tested, results support the above principles and suggest directions for future transfer studies.

Keywords Non-market valuation · Stated preference · Benefit transfers · Transfer errors · Methodology · Water quality

JEL Classification Q51 · Q15 · Q26 · Q24 · Q28

1 Introduction

Decision making is an essential yet costly undertaking and resource constraints inevitably mean 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 associated with 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 has placed great emphasis upon the development and testing of value transfer methods (e.g., Desvousges et al. 1992; Bergland et al. 1995; Brouwer and Spaninks 1999; Zandersen et al. 2007; Johnston and Duke 2009).[1] These methods can be broadly categorised into two types (Navrud and Ready 2007a). The simplest approach is to transfer 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 value functions from study site data. These are then used to predict new values for policy sites. 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 those at policy sites. However, 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.

One of the major objectives of this paper is to develop and test simple rules for conducting benefit transfers based upon the hypothesis that univariate approaches will be more appropriate for transferring between relatively similar sites, while transferable value functions will yield lower errors for transfers between less similar sites. The intuition behind such a hypothesis is also straightforward with the same driver responsible for these different outcomes. Value functions explicitly incorporate differences between sites. Where these differences are

[1] 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).

relatively large a *well specified* value function will reflect and incorporate this heterogeneity, so providing a better estimate of the value of a policy site than is afforded by a simple univariate mean transfer. However, when transferring between similar sites the incorporation of differences inherent in the value function approach may generate higher degrees of error than the simple transfer of mean values between those relatively homogeneous sites.

An immediate question which such a hypothesis generates is how an analyst would determine (a-priori and without survey evidence from policy sites) whether sites where similar or not. We assess this through an examination of secondary source data regarding the characteristics of sites and their surrounding populations, examining whether such information is sufficient to determine similarity and hence the appropriate choice of transfer method.

In essence our study tries to make sense of the conflicting evidence available in the benefit 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; Brouwer and Bateman 2005). However, the paper has further related objectives. We argue that an additional reason for these diverse results is misspecification of empirical value functions for transfer purposes. We do not claim that the value functions used previously are statistically erroneous, but rather argue that what is appropriate for statistical specification (maximisation of explained variance) may be inappropriate for function transfer purposes (minimisation of transfer error). We contend that many value function transfer exercises have failed because they have employed ad-hoc, empirically driven specification of utility functions which fit the data of study sites well but appear over-parameterised when applied out of sample to policy sites. As an alternative we suggest that value functions estimated for transfer purposes should instead draw upon the common drivers of preferences reflected in economic theory as such functions should avoid the problems of over-parameterisation by containing only those variables which are applicable to all sites.

Theory suggests that the utility of improvements to a spatially confined, environmental resource should be determined by the change in provision, characteristics of the site (e.g. its distance from the home of the valuing individual), the availability of substitutes (e.g., again measured as the distance from the individual's household to substitute sites) and characteristics of the valuing individual common to all utility functions (e.g., the individual's income). Of course the value of improvements at a given study site may also be influenced by context specific factors unique to (or of particular relevance to) that site. While the inclusion of such contextual variables may improve the degree to which a value function explains values at the study site, by including such factors within a transferred function the analyst implicitly assumes that the relevance of that variable holds for the policy site. To the extent that this is not the case, so the value function transfer will generate error when applied out of sample to policy sites. Because of the multiplicative nature of a value function it seems clear that such error has the potential to be substantial. We test the contention that value functions specified to only include those variables which economic theory expects to be common to all contexts will generate lower transfer errors than functions which include ad-hoc (non-theoretic) variables, even when the latter functions provide higher degrees of statistical fit at the study site.

The paper also offers a number of further, more minor, contributions.[2] First, as part of the specification of economic-theoretic value functions we utilise geographic information systems (GIS) to assess the impact of the location of both policy and substitute sites upon values. The measures obtained reveal decay in values as the distance between the policy site (where an improvement is planned) and the survey respondent's home increases. In contrast

[2] As discussed in greater detail subsequently in Bateman et al. (2009c) we also consider the incorporation of ordering effects within benefit transfer studies.

values increase when the distance to substitutes increases. Both relationships conform directly with prior-theoretic expectations and have long been recognised as a feature of valuations of environmental goods (Sutherland and Walsh 1985).[3] Second, we incorporate variation in the scope of the good within our transfer analyses; a facet which should enhance the applicability of derived values to differing policy situations. A final additional contribution is the introduction of a novel tool for conveying changes in the quality of the environmental good under consideration (water quality). Building upon ecological assessments of the impacts of diffuse water pollution upon aquatic life and general ecosystem quality, we develop a new water quality ladder incorporating information pertinent to both use and non-use values and presented in a readily comprehended visual form.

The remainder of this paper is organized as follows. The next section provides further details regarding the motivation for our focus upon site similarity when choosing transfer methodology. This section also clarifies the key principles and expectations which theory provides for undertaking and assessing valid and robust value function transfers. Section 3 discusses our case study test of the arguments summarised above. We conduct a set of new studies, deliberately designed to be practically identical in terms of the good under evaluation so as to provide a clean test of the issues under consideration. We recognise that this abstracts from the commonplace real world situation where decision makers are faced with a somewhat haphazard array of prior studies to draw upon for transfer purposes. While some might see this abstraction as a weakness, we contend that such standardisation is no more than is typical of much experimental design work which permits the analyst to examine the impact of adjusting certain methodological parameters. Our application concerns a common design, contingent valuation (CV) assessment of willingness to pay (WTP) for improvements in the quality of open waters in Europe; a topic chosen because of its policy relevance given the ongoing implementation of the EU Water Framework Directive (WFD; European Parliament 2000) which requires improvements in European rivers. Section 3 also introduces the case study areas which are distributed across five European countries chosen to include both relatively similar and dissimilar sites. Test procedures are specified to examine our central arguments. As part of this discussion we also consider the importance of allowing for distance decay, substitution and scope effects within valuation studies and overview our new water quality ladder for conveying information on water quality to survey respondents. Section 4 reports results. This opens with a consideration of measures of case study similarity and reports findings supporting the central contention that, when sites are relatively similar, simple mean value transfers minimise transfer errors but that value function methods are required when transferring across dissimilar sites. Results also show that, in the latter case, value function transfer errors are lower when those functions are specified to only include those generic drivers of utility highlighted by economic theory rather than transferring ad-hoc and possibly over-parameterised, statistical best-fit functions. Section 5 concludes with a discussion of the general implications arising from this research.

2 Expectations Based Principles for Benefit Transfers

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

[3] We are grateful to an anonymous reviewer for highlighting that in some cases, for instance when a proposed scheme implies a loss of local jobs, increasing distance may not have a negative effect on WTP for the environmental good under valuation. Indeed distance relationships may be positive, negative or insignificant depending upon the precise good and context under consideration.

simple mean value transfers. In essence the present paper adds certain qualifiers to the Pearce et al. argument. In particular this may not hold for transfers across relatively similar sites. Mean values smooth out the variation which inevitably arises when a sample survey is undertaken. By comparison, value functions give a greater reflection of the variability of a sample. Standard fit-maximising statistical procedures will seek to capture all of the variation in a sample, even when some of this is due to relatively small sub-samples. While this is optimal within sample, when extrapolated out of sample to estimate for sites which are apparently similar, such functions may appear relatively over-parameterised and generate larger errors than would arise through simple mean value transfers.

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 are likely to be so gross that simple mean transfers will yield substantial errors. In contrast, now the value function approach has the ability to adjust for the gross differences between dissimilar sites; be they physical or socioeconomic/demographic in nature. Simple mean value transfers will be unable to make such adjustments here and in principle are liable to yield larger errors than value function approaches. However, here we encounter a second caveat to the Pearce et al. argument; that value functions need to be specified for the purposes of transferral between study and policy sites rather than for obtaining statistical best-fit at study sites alone.

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 we begin to see a failure of the assumption that parameter estimates from study sites will hold for policy sites. Due to the multiplicative role of coefficients, such failure can result in major transfer errors. Our contention is that statistical best-fit guidelines for function specification have to be abandoned when the purpose of that model building is for transfers.[4] Instead the focus should be restricted to generic factors. We argue that 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? From basic microeconomics principles we can identify a number of factors which should influence preferences and WTP. Within the context of the 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 an open-access good this mainly relates to the proximity of the good to the respondent's home); the availability of substitutes (again a spatial relationship) and the individual's income constraints. All of these variables can be assessed from secondary data (e.g. digital map data for the accessibility of sites and their substitutes, census data for incomes, etc.) for both survey and policy sites.[5] These factors provide a rich source of theoretically consistent variables

[4] This argument also frequently arises in the econometric literature when the issue is to define the function to use for prediction. Whereas an over-parameterised model performs better in-sample, it is likely to produce poor predictive values out-of-sample. For out-of-sample prediction, a simpler model may hence be preferable. Similarly, in benefit transfer exercises the main question is how to specify a 'simpler' model that will lead to minimal transfer errors.

[5] Other factors such as usage are also important but are not available from secondary data and therefore cannot be directly applied to the prediction of values at policy sites irrespective of the method chosen. Note, however, that usage is well proxied by accessibility and is therefore effectively controlled for within our subsequent function transfer analysis.

for inclusion in our theory driven value functions and we consider each in turn within the remainder of this section. Furthermore, their inclusion within our value functions means that we have a number of clear expectations against which we can validate our findings.

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 reasonable degree of increase in WTP. 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),[6] 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. Therefore, while we design our case study to allow us to include changes in provision within the transferable value function, we do not consider a finding of significant scope sensitivity either a necessary or sufficient test of study validity.

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 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. We therefore include such variables within our transferable value function.

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 respondents and generates variables which are not available to the policy maker wanting to estimate values for unsurveyed sites. Therefore we follow the approach of Jones et al. (2002, 2010) in applying a GIS to generally available, comprehensive coverage, map data to calculate distances to multiple potential substitutes. Jones et al. (2010) allow the data to determine which are the significant substitutes which is an appealing strategy. However, findings show, not surprisingly, that the nearest similar substitute exerts particular influence over visitation behaviour and values. Given this result we adopt a simple approach by including the distance to the nearest substitute within our transferable value function.

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

[6] In fact, many alternative explanations might justify statistically insignificant scope effects (see, for example, Rollins and Lyke 1998; Powe and Bateman 2004; Heberlein et al. 2005; Carson and Groves 2007; Amiran and Hagen 2010).

constraints it may impose on WTP. We might expect, *ceteris paribus*, that those with higher incomes will have higher stated 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 are included in our transferable value functions.

The central argument of our paper is that a value function, specified to only include generic, theory derived variables, will yield lower transfer errors than a function which improves its statistical fit to the survey site data by including ad-hoc, potentially context specific variables. To provide a Popperian falsifiable test of this hypothesis we also examine the influence upon valuation transfers that inclusion of variables not suggested by economic expectations may have. Clearly we could identify extremely ad-hoc variables that could not reasonably apply to more than one or two study sites. However, such an artificial test would not validate our argument. Instead we include a small number of variables, also available from secondary source of data, which have appeared as significant predictors of values within the literature. If our central hypothesis holds then the inclusion of such ad-hoc variables should increase transfer errors over those associated with functions specified purely upon economic expectations.

3 Developing and Implementing the Common Design

The study design followed a set of valuation design principles set out in Bateman et al. (2002a). Initial concerns for study design were to identify a public good and case study 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 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. To test out our central hypotheses regarding the importance of site similarity in selecting the most appropriate method for transfers, we sought to select case study sites from both similar and dissimilar contexts. With this objective in mind, research collaboration was agreed between 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 our methodology.

While we deliberately sought variation in the study site contexts, in order to ensure good experimental control over our hypothesis tests, 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; Johnston et al 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 WFD 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.[7] 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 SP valuation 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 (Hutchins et al. 2006). To some considerable degree the pathways linking pollution to ecological impact are 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.

It is well known that 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. 2005; Fagerlin et al. 2005; Zikmund-Fisher et al. 2005; Bateman et al. 2009a) 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)]. 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 Fig. 1.[8] 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

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

[8] See the online version of this paper for a colour reproduction of this Figure or download Hime et al. (2009) at http://www.uea.ac.uk/env/cserge/pub/wp/edm/edm_2009_01.pdf

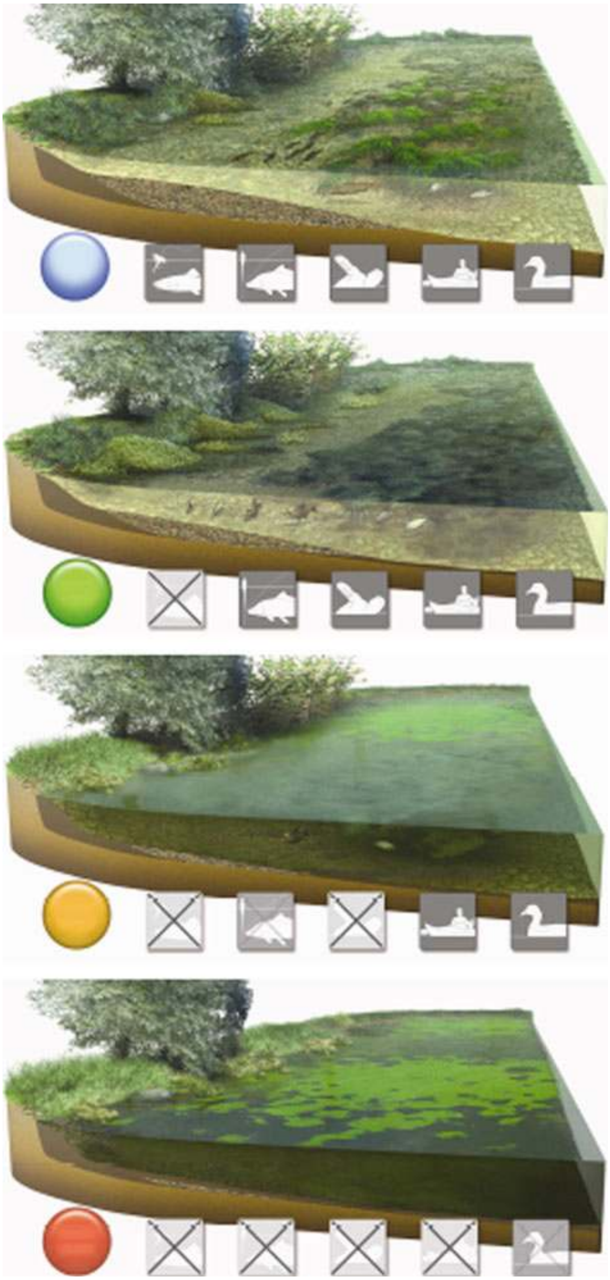

Fig. 1 The generic water quality ladder Source: adapted from Hime et al. (2009): Copyright protected

ladder while the quality of improved stretches was specified as attaining the blue level. 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

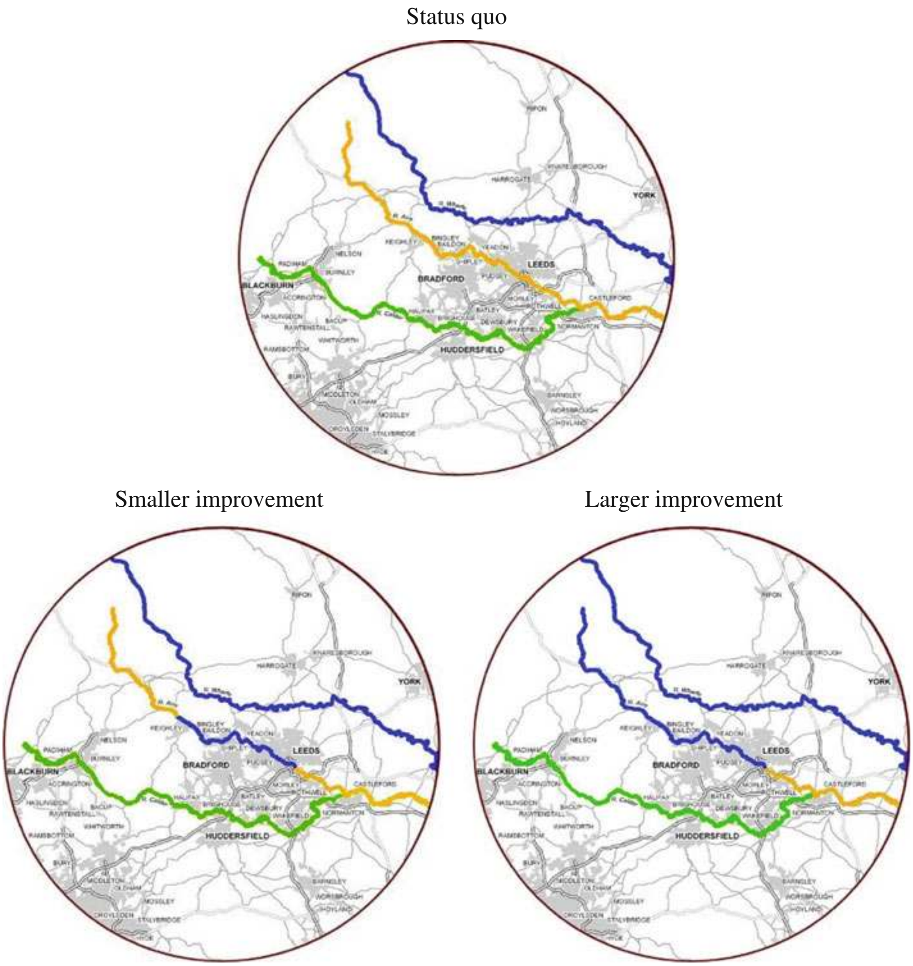

Fig. 2 Maps from the UK case study depicting status quo provision of river water quality and smaller and larger improvements

(which we term the Small improvement).[9] 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 Fig. 2 illustrating maps from the UK case study.

[9] We fully recognise that this replication of a single design across all of the study sites provides a far greater degree of standardisation than would commonly be the case for most practical benefit transfer exercises where the analyst typically attempts to derive values for a policy site on the basis of prior studies of a set of somewhat heterogeneous study sites. Indeed, such standardisation is deliberately similar to experimental control methods. Nevertheless, a common design approach is ideal for isolating focal effects such as those considered here and we see no obvious reason why the findings taken from our study should not assist in the conduct of real world benefit transfer exercises.

Testing for scope sensitivity and diminishing marginal WTP involves examination of valuation changes between the Small and Large improvement. 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 improvement. In order to control for possible ordering effects, a split sample approach is adopted where some individuals are randomly allocated to initially face the valuation question concerning the Small improvement while the reminder first value the Large improvement. Using an 'exclusive list' question format (Bateman et al. 2004), immediately after providing their first WTP answer respondents are asked to state a second WTP answer for a change compared to the same baseline scenario as for the first WTP-question. By ensuring both valuations are made from the common status quo level then, in theory, we avoid the sequencing problems highlighted by Carson et al. (1998) of the non-comparability of values for a given good made from different baselines.[10]

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 between countries and the impact upon the statistical efficiency of WTP estimates of different payment card levels (see Bateman et al. 2009c for details including the full text of the valuation question).[11] The WTP question was prefixed by a standard budget constraint reminder.

The common questionnaire also contained uniform questions regarding respondents' household income for incorporation within our transferable value function. Given that we wish to test a theoretically specified function against one derived from statistical principles, the questionnaire also included a variety of other socio-economic and demographic characteristics such as respondent age [12] 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

[10] Of course this assumes that respondents genuinely do readjust their baseline back to the status quo for the second valuation question (effectively ignoring the first question). The theoretical exposition of Carson and Groves (2007) as well as our own empirical work (Bateman et al. 2009b) suggests that this may not necessarily occur as respondents may infer information from the initial question and respond strategically thereafter. In further analysis (detailed in Bateman et al. 2009c) we show that this is indeed the case for our application. However, given that such effects are consistent with the expectations of Carson and Groves and were found not to have any material impact upon our central results, we omit them from the present paper which focuses upon the value transfer issue.

[11] 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 the Appendix to this paper.

[12] Additional questions concerned usage of water and other outdoor recreation resources, respondents' motivations for their WTP response, etc. Only minor variations of procedure were allowed for across studies. For example, in Lithuania, respondents faced prior valuation questions regarding changes in the hydromorphology of the case study (Neris) river. In the UK all respondents undertook a series of questions on the various sites they had visited in the past twelve months (for subsequent revealed preference analysis). Arguably these may have impacts upon the valuation responses reported in the present paper.

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.

A sampling strategy was developed to ensure that the data contained a high degree of variation in terms of the distance to the improvement site and to substitutes. In essence this strategy considered a regular grid of potential interview locations around the study site, assessing each location in terms of its distance (to site and substitutes), socioeconomic and 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 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.[13]

4 Results

4.1 Descriptive Statistics and the (Dis)Similarity of Study Sites

All surveys were undertaken during 2008 and a total of 3,589 questionnaires were completed across our five study countries.[14] Response options and data coding was common across studies and data pooled into a single analysis. Following guidelines for international value transfers (Navrud and Ready 2007b), WTP responses (and income data) were corrected for differences in purchasing power parity (PPP) between countries using indices from the World Development Indicators (World Bank 2008) and then converted into 2008 Euros. Table 1 presents descriptive statistics of each sample together with WTP sums disaggregated by size of provision change.

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 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[15] (for the physical characteristic descriptors) to allow assessments of similarity for

[13] Surveys were conducted online in Belgium, Denmark and Norway, and face-to-face in Lithuania and the UK with response rates ranging from 12% in the online Belgian survey up to 55% in the Lithuanian face-to-face survey. Differences in response rate by survey mode have been noted by Marta-Pedroso et al. (2007) although they also suggest that differences in values are not significant. Impacts arising from response rate differences should be accounted for through the country specific dummy variables included within our regression analyses. Care was taken to ensure that all studies sampled respondents from both areas near to and distant from the resources being valued so as to capture the shape of any distance-decay relationships. Building upon empirical findings from a similar water valuation context (Bateman et al. 2006), a simple rule of thumb of a 30 km minimum radius of sampling area was applied in all countries. This should both capture distance decay and avoid the problem of declining credibility of the contingent market if very distant groups are interviewed.

[14] 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 Bateman et al. (2009c).

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

Table 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;	9531	40877	34854	24884	26686	28310
€ PPP (s.d. in brackets)	(7823)	(19002)	(17708)	(11452)	(16709)	(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)						
Average WTP—small	6	47	25	42	19	31
	(23)	(66)	(38)	(82)	(29)	(61)
Average WTP—large	8	48	36	47	26	37
	(38)	(70)	(52)	(86)	(35)	(66)
Protest bids (% of country sample)	8	5	2	12	2	7

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

unsurveyed policy sites (although given the representativeness of our samples we report their values for convenience). The first two rows of this section consider distance from respondents' home address to the improvement site and their nearest substitute site. Neither of these statistics suggest 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 mean 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 now expect that mean value transfers would outperform value function transfers for the four remaining similar countries. Subsequently we formally test both of these hypotheses.

The remaining rows of this section of Table 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 suggest 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 1 overviews our WTP valuation results. Here the last 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.[16]

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 benefit transfer we would not have a-priori values for the policy sites. What seems apparent 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 mean value transfers can or cannot be relied upon. Subsequently we formally test these inferences.

A further point to note in Table 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 note double the latter. As discussed earlier in this paper, this is not of itself an anomalous result as it is perfectly feasible that respondents may have a diminishing marginal WTP for additional improvements once an initial length of high quality river has been provided.

4.2 Specification and Estimation of Value Functions

In order to compare simple mean 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 with a 'statistically-driven' model which supplements the former with ad-hoc variables, generally available from secondary data, 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.

If the statistically driven model is over-parameterised for use out of sample upon unsurveyed policy sites then it should yield higher transfer errors than the full theory driven model when applied to our full set of similar and dissimilar sites. However, both may generate higher errors than a simple mean value transfer when applied to a reduced set of similar sites. To examine this, in the next subsection we transfer these various functions, calculate transfer errors and contrast these with those arising from simple mean value transfers.

[16] We retain the 4% of WTP responses which 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. 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. 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).

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 models and report non-parametric results elsewhere (Bateman et al. 2009c).

As our data contains both non-zero and valid zero bids, we have a WTP distribution which is censored. Therefore, we specify a panel Log Normal Tobit model which allows both censoring for non-negative dependent variables and the fact that each respondent provides us with two WTP answers. The structural equation for such a random effects Log-Normal Tobit model is:

$$y_{it}^{*} = \exp\left(X_{ikt}\boldsymbol{\beta}_{k} + w_{it}\right) \tag{1}$$

where y^{*} is a latent variable observed for values greater than some threshold (typically zero) and censored otherwise, k indexes the number of independent variables included in the model, i indexes individual respondents and t indexes their repeated responses. The random disturbances can be disentangled as $w_{it} = \mu_i + \varepsilon_{it}$ where $\varepsilon_{it}|X_{ikt}, \mu_i$ is i.i.d $N(0, \sigma_{\varepsilon}^{2})$ and the random term μ_i is normally distributed and assumed independent from X_{ikt}. 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 $\log(y_{it}^{*}) \geq \gamma$ and zero when $\log(y_{it}^{*}) < \gamma$, where γ is a non-stochastic constant. In many empirical applications, the censoring threshold is assumed to be zero (censoring all negative values). However, Carson and Sun (2007) argue that a positive threshold may be relevant if respondents hold a non-zero expected cost and that this would lead to bias if a zero threshold were imposed (see also Sigelman and Zeng 1999). Accordingly we tested both zero and positive thresholds, finding that, at least in our application, the latter makes little impact upon estimated models and transfer errors values.[17] Given this, we focus upon more conventional zero threshold Log-Normal Tobit models obtained by specifying the dependent as $\log(y_{it} + 1)$.[18]

The specification of Eq. (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.[19] Table 2 reports the resulting models.[20]

The model in Table 2 is globally significant and all individual parameters' p-values are 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-

[17] Our analysis examined the effects of using various values from the lower end of the positive WTP distribution as a truncation point; these values reflecting possible expected cost responses. Results from these positive thresholds are available on request from the authors but did not radically alter value estimates (although obviously truncating progressively larger portions of responses would eventually impact significantly upon parameter estimates).

[18] As pointed out by an anonymous reviewer, adding 1 to the y value, makes this a "shifted" or "translated" Log-Normal model. The error structure of such a model is justified by the number of respondents having trivial WTP.

[19] This model was estimated using the STATA10 package.

[20] Results for further specifications are reported in Bateman et al. (2009c).

Table 2 Results from different model specifications using random effects Log Normal Tobit panel data models

Variable	All theory driven variables except income Coefficient (s.e.) [*z*-value]	Full theory-driven model Coefficient (s.e.) [*z*-value]	Statistically-driven model (including ad-hoc empirical regularities) Coefficient (s.e.) [*z*-value]
Large improvement	0.203 (0.018) [11.37]	0.201 (0.018) [11.33]	0.203 (0.018) [11.41]
Distance to the improvement site (km)	−0.008 (0.002) [−4.13]	−0.009 (0.002) [−4.33]	−0.008 (0.002) [−3.89]
Distance to nearest substitute site (km)	0.005 (0.0009) [5.56]	0.005 (0.0009) [5.59]	0.005 (0.0008) [5.57]
Income (net household income in € per year)	–	0.00002 (4.87e-06) [4.87]	0.00002 (4.85e-06) [4.74]
Age of respondent (in years)	–	–	−0.013 (0.0029) [−4.48]
Urban (respondent lives in urban area = 1; otherwise = 0)	–	–	0.269 (0.095) [2.84]
Norway	0.0006 (0.129) [−0.002]	0.055 (0.129) [0.45]	0.081 (0.132) [0.61]
UK	−0.550 (0.156) [−3.52]	−0.622 (0.156) [−3.99]	−0.604 (0.155)[−3.89]
Belgium	0.893 (0.131) [6.80]	0.676 (0.138) [4.90]	0.711 (0.142) [5.01]
Lithuania	−2.522 (0.155) [−16.30]	−2.263 (0.163) [−13.89]	−2.266 (0.164) [−13.81]
Constant	2.180 (0.111) [19.61]	1.74 (0.143) [12.12]	2.170 (0.228)[9.52]
Sigma u	2.116 (0.037) [57.90]	2.105 (0.036) [57.93]	2.097 (0.036) [57.96]
Sigma ε	0.525 (0.008) [62.99]	0.525 (0.008) [62.98]	0.525 (0.008) [62.89]
Rho	0.94	0.94	0.94
No. of observations	5466	5466	5466
Number of censored observations	1455	1455	1455
Log-Likelihood	−7657	−7644	−7630
AIC	15334	15311	15286
Wald χ^2 (K = restriction for overall significance)	824 (7)	852 (8)	899 (10)
p value	0.000	0.000	0.000

The country dummy for Denmark is omitted making this the baseline from which country departures are to be interpreted. The dummy for the Small improvement is omitted making this the baseline from which the *Large improvement* departures are to be interpreted

Rho = $\text{var\_}\mu/(\text{var\_}\mu + \text{var\_}\varepsilon)$ and represents the percent contribution to total variance of the panel-level variance component

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 2, scope sensitivity to changes in the quantity of improvement provided are inspected by assigning the Small improvement as our base case WTP values. As can be seen, the values accorded to the Large improvement are very larger than this base case suggesting clear scope sensitivity.[21] Given this result, in our subsequent function transfers we allow for the scope sensitivity difference between the Small and Large improvement.

The results in Table 2 also report significant distance decay; as the distance between the respondent's home and the improvement site increases so WTP decays. This conformity with prior expectations is 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 statistically-driven model extends the former analysis by adding in two ad-hoc variables; age and urban residence. Economic theory has no prior expectations regarding these variables but they are generally available from secondary data and so often appear within benefit transfer studies. 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.[22] 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 result in over-parameterisation for transfer purposes and 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 μ_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 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.[23]

4.3 Value Transfer and Error Analyses

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

Mean 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 being used as the 'policy' site. Transfer errors are then calculated as an absolute percentage by comparing the mean value from the study sites with the actual mean of the policy site. We undertake this procedure for both the small and large improvement.

Value function transfers again compare the observed value of the 'policy' site[24] 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 2) on data from

[21] Non-parametric findings reported in Bateman et al. (2009c).

[22] Possible correlation between age, income and living in an urban area was tested and was found to be extremely low.

[23] Transfer errors remained reasonably stable to the choice of either panel or non-panel specification with the former being preferred in the present paper due to their replication of the data structure and superior fit.

[24] Some analysts compare value function estimates from study sites with the mean 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 mean value comparisons here (directly comparable with the simple univariate transfers).

Table 3 Transfer errors (%) from mean value and function transfer methods: dataset including the dissimilar (Lithuanian) site

WTP measure	Lithuania	Belgium	Denmark	Norway	UK	Average errors (weighted)[a]
Mean value transfers						
Small improvement	500	43	32	36	74	137(107)
Large improvement	413	31	3	32	46	105(81)
Average error	465	37	17	34	60	121
Value function transfers						
Reduced theory-driven model						
(Provision change, distance to site and distance to substitute)						
Small improvement	1	74	48	86	52	52(58)
Large improvement	6	71	66	86	65	59(64)
Average error	3	73	57	86	59	56
Full theory-driven model						
(Provision change, distance to site, distance to substitute and income)						
Small improvement	13	48	22	78	21	37(42)
Large improvement	7	39	48	78	41	43(48)
Average error	10	44	35	78	31	40
Statistically-driven model						
(Provision change, distance to site, distance to substitute, income, age and urban)						
Small improvement	6	71	61	90	56	57(63)
Large improvement	0	65	74	90	67	59(65)
Average error	3	68	68	90	62	58

Transfer errors are calculated by comparing the predicted WTP with the mean WTP values estimated at each site

[a] Average errors are given outside the parentheses. Figures within the parentheses are average errors weighted by relative sample size

the study sites (i.e. omitting data from the policy site), then applying the coefficients[25] to the values of the predictors at the policy site to yield a function transfer value for that policy site. Defining this value as $W\hat{T}P_{k|s}$ 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 $\left(\frac{W\hat{T}P_{k|s} - \overline{WTP}_k}{\overline{WTP}_k}\right)\%$.

Table 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 mean value transfers, disaggregated into values for the Large and Small improvement. We can see that this mix of similar and dissimilar sites yields high levels of overall error (shown in the final column) when the mean value approach is applied with an overall raw error rate of 121%. 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 mean 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 (although, of course, this would not be known in a real world analysis where policy site valuations were unavailable). Subsequently we investigate the impact of restricting our analysis to just the similar countries.

The remainder of Table 3 details results for our various function transfer analyses. This starts with the 'reduced theory-driven' model containing all of those variables suggested by

25 Given the Log-Normal Tobit specification, the parameters to be used for function transfer must be adjusted for censoring and log transformation. Discussions of this adjustment process can be found in Halstead et al. (1991), Haab and McConnell (2003), Brouwer and Bateman (2005) and Carson and Sun (2007).

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 mean value approach (down to 56%). This error reduction is further improved when we add in the income variable to specify our full theory-driven model (down to 40%; one-third of the rate generated by simple mean transfers across both similar and dissimilar countries). 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 analysis.[26] We now test our hypothesis that the theory driven model, although not providing quite such a good fit to individual study site data as the statistically driven model (recall the results Table 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 paper.

The results presented in Table 3 support the hypothesis that, when faced with a heterogeneous set of sites, analysts should prefer function transfer over mean value transfer (and within the former should restrict the specification of models to those variables available from secondary data regarding which economic theory has clear expectations). However, the individual country results show that mean value transfers can yield both high rates of error when transfers from similar countries are applied to a dissimilar country (Lithuania) and low rates when fairly 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.

Comparison of Tables 3 and 4 clearly shows that the exclusion of the dissimilar Lithuanian sites dramatically reduces the rate of error associated with mean value transfers (from 121% to just 37%). Considering the remainder of Table 4 we now see that, with the dissimilar country excluded, a shift to using value function transfers actually increases errors. Even the full theory driven model still yields a transfer error rate which is nearly three quarters higher than that afforded by simple mean value transfers. These findings support 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.

5 Conclusions

The central objective of this study was to develop principles for choosing between the different methodologies available for conducting benefit transfers. Our central hypothesis argued that the crucial issue determining methodological appropriateness concerns the degree of heterogeneity between the various sites across which transfers are to be undertaken. To test this we conducted a number of common design studies across various similar and dissimilar sites. We fully acknowledge that these studies and the transfers they facilitate are abstracted from the typical situation facing decision analysts undertaking field transfers. That typical situation can be characterised as one of noise with analysts having to pull together often disparate studies conducted on somewhat differently defined goods, valued using at best

[26] Of course one simple rule of thumb for such transfers could simply be to focus upon the main source of difference across sites. However, as shown in this case study, this will probably be somewhat inferior to the full theory driven approach.

Table 4 Transfer errors (%) from mean value and function transfer methods: dataset excluding the dissimilar (Lithuanian) site

WTP measure	Belgium	Denmark	Norway	UK	Average errors (weighted)[a]
Mean value transfers					
Small improvement	43	32	36	74	46(43)
Large improvement	31	3	32	46	28(28)
Average error	37	17	34	60	37
Value function transfers					
Reduced theory-driven model					
(Provision change, distance to site and distance to substitute)					
Small improvement	79	55	90	62	71(75)
Large improvement	75	71	89	72	77(78)
Average error	77	63	89	67	74
Full theory-driven model					
(Provision change, distance to site, distance to substitute and income)					
Small improvement	68	41	85	48	60(65)
Large improvement	62	61	84	62	67(69)
Average error	65	51	85	55	64
Statistically-driven model					
(Provision change, distance to site, distance to substitute, income, age and urban)					
Small improvement	69	42	88	46	61(66)
Large improvement	63	61	88	60	68(71)
Average error	66	51	88	53	64

Transfer errors are calculated by comparing the predicted WTP with the mean WTP values estimated at each site

[a] Averages are given outside the parentheses. Figures in the parentheses are average errors weighted by relative sample size

differently designed studies and at worst entirely different techniques. That this state of affairs exists does not, to us, seem a good reason to replicate such a situation. Indeed we have argued elsewhere that a pre-requisite to high quality transfers is the commissioning of a set of uniform studies designed specifically for the purposes of transferral (Bateman et al. 2010). Furthermore, the minimisation of external noise is an accepted basis for most experimental work, including that intended to inform methodological development.

Our study findings suggest some broad and pragmatic guidelines for the future application of value transfers. Essentially we conclude that, where transfers involve broadly similar provision changes of similar goods across similar contexts, then simple mean value transfers are likely to give defensible welfare estimates. However, if any one of those similarities fails to hold then it is likely that value function transfers will reduce errors relative to means, particularly those functions are constructed from general economic theoretic principles to contain only those variables about which we have clear, prior expectations.[27] While more complex and sophisticated functions may cater better for the context specific factors which determine value at any given study site, there is no guarantee that such contextual conditions will apply at policy sites. The different operation (or even omission) of contextual factors between sites is liable to generate substantial transfer errors.

[27] Given the important role which prior information is accorded within such guidelines, one can obviously see a place for Bayesian techniques within value transfer analyses. Indeed Leon-Gonzalez and Scarpa (2008) employ a Bayesian approach to show that transfer errors are lower even when applied to data from a relatively small set of sites selected to have similar characteristics than when a larger set of more heterogeneous sites. However, while this finding accords with our own, we would go further to suggest that where a set of clearly similar sites can be established, then simple mean value transfers may be optimal.

While we hope that such simple guidelines will provide a positive contribution to applications of value transfer, they certainly do not conclude that debate. In particular the issue of what constitutes an acceptable similarity of provision change, good and context is something which requires further consideration. We deliberately engineered the present exercise to ensure that the provision change and good were practically identical across applications such that only the socio-economic context varied. However, it remains an open empirical question as to what variation in provision change, good and context is allowable before reliance upon simple mean transfers becomes unsafe. Nevertheless, we hope that the transfer protocol mapped out within the present paper will provide a useful framework both for developing practical value transfer guidance and for focusing future research in this area.

Acknowledgments This study was carried out as part of the EU DG Research funded AquaMoney project (SSPI-022723, see http://www.aquamoney.org). Additional funding for the UK study was supplied by the ESRC through the SEER projects (RES-060-25-0063, see http://www.cserge.ac.uk) and Eftec. Ian Bateman is also Adjunct Professor in the Department of Agricultural and Resource Economics at the University of Western Australia, Perth and the Department of Economics, University of Waikato Management School, New Zealand. The useful comments of the wider project group are gratefully acknowledged. We also thank Luke Brander, carlofezzi, Joseph Herriges and Cathy Kling for valuable suggestions regarding this work and are very grateful to an anonymous reviewer for a set of extremely insightful comments on earlier drafts of this paper. Remaining errors are the sole responsibility of the authors.

References

Amiran EY, Hagen DA (2010) The scope trials: variation in sensitivity to scope and WTP with directionally bounded utility functions. J Environ Econ Manag 59:293–301. doi:10.1016/j.jeem.2009.06.003

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

Banerjee S, Murphy JH (2005) The scope test revisited. Appl Econ Lett 12(10):613–617

Barton DN (2002) The transferability of benefit transfer: contingent valuation of water quality improvements in Costa Rica. Ecol Econ 42:147–164

Bateman IJ, Langford IH, Turner RK, Willis KG, Garrod GD (1995) Elicitation and truncation effects in contingent valuation studies. Ecol Econ 12(2):161–179. doi:10.1016/0921-8009(94)00044-V

Bateman IJ, Carson RT, Day BH, Hanemann WM, Hanley N, Hett T, Jones-Lee M, Loomes G, Mourato S, Özdemiroğlu E, Pearce DW, Sugden R, Swanson J (2002) Economic valuation with stated preference techniques: a manual. Edward Elgar Publishing, Cheltenham

Bateman IJ, Jones AP, Lovett AA, Lake I, Day BH (2002) Applying geographical information systems (GIS) to environmental and resource economics. Environ Resour Econ 22(1–2):219–269. doi:10.1023/A:1015575214292

Bateman IJ, Jones AP (2003) Contrasting conventional with multi-level modelling approaches to meta-analysis: an illustration using UK woodland recreation values. Land Econ 79(2):235–258. doi:10.3368/le.79.2.235

Bateman IJ, Cole M, Cooper P, Georgiou S, Hadley D, Poe GL (2004) On visible choice sets and scope sensitivity. J Environ Econ Manag 47:71–93. doi:10.1016/S0095-0696(03)00057-3

Bateman IJ, Day BH, Georgiou S, Lake I (2006) The aggregation of environmental benefit values: welfare measures, distance decay and total WTP. Ecol Econ 60(2):450–460. doi:10.1016/j.ecolecon.2006.04.003

Bateman IJ, Day BH, Jones AP, Jude S (2009) Reducing gains/loss asymmetry: a virtual reality choice experiment (VRCE) valuing land use change. J Environ Econ Manag 58:106–118. doi:10.1016/j.jeem.2008.05.003

Bateman IJ, Day BH, Dupont D, Georgiou S (2009) Procedural invariance testing of the one-and-one-half-bound dichotomous choice elicitation method. Rev Econ Stat 91(4):806–820. doi:10.1162/rest.91.4.806

Bateman IJ, Brouwer R, Ferrini S, Schaafsma M, Barton DN, Dubgaard A, Hasler B, Hime S, Liekens I, Navrud S, De Nocker L, Ščeponavičiūtė R, Semėnienė D (2009c) Making benefit transfers work: transferring theoretically derived, spatially explicit, value functions for the non-market benefits of water quality

improvements across Europe, CSERGE Working Paper, Centre for Social and Economic Research on the Global Environment, University of East Anglia, UK
Bateman IJ, Brouwer R, Cranford M, Hime S, Ozdemiroglu E, Phang Z, Provins A (2010) Valuing Environmental Impacts: Practical Guidelines for the Use of Value Transfer in Policy and Project Appraisal, Main Report to the Department for Environment, Food and Rural Affairs, Economics for the Environment Consultancy (eftec), London
Bergland O, Magnussen K, Navrud S (1995) Benefit transfer: testing for accuracy and reliability. Discussion Paper, #D-03/1995, Department of Economics and Social Sciences. Agricultural University of Norway
Brouwer R (2000) Environmental value transfer: state of the art and future prospects. Ecol Econ 32:137–152
Brouwer R, Bateman IJ (2005) Benefits transfer of willingness to pay estimates and functions for health-risk reductions: a cross-country study. J Health Econ 24:591–611
Brouwer R, Spaninks FA (1999) The validity of environmental benefits transfer: further empirical testing. Environ Resour Econ 14(1):95–117
Carson RT, Mitchell RC (1993) The value of clean water: the public's willingness to pay for boatable, fishable and swimmable quality water. Water Resour Res 29(7):2445–2454
Carson R, Hanemann M, Kopp R, Krosnick J, Mitchell R, Presser S, Rudd P, Smith VK, with Conaway M, Martin K (1996) Was the NOAA panel correct about contingent valuation? Resources for the Future Discussion Paper 96-20, Washington, DC
Carson RT, Flores NE, Hanemann WM (1998) Sequencing and valuing public goods. J Environ Econ Manag 36:314–323
Carson RT, Groves T (2007) Incentive and informational properties of preference questions. Environ Resour Econ 37(1):181–210
Carson RT, Sun Y (2007) The Tobit model with a non-zero threshold. Economet J 10:488–502
Champ AP., Boyle K., Tom Brown C (eds) (2003) A primer on non-market valuation, the economics of non-market goods and services: vol 3. Kluwer, Dordrecht
Chattopadhyay S (2003) A repeated sampling technique in assessing the validity of benefit transfer in valuing non-market goods. Land Econ 79(4):576–596
Covey J, Loomes G, Bateman IJ (2007) Valuing risk reductions: testing for range biases in payment card and random card sorting methods. J Environ Plan Manag 50(4):467–482. doi:10.1080/09640560701401986
Desvousges WH, Naughton MC, Parsons GR (1992) Benefit transfer: conceptual problems in estimating water quality benefits using existing studies. Water Resour Res 28(3):675–683
Desvousges WH, Smith VK, Fisher A (1987) Option price estimates for water quality improvements: a contingent valuation study of the Monongahela River. J Environ Econ Manag 14:248–267
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 PA (2005) Reducing the influence of anecdotal reasoning on people's health care decisions: is a picture worth a thousand statistics? Med Decis Mak 25:398–405
Haab TC, McConnell KE (2003) Valuing environmental and natural resources: the econometrics of nonmarket valuation. Edward Elgar Publishing, Massachusetts
Halstead JM, Lindsay BE, Brown CM (1991) Use of the Tobit model in the contingent valuation: experimental evidence from the Pemigewasset wilderness area. J Environ Manag 33:79–89
Heberlein TA, Wilson MA, Bishop RC, Schaeffer NC (2005) Rethinking the scope test as a criterion for validity in contingent valuation. J Environ Econ Manag 50:1–22
Hime S, Bateman IJ, Posen P, 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
Hutchins MG, Deflandre A, Boorman DB (2006) Performance benchmarking linked diffuse pollution and in-stream water quality models. Archiv fur Hydrobiol Suppl 161(17):133–154
Johnston RJ, Besedin EY, Iovanna R, Miller CJ, Wardwell RF, Ranson MH (2005) Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: a meta-analysis. Can J Agricul Econ 53:48–221
Johnston RJ, Ranson MH, Besedin EY, Helm EC (2006) What determines willingness to pay per fish? A meta-analysis of recreational fishing values. Marine Resour Econ 21:1–32
Johnston RJ, Duke JM (2009) Willingness to pay for land preservation across states and jurisdictional scale: implications for benefit transfer. Land Econ 85(2):217–237
Jones AP, Bateman IJ, 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.

Jones AP, Wright J, Bateman IJ, Schaafsma M (2010) Estimating arrival numbers for informal recreation: a geographical approach and case study of British woodlands. Sustainability 2(2):684–701. doi:10.3390/su2020684

Leon-Gonzalez R, Scarpa R (2008) Improving multi-site benefit functions via Bayesian model averaging: a new approach to benefit transfer. J Environ Econ Manag 56(1):50–68

Lindhjem HL, Navrud S (2008) How reliable are meta-analyses for international benefit transfers? Ecol Econ 66:425–435

Marta-Pedroso C, Freitas H, Domingos T (2007) Testing the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews. Ecol Econ 62:388–398

Mitchell RC, Carson RT (1989) Using surveys to value public goods: the contingent valuation method. Resources for the Future, Washington

Moeltner K, Boyle KJ, Paterson RW (2007) Meta-analysis and benefit transfer for resource valuation–Addressing classical challenges with Bayesian modeling. J Environ Econ Manag 53:69–250

Muthke T, Holm-Mueller K (2004) National and international benefit transfer testing with a rigorous test procedure. Environ Resour Econ 29:323–336

Navrud S, Ready R (2007) Review of methods for value transfer. In: Navrud S, Ready R (eds) Environmental value transfer: issues and methods. Springer, Dordrecht

Navrud S, Ready R (eds) (2007) Environmental value transfer: issues and methods. Springer, Dordrecht

Pearce DW, Whittington D, Georgiou S (1994) Project and policy appraisal: integrating economics and environment. OECD, Paris

Peters E, Slovic P, Hibbard J (2005) Bringing meaning to numbers: evaluability and affect in choice. University of Oregon, Eugene

Powe NA, Bateman IJ (2004) Investigating insensitivity to scope: a split-sample test of perceived scheme realism. Land Econ 80(2):258–271

Rollins KS, Lyke AJ (1998) The case for diminishing marginal existence values. J Environ Econ Manag 36:324–344

Rowe RD, Schulze WD, Breffle W (1996) A test for payment card biases. J Environ Econ Manag 31:178–185

Sigelman L, Zeng L (1999) Analyzing censored and sample-selected data with Tobit and Heckit models. Polit Anal 8(2):167–182

Sutherland RJ, Walsh RG (1985) Effect of distance on the preservation value of water-quality. Land Econ 61(3):281–291

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 TP, Poe GL, Powell JR (2001) Assessing the accuracy of benefits transfers: evidence from a multi-site contingent valuation study of ground water quality. In: Bergstrom J, Boyle K, Poe G (eds) The economic value of water quality. Edward Elgar, Northampton

WATECO (2004) Economics and the environment—the implementation challenge of the water framework directive. Luxembourg: European Commission

World Bank (2008) Global purchasing power parities and real expenditures. World Bank, Washington

Zandersen M, Termansen M, Jensen FS (2007) Testing benefits transfer of forest recreation values over a twenty-year time Horizon. Land Econ 83(3):412–440

Zikmund-Fisher BJ, Fagerlin A, Ubel PA (2005) What's time got to do with it? Inattention to duration in interpretation of survival graphs. Risk Anal 25(3):589–595