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

Annex 2 – Assessing the Quality of Primary Valuation Studies

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

QUALITY ASSURED FIRM ISO 9001
THE BRITISH ACCREDITATION BUREAU
REGISTRATION NUMBER 183887

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

Study team (in alphabetical order):

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

Acknowledgements:

The study team would like to thank Prof. Stale Navrud (Norwegian University of Life Sciences), Prof. Ken Willis (University of Newcastle upon Tyne) and members of the Steering Group for their comments on the previous versions of the Guidelines, Case Studies and 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.

ANNEX 2: ASSESSING THE QUALITY OF PRIMARY VALUATION STUDIES

- *The quality of a primary study is influenced by how it is designed – in particular its coverage of the policy context, the good and the change, its definition of the affected population and sampling (where relevant) and whether the results are valid and robust.*
- *The criteria for all these are discussed here. This section should be read in conjunction with the Glossary of Econometric Terminology (**Annex 3**).*

Introduction

This section discusses how the value transfer analyst can judge whether a given primary study is of a sufficient quality to use in value transfer. Three main questions need to be answered for this:

1. Is the coverage of the policy context, the good and the change of sufficient quality?
2. Is the affected population defined correctly and, where relevant, sampled sufficiently?
3. Are the results valid and robust?

The following presents the criteria to use when judging whether these questions can be answered positively for a given primary valuation study. Some of these criteria apply generally (Section I below), while others are specific to individual economic valuation methods (Section II). The Glossary of Econometric Terminology (**Annex 3**) should also be consulted as it defines the terminology used in this Annex.

I. Generally applicable criteria for good quality studies

Generally applicable rules of thumb for assessing study quality relate to the questions (1) and (2) above.

The policy context, the good and the change

A good quality study:

- Reports details of the policy context within which the economic valuation is undertaken and its purpose for policy use;
- Defines the good valued;
- Makes it clear which aspects of the good (and ecosystem services associated with it) are covered by the study (e.g. whether the market and/or the non-market; and use and/or non-use aspects of the good are valued);
- Uses high quality scientific and economic data to define the change valued. In most cases the change information will be the product of a multi-disciplinary approach to impact assessment depending on the characteristics of the good. Scientific data should come from a reliable source and should ideally be specific to the study site. In some cases scientific studies undertaken elsewhere could be used (e.g. dose-response relationships for various airborne pollutants and

human health risks could be taken from another country). If this is inevitable due to data and time restrictions, then the same scrutiny shown over the economic value transfer should be applied to such (scientific) information transfers (e.g. whether the environmental and human conditions are sufficiently similar between the original scientific study and those in the primary valuation study); and
- Explains the change valued in terms that can be clearly understood by the respondents (for methods that use surveys).

Affected population and sampling

A good quality study:

- Clearly and correctly defines the affected population. This is required to judge whether sampling is representative.
- Distinguishes between users and potential non-users, even if the latter is not captured in the study.
- Captures a sample that is (a) large enough for the purposes of the study and (b) representative of the affected population. How large is 'large enough' is context dependent but the study results hold clues as to whether the sample was sufficient. Some studies caveat their results (especially when some obvious explanatory variables come out as insignificant) saying that the samples are small. Samples less than 100-200 observations are likely to be too small for a robust stated preference survey[1].

In particular for spatially distributed goods, the use of GIS for population data (numbers, socio-economic data) would be preferred especially as the use of GIS in economic valuation studies increases. In particular, if the primary study detects a distance-decay function, use of GIS (to aggregate across the population and to hold socioeconomic and other data about the population) will be preferred.

Results

A good quality study:

- Defines the value function to include all explanatory variables that are likely (on the basis of economic theory and empirical evidence) to influence the economic value;
- Tests different econometric models to find the best fitting model;
- Reports principal descriptive statistics and data that will help explain and transfer the unit value results; and
- Reports the year the data and estimated values apply to.

Recent research (see the ***Technical Report***) finds that the value functions which focus on variables that are predicted by economic theory and empirical evidence to influence value estimates lead to lower transfer errors. These factors are typically limited to:

[1] This is particularly the case for contingent valuation studies (and to some extent travel cost studies, which are revealed preference surveys). Note that choice modelling studies typically generate a greater number of data observations than the sample size; this should be taken into account when considering the sample size of such studies.

- The change in the provision of the good; i.e. the scope of the change
- The costs faced in using/accessing the good (for non-market goods this may primarily by related to proximity to the individual as revealed by a distance-decay relationship)
- The availability of substitutes
- Income (or proxy variables for this)

Reporting these functions (as well as those that give the best fit to the data specific to the context of the primary study) should be a criterion for a good quality study.

II. Valuation method specific conditions for good quality studies

To ensure good quality design and results, primary valuation studies should follow the best practice guidelines. **Box A2.1** lists a selection of best practice guidelines for each valuation method.

Box A2.1: Existing guidelines and manuals for economic valuation

A number of UK Government documents provide guidance for undertaking economic valuation of non-market goods and services, including:

- Defra (2007) *An introductory guide to valuing ecosystem services*
- eftec (2006) *Valuing our natural environment*, report to Defra
- eftec (2009) *Economic valuation of upland ecosystem services*, report to Natural England
- Bateman et al (2002) *Economic valuation with stated preference techniques: a manual*

The *Green Book* also surveys further reading for various non-market impacts, including valuing time, health and environmental impacts. Guidance on technical aspects of economic valuation is provided by numerous references. Key references include:

- Bockstael and McConnell (2006) *Environmental and resource valuation with revealed preferences: A theoretical guide to empirical models*
- Champ et al. *(2003) A primer on non-market valuation*
- Freeman (2003) *The measurement of environmental and resource values*
- Haab and McConnell (2002) *Valuing environmental and natural resources*
- Hensher et al. (2005) *Applied choice analysis: a primer*
- Kanninen (2006) *Valuing environmental amenities using stated choice studies: a common sense approach to theory and practice*
- Louviere et al (2000) *Stated choice models*
- Ward and Beal (2000) *Valuing nature with travel cost models: a manual*

Full citations are provided in the reference section of the Value Transfer Guidelines.

The rest of this section presents the factors that should be taken into account when assessing the quality of a study. A study that takes note of and reports on these factors can be deemed to be a good quality study.

Market prices and price proxies

A good quality study:

- Ensures that the good and the change are defined correctly. Depending on the method, assumptions also need to be made about the complements (e.g. for avertive expenditure) and substitutes (e.g. cost of alternatives).
- Provides the references to data sets and other sources used to demonstrate the reliability of the data.
- Reports how real (or shadow) prices are calculated – ideally by identifying and netting out the taxes and subsidies.
- Tests the robustness of the results through comparison with other data from actual markets and using statistical / econometric analysis tests (see Annex 3).

Revealed preference methods – hedonic property pricing

A good quality study:

- Selects the explanatory variables for the hedonic function that explains the variation in property prices. The good that is valued can be included in the function directly (e.g. air quality measurements and noise levels can be continuous variables; or existence of a park in the area of the study as a dummy variable). In some contexts, a proxy for a good can be used – for example, the visual disamenity of a transport scheme can be included in the property price function as the distance between the individual property and the road and whether the road can be seen from a given property.

- Ensures all other potential explanatory variables that are not directly related to the study good are also included in the property price function. These include characteristics of the property, the characteristics of the neighbourhood and the socioeconomic characteristics of the population in the area. The exclusion of these variables could result in the good being over-valued.

- Explicitly states whether the first or second stage of hedonic pricing is implemented. The first stage estimates the implicit price for the good being valued. The second stage estimates the demand curve with respect to the quantity of the good being valued (between the initial and final level of its provision). Most studies only estimate the first stage model.

- Tests the validity of the results can be assessed either by comparing the results of a given study to other studies that cover similar geographical, policy and good contexts and use hedonic property pricing or other methods. This is the so-called convergent validity (see below).

Revealed preference methods – discrete choice models and travel cost method

A good quality study:

- Selects the relevant explanatory variables including the characteristics of the site visited, the characteristics of the visitors and travel cost. The variables could be defined for individuals or for (geographical) zones in different distances from the valued site.

- Collects data on place of residence of visitor, demographics, attitudinal information, frequency and length of visit to site that is valued and the substitute sites and information about the trip (e.g. purpose, length, time spent travelling to and from the site, all relevant travel costs etc.). For the random utility version of travel cost (i.e. discrete choice models and also termed 'multi-site

recreation demand models'), which compares the demand for different sites with different characteristics, data should be also be collected on the characteristics of each site to be compared.

- Describes the affected population, give references to data sources for the population and report the sampling method, size and representativeness.

Stated preference methods – contingent valuation and choice modelling

A good quality study:

- Designs the survey questionnaire through the reiterative testing steps including focus groups, cognitive interviews, and pilot survey(s). If a study did not follow all of these testing steps and improved the design after each step, the reasons for omissions must be explained.

- Includes questions on the attitudes and opinions, valuation scenario, follow-up questions to probe the motivations behind willingness to pay or willingness to accept responses and socio-economic characteristics of the respondents. Other sections could be added depending on the purpose and scope of the study but the ones listed here should be included, or their exclusion explained.

- Tests the validity of the results. Since, there are usually no actual markets for the good or aspects of the good valued using stated preference methods, the quality of the results cannot be compared to data from actual markets. This is why the development of the methods involved defining the reasons why the estimated values could diverge from 'real' values (the so-called biases) and devising tests to assess the validity of the results (to test the absence of such biases). Validity can be assessed in different dimensions: content, construct and convergent validity.

  Content validity

  At the simplest level, the face / content validity of a CV survey instrument can be assessed: Does the survey instrument present the 'correct' good in a manner that is likely to be understandable to respondents from a wide range of different backgrounds and education levels? Is the method of providing the good and collecting payment for its provision plausible?

  Table A2.1 presents the likely biases that stated preference study results can suffer from and how good quality studies should account for these. It is, in other words, a template for content validity check. A good stated preference study is one that tests these biases in its questionnaire design and econometric analysis.

  The definition of affected population and sampling are also contributors to content validity. A good study is one that describes the affected population, gives references to data sources for the population and reports the sampling method, size and representativeness.

Table A1.1: Potential biases in stated preference methods and how a good quality study should address them				
General type of bias	***General rules for addressing this type of bias***	***Specific biases***	***Description of the specific bias***	***Specific rules for addressing this type of bias***
Reference Point	Present information in the questionnaire neutrally compared to the real world context. Attempt to understand reference point through survey design (survey) and include in value function (analysis).	Anchoring	Quantitative cues in the wording of the questionnaire as to what the WTP or WTA 'should' be.	Do not provide quantitative cues in the questionnaire (e.g. cost of an investment, likely increase in future bills).
		Framing	Equivalent lotteries presented differently (i.e. average win of £50 is preferred to 50% chance of winning £100, even though statistically equivalent)	Frame the valuation questions as it would be presented in the real world (e.g. if there is a 50% chance of winning £100, do not state that £50 is won on average)
		Prominence	Different weights given to different aspects of question, based on survey format.	Format the valuation question as it would be presented in the real world (e.g. based on store purchase if discussing labelled products). Also allow survey to be flexible in, for example, ordering of potential responses so that there is no preference given to the answer at the top of a list of pre-coded responses.
		Saliency	Different weights given to different aspects of question, based on respondent irrationality.	Format the valuation as it would be presented in the real world.
		Asymmetry	Individuals weigh losses more heavily than gains.	Use a value function that accounts for asymmetry if it is an issue in the study (e.g. use a function that is different for WTP and WTA if comparing the two).

Table A1.1: continued

General type of bias	*General rules for addressing this type of bias*	*Specific biases*	*Description of the specific bias*	*Specific rules for addressing this type of bias*
Availability	Humans do not process information perfectly and so act with (or as if there is) a lack of information.	Reference point	Not all respondents have the same reference point for what is 'normal', therefore they may be valuing different changes when valuing a change from normal.	Include questions in survey to determine reference point and include in analysis (e.g. question on 'how much do you pay now?') or clearly state what the reference point is that respondents should be working from.
		Status quo/endowment	Respondents tend to prefer the norm over any different situation.	Design study so respondents choosing status quo out of protest can be identified and removed (e.g. test questions and extra scenario wording before the WTP / WTA question, or in follow up questions).
		Availability	Respondents rely too much on readily retrieved information and so may fail to account for information that is equally important, but harder to find or understand.	Make difficult information easier to understand when compared to simple information (e.g. it may be useful to simplify the technical description of environmental impact to a more easily compared, qualitative description).
Superstition	Consumers include irrelevant, subjective beliefs in information processing.	Certainty	Sure outcomes preferred over uncertain ones (i.e. 100% chance of winning £50 is preferred to 50% chance of winning £100)	Ask follow-up questions and account for this bias in analysis (e.g. follow-up question of 'why did you choose x?' with a potential response of 'because it was certain').
		Focal	Quantitative information viewed as categorical (i.e. 10 compared to 1,000 is viewed by individual as small and large).	Set cues to educate respondent on how much bigger or smaller quantitative values are (e.g. throughout survey state 1,000 and large together so they are constantly equated and categorical response is still relevant to quantitative cue).

Table A1.1: continued

General type of bias	*General rules for addressing this type of bias*	*Specific biases*	*Description of the specific bias*	*Specific rules for addressing this type of bias*
Superstition (*continued*)	Consumers include irrelevant, subjective beliefs in information processing.	Isolation	Elements are evaluated separately.	Elements are evaluated separately.
		Recency	Recent events more easily remembered (i.e. responses will be more highly influenced by recent events still fresh in respondents' minds than older, but perhaps more relevant events).	Survey should remind people of important events and norms and, if possible, avoid discussion of less relevant but more recent events or coincidental events (e.g. a recent accidental oil spill will stick in people's minds and influence responses in relation to marine/coastal environment; the survey should thoroughly remind respondents of what the usual state is).
		Regression (aka return to normal)	Individuals mentally weight coincidences heavily, often mistaking them for normal occurrences. Return to the normal state is underestimated (e.g. oil spill is considered norm if happened a couple times in same area when actually the vast majority of time there is no oil spill).	Recent events more easily remembered.
		Representativeness	High *conditional* probabilities cause overestimates of *unconditional* probabilities (i.e. if the probability of A given B is high, individuals assume probability of A is high even if probability of B is low).	Represent cumulative frequencies rather than separate frequencies, which leaves respondent to calculate incorrectly (e.g. do not state there is a 90% chance of A if B occurs and there is a 10% chance of B, but rather state there is a 9% chance of A occurring).
		Credulity	Respondents too readily infer causation before there is enough evidence to do so.	Remind respondents that correlation does not ensure causation.
		Disjunctive	Respondents may fail to reason through cause and effect relationships, never arriving at the logical conclusion.	Make outcomes very clear (e.g. do not leave it to respondents to reasons that there will be more environmental degradation under poor policy decisions, but illustrate how that is the case).

Table A1.1: continued

General type of bias	*General rules for addressing this type of bias*	*Specific biases*	*Description of the specific bias*	*Specific rules for addressing this type of bias*
Process	Humans have a limited computational ability and often adopt boundedly rational heuristics (i.e. simple rules of logic).	Superstition	Causation attached to coincidence and "quasi-magical" powers attached to opponents (i.e. respondents reject coincidence and prefer to think that there is some underlying, unknown connection between unrelated events).	Do not present unrelated events together unless absolutely necessary, and when presented together clarify that there is no relationship (e.g. do not allow respondents to believe a connection between environmental events and political parties when there is not policy change connecting the two).
		Suspicion	Individuals may mistrust or be suspicious of offers, particularly when speaking about monetary valuations.	Make the valuation construct as realistic as possible (e.g. provide an idea of how payments may be made by individuals, such as through additional council tax to pay for local environmental improvements).
		Rule-driven	An individual's values may be derived by principles, analogies and exemplars rather than logical understanding of what would best serve their needs and preferences.	If there is concern that responses may be rule-driven, include questions to understand these effects and include in analysis (e.g. include questions on beliefs and attitudes in survey).
		Temporal	Short-term view is discounted too sharply relative to long term view meaning respondents may choose a short term gain that is smaller than a long term gain (even accounting for normal market discounting).	Carefully remind respondents that long-term gains, although further away may be larger overall (e.g. explain using market discount rates how much the short term and long term gains are worth and compare them).
Projection	Design survey to be as anonymous and politically neutral as possible (survey). Present information in the questionnaire neutrally compared to the real world context.	Misrepresentation	Respondents present values that promote real/perceived strategic advantage.	Use language to suggest that values will not influence policy (e.g. language on how this survey if only 'academic' or 'informational').
		Projection/compliance	Respondents present values that are influenced by a desire to reinforce internally or project to others a specific self-image.	Allow respondents to answer survey in a neutral and comfortable setting (e.g. computer surveys preferred over human administered, remove any indication of sponsorship of survey).
		Anchoring	Quantitative cues in the wording of the questionnaire as to what the WTP or WTA 'should' be.	Do not provide quantitative cues in the questionnaire (e.g. cost of an investment, likely increase in future bills).

Table A1.1: continued				
General type of bias	***General rules for addressing this type of bias***	***Specific biases***	***Description of the specific bias***	***Specific rules for addressing this type of bias***
Projection *(continued)*	Design survey to be as anonymous and politically neutral as possible (survey). Present information in the questionnaire neutrally compared to the real world context.	*Framing*	Equivalent lotteries presented differently (i.e. average win of £50 is preferred to 50% chance of winning £100, even though statistically equivalent).	Frame the valuation questions as it would be presented in the real world (e.g. if there is a 50% chance of winning £100, do not state that £50 is won on average).
		Prominence	Different weights given to different aspects of question, based on survey format.	Format the valuation question as it would be presented in the real world (e.g. based on store purchase if discussing labelled products. Also allow survey to be flexible in, for example, ordering of potential responses so that there is no preference given to the answer at the top of a list of pre-coded responses.

The following are indicators of low content validity of a study:

- Inadequate sample size or poor coverage of the relevant population;
- Non-stratified or biased sampling methods (where the representativeness of the sample is an issue; e.g. for aggregation purposes);
- High survey or item non-response rates;
- Large numbers of 'protest' bids;
- Prevalence of free-riding behaviour;
- High numbers of outlier (i.e. infeasibly large) WTP or WTA responses;
- Inadequate responsiveness to the scope of the good in question;
- The valuation scenario and corresponding valuation task is poorly understood and/or has low credibility;
- The provision change description is poorly understood and/or has low credibility;
- Low explanation of responses in terms of theoretical or other expectations; and
- Survey respondents provide answers which indicate that strategic behaviour may have affected responses.

Construct validity

The other major type of validity is construct validity. This is an assessment of whether the value estimate is related in particular ways to factors that are predicted to influence the value, based on economic theory and empirical regularities in the form of associations with other variables, which seem intuitively correct and which hold across a large number of studies.

The factors that are found to influence economic value by economic theory and empirical evidence are presented in **Table A2.2**. A good quality study should collect data for each of these factors or explain the reasons for omission. Data collection could be from secondary sources (including scientific or other change data) or from the questionnaire (by including questions or designing different versions of the valuation scenario to cater for the variations in these factors).

Convergent validity

Another form of construct validity occurs when there are multiple ways to measure the same underlying quantity. In such cases, the different measurements can be compared in a test of convergent validity; i.e. a process of using different valuation techniques to see if they produce similar answers or answers that vary in a predicted manner. A common example is the use of travel cost and contingent valuation methods together.

Convergent validity is useful but could be a misleading criterion to judge the quality of a given primary valuation study if used on its own. Primary valuation studies that produce similar results could be consistently wrong. Without other criteria to check validity it is not possible to if consistency is a sign of validity or not. This is why meta-analysis – rather than superficial comparison of the results – is preferred.

Table A2.2: Factors influencing economic value	
Factors	***Direction of influence****
Factors related to the good	
Size of the good - could be tested through different levels of the good in the status quo	-: generally, the more of the good there is the less the marginal value of a unit change, but: 1) this depends on relative location. 2) there may be threshold effects, e.g. minimum quantity of good for providing some service, or other discontinuities.
Availability of substitutes – could be varied in the wording of the valuation scenario. At the very least the availability of substitutes in the status quo should be presented	-: generally the more substitutes there are the less the marginal value for a change is likely to be. The relationship applies to the substitutes for the good, individual attributes of the good, individual ecosystem services provided by the good, or particular human uses of the good
The scale (direction and size) of change – different scales of change could be tested through different versions of the questionnaire (but smaller surveys may not have sufficient sample to do so)	+ for loss: as the size of the loss increases so does the marginal value - for gain: as the size of the gain increases the marginal value decreases In both cases, thresholds can intervene and introduce discontinuities in the total and marginal values
Price of the good (e.g. bid levels in contingent valuation or the price attribute in choice experiment)	- for WTP: the higher the price the lower the likelihood of respondent accepting the price + for WTA: the higher the price the higher the likelihood of respondent accepting the price
Factors related to individuals	
Income (or socio-economic group as a proxy)	+: the more income the higher the marginal value per person, generally (there can be exceptions for cheaper, less valuable services, where increasing incomes allow consumers to "switch" to more expensive, more valuable services, resulting in reduced values for the cheaper services)
Average age	+/-: no prior expectations on the direction of influence
Education	+: the higher the education the higher the environmental awareness and the higher the value placed on the environment
Gender	+/-: no prior expectations on the direction of influence
Familiarity with the good– depends on where, which land etc.	+/-: no prior expectations on the direction of influence

Notes: *+ and – refer to the sign of the coefficient in the valuation function – see Annex 3 for further details on functional forms etc.