Accent

PJM economics

Outcome Delivery Incentive Research: Design of Methodology

Academic Literature Review

November 2021

Contact: Paul Metcalfe
E-mail: Paul.metcalfe@pjmeconomics.co.uk
Telephone: +44 (0) 7786 656834

File name: 3524rep03_AcademicReview_v1

Registered in London No. 2231083
Accent Marketing & Research Limited
Registered Address: 30 City Road, London, EC1Y 2AB

Contents

Executive Summary

Introduction

Accent and PJM economics have been commissioned by Ofwat and CCW to develop a methodology for obtaining the customer evidence to support ODI rate setting for common PCs at PR24 (Stage 1), and to develop and test the materials based on this methodology (Stage 2).

This document forms one part of an initial desk review for the study which, along with a consultation with companies and stakeholders, is intended to lay a solid foundation for the development of the methodology options that will form the principal content of the Stage 1 report.

It looks at two related issues:

- The range of different non-market valuation methods used in the academic literature, with particular focus on their suitability to value water-related attributes; and,
- Best practice guidance for stated preference studies.

At the present, interim, stage, the intention is that this review, which has been conducted in a relatively short space of time, should lay the groundwork for more detailed, water-specific, guidance that will be developed for the final Stage 1 methodology report. Accordingly, it is not intended to be exhaustive nor meticulous at this stage.

Non-market valuation methods

For cases where market or generic prices are unavailable, valuation methods are typically grouped into three categories: Stated preference, Revealed preference, and Subjective wellbeing approaches.

Stated preference approaches consist of surveys asking individuals (directly or indirectly) for their preferences regarding a non-market good. These approaches are based on hypothetical scenarios. This allows the valuation of a range of possible future changes, but it can also be a limitation due to the hypothetical nature of the choices made by survey participants.

Among stated preference approaches:

- Contingent valuation consists of asking participants directly for their willingness to pay/accept for a non-market good. The method is relatively simple but cannot value specific attributes within a policy package or estimate trade-offs between those attributes.
- Choice experiments allow for the valuation of multiple attributes (often the case of improvement packages implemented by water companies). However, survey

participants may ignore some attributes or be influenced by the range of the levels that those attributes assume in the experiment.

- Other stated preference methods (menu choice experiments, impact-weighted valuation, best-worst scaling, contingent ranking, contingent rating, paired comparisons, and stated preference exercises using citizen juries) have also been used for non-market valuation, either in place of, or to augment or complement, choice experiments.

Revealed preference approaches infer preferences from the individuals' choices in real-world markets. Unlike stated preference approaches, revealed preference approaches do not rely on hypothetical scenarios. However, this can be a limitation, as the real world does not replicate all the possible combinations of changes in all relevant attributes that can be changed by policies or business decisions. Revealed preference approaches also cannot easily estimate non-use values, and tend to require large datasets. Among revealed preference approaches:

- Travel cost methods infers the value of a non-market good by the travel costs users incur to access the site where they consume that good. As such, it is mainly used to value aspects of outdoor recreation (including water quality)

- Hedonic analysis infers the value of a non-market good by the choices that individuals make in markets that incorporate that good (usually property markets). It has been used to value flood risk and water quality in water bodies near homes. It assumes that markets capitalise the value of the non-market good, markets are competitive and in equilibrium, individuals are rational and fully-informed, and there are no transaction costs.

- Averting behaviour methods involve analysis of averting behaviour (e.g., purchases of bottled water, filters, etc) to value water-related issues faced by households (e.g., poor quality tap water).

Stated preference and revealed preference methods can be combined, either by scaling estimates of one method to another, or by combining datasets, to potentially derive more robust valuations.

Wellbeing approaches infer values from models of subjective wellbeing, bypassing the measurement of individual preferences. There are several applications to value water-related attributes that affect subjective well-being (e.g. drought, floods, water pollution, roadworks).

Value transfer consists of applying in a certain location/time values obtained in a different location/time. It is a convenient technique, as it does not require collection of new primary valuation data, but is limited by differences in the non-market good, the context for valuation and the population, between the target and the original study.

Best practice guidance on stated preference studies

This report provides a summary review of best practice guidance on stated preference studies. At the present, interim, stage, the intention is that this should serve as the

foundation for more detailed, water-specific, guidance that will be developed for the final Stage 1 methodology report.

The UKWIR 2011 guidelines for carrying out WTP surveys in the water industry provide detailed guidance but are now 10 years old. More recent guidance is available from academic studies and documents published by national and international governmental organisations. The main points in common in these guidelines are as follows:

- The choice between contingent valuation and choice experiments depends on the type of information needed by decision-makers and on people's perceptions of the problem. In both cases, the scenario should be clear and plausible. The number of options, attributes, and levels in choice experiments should exhaust the problem, and the experimental design should yield efficient, unbiased, and robust estimates. However, these considerations should be balanced against a need to reduce choice complexity and participant fatigue.

- Preliminary qualitative research should be used to identify the relevant options, attributes, and attribute levels and to test choice questions. Pilot surveys should be used to assess problems (e.g. non-response, anomalous choice behaviour), and estimate preliminary models to test hypotheses and calibrate the experimental design.

- The payment vehicle should be relevant to participants and non-voluntary. Regular (rather lump sum payments) are recommended.

- The questionnaire should probe for the reasons for protest and strategic answers, and non-trading behaviour. It should be ascertained if these issues are individual-specific or systematic across the sample.

- Each survey mode (online, telephone, email, face-to-face) has its own advantages and disadvantages and there is no single best method.

- The sample should be aligned with the population (and the population aligned with the group of individuals potentially affected by the change in the provision of the non-market good). Sampling should be random, with possible quotas for certain groups or areas. Increased effort in recruitment, contacting participants in advance, and financial incentives, can increase response rates. Low non-response among some groups should be identified and addressed.

- Choice models should conform to utility theory and account for preference heterogeneity. Stated preference studies should report estimates of dispersion of the value estimates, along with their central value.

- Studies should include checks and tests of construct validity, examining values in light of the theory, preliminary analyses, and results from previous studies valuing the same non-market good. Studies should also include evaluations of content validity, by identifying behavioural anomalies and asking participant and interviewer feedback.

- All steps should be documented, and all hypotheses should be explicit. Reports should mention the conditions under which the values obtained can be used in value transfer, meta-analyses, study replication, and to support political or business decisions. Data should be made available, where possible, in an anonymised format.

1 Introduction

Accent and PJM economics have been commissioned by Ofwat and CCW to develop a methodology for obtaining the customer evidence to support ODI rate setting for common PCs at PR24 (Stage 1), and to develop and test the materials based on this methodology (Stage 2).

This document forms one part of an initial desk review for the study which, along with a consultation with companies and stakeholders, is intended to lay a solid foundation for the development of the methodology options that will form the principal content of the Stage 1 report.

It looks at two related issues:

- The range of different non-market valuation methods used in the academic literature, with particular focus on their suitability to value water-related attributes; and,
- Best practice guidance for stated preference studies.

At the present, interim, stage, the intention is that this review, which has been conducted in a relatively short space of time, should lay the groundwork for more detailed, water-specific, guidance that will be developed for the final Stage 1 methodology report. Accordingly, it is not intended to be exhaustive nor meticulous at this stage.

The report is structured as follows.

- Section 2 reviews the literature on non-market valuation methods. The methods are reviewed using a similar structure:
  - identification of the main features of the method
  - presentation of examples of applications (in the water industry and other domains)
  - discussion of strengths and weaknesses.
- Section 3 reviews best practice guidance for stated preference studies, covering the main steps of these studies (survey development and implementation, value elicitation, data analysis, validity assessment, and reporting).
- Section 4 concludes the report.

2 Non-Market Valuation Methods

2.1 Introduction

In many appraisals, including most if not all in the water sector, significant costs or benefits arise from an intervention's impact on non-market goods such as environmental quality, health, safety or the risks of a network service failure. The techniques of non-market valuation are applied in these cases, to provide contributing evidence to a specific CBA appraisal or to provide generic evidence, e.g. on the value of a prevented fatality, that is applicable to a range of appraisal contexts.

Non-market valuation has been applied in the context of the UK water company price reviews since 1994. (See Willis and Sheldon 2021 for a review of the history of valuation in the England and Wales water sector.)

It is also common in other domains. For example, it is routinely used to value natural resources and environmental equality, the benefits of improved health, and improvements in travel time and road safety. Accordingly, the academic literature on non-market valuation is extensive.

For cases where market or generic prices are unavailable, valuation methods are typically grouped into three categories:

- Stated preference,
- Revealed preference, and
- Subjective wellbeing approaches.

Combined stated and revealed preference studies are also sometimes undertaken and, additionally, value transfer is a widely used technique in cases where primary valuation is not considered to be cost-effective.

The remainder of this section reviews the methods used in the literature under each of these broad categories.

2.2 Stated preference approaches

Stated preference approaches involve using surveys to ask individuals for their preferences regarding the provision of non-market goods, in hypothetical scenarios. Valuations, in the form of willingness to pay for improvements or willingness to accept deteriorations in exchange for lower bills, can be derived from an analysis of responses.

A variety of stated preference methods have been used in the academic literature, as reviewed in the remainder of this section.

2.2.1. Contingent valuation

Description

Contingent valuation methods consist of surveys asking individuals whether they would be willing to pay/accept a certain amount for a hypothetical policy outcome (i.e. a change in the provision of a non-market good). The method captures the value of a single policy outcome in its entirety, i.e. not disaggregated by its characteristics.

Survey questionnaires first define the hypothetical outcome (i.e., the change in the provision of the non-market good) and the payment vehicle (i.e., how individuals are expected to pay/accept). Usual payment vehicles include changes in taxes (at the national or local level) and changes in the prices of a market good. Contingent valuation questions then ask participants to choose a certain positive/negative outcome and make/receive a payment or choose not to have that outcome and not making/receiving a payment.

Elicitation of values is often achieved through a bid sequence, with participants being asked multiple questions, varying in the amount they are asked to pay/accept. The result is the maximum willingness to pay or minimum willingness to accept of each participant.

Questionnaires also include questions about the participants' demographic and socio-economic characteristics, and perceptions and usual behaviour regarding the payment vehicle and the outcomes in question. These questions help researchers understanding the motivations of participants and how they explain willingness to pay/accept.

The maximum willingness to pay (or minimum willingness to accept) is then modelled, to determine the factors influencing willingness to pay/accept, and to derive mean or median values across all participants.

Contingent valuation can be combined with the analysis of perceptions and attitudes of survey participants, to better understand preferences. For example, Cooper et al (2004) and Aldrich et al (2007) used Likert scales to measure perceptions/attitudes and then statistical methods to synthesise the results into a small number of dimensions. These dimensions were then used in models that explain willingness to pay, obtained with contingent valuation.

There is a long history of applications of contingent valuation, since the 1960s. In fact, contingent valuation was for several decades the main method for non-market valuation. However, the method has often been criticised on the grounds that it is subject to hypothetical bias and/or because responses fail to conform to the tenets of economic theory, such as plausible sensitivity to the scope of the policy/improvement offered (Diamond and Hausman 1994; Hausman 2012; McFadden and Train 2017).

Applications

There is a substantial volume of literature on contingent valuation applications to value water quality and wastewater services. For example, the method has been applied to value water supply for domestic consumption (Genius et al 2008, Justes et al 2014) and

irrigation (Mesa-Jurado et al 2012), and water quality for recreational purposes such as angling (Laitila and Paulrud 2006, Toivonen et al 2004), bathing/swimming (Georgiou et al. 1998; Machado and Mourato 2002) and multiple purposes (Magat et al 2000, Ferrini et al 2014).

Other important areas of application include environmental quality (e.g., air quality), natural resources, outdoor recreation, and health.

Strengths and weaknesses

Contingent valuation is a relatively simple method. Unlike choice experiments, it does not require the definition of a set of attributes and levels, and requires less complex experimental designs and econometric models. Contingent valuation can also estimate non-use values and values of future changes.

However, the method has several weaknesses (Table 1). It cannot be easily used to value policy outcomes involving multiple attributes – the case, for example, of changes in water services, which tend to affect several characteristics of the service simultaneously (e.g. taste, odour). Isolating the valuation of these characteristics would lead to invalid answers, as participant could value one characteristic based on other possible simultaneous changes. Alternatively, multiple independent contingent valuation surveys would be required to value each attribute, resulting in high survey costs.

Several problems can arise due to the hypothetical nature of the outcome in question: the scenario shown (i.e. the description of the outcome and of the payment vehicle) may not be meaningful, plausible, or understandable for participants. Answers may also be invalid because participants state willingness to pay values that are either too high (as a strategy to ensure the policy is implemented) or too low or zero (as protest against the need to pay to derive the benefits of the outcome in question).

Table 1: Contingent valuation: strengths and weaknesses

Strengths	Weaknesses
■ Simple ■ Can estimate non-use values ■ Can estimate values of future changes	■ Difficult to use for the valuation of outcomes involving multiple attributes ■ Scenario may not be meaningful, plausible, or understandable for participants ■ Potential for hypothetical bias ■ Can be susceptible to protest answers (where stated willingness to pay is zero despite good/policy being valued by respondent) ■ Responses can be insufficiently sensitive to the scope of the policy/improvement offered ■ Can lead to unrealistically large differences between willingness to pay and accept

Note: The number of bullet points on each side does not necessarily indicate the relative strengths of the method

Additionally, many studies have shown that contingent valuation estimates can be insufficiently sensitive to the scope of the policy/improvement offered. This finding is particularly significant in studies that value small risk reductions (Hammitt and Graham 1999; Metcalfe and Sen 2021).

The method can also produce implausibly large differences between willingness to pay and willingness to accept.

2.2.2. Choice experiments

Description

Choice experiments involve asking individuals to make hypothetical choices among alternatives. Each alternative is defined by a set of characteristics (defined as alternative values of a set of attributes) (Figure 1). The method captures the values of changes in several attributes, in the context of a policy or a business plan program.

Figure 1: Example of choice experiment question

WATER RELIABILITY
Residential Questionnaire

	PACKAGE A	PACKAGE B
Number of times water is unavailable to your home:	1 time per year	2 times per year
Length of time that water is unavailable to your home each time that it goes off:	8 hours	5 hours
Time of day that water is unavailable to your home each time that it goes off:	Over the weekend	Mon-Fri sometime after 8am
Prior notification that water will be unavailable to your home:	1 day	2 days
Response to phone inquiries in the event of water becoming unavailable to your home:	You get straight through to a PERSON - you are not put on hold and there is no machine directing you to press buttons	You get straight through to a PERSON - you are not put on hold and there is no machine directing you to press buttons
Total Water & Sewerage bill for the year:	**$800**	**$850**

YOUR DECISION: If these were the only 2 options available to you, which option would you choose: Package A or Package B ?

Source and description: From Hensher et al. (2005). The question was extracted from a choice experiment where participants chose from two options, each characterised by different levels of five attributes of water reliability (defined on the left-side), plus the water/sewerage bill.

Choice experiments involve participants making a trade-off between having more or less of certain goods (as defined by attribute levels) and making or receiving a payment. Willingness to pay/accept for the good is the trade-off between money and the provision of the good. In practice, the estimation of willingness to pay/accept involves:

- Estimating a model where the probability of choosing an option depends on the attribute levels of that option (including the cost attribute)
- Estimating willingness to pay for an attribute as the ratio of the coefficient of that attribute and the coefficient of cost.

As with contingent valuation, choice experiments can also be combined with the analysis of perceptions and attitudes of survey participants, to better understand preferences. For example, recently, Hampson et al (2021) combined a choice experiment with Q-methodology (which synthesises viewpoints of a group of individuals), estimating a model which related viewpoints to the attribute levels of the choice experiment.

Choice experiments can also be augmented with other stated preference techniques, to derive valuations of a wider range of attributes or levels (See Section 2.2.3), or combined with revealed preference methods, to derive more robust estimates (See Section 2.4).

The use of choice experiments for non-market valuation has increased in recent years, following developments in econometric models. For example, mixed logit models can account for heterogeneity of preferences across individuals.

Applications

The choice experiment method has been used many times to value multiple water-related attributes (e.g. Hensher et al 2005, Willis et al 2005), drought restrictions (Hensher et al 2006), river flood risk (Brouwer et al. 2016), irrigation water storage infrastructure (Kim et al 2016), general river water quality (Glenk et al 2011), marine water quality (Pakalniete et al. 2017), quality of water for recreational purposes such as fishing (Lee et al 2013, Thangavelu et al 2017), bathing in beaches (Meyerhoff et al 2010, Hynes et al 2013, Penn et al. 2016), and bathing in rivers (Morrisson and Bennett 2004, Perni et al 2012). In most of these studies, water quality was valued as an attribute among other attributes related to the local environment (e.g., biodiversity, green areas).

Applications are very common in many other domains, such as transport, health, and marketing.

Strengths and weaknesses

Table 2 outlines the strengths and weaknesses of using choice experiments for non-market valuation. The use of hypothetical scenarios allows for the estimation of non-use values and values of future changes. However, at the same time, it can lead to hypothetical bias, i.e. survey participants may not take into account all the constraints they would face in a real world situation (Haghani et al 2021), which tends to lead to an overestimation of willingness to pay (Little and Berrens 2004, Murphy et al 2005).

Unlike contingent valuation, choice experiments can be used to derive values for multiple attributes. However, not all policy outcomes can be framed in terms of individual attributes. Describing options in terms of attributes and levels may also oversimplify the policy scenario in question. In addition, participants may not understand the attributes and levels shown.

The complexity of the choice scenarios may also lead participants to make less informed or less rational choices. For example, they may adopt simple heuristics, ignoring some attributes (Hensher et al 2005) or non-trading behaviour (always choosing the same option, regardless of the attribute values shown)

The method assumes that participants choose among options based on the absolute values of the attributes. However, the range of attribute values shown in the experiment may influence choices (Metcalfe and Sen 2021, Bestard and Font 2021). This is particularly the case where the attribute is the risk of some physical or environmental harm. This may happen if preferences are not exogenous to the exercise, but rather constructed or adapted as participants answer the sequence of questions shown (see, e.g., Ariely et al. 2003).

Table 2: Choice experiments: strengths and weaknesses

Strengths	Weaknesses
■ Can estimate non-use values ■ Can estimate value of future changes ■ Can derive values for multiple attributes	■ Hypothetical bias ■ Not all policy outcomes can be framed in terms of attributes ■ Describing options in terms of attributes and levels may oversimplify the problem ■ Scenarios may not be meaningful, plausible, or understandable for participants ■ Complexity of choice situations may lead to simple heuristics or non-trading behaviour ■ Choices may be insufficiently sensitive to the range of values shown

Note: The number of bullet points on each side does not necessarily indicate the relative strengths of the method

2.2.3. Other stated preference methods

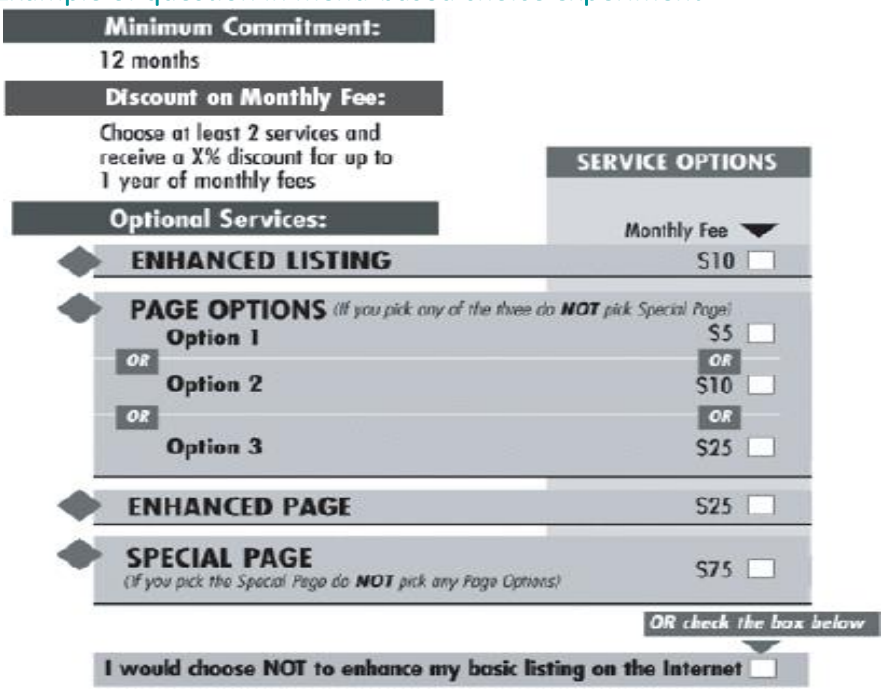
Menu-based choice experiments

Menu-based choice experiments are surveys where respondents choose from a complex set of options, presented in menus. Participants construct their own package of attribute levels from a menu where each level is associated with a certain cost.

This technique is suitable to model choice situations where individuals make multiple choices simultaneously. This is the case of consumer choices in markets where the purchase is typically completed via a menu, such as food, rail tickets, and customised laptops.

Modelling menu-based choices requires complex model specifications, because the items in a menu are not independent. The utility of one menu item is related to the set of items that have also been chosen. This may explain the relatively small number of applications of this method. Liechty et al (2001) was an early example, modelling the choice for purchasing advertisement space on internet pages (Figure 2). More recently, Kamakura and Kwak (2020) modelled the choice for business-to-business information services as a menu-based choice, but did not estimate willingness to pay.

Figure 2: Example of question in menu-based choice experiment



Source and description: From Liechty et al (2011). Menu shows various options for purchasing advertisement space on internet pages

To our knowledge, the approach has not been used for non-market valuation within the academic literature. However, some water companies used menu-based questions, or sliders, in surveys done in the context of the 2019 Price Review (see accompanying Industry Review report)

Impact-weighted valuation

Chalak and Metcalfe (2021) has recently developed a new approach within the same random utility framework as choice experiments that mitigates the problem of sensitivity to the range of values shown in choice experiments where there is a need to trade off multiple small risks of service failures. This new approach has two steps:

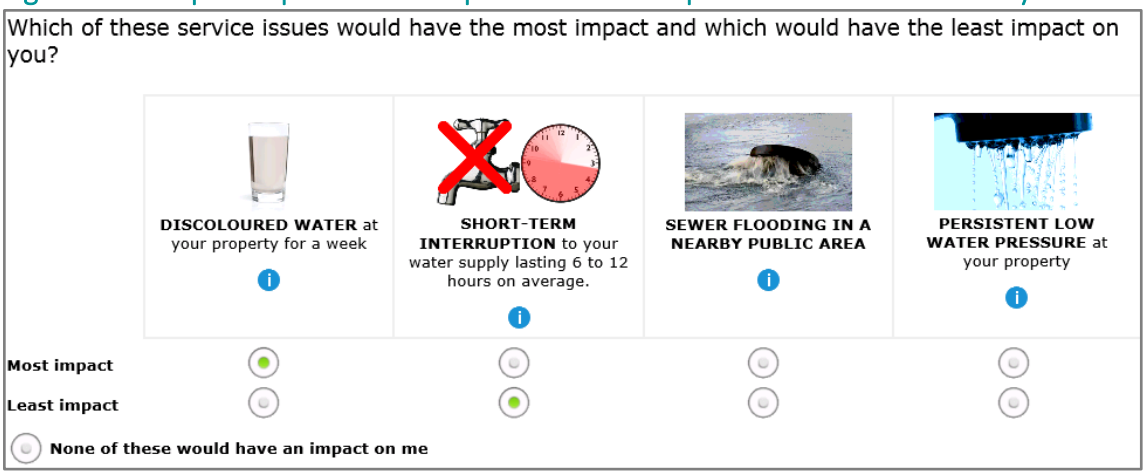
- Estimate the relative impact that the different attributes have on survey participants (based on input from survey participants – see Figure 3)
- Use the relative impacts to apportion the overall package value into values of changes in individual attributes.

This approach was used to value a range of water-related attributes in the context of the 2019 Price Review process for a water company (Welsh Water).

The approach is focused on cases where service levels reflect the risks of different types of service failure, as is the case in water and wastewater services valuation for price reviews. It can produce values per unit of service failure that are proportional to the impact of the attribute, not depending on differences in the scopes of changes shown in the options. Another advantage of the method, compared with standard choice experiments, is a less complex scenario, without a large number of small changes in attribute levels.

The approach is not suitable for attributes with a positive impact, or for a mix of attributes with positive and negative impacts. In addition, the overall size of the estimated values is still dependent on the range of attribute levels. The method also imposes a rational structure to choices, which may not correspond to the real choices individuals would make between packages of attributes.

Figure 3: Example of question on impact within an impact-based valuation study



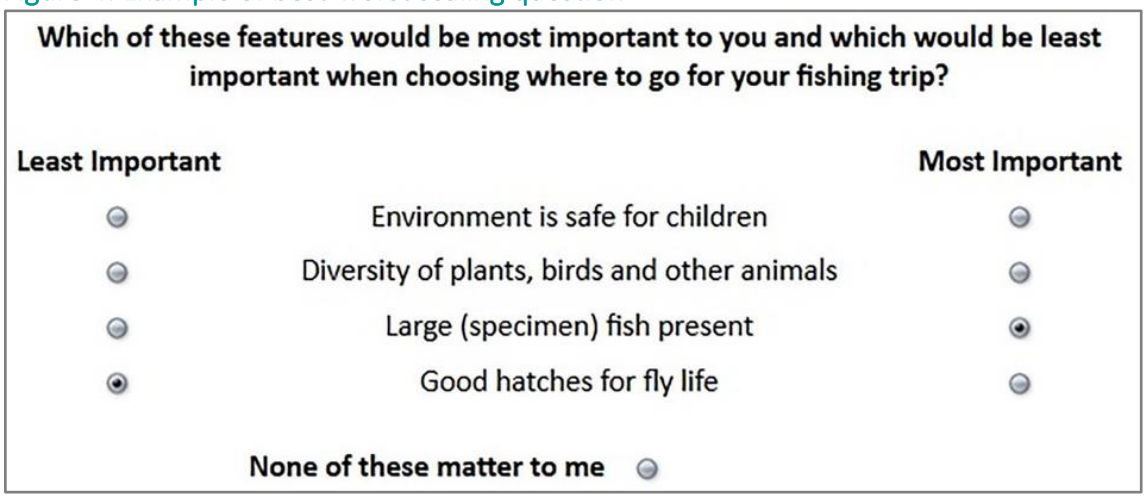
Source and description: From Chalak and Metcalfe (2021). The question asks for the participants' self-assessed impact of a set of water-related attributes

Best-worst scaling

Best-worst scaling consists of surveys asking individuals to choose their most and least preferred options among a subset of a large set of options. Participants are usually asked multiple questions, each showing a different subset of options. There are three variants of this method:

- Type 1 ("Object case"): the options are expressed in terms of levels of an attribute (e.g. agreement with statements)
- Type 2 ("Profile case"): the options are expressed in terms of levels of attribute
- Type 3 ("Multi-profile case") (also known as best-worst discrete choice experiments): the options are expressed in terms of levels of various attributes – see Figure 4.

Figure 4: Example of best-worst scaling question



Source and description: From Anciaes et al (2020). The question was part of a best-worst scaling exercise with 18 attributes related to freshwater angling (only 4 being shown in each question). This exercise was used to derive the relative importance of each attribute. This information was then linked to a choice exercise containing two common attributes, to derive values for all 18 attributes.

Each choice can be decomposed as a series of choices:

- Choice of the most preferred option over all other options
- Choice of each of the 'middle' options (i.e. not the most preferred or the least preferred) over the least preferred option, in a pairwise comparison.

The choices are then modelled as a function of dummy variables representing the different items. The model can be used to estimate odds ratios measuring the relative likelihood that an option is chosen with, versus without, a given attribute. The odds ratios can be rescaled so that they represent the relative importance of each item on a scale from 0 to 1 where 1 is the most preferred item (i.e. the one with the highest odds ratio).

The method does not allow valuation per se, but it can be used to derive the relative values of individual attributes, with the absolute value of a package (i.e. changes in various attributes) being determined by contingent valuation or choice experiments.

Another possibility is to derive absolute values for a package (with a given number of attributes) using a choice experiment and then estimating the absolute value of an additional set of attributes. This approach was used by Anciaes et al. (2020) to derive importance ratios for a set of 18 attributes related to freshwater angling, which were valued via two common attributes valued in a choice experiment (Figure 4).

Best-worst scaling methods have the advantage of producing many observations per participant, allowing for the estimation of robust models even with relatively small samples. It also reduces cognitive burden (compared with choice experiments), as there is no need to consider all options at once

On the other hand, the choices may not be consistent as the participant answers the sequence of questions. Recently, Geržinič et al (2021) showed that participants use

different decisions rules for selecting the best option (they maximise utility) and the worst option (they minimise regret).

Other choice modelling techniques

There are survey-based choice modelling techniques other than choice experiments that can be used to complement valuations produced with choice experiments or contingent valuation. These include:

- Contingent ranking: participants rank a set of options
- Contingent rating: participants score a set of options
- Paired comparisons: participants score pairs of options

In all three cases, the options are defined as a series of attribute levels.

As with best-worst scaling, these methods do not allow valuation per se, but can be used to derive the relative values of individual attributes, with the absolute value of a package determined by contingent valuation or choice experiments. They can also be used to derive absolute values for a package (with a given number of attributes) using a choice experiment and then estimating the absolute value of an additional set of attributes.

Stated preference using citizen juries

Citizen juries are a type of participatory planning approach where a small number of individuals, representing the general public, make a decision with regards to a public policy. They are relatively common in health policy (e.g. Street et al 2014). Jury members discuss the issue over several sessions, are presented with evidence, and can question those presenting that evidence.

Citizen juries could be used to make a decision on the value of a non-market good. Some past applications were group-based contingent valuation exercises. As an example, Macmillan et al (2002) used a citizen jury to value the wild goose conservation and compared the values with those obtained from survey-based contingent valuation. The citizen juries involved two 1h meetings held 1 week apart. Estimated mean willingness to pay was lower for the citizen juries than for participants answering the contingent valuation survey. Lienhoop and Macmillan (2007) used a similar citizen jury approach to value hydro scheme developments in a natural area. The approach performed better than contingent valuation: non-response rate was lower, participants were more engaged, and willingness to pay was more easily explained by socio-economic variables.

Other applications were group-based choice experiments. The study of Álvarez-Farizo et al (2007) valued several water-related attributes (water habitats, river water quality, urban and irrigation supply), comparing individual choices prior to group discussion, individual choices after group discussion, and collective choices after group discussion – the differences were small.

Using citizen juries for valuation of non-market goods potentially allows for better-informed choices than survey-based approaches, as participants are usually provided

with information, can make questions, and discuss with each other. This can also reduce protest answers (Szabó 2011). Citizen juries also give participants more time to think, compared to contingent valuation or choice experiment surveys.

The format is also suited for the achievement of consensus regarding issues that affect society as a whole – participants can better position themselves as citizens, rather than consumers (the implicit hypotheses of other stated preference methods). This may also solve the issues caused by the relationship between willingness to pay and ability to pay. On a practical level, the process of arriving at a value, during citizen jury sessions, also provides information on the attitudes and motivations of participants that determine their preferences and valuations of the non-market good.

However, the method suffers from lack of guidelines on how to obtain values and a lack of theoretical foundations from an economic perspective. As noted above, it is suited to derive social willingness to pay/accept, which is not the aggregate of individual values. As such, the use of citizen juries will tend to arrive at different values than stated preference approaches due to differences in the process of valuation and the type of value elicited (Spash 2007). In addition, citizen juries typically use a small sample, which may not be representative of the population, or allow for the estimation of robust values.

Another weakness of this method is that valuations are influenced by social pressures to conform (within the group). Vargas et al (2017) found that the impacts of social conformity are greater in the case of collective decisions, rather than individual ones. Previous (pre-group discussion) individual preferences also influence the diversity of arguments exchanged during discussion and post-discussion willingness to pay (Völker and Lienhoop 2007).

2.3 Revealed preference approaches

2.3.1. Travel cost method

Description

The travel cost method infers the value of a non-market good through the travel costs incurred by the users to access the site where they consume the good. The assumption is that individuals balance the utility of visiting the site with the disutility of the costs to access the site. The utility of visiting the site depends on the quality of the non-market good individuals consume there. Travel costs include:

- Monetary expenditure, i.e. public transport fares or car use costs (fuel and other). Fuel costs can be defined as a function of distance
- Opportunity costs of travel time – these can be defined as the travel time multiplied by a unit value of travel time (for example, based on wage rates)

The approach relies on surveys, asking participants about frequency of visiting a site or sites, as well as residence location, demographic, and attitudinal questions.

Single site models

The travel cost method can be used for a single site, modelling trip frequency as a function of travel cost. Travel cost can be asked directly to survey participants or estimated from travel distance and time. These can be estimated in a geographic information system, based on location of residences and the site, and assuming certain hypotheses about route choice and travel speed.

Single-site models are usually based on individual behaviour. However, early efforts have also used models aggregated by area (i.e. number of trips per person as a function of average travel costs).

In many cases, surveys are conducted at the recreational site, asking participants about the origin of their trip and the frequency of visiting the site, among other questions.

The method then consists of three steps:

- Calculate travel costs from the residence location of all respondents to the site. Travel costs are usually estimated from travel distance and time
- Estimate a model where number of trips depend on travel costs to access the site, plus individual-level variables. Most recent applications have used count data models (Poisson, negative binomial), which take into account that the number of visits is non-negative. If sampling is conducted on-site, then models also need to be truncated so that number of trips assumes only positive numbers (as all participants made at least one trip).
- Estimate the recreation value as consumer surplus, i.e. the area between the demand curve (the plotted relationship between number of trips and travel cost) and the observed travel costs.

Site choice models

The travel cost method can also be used to model site choice as a function of the characteristics of all sites available, and the travel costs to access the sites – a model known as random utility model (Bockstael et al 1987). In this case, willingness to pay for a certain site characteristic can be derived as the trade-off between changes in that characteristic and travel cost.

The method consists of five steps:

- Collect data on the attributes of all sites (e.g. water quality, facilities). Some variables may be objective (i.e. indicators of water or air pollution), other variables are subjective, based on ratings of visitors to the site (e.g. cleanliness, quality of facilities).
- Define, for each respondent, a choice set of possible sites to visit. These may be all the sites in a defined area, or those at a certain maximum distance.
- Calculate travel costs from the residence location of all respondents to all sites in their choice set. Travel costs are usually estimated from travel distance and time.

- Estimate a model where the probability of choosing a given site depends on the site attributes and the travel costs to access the site.
- Estimate willingness to pay for a site attribute as the ratio of the coefficient of that attribute and the coefficient of travel cost.

Applications

The travel cost method has been mainly used to value outdoor recreation, including water-based recreation. Single-site applications may consider generic outdoor activities (Huhtala and Lankia 2012) or a specific activity such as fishing (Wallentin 2016, Curtis and Stanley 2016). Site choice analyses have been conducted to value attributes of beaches (Lew and Larson 2005), rivers (Melstrom et al 2015, Bateman et al 2016), or both (Anciaes 2021).

The travel cost method has also been used to value general outdoor recreation in natural areas (Hanauer and Reid 2017, Sinclair et al 2020, Cetin et al 2021), and the negative impact of pollution (e.g. visual pollution) on recreation (Kipperberg et al 2019).

Strengths and weaknesses

As with other revealed preference methods, the travel cost method has the advantage of using data based on choices that individuals made in the real-world. However, it is restricted to the valuation of non-market goods consumed at particular sites (hence all the applications have been for outdoor recreation).

The literature has identified several other weaknesses (Table 3). The method is based on past behaviour (visits to sites), so it has limited ability to value future changes, since the real world does not replicate all the possible combinations of changes in all relevant attributes. In addition, it only estimates use value, derived from visiting a site and consuming the non-market good. It does not estimate non-use values (i.e. option, altruistic, bequest, and existence values). Travel cost on-site surveys also typically sample visits, and not visitors, possibly leading to bias towards the behaviour of frequent visitors.

It can also be difficult to assign a travel cost to a specific recreational trip. Individuals may travel to one location for many purposes (recreational and non-recreational), so the travel cost to visit one recreational site can be an inaccurate indicator of the benefit of visiting that site.

Site choice models also require consistent data on all attributes of all sites. Furthermore, the data on some relevant attributes may not be available (for any site). The method also relies on assumptions to estimate travel cost to all pairs of residence locations and sites: for example, the choice of a unit value of travel time. Finally, the calculation of travel distance and/or time to all pairs of residence location and sites is computationally intensive and time consuming, when the number of sites is large.

Table 3: Travel cost method: strengths and weaknesses

Strengths	Weaknesses
■ Based on real-world choices ■ Useful to specifically estimate values revealed by recreational activity	■ Restricted to non-market goods consumed at particular sites ■ Limited ability to value future changes ■ Only estimates use value ■ Samples visits, not visitors – possible bias towards frequent visitors ■ Trips may have more than one purpose ■ Requires consistent data on all attributes in all sites ■ Data on some relevant attributes may not be available ■ Reliance on several assumptions in the calculation of travel cost ■ Calculation of travel distance/time can be computationally intensive and time consuming

Note: The number of bullet points on each side does not necessarily indicate the relative strengths of the method

2.3.2. Hedonic approaches

Description

Hedonic approaches infer the value of a non-market good from the choices that individuals make in markets that incorporate that good. The hypothesis is that the non-market good is implicitly traded as one of several components of the market good.

The most common type of applications uses residential or commercial property prices/rents. These applications rely on the hypothesis that a property is a bundle of attributes, including tangible ones (e.g. property size) and intangible ones (e.g. environmental quality) and that the marginal value of each attribute can be isolated.

Econometric models are estimated to model the market value of a good (for example, property prices) as a function of the attributes hypothesised to influence that value, including non-market attributes.

A second step (not included in most studies) is the construction of the demand curve for the non-market good.

Applications

Hedonic approaches have been extensively applied in the valuation of water quality. A recent meta-analysis by Chen and Hua (2019) found 30 studies around the world valuing river water quality alone. Several other studies have looked at other types of water bodies (e.g. lakes, estuaries). Leggett and Bockstael (2000) estimated the impact of water quality on the value of residences on the waterfront of an estuary, isolating the impact of water quality and other attributes (e.g. visual amenity). Moore et al (2020) found that the

attribute of lake water quality that home buyers value is water clarity. The study of Poor et al (2007) looked at the impact of ambient water quality for an entire local watershed (i.e. not restricted to properties at or near waterfronts).

Some studies integrate hedonic approaches with spatial analysis, to capture how the impact of water quality on property prices varies with distance. Examples include Walsh et al (2011) and Artell (2014).

Hedonic approaches have also been used to value other water-related attributes such as sewage odour and flood risk (Hirsch and Hahn 2018, Beltrán et al 2019).

Hedonic approaches are also common to value other local aspects, including access to transport (Seo et al 2019, Yang et al 2020), noise (Winke 2017), air pollution (Beyer et al. 2009, Mei et al 2020), and natural amenities (Gibbons et al 2014).

The main type of application not using property prices is the value of occupational health and safety, assumed to be implicitly traded in labour markets.

Strengths and weaknesses

The main strength of hedonic approaches corresponds to the main weakness of stated preference approaches (Table 4): hedonic approaches are based on real-world data (actual choices made by individuals), eliminating possible hypothetical bias. The method can also be used for the valuation of multiple attributes.

On the other hand, real-world data also does not include all possible attributes that individuals may consider in their future options. In addition, the method has limited ability to value future changes and only estimates use value, derived from purchasing a market good in order to consume the non-market good. It does not estimate non-use values.

The method also relies on several assumptions: markets capitalise the value of the non-market good, markets are competitive and in equilibrium, individuals are rational and fully-informed (about all the attributes of the market good), and there are no transaction costs or other barriers to choosing the market good that suits one's preferences (e.g. costs of moving from one property to another).

In order to estimate robust models, the method requires large datasets, with many transactions made by many individuals. This data may be difficult to acquire. The property dataset also requires data on many attributes (e.g. distance to various environmental attributes, views from the property), some of which may not be available or require time-consuming collection and processing. The lack of relevant variables leads to omitted variable bias in the models.

Hedonic models can also be difficult to estimate. The value of individual attributes may not be separable from the value of other attributes, due to multicollinearity, i.e. some of the relevant attributes being correlated (such as visual intrusion and noise caused by busy roads). In many cases, there is also no clear delimitation of the relevant spatial extent of

the market and time period for the set of transactions included in the data. The spatial extent of the market may even differ from individual to individual.

Hedonic models relating property prices to property attributes are normally used to estimate the value of marginal changes in those attributes. The estimation of non-marginal changes is more complex, requiring reconstructing the full demand curve.

Table 4: Hedonic approaches: strengths and weaknesses

Strengths	Weaknesses
■ Based on real-world choices ■ Can be used for the valuation of multiple attributes	■ Real-world data does not include all possible combinations of attributes ■ Limited ability to value future changes ■ Only estimates use value ■ Relies on several assumptions about markets and individuals ■ Requires large datasets, with many choices made by many individuals ■ Data on some relevant attributes may not be available ■ Models can be difficult to estimate ■ Spatial extent of the market is not clear ■ Difficult to estimate non-marginal change.

Note: The number of bullet points on each side does not necessarily indicate the relative strengths of the method

2.3.3. Averting behaviour

Description

Averting behaviour (or defensive expenditures) approaches infer the value of a non-market good by the expenditures individuals make to avoid or mitigate the negative aspects of that good (or in other words, to avoid a non-market 'bad'). The level of these expenditures is assumed to be a proxy for the value of reducing the provision of the non-market bad (i.e. a proxy for the willingness to pay to reduce that provision).

Averting behaviour can sometimes involve the use of non-market goods, rather than the consumption of market goods. This is the case when individuals spend time to avert the negative effect. As an example, to avoid consuming poor quality tap water, individuals can spend time collecting spring or underground water (Um et al 2002). These time expenditures can be valued using non-market valuation methods (i.e. deriving value of travel time using stated preference methods).

The method is much less common than travel cost and hedonic methods, which may be because of the absence of averting behaviour in many cases, or because of the method's limitations (see Strengths and Weaknesses section below).

Applications

Several studies have estimated expenditures to mitigate the effects of poor quality drinking water and how those expenditures depend on levels of water quality. Some studies considered only 1-2 types of expenditures (e.g. bottled water, filter equipment) (Abraham et al 2000, Lavee 2010). Lanz and Provins (2016) considered a wider range of expenditures that individuals incur on products that mitigate hardness, taste, smell, and appearance of tap water Expenditures included water softener devices, softening tablets, descaling agents, filtering devices, purchase of bottled water, and addition of squash or cordial. Um et al (2002) analysed some of these expenditures (costs of bottled water and filtering systems) but also added transport and time costs of alternatives to tap water, such as time costs of boiling water, and transport and time costs of using spring water or underground water.

Other water-related applications include expenditures in the context of groundwater contamination (e.g. purchases of bottled water, home water treatment systems, hauling water, boiling water) (Abdalla et al 1992).

The method has also been used to value health and safety, e.g. expenditure of bicycle helmets (Jenkins et al 2001). It is also relatively common in the valuation of local environmental quality. This includes expenditures on double-glazed windows to avoid exposure to noise.

Strengths and weaknesses

As with other revealed preference methods, the analysis of averting behaviour has the advantage of using information from choices made by individuals in the real world (and not in hypothetical scenarios)

However, the method has several weaknesses (Table 5). As with other revealed preference methods, averting behaviour methods cannot value future changes and only estimates use value, derived from purchasing a market good in order to reduce the negative effects of a non-market good. It does not estimate non-use values.

A theoretical weakness is that defensive expenditures represent only the lower bound of willingness to pay to mitigate the negative effect (Bartik 1988). Individuals will only incur the expenditures if the value of reducing the negative effect is higher than the value of the expenditures. Some studies have confirmed that defensive expenditures are lower than willingness to pay derived with contingent valuation methods (Wu and Huang 2001, Rosado et al 2006). However, one recent study found that averting expenditures were higher than willingness to pay for good quality tap water, derived from contingent valuation (Orgill-Meyer et al 2018) – this was explained by the authors as reflecting lack of household confidence in the quality of water.

Another weakness of the averting behaviour method is assuming that individuals are fully informed about the effect they are avoiding/mitigating, and the risks involved.

The averting behaviour may also not be fully effective in mitigating the negative effect. Furthermore, it can have benefits and costs for the individual other than the mitigation

of the negative effect. For example, double-glazing windows to reduce noise may also lead to energy savings. In theory, the estimated value of the negative effect should be net of these other benefits and costs – but it is not easy to model these multiple benefits and costs, due to lack of data.

In addition, as noted previously, not all averting behaviour involves monetary expenditures. The estimation of non-monetary expenditures (i.e. time) involves additional models, for which data may not be available.

Other modelling problems are related to the existence and nature of the link between the negative effect and the averting expenditures. This link is probably mediated by the perceptions that individuals have of the negative effect. This problem was addressed by Lanz and Provins 2017, who disentangled the links between objective level of service, perceived level of service, and averting expenditures. However, this requires complex 2-stage modelling procedures.

Table 5: Averting behaviour: strengths and weaknesses

Strengths	Weaknesses
■ Based on real-world choices	■ Limited ability to value future changes ■ Only estimates use value ■ Expenditures represent only the lower bound of willingness to pay ■ Assumes that individuals are fully informed about the effect they are avoiding/mitigating ■ Expenditures may not be effective in mitigating the effect ■ Averting behaviour may have other benefits (and costs) for the individual ■ May require estimating non-monetary expenditures (i.e. time used) ■ Negative effect and averting expenditures are mediated by perceptions

Note: The number of bullet points on each side does not necessarily indicate the relative strengths of the method

2.4 Combination of stated and revealed preference methods

Stated preference and revealed preference method are often combined, as the main weaknesses of one method are strengths of the other.

One possibility is to scale the stated preference estimates to the revealed preference estimates. For example, Anciaes et al (2020) estimated the value of improvements in the quality of freshwater angling sites in England, combining the results of a choice experiment and a travel cost model. The results of the choice experiment were scaled to the ones from the travel cost model and then calibrated to real-world data.

Another possibility is to pool stated preference and revealed preference datasets in a single model. Whitehead and Lew (2020) found that this approach produces models that have a better fit to the data, while also reducing bias in the stated preference and revealed preference estimates. The method can also be used to derive estimates of both use and non-use values (Eom and Larson 2006).

There are several examples of joint estimation of travel cost data and stated preference data, using pooled datasets. The studies of Cameron (1992) and Adamowicz et al. (1994) were early examples, combining the travel cost method with contingent valuation and choice experiments, respectively. Rosado et al (2006) is a rare example of combination of averting behaviour data with stated preference data (from a contingent valuation exercise), to value drinking water quality.

A third possibility is to combine the travel cost method with a contingent behaviour approach, i.e. asking participants about their future behaviour under different scenarios (Hanley et al 2003, Lankia et al 2019). This allows for the estimation of values for changes outside the observed range of attribute levels.

2.5 Wellbeing approaches

Description

Wellbeing approaches infer values from the estimated relationships between subjective wellbeing, income, and provision of a non-market good.

Subjective wellbeing denotes individuals' perceptions, reported in surveys, about their own wellbeing, or happiness. Most applications have worked with a concept of evaluative subjective wellbeing (i.e. how satisfied individuals are with their life in general or with specific aspects), with the survey question being a variant of "*Overall, how satisfied are you with your life these days?*". There are almost no applications using the more philosophical concept of eudemonic wellbeing (i.e. the idea that one's life has meaning and one is fulfilling their potential) or experienced wellbeing (feeling of wellbeing at a particular point or period in time).

Wellbeing approaches require surveys with questions about aspects of subjective wellbeing, other demographic and socio-economic data (including income), and levels of consumption of some non-market good or service (e.g. environmental quality). The analysis then consists of three steps:

- Estimate a composite indicator of subjective wellbeing
- Estimate a model explaining subjective wellbeing as a function of income, other individual characteristics, and levels of the intangible good
- Estimate the value of a change in the provision of the intangible good as the change in income that would have the same effect on wellbeing and the change in the non-market good.

Wellbeing approaches are becoming more common, following the trend for a stronger focus of public policies on aspects such as wellbeing, health happiness, quality of life, and liveability (see for example APPG 2019).

However, there is still relatively little guidance and unresolved issues (as detailed in the Strengths and Weaknesses sub-section below). Dolan et al (2011) in the UK, and OECD (2013), internationally, produced some recommendations on how to measure subjective well-being for public policy, but not focusing on valuation. SITF (2021) recently published a report on the use of wellbeing approaches, as supplementary guidance for the HM Treasury's guidance on appraisal/evaluation of public policies (Green Book).

Applications

Wellbeing approaches have been used to value water-related issues, including drought (Carroll et al 2009), water pollution (Israel and Levinson 2003), and large-scale floods (Luechinger and Raschky 2009, Fernandez et al 2019).

Recently, Fujiwara et al (2021) used this approach to value the impact of small-scale water flooding incidents for Anglian Water customers. The use of this approach allowed for comparisons between the value of different attributes (i.e. producing results that can be compared to those obtained by stated preference approaches). For example, the cost per incident for flooding was considerably higher than for roadworks (that come as a consequence of accessing underground water infrastructure) and that the cost per property was higher for internal sewer flooding than external sewer flooding and internal water flooding.

There are also several applications in the valuation of other aspects of local environmental quality. This includes air pollution (Welsch 2007, Luechinger 2009), airport noise (Van Praag and Baarsma 2005), scenic amenity (Ambrey and Fleming 2011), green spaces (Tsurumi and Managi 2015), forest fires (Jones 2017), and ecosystem diversity (Ambrey and Fleming 2014). Wellbeing approaches have also been used to value health (Brown 2015, McNamee and Mendolia 2019), and cultural goods (Saz-Salazar et al 2017).

Strengths and weaknesses

Table 6 outlines the main strengths and weaknesses of wellbeing approaches. The main strength is not being subject to hypothetical bias, as the method is based on the individuals' perceptions and experiences, and not on choices made under hypothetical scenarios – an advantage, given that in stated preference studies, individuals are not always fully aware of how the non-market good in question affects their lives. Revealed preference methods also assume the observed choices are made by rationally and fully-informed individuals (i.e. individuals were aware of the implications of their choices), an assumption that is not always met. Wellbeing approaches bypass this assumption.

Wellbeing approaches are also well-suited to value policy outcomes that cannot easily be expressed as a set of attributes.

One weakness is that the method primarily estimates use value (implicit in subjective wellbeing) but not non-use values (i.e. option, altruistic, bequest, and existence values).

It also cannot be easily used to value future changes, as it is based on present or past levels of subjective wellbeing. Furthermore, the method has limited ability to value small, marginal, changes in non-market goods (which have a minor impact on wellbeing).

The method also cannot account for confounding factors that are not possible to control within the study (as it does not rely on an experimental design, unlike stated preference approaches). It is also difficult to rule out reverse causality in models explaining subjective wellbeing. The level of the non-market good (i.e. environmental quality) explains wellbeing. But levels of wellbeing can also influence the consumption of the non-market good (for example, by affecting people's behaviour with regards to residence location choice, daily mobility patterns, relations with others, etc.).

Finally, the measurement of wellbeing poses its own challenges. Participants' answers to wellbeing questions may be influenced by temporary conditions (e.g. weather), not reflecting overall life satisfaction. Participants may also not be able to translate life satisfaction into a number on a scale. The scale may also be regarded as either too wide or too narrow for the participant, leading to unreliable responses.

Table 6: Wellbeing approaches: strengths and weaknesses

Strengths	Weaknesses
■ Based on the individuals' perceptions and experiences - no hypothetical bias ■ Can value policy outcomes that cannot easily be expressed as a set of attributes	■ Does not account for non-use values ■ Limited ability to value future changes ■ Limited ability to value small changes (which have a minor impact on wellbeing) ■ No experimental design: cannot account for all confounding factors ■ Reverse causality: subjective wellbeing is not always a consequence. It can be a cause ■ Difficult to accurately measure subjective wellbeing ■ Subjective wellbeing, as measured, is not necessarily what people would choose to maximise, so results can conflict with what people would choose for themselves

Note: The number of bullet points on each side does not necessarily indicate the relative strengths of the method

2.6 Value transfer

Description

Value transfer (or benefits transfer) consist of valuing a non-market good in a certain location or time using values obtained by a previous study, valuing the same good in a different location and/or time. The approach has four steps

- identify a unit value for the non-market good from a suitable previous study

- collect data on the baseline levels of provision of the non-market good being valued and assume a certain level of future change.
- apply the unit values of the previous study
- aggregate the values to the relevant population consuming the non-market good, or affecting by some of its characteristics. This is achieved by multiplying the unit values by the population size.

The simplest way to apply the values of the previous study (i.e. step 3 in list above) is unit transfer, i.e. applying the mean or median value of the previous study. This approach relies on the similarity between the context of the original and new studies, as this context may be related with the values. This context encompasses:

- the market good (its various attributes, the use people make of it, and the availability of alternatives)
- the characteristics of the population (e.g. age structure, income).
- the scale of the change in provision of the market good (the original study may have been designed to assess small changes while the new study needs to assess large changes, or vice-versa).

To overcome differences in context, the mean/median values can be adjusted to account for differences in the characteristics of the population. A common example is adjusting for average income in the area/time the original values were estimated and in the area/time of the new study. This involves multiplying the original estimates by a function of the average income in the two areas/times.

An extension of this method is to transfer a value function, rather than simply a mean or median value. A value function is a relationship between the value and a series of explanatory variables, including the characteristics of the market good and of the population. This relationship can be imported directly from a single study, or be the result of a meta-analysis of results of several studies. In the latter case, the characteristics of the studies can be included as an additional explanatory variable of the value. The value that applies in the new study can then be calculated as the prediction of the value function using data for the characteristics of the market good and the population in the area/time in question in the new study.

Transfer errors (i.e. the extent to which values transferred from previous studies differs from values obtained by a new study conducted in the site in question) can be reduced by adjustments to socio-economic characteristics of the population (Brouwer et al 2015). In general, distance between the sites tends to increase transfer errors (Kaul et al 2013). Transfer errors also vary with the type of method used in the original study: contingent valuation has been found to produce smaller transfer errors than travel cost methods (Ferrini et al 2014) and choice experiments (Kaul et al 2013). Differences between methods may be even larger than differences between study areas (Ferrini et al 2014).

There is relatively little guidance on how to use value transfer. The main existing guidance is now more than 10 years old: a 2007 book (Navrud and Ready 2007) and a 2010 report

for DEFRA (EFTEC 2010). This guidance emphasises the need to clearly define three main aspects:

- the non-market good to be valued and the relevant population
- the validity and reliability of the values to be transferred from other studies
- the level to which the non-market good valued (and the change in its provision) in the original study match the ones in the new study. This should take into account location, time, and characteristics of the population

Applications

The method is used routinely by practitioners in several fields. Common applications include estimating the value of greenhouse emissions, travel time savings, health, and the value of a statistical life. There are several databases of non-market values for routine use in policy appraisal at a national/regional level. Examples include the UK Department of Transport's TAG data book[1] and the Environmental Valuation Reference Inventory[2], developed by Environment and Climate Change Canada.

In academic studies, value transfer has been used mainly to assess the method itself. In the literature valuing water-related attributes, Hanley et al (2006) and Brouwer et al (2016) assessed transferability of benefits of river restoration, and Ferrini et al (2014) assessed the transferability of values for river water quality.

Strengths and weaknesses

Table 7 shows the main strengths and weaknesses of using value transfer for non-market valuation. Value transfer is a convenient method when resources are limited to collect and analyse data, saving time, money, and effort.

However, the method relies on the quality (i.e. the validity and reliability) of the original study. The context of the non-market good may also be different from context of the good valued in the original study: individuals may have different preferences. This leads to transfer errors when transferring values across countries, even when the non-market good is the same (Rozan 2004, Brouwer et al 2016). Meta-analyses of studies comparing values estimated in a site and those transferred from a different site typically find large means absolute errors (e.g. around 40% in the studies of Ready et al 2004 and Kaul et al 2013).

Table 7: Value transfer: strengths and weaknesses

Strengths	Weaknesses
■ Saves time and money (no need for collect/analyse datasets)	■ Relies on the quality of the original study. ■ Context of application may be different from context of original study

Note: The number of bullet points on each side does not necessarily indicate the relative strengths of the method

[1] https://www.gov.uk/government/publications/tag-data-book
[2] https://www.evri.ca/en

3 Best Practice Guidelines for Stated Preference Studies

3.1 Introduction

This chapter proves a summary review of best practice guidance on stated preference studies. At the present, interim, stage, the intention is that this should serve as the foundation for more detailed, water-specific, guidance that will be developed for the final Stage 1 methodology report.

Section 3.2 is an overview of sources of guidance for stated preference studies. The rest of the chapter synthesises guidance by stage of application of stated preference approaches: survey development and implementation (Section 3.3), value elicitation (Section 3.4), data analysis (Section 3.5), validity assessment (Section 3.6) and reporting (Section 3.7). We follow the structure of Johnston et al. (2017), as a recent authoritative academic source, synthesizing the guidance in that paper, complementing it with other sources, and updating it with material published since 2017.

3.2 Sources

The 2011 guidelines for the water industry (UKWIR 2011) provide detailed guidance but are now 10 years old. The present report updates this guidance with recent developments, from various sources, including academic studies/reports and documents published by governmental organisations.

Several academic studies and reports provide guidance for stated preference studies. Early examples include manuals by Arrow et al. (1993), Bateman et al (2002) and Champ et al (2003). The paper by Johnston et al. (2017) and the book by Mariel et al (2021) include more recent developments and updated guidance. The book by Champ et al (2017) expands and updates the 2003 guidance. Several other papers focus on specific issues.

There are also several relevant documents published by governmental organisations. In the UK, this includes:

- The Green Book, published by HM Treasury[3], provides guidance on appraisal and evaluation of public policies, including the application of stated preference methods

[3] https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-governent

in specific domains: environment/natural, land values, energy efficiency and greenhouse gases, life and health, and travel time.

- ENCA (Enabling a Natural Capital Approach)[4] supplements the Green Book guidance on the non-market valuation of natural capital.

- TAG (Transport Analysis Guidance)[5] provides guidance on non-market valuation of impacts of transport policies (e.g. noise, air quality, and greenhouse gases).

Other countries have also issued guidance for valuing non-market goods and environmental impacts of public policies (e.g. Australia's Productivity Commission report on the valuation of environmental impacts (Baker and Ruting 2014).

International examples include guidance from the World Bank on the estimation of the cost of environmental degradation (Bolt et al 2005) and several documents with guidelines within the context of cost-benefit analysis, e.g. the European Commission guidance to cost-benefit analysis of investment projects (EC 2015), a report for the OECD with guidelines of cost-benefit analysis and the environment (OECD 2018), and guidelines by the Inter-American Development Bank (Dixon 2013).

The following sections synthesise the main points that these documents have in common.

3.3 Survey development and implementation

Survey mode and recruitment

Stated preference surveys can be conducted online, by telephone, by mail, or face-to-face (in public places or the participants' homes). All modes have advantages and disadvantages, and guidance documents do not recommend one mode above the others.

Online surveys are convenient and relatively inexpensive but may lead to the underrepresentation of some groups in the sample because of poorer internet access and level of computer illiteracy among those groups. Recruitment for online surveys is often through commercial panels, i.e. individuals join a panel and then are contacted to participate in specific surveys. However, this may lead to selection bias, if the characteristics of the panel differ from the characteristics of the population. Online surveys may also suffer from reduced engagement from the participants, as the survey competes for attention with email and website notifications in the participant's device.

Telephone-based surveys are not suitable for surveys where participants are asked to answer multiple questions, each with multiple options and attributes. Recruitment is also increasingly difficult, due to refusals and individuals blocking unknown numbers. It is also not possible to link mobile phone numbers to specific areas of residence, reducing the scope for implementing sampling quotas.

[4] https://www.gov.uk/government/publications/enabling-a-natural-capital-approach-enca-guidance/enabling-a-natural-capital-approach-guidance
[5] https://www.gov.uk/guidance/transport-analysis-guidance-tag

Face-to-face surveys have higher costs and may introduce bias due to "interviewer effects" (Leggett et al 2003).

Recruitment is often based on lists of addresses. This can be used for all types of survey modes. Participants are first contacted by mail and asked to answer the survey by mail, telephone, online, or face-to-face.

Sampling

The characteristics of a stated preference sample should be aligned with those of the respective population. In turn, the population should correspond to the group of all individuals potentially affected by the policy in question (which may not necessarily live within a defined administrative jurisdiction).

The sampling process should be random. When recruitment is based on postal address, sampling individuals inside households should also be random. However, quotas can be used for certain groups (e.g. age groups) or areas, so that the characteristics of the sample match those of the population.

Response rates

Low response rates may introduce non-response bias in the analysis, if the group of non-respondents differs from the group of respondents in terms of preferences, or characteristics affecting preferences.

Increasing response rates can be achieved by increased effort in recruitment, contacting participants in advance of the survey or contacting them several times. It can also be achieved by providing financial incentives (e.g. vouchers). However, these incentives may impact sample composition and results.

Demographic or socio-economic patterns in non-response rates should be identified (in pilot surveys) and procedures implemented to address low non-response rates among some groups.

Non-responses to specific questions reduce the dataset available for analysis and can introduce bias if the group of non-respondents differs from the group of respondents with regards to preferences or other relevant characteristics. This is usually the case of income questions. Possible solutions to increase response rates include asking income questions towards the end of the survey and to use categories (rather than asking absolute numbers).

Type of stated preference method

The choice between contingent valuation and choice experiments depends on whether the policy outcome in question can be framed in terms of a single or multiple attributes. This depends on:

- the type of information needed by decision-makers.

- people's perceptions (i.e. if they perceive the outcome in terms of a single or multiple attributes) and preferences (i.e. if they have separate preferences for each attribute). This can be ascertained through preliminary qualitative analysis)

Options

The survey questionnaire should provide a clear explanation of the scenario, including time, location, and how changes will occur (i.e. either the result of a specific policy intervention or not).

The options should be plausible and understandable by participants. Participants should easily judge the impacts of the options for them and their household. The wording of the questions should be understood by participants: jargon should be avoided.

The number of options should describe accurately the set of relevant options faced by political or business decision-makers, but should not be as high as to increase the complexity of the choice option for participants.

If applicable, some of the information can be presented using (easy to understand) photos, maps, or illustrations.

Attributes and levels

Attributes are the characteristics of each option. They can assume different levels, from question to question. In contingent valuation studies, levels are the bid amounts.

The set of attributes and levels shown in the various questions should be based on information from the results of previous studies or from preliminary qualitative research (e.g. focus groups, interviews). All relevant attributes should be included: omitting one may lead to bias in the participants' choices. This need to be balanced against parsimony: participants may not be able to make informed choices when options have too many attributes. This may lead them to develop simple heuristics when making choices (e.g. systematically ignoring some attributes).

Participants should be able to understand how the set of attributes impact on their utility.

The set of levels should cover all the policy-relevant levels, but also allow for the estimation of different models (linear or non-linear) and of interaction effects. All levels should be plausible for the participant. They should also be precise and understandable by participants. Johnston et al (2012) suggests avoiding qualitative terms (e.g. "high", "medium", "low"), unless they are obvious to all participants.

Experimental design

In contingent valuation models, experimental design consists of the sequences of values presented in the bid games. In choice experiments, the experimental design consists of the sets of attribute levels shown in each question to each participant. The sets can be grouped into blocks, shown to different groups of participants.

The experimental design should lead to model estimates that are efficient (with minimum variance), unbiased, and robust to different model specifications.

At the same time, the choice situations should not be too complex, i.e. the participant should be able to consider all attributes shown. Complex questions, or too many questions, lead to participant fatigue and choice inconsistency (De Shazo and Fermo 2002).

The experimental design should be tested in preliminary research (how participants react to the mix of attributes shown in each question) and pilot surveys (whether choices consider all attribute levels). In addition, model coefficients estimated with pilot data should be used to calibrate the experimental design.

Preliminary qualitative research

The design of stated preference questionnaires should be informed by preliminary research. This usually consists of qualitative methods (e.g. interviews, focus groups), with the aims of

- identifying the relevant options, attributes, and attribute levels for the stated preference study.
- testing choice questions

Johnstone et al. (2017) recommends a minimum of 4-6 focus groups, but more for new or difficult-to-quantify non-market goods.

Participants in focus groups and interviews should be representative of key groups (e.g. gender, age groups, socio-economic groups, regions) in the target population.

The qualitative research could include open-ended questions to collect participants' perceptions, attitudes, and preferences. It can also include more structured questions, including showing to participants' examples of the choice questions in the stated preference exercise. In the latter case, feedback should be sought on the clarity of choice scenarios, mix of different attributes in the question, language, and visual representation of the options (if applicable).

Qualitative research can also include interviews with experts.

Pilot surveys

Pilot surveys are surveys with small samples with the purpose of testing the appropriateness of the survey questionnaire, including the stated preference component. More than one pilot survey can be conducted, prior to the main research.

There are no guidelines on sample size for pilot surveys. The sample should be representative of the population.

The analysis of pilot surveys should include checks on:

- survey response rates
- non-response to some questions, protest answers, and non-trading behaviour (participants always choosing the same option in all questions)
- participant feedback on the questionnaire and interviewer feedback (where available) on the engagement the participant has shown
- suitability of the range of attribute levels shown (e.g. whether options with high or low levels are never chosen)
- how the issues above differ by the demographic and geographic profile of participants

The main hypotheses of the survey should be tested using the data from the pilot survey. This allows for the identification of anomalies (e.g. model coefficients with the "wrong" sign). Validity checks can also be performed (see Section 3.6). In addition, as noted previously, results of models estimated with pilot data should also be used to calibrate the experimental design.

The identified issues can be corrected in the main stage of the survey. This may include changes to the experimental design (in the case of choice experiments).

The tests implemented, issues found, and changes made, should be fully documented.

3.4 Value elicitation

The main issue with value elicitation in stated preference studies is the definition of the payment vehicle, i.e. the mechanism through which survey participants pay/accept the payment as trade-off for a larger/smaller account of provision of the non-market good. There is extensive evidence that the type of payment vehicle influences the participants' choices.

The payment vehicle should be relevant to participants (i.e. they can realistically hypothesise that they can be asked to pay or accept the payment in question). It should also be binding (i.e. non-voluntary) – voluntary payments can cause free riding. This should be tested in preliminary qualitative analysis.

Taxes are often used as payment vehicle in stated preference studies. However, not all individuals pay tax, or they may be subject to different tax regimes.

The timing of the payment may also influence participants 'choices. The timing depends on the context of the provision of the non-market good being valued. However, regular payments (rather than lump sum payments) are usually recommended. Egan et al (2015) gives three reasons for this recommendation:

- Participants do not need to perform complicated present value (discounting) calculations if the duration of the payments matches the duration of consumption of the non-market good
- Participants are less likely to face binding mental budget constraints with annual payments than with a large lump sum payment
- Results from contingent valuation surveys with annual payments better match those from travel cost studies

3.5 Data analysis

Modelling

The specification used to model participants' choices should conform to utility theory (e.g. individuals maximise utility subject to a budget constraint).

Models should also account for preference heterogeneity – not doing so may lead to bias in model coefficients. If preferences differ from group to group, a simple solution is to interact variables representing attributes with dummy variables representing groups. If preferences differ by individual, solutions include models where parameters are a function of observed variables (e.g. age, gender, location), or models where parameters are assumed to be random (e.g. mixed logit models). The type of heterogeneity can be identified through information gathered from previous studies, preliminary qualitative research, or tests of different specifications using pilot survey data.

Data analysis should include tests of the robustness of the model specification used.

Behavioral anomalies

Behavioural anomalies are typically defined as preferences expressed by survey participants that do not correspond to their true preferences. However, there are also questions around whether 'true preferences' exist in an objective sense, or whether they are inherently context-specific.

Protest answers are those where participants state they are not willing to pay any amount for the non-market good in questions despite having a positive value for the good. This could be, for example, because they feel the water company makes too much profit already and so water bills should not rise in order to pay for service improvement.

Strategic answers are those where participants answer not according to their preferences (and behaviour intentions) but according to what they believe would lead to their desired policy outcome. For example, participants may express a high willingness to pay in the hope that this contributes to their water company improving service levels. The survey questionnaire should include questions probing for the reasons for protest and strategic answers.

Non-trading behaviour happens when participants always choose the same option, regardless of the attribute levels shown. Non-trading behaviour may be identified within the survey questionnaire, in questions prior to the stated preference exercise, by showing extreme scenarios. A choice for an extreme scenario may signal that the participant is willing to choose any other less extreme scenario, identifying the participant as a non-trader. These participants can be routed away from the stated preference exercise and be asked to answer a different set of questions, focusing on reasons for non-trading behaviour.

Behavioural anomalies can be identified in preliminary qualitative analysis or pilot surveys. The analysis should ascertain if anomalies are individual-specific or systematic across the sample. In the latter case, the analysis should identify possible reasons, as well as the impact on value estimates. The sampling procedure and the questionnaire used in the main stage of the survey should be adjusted to minimise the anomalies.

Value estimation and aggregation

Willingness to pay is usually derived from models of choices made by participants. This is straightforward for simple models, with fixed coefficients (i.e. not random): willingness to pay can be derived as the ratio between the coefficients representing an attribute and the cost coefficient. It is more complex in the case of mixed logit models, where coefficients are random. Assuming a fixed cost coefficient is usually not a realistic option. An alternative is to model utility in willingness to pay space, i.e. to use willingness to pay as the dependent variable.

Stated preference studies should report estimates of dispersion of the value estimates, along with their central value.

The aggregation of values from the sample to the population should ensure that the population corresponds to the group of individuals affected by the policy in question, as mentioned previously.

3.6 Validity assessment

The validity of the values estimated from stated preference studies is the degree to which those values correspond to the individuals' true values. This encompasses three aspects: construct validity, content validity, and criterion validity.

Construct validity

Construct validity is the degree to which the values are consistent with theory and with empirical evidence.

Stated preference studies should include a variety of checks and tests of construct validity, focusing mainly on issues that are well-known in the literature (having been identified in previous studies) or issues that arose in pilot studies or preliminary qualitative research.

Consistency with the theory can be checked by examining values in light of economic theory: individuals make rational choices and maximize utility. In addition, in most cases, utility increases with the amount of the non-market good that is consumed (but that increase is progressively smaller).

These assumptions should be reflected in the model results. The cost coefficient should be significantly negative. The payment reduces the individuals' utility, so an option with a lower cost should be more preferred than one with higher cost, all else equal. Decreasing marginal utility of income can also be tested by including interactions of income (if available) and the cost coefficient in the model.

If attribute levels are entered in models as dummy variables, the coefficients representing increased provision the non-market good should also be higher than those representing lower provision. If attribute levels are entered as numerical variables, then they should have positive coefficients (or negative, in case of a non-market 'bad').

Another common validity check is to test scope sensitivity: higher provision of a non-market good should yield higher willingness to pay. This can be tested by comparing mean values from separate samples (Heberlein et al 2005).

Construct validity can also be assessed by examining the values in light of information obtained from the pilot stated preference study or from preliminary qualitative analysis. The questionnaire can also include rating or ranking exercises, which can be compared with the main choice experiment.

Construct validity encompasses convergent validity: whether values are consistent with those obtained with other studies valuing the same non-market good. This includes previous stated preference studies using the same method, other types of stated preference methods, or revealed preference studies. As shown in several places in Section 2 of this document, different methods often yield different values for the same market good.

Content validity

Content validity is the degree to which the values were estimated with appropriate methods, i.e. methods conducive for estimating the individual's true values. It includes considering all the issues mentioned in Sections 3.3-3.5 of this document, related to the procedures adopted during the survey design and implementation, value elicitation, and data analysis stages of the study.

Stated preference studies should include a variety of evaluations of content validity, again focusing on well-known issues or those that arose in pilot studies or preliminary qualitative research.

Content validity may be checked by identifying anomalies or unexpected patterns, in the way participants answered the questions (e.g. always choosing the same option, high frequencies of the "none of these options" answer, not finishing the exercise).

Content validity can also be evaluated by asking feedback from survey participants on the appropriateness of the stated preference exercise, at the end of the questionnaire. Examples of these questions include: whether the participant understood the scenarios shown, if the attribute levels shown were realistic, and if they were able to make comparisons between options. These questions can be followed by open ended questions asking for reasons.

Follow-up questions can also be placed after the first question(s) in the stated preference exercise, asking for reason for the choice.

In the case of face-to-face or telephone surveys, feedback can also be asked from interviewers, e.g. if they thought participants understood what they were asked to do in the questions, the amount of thought the participant put into responding, and the participant's ability to maintain concentration.

Criterion validity

Criterion validity is the degree to which the estimated values correspond to a presumed true value, obtained in non-hypothetical settings, or through methods that produce better approximations of the true value than stated preference methods. These may be experiments involving actual market transactions (Bishop and Heberlein 1979). Applications of these methods are rare, and when they exist, they can actually replace the need for a stated preference study. For this reason, criterion validity checks are rare in stated preference studies.

3.7 Reporting

Stated preference studies should fully document all the steps described in previous sections, i.e. survey development and implementation, value elicitation, data analysis, and validity assessment.

All the methodological steps of the process should be documented. This includes sampling and recruitment procedures, information used to choose attributes and levels (from previous studies or preliminary qualitative research), type of experimental design. The hypotheses of the study should be explicit

The explanation of all methodological details may be lengthy: in this case, it can be added as appendices to the main report or paper.

Reporting of the results should also describe the conditions under which the values obtained can be used in value transfer exercises, or for supporting political or business decisions. It should also facilitate the use of results in meta-analyses and systematic reviews, as well as validity assessment or replication of the study by other researchers.

Where possible, data should be archived and made available open access in secure repositories, to allow for replication of results, or further analyses by other researchers. Data should be anonymised so that it is impossible to match records with individuals.

4 Conclusions

This report has provided a summary overview of the academic literature on different types of non-market valuation methods and best practice guidance for stated preference research. The main conclusions of this review are as follows.

Non-market valuation methods

Stated preference approaches consist of surveys asking individuals (directly or indirectly) for their preferences regarding a non-market good. These approaches are based on hypothetical scenarios. This allows the valuation of a range of possible future changes, but it can also be a limitation due to "hypothetical bias" in the choices made by survey participants. The two main stated preference method are contingent valuation and choice experiments:

- Contingent valuation consists of asking participants directly for their willingness to pay/accept for a non-market good. The method is relatively simple and was for several decades the main method for non-market valuation.

- Choice experiments have been used extensively to value water-related attributes. They allow for the valuation of multiple attributes, and so are particularly suitable to value improvement packages of water companies, which typically involve changes in several characteristics of the water and wastewater services. The method is also popular in other domains (e.g. environment, transport, health, and marketing). However, there are unresolved methodological issues, e.g.: participants may ignore some attributes or be influenced by the range of attribute levels shown.

- There is a range of other stated preference methods. Some are extensions of choice experiments (menu choice experiments, and impact-weighted valuation), which can be used for non-market valuation (but still have few or no applications). Others (best-worst scaling, contingent ranking, contingent rating, paired comparisons) can be used for non-market valuation when combined with contingent valuation or choice experiments. Others (citizen juries) use or adapt existing valuation methods in a different setting (e.g. a group discussion)

Revealed preference approaches infer preferences from the individuals' choices in real-world markets. These methods have the advantage of not relying on hypothetical scenarios, i.e. they are based on choices that individuals made in the real world. This can also be a limitation, as the real world does not replicate all the possible combinations of changes in all relevant attributes that can be changed by policies or business decisions. Revealed preference methods also cannot easily estimate non-use values and require large datasets (which often have missing data on some variables). Real-world data is also often subject to limitations such as omitted variable or missing data biases.

There are three main revealed preference methods:

- The travel cost method infers the value of a non-market good by the travel costs of the users to access the site where they consume that good. As such, it is mainly used to value aspects of outdoor recreation (including water quality) – it is of limited use to value cases where the consumption of a non-market good does not require travelling (e.g. water quality or flood risk at home).

- Hedonic analysis infers the value of a non-market good by the choices that individuals make in markets that incorporate that good (usually property markets). It has been used to value flood risk and water quality in water bodies near homes. However, the method relies on several assumptions: markets capitalise the value of the non-market good, markets are competitive and in equilibrium, individuals are rational and fully-informed, and there are no transaction costs.

- The analysis of averting behaviour (i.e. individuals' expenditures to reduce the negative effects of non-market goods) is a less common revealed preference method. It is suitable to value water-related issues faced by households, as they can adopt a range of averting behaviours. There are several examples of modelling household expenditures to reduce consumption of tap water perceived to be of bad quality.

The main strengths of stated preference methods are weaknesses of revealed preference method and vice-versa. For this reason, the two methods are often combined in non-market valuation, either by scaling estimates, or by combining datasets, to derive more robust valuations.

Wellbeing approaches infer values from models of subjective wellbeing, assuming that this depends on both income and the provision of non-market goods (among other variables). These approaches are becoming more popular, as they bypass the need to estimate individual preferences (which are not always rational and well-informed). There are several applications to value water-related attributes that affect subjective well-being (e.g. drought, floods, water pollution). As with revealed preference approaches, wellbeing approaches cannot easily estimate non-use values.

Value transfer consist of applying in a certain location/time values obtained (with one of the methods mentioned above) in a different location/time. It is a convenient technique, as it does not require collection of new data, and it is routinely used by practitioners in some fields, given the development of databases of values for policy appraisal. The method is limited by differences in the non-market good and in the population in the original study.

Best practice guidance on stated preference studies

The 2011 guidelines for the water industry provide detailed guidance but are now 10 years old. More recent guidance is available from academic studies and documents published by national and international governmental organisations. The main points in common in these documents are as follows:

- The choice between contingent valuation and choice experiments depends on the type of information needed by decision-makers and on people's perceptions of the problem. In both cases, the scenario should be clear and plausible. The number of

options, attributes, and levels in choice experiments should exhaust the problem, and the experimental design should yield efficient, unbiased, and robust estimates. However, these considerations should be balanced against a need to reduce choice complexity and participant fatigue.

- Preliminary qualitative research should be used to identify the relevant options, attributes, and attribute levels and to test choice questions. Pilot surveys should be used to assess problems (e.g. non-response, anomalous choice behaviour), and estimate preliminary models to test hypotheses and calibrate the experimental design.

- The payment vehicle should be relevant to participants and non-voluntary. Regular (rather lump sum payments) are recommended.

- Choice models should conform to utility theory and account for preference heterogeneity. Stated preference studies should report estimates of dispersion of the value estimates, along with their central value.

- The questionnaire should probe for the reasons for protest and strategic answers, and non-trading behaviour. It should be ascertained if these issues are individual-specific or systematic across the sample.

- Each survey mode (online, telephone, email, face-to-face) has its own advantages and disadvantages.

- The sample should be aligned with the population (and the population aligned with the group of individuals potentially affected by the change in the provision of the non-market good). Sampling should be random, with possible quotas for certain groups or areas. Increased effort in recruitment, contacting participants in advance, and financial incentives, can increase response rates. Low non-response among some groups should be identified and addressed.

- Studies should include checks and tests of construct validity, examining values in light of the theory, preliminary analyses, and results from previous studies valuing the same non-market good. Studies should also include evaluations of content validity, by identifying behavioural anomalies and asking participant and interviewer feedback.

- All steps should be documented, and all hypotheses should be explicit. Reports should mention the conditions under which the values obtained can be used in value transfer, meta-analyses, study replication, and to support political or business decisions. Data should be made available, where possible, in an anonymised format.

References

Abdalla, C., Roach, B., Epp, D. (1992) Valuing environmental quality changes using averting expenditures: an application to groundwater contamination. Land Economics, 68, 163-169.

Abrahams, N A., Hubbell, B J., Jordan, J L. (2000) Joint production and averting expenditure measures of willingness to pay: do water expenditures really measure avoidance costs? American Journal of Agricultural Economics 82, 427-437.

Adamowicz, W L., Louviere, J., Williams, M. (1994) Combining revealed and stated preference methods for valuing environmental amenities. Journal of Environmental Economics and Management 26, 271-292.

Álvarez-Farizo, B., Hanley, N., Barberán, R., Lázaro, A. (2007) Choice modeling at the "market stall": individual versus collective interest in environmental valuation. Ecological Economics 60, 743-751.

Ambrey, C L., Fleming, C M. (2011) Valuing scenic amenity using life satisfaction data. Ecological Economics 72, 106-115.

Ambrey, C L., Fleming, C M. (2014) Valuing ecosystem diversity in South East Queensland: a life satisfaction approach. Ecological Economics 115, 45-65.

Anciaes, P. (2021) Revealed preference valuation of beach and river water quality in Wales. Journal of Environmental Economics and Policy, DOI:10.1080/21606544.2020.1864778

Anciaes, P., Metcalfe, P., Sen, A. (2020) A combined SP-RP model to estimate the value of improvements in freshwater angling in England, Journal of Environmental Economics and Policy, 9:2, 167-187.

APPG (All-Party Parliamentary Group on Wellbeing Economics) (2019) Spending Review to Increase Wellbeing - An open letter to the Chancellor., https://wellbeingeconomics.co.uk/wp-content/uploads/2019/05/Spending-review-to-ncrease-wellbeing-APPG-2019.pdf

Ariely, D., Loewenstein, G., Prelec, D. (2003) "Coherent arbitrariness": Stable demand curves without stable preferences. Quarterly Journal of Economics, 118, 73-105.

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

Artell, J. (2014) Lots of value? A spatial hedonic approach to water quality valuation, Journal of Environmental Planning and Management 57, 862-882.

Atkinson, G., Braathen, N A., Groom, B., Mourato, S. (2018) Cost-Benefit Analysis and the Environment – Further Developments and Policy Use. OECD Publishing, Paris., http://dx.doi.org/10.1787/9789264085169-en

Baker, R., Ruting, B. (2014), Environmental Policy Analysis: A Guide to Non-Market Valuation. Productivity Commission. Staff Working Paper. Productivity Commission, Melbourne., https://www.pc.gov.au/research/supporting/non-market-valuation

Bartik, T. (1988) Evaluating the benefits of no-marginal reductions in pollution using information on defensive expenditures. Journal of Environmental Economics and Management 15, 111-127.

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

Bateman, I., Agarwala, M., Binner, A., Coombes, E., Day, B., Ferrini, S., Fezzi, C., Hutchins, M., Lovett, A., Posen, P. (2016) Spatially explicit integrated modeling and economic valuation of climate driven land use change and its indirect effects. Journal of Environmental Management 181, 172-184.

Bayer, P., Keohane, N., Timmins, C. (2009) Migration and hedonic valuation: the case of air quality. Journal of Environmental Economics and Management 58, 1-14.

Beltrán, A., Maddison, D., Elliott, R. (2019) The impact of flooding on property prices: a repeat-sales approach. Journal of Environmental Economics and Management 95, 62-86.

Bestard, A B., Font, A R. (2021) Attribute range effects: preference anomaly or unexplained variance? Journal of Choice Modelling 41:100321.

Bishop, R C., Heberlein, T A. (1979) Measuring values of extramarket goods: are indirect measures biased? American Journal of Agricultural Economics 61, 926-930.
Bockstael, N E., Hanemann, W M., Kling, C L. (1987) Estimating the value of water quality improvements in a recreational demand framework. Water Resources Research 23, 951-960.

Bolt, K., Ruta, G., Sarraf, M. (2005) Estimating the cost of environmental degradation. World Bank Environmental Department Papers 106.

Brouwer, R., Bliem, M., Getzner, M., Kerekes, S., Milton, S., Palarie, T., Szerényi, Z., Vadineanu, A., Wagtendonk, A. (2016) Valuation and transferability of the non-market benefits of river restoration in the Danube River Basin using a choice experiment. Ecological Engineering 87, 20-29.

Brouwer, R., Martin-Ortega, J., Dekker, T., Sardonini, L., Andreu, J., Kontogianni, A., Skourtos, M., Raggi, M., Viaggi, D., Pulido-Velazquez, M., Rolfe, J., Windle, J. (2015) Improving value transfer through socio-economic adjustments in a multicountry choice experiment of water conservation alternatives. Australian Journal of Agricultural and Resource Economics 59, 458-478.

Brown, T T. (2015) The subjective well-being method of valuation: an application to general health status. Health Services Research 50, 1996-2018.

Cameron, T. (1992) Combining contingent valuation and travel cost data for the valuation of non-market goods. Land Economics 68, 302-317.

Carroll, N., Frijters, P., Shields, M A. (2009) Quantifying the costs of drought: new evidence from life satisfaction data. Journal of Population Economics 22, 445-461.

Cetin, N I., Bourget, G., Tezer, A. (2021) Travel-cost method for assessing the monetary value of recreational services in the Ömerli Catchment. Ecological Economics 190: 107192.

Chalak, A., Metcalfe, P. (2021) Valuing water and wastewater service improvements via impact-weighted numbers of service failures. Forthcoming in Journal of Environmental Economics and Policy.

Champ, P A., Boyle, K J., Brown, T C. (Eds.) (2003) A primer on nonmarket valuation. Kluwer Academic Publishers, Dordrecht.

Champ, P A., Boyle, K J., Brown, T C. (Eds.) (2017) A primer on nonmarket valuation – 2nd edition. Springer, Dordrecht.

Chen, W Y., Li, X., Hua, J. (2019) Environmental amenities of urban rivers and residential property values: a global meta-analysis. Science of The Total Environment 693: 133628.

Chrzan, K., Peitz, M. (2019) Best-Worst Scaling with many items. Journal of Choice Modelling 30, 61-72.

Cooper, P., Poe, G L., Bateman, I J. (2004) The structure of motivation for contingent values: a case study of lake water quality improvement. Ecological Economics 50, 69-82.

Curtis, J., Stanley, B. (2016) Water quality and recreational angling demand in Ireland. Journal of Outdoor Recreation and Tourism 14, 27-34.

De Shazo, J R., Fermo, G. (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency. Journal of Environmental Economics and Management 44, 123-43.

Del Saz-Salazar, S., Navarrete-Tudela, A., Alcala-Mellado, J., Del Saz-Salazar, D C. (2017) On the use of life satisfaction data for valuing cultural goods: a first attempt and a comparison with the contingent valuation method. Journal of Happiness Studies 20, 119-140.

Diamond, P A., Hausman, J A. (1994) Contingent valuation: is some number better than no number? Journal of Economic Perspectives 8, 45-64.

Dixon, J A. (2013) Economic Cost-Benefit Analysis (CBA) of Project Environmental Impacts and Mitigation Measures: Implementation Guideline. IDB Technical Note IDB - TN – 428, IDB., https://publications.iadb.org/en/economic-cost-benefit-analysis-cba-project-environmental-impacts-and-mitigation-measures

Dolan, P., Layard, R., Metcalfe, R. (2011) Measuring subjective well-being for public policy: recommendations on measures. Centre for Economic Performance Special Paper No. 23, London School of Economics and Political Science, London, http://eprints.lse.ac.uk/47518/1/CEPSP23.pdf

EC (European Commission) (2015) Guide to Cost-Benefit Analysis of Investment Projects,https://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/cba_guide.pdf

EFTEC (2010) Valuing environmental impacts: practical guidelines for the use of value transfer in policy and project appraisal. Report for DEFRA., https://www.cbd.int/financial/values/unitedkingdom-tech-report.pdf

Egan, K J., Corrigan, J R., Dwyer, D F. (2015) Three reasons to use annual payments in contingent valuation surveys: convergent validity, discount rates, and mental accounting. Journal of Environmental Economics and Management 72, 123-36

Eom, Y-S., Larson, D M. (2006) Improving environmental valuation estimates through consistent use of revealed and stated preference information. Journal of Environmental Economics and Management 52, 501-516.

Fernandez, C J., Raitzer, D., Ginting, E. (2019) Potential use of the life satisfaction approach to value nonmarket goods and services. Asian Development Bank Economics Working Paper Series 569., https://www.think-asia.org/bitstream/handle/11540/9605/ewp-569-life-satisfaction-approach-nonmarket-goods-services.pdf?sequence=1

Ferrini, S., Schaafsma, M., Bateman, I. (2014) Revealed and stated preference valuation and transfer: a within-sample comparison of water quality improvement values. Water Resources Research 50, 4746-4759.

Fujiwara, D., Houston, R., Keohane, K., Maxwell, C., Van Emmerik, I. (2021) Applying the wellbeing valuation method to value the costs of roadworks and flooding, Journal of Environmental Economics and Policy, DOI: 10.1080/21606544.2021.1938688

Genius, M., Hatzaki, E., Kouromichelaki, E M., Kouvakis, G., Nikiforaki, S., Tsagarakis, K P. (2008) Evaluating consumers' willingness to pay for improved potable water quality and quantity. Water Resources Management 22, 1825-1834.

Georgiou, S., Langford, I H., Bateman, I J., Turner, R K. (1998) Determinants of individuals' willingness to pay for perceived reductions in environmental health risks: a case study of bathing water quality. Environment and Planning A: Economy and Space 30, 577-594.

Geržinič, N., Van Cranenburgh, S., Cats, O., Lancsar, E., Chorus, C. (2021) Estimating decision rule differences between 'best' and 'worst' choices in a sequential best worst discrete choice experiment. Journal of Choice Modelling 41: 100307.

Gibbons, S., Mourato, S., Resende, G. (2014) The amenity value of English nature: a hedonic price approach. Environmental and Resource Economics 57, 175-196.

Glenk, L., Lago, M., Moran, D. (2011) Public preferences for water quality improvements: implications for the implementation of the EC Water Framework Directive in Scotland. Water Policy 13, 645-662.

Haghani, M., Bliemer, M C J., Rose, J M., Oppewal, H., Lancsar, E. (2021) Hypothetical bias in stated choice experiments: Part I. Macro-scale analysis of literature and integrative synthesis of empirical evidence from applied economics, experimental psychology and neuroimaging. Journal of Choice Modelling 41: 100309.

Hammit, J. K., Graham, J. D. (1999) Willingness to pay for health protection: Inadequate sensitivity to probability? Journal of Risk and Uncertainty, 8, 33-62.

Hampson, D I., Ferrini, S., Turner, R K. (2021) Assessing subjective preferences for river quality improvements: combining Q-methodology and choice experiment data, Journal of Environmental Economics and Policy, DOI:10.1080/21606544.2021.1879682

Hanauer, M M., Reid, J. (2017) Valuing urban open space using the travel-cost method and the implications of measurement error. Journal of Environmental Management 198, 50-65.

Hanley, N., Bell, D., Alvarez-Farizo, B. (2003) Valuing the benefits of coastal water quality improvements using contingent and real behaviour. Environmental and Resource Economics 24, 273-285.

Hanley, N., Wright, R E., Alvarez-Farizo, B. (2006) Estimating the economic value of improvements in river ecology using choice experiments: an application to the Water Framework Directive. Journal of Environmental Management 78, 183-193.

Hausman, J. (2012) Contingent valuation: from dubious to hopeless. Journal of Economic Perspectives 26, 43-56.

Heberlein, T A., Wilson, M A., Bishop, R C., Schaeffer, N C. (2005) Rethinking the scope test as a criterion for validity in contingent valuation. Journal of Environmental Economics and Management 50, 1-22.

Hensher, D A., Rose, J., Greene, W H. (2005) The implications on willingness to pay of respondents ignoring specific attributes. Transportation 32, 203-222.

Hensher, D., Shore, N., Train, K. (2005) Households' willingness-to-pay for water service attributes. Environmental and Resource Economics 32, 509-531.

Hensher, D., Shore, N., Train, K. (2005) Water supply security and willingness to pay to avoid drought restrictions. The Economic Record 82, 56-66.

Hirsch, J., Hahn, J. (2018) How flood risk impacts residential rents and property prices: empirical analysis of a German property market, Journal of Property Investment and Finance 36, 50-67.

Huhtala, A., Lankia, T. (2012) Valuation of trips to second homes: do environmental attributes matter? Journal of Environmental Planning and Management 55, 733-752.

Hynes, S., Tinch, D., Hanley, N. (2013) Valuing improvements to coastal waters using choice experiments: an application to revisions of the EU Bathing Waters Directive. Marine Policy 40, 137-144.

Israel, D., Levinson, A. (2003) Examining the relationship between household satisfaction and pollution. Presented at the Eastern Economics Association Meeting, 23 February 2003.

Jenkins, R R., Owens, N., Wiggins, L B. (2001) Valuing reduced risks to children: the case of bicycle safety helmets. Contemporary Economic Policy 19, 397-408.

Johnston, R J., Boyle, K J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T A., Hanemann, W M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., Vossler, C A. (2017) Contemporary guidance for stated preference studies. Journal of the Association of Environmental and Resource Economists 4:2, 319-405

Johnston, R J., Schultz, E T., Segerson, K., Besedin, E Y., Ramachandran, M. (2012) Enhancing the content validity of stated preference valuation: the structure and function of ecological indicators. Land Economics 88, 102-120.

Jones, B A. (2017) Are we underestimating the economic costs of wildfire smoke? An investigation using the life satisfaction approach. Journal of Forest Economics 27, 80-90.

Justes, A., Barberán, R., Farizo, B A. (2014) Economic valuation of domestic water uses. Science of The Total Environment 472, 712-718.

Kamakura, W A., Kwak, K. (2020) Menu-choice modeling with interactions and heterogeneous correlated preferences. Journal of Choice Modelling 37: 100214.

Kaul, S., Boyle, K J., Kuminoff, N V., Parmeter, C F., Pope, J C. (2013) What can we learn from benefit transfer errors? Evidence from 20 years of research on convergent validity. Journal of Environmental Economics and Management 66, 90-104.

Kim, H N., Boxall, P C., Adamowicz, W L. (2016) Analysis of the impact of water quality changes on residential property prices. Water Resources and Economics 16, 1-14.

Kipperberg, G., Onozaka, Y., Bui, L T., Lohaugen, M., Refsdal, G., Sæland, S. (2019) The impact of wind turbines on local recreation: Evidence from two travel cost method – contingent behavior studies. Journal of Outdoor Recreation and Tourism 25, 66-75.

Kountouris, Y., Remoundou, K. (2011) Valuing the welfare cost of forest fires: a life satisfaction approach. Kyklos 64, 556-78.

Laitila, T., Paulrud, A. (2006) A multi-attribute extension of discrete-choice contingent valuation for valuation of angling site characteristics. Journal of Leisure Research 38, 133-142.

Lancsar, E., Louviere, J., Donaldson, C., Currie, G., Burgess, L. (2013) Best worst discrete choice experiments in health: methods and an application. Social Science and medicine 76, 74-82.

Lankia, T., Neuvonen, M., Pouta, E. (2019) Effects of water quality changes on the recreation benefits of swimming in Finland: combined travel cost and contingent behavior model. Water Resources and Economics 25, 2-12.

Lanz, B., Provins, A. (2016) The demand for tap water quality: survey evidence on water hardness and aesthetic quality. Water Resources and Economics 16: 52-63.

Lanz, B., Provins, A. (2017) Using averting expenditures to estimate the demand for public goods: combining objective and perceived quality. Resource and Energy Economics 47, 20-35.

Lavee, D. (2010) Averting expenditures and valuation of damages: two methods for assessing the benefits of water filtration in Israel. Water Policy 12, 290-303.

Lee, D. E., Hosking, S G., Preez, M. (2013) A choice experiment application to estimate willingness to pay for controlling excessive recreational fishing demand at the Sundays River Estuary, South Africa. Water SA 44, 39-48.

Leggett, C G., Bockstael, N E. (2000) Evidence of the effects of water quality on residential land prices. Journal of Environmental Economics and Management 39, 121-144.

Leggett, C G., Kleckner, N S., Boyle, K J., Dufield, J W., Mitchell, R C. (2003) Social desirability bias in contingent valuation studies administered through in-person interviews. Land Economics 79, 561-75.

Lew, D. L., Larson, D M. (2005) Valuing recreation and amenities at San Diego County beaches. Journal of Coastal Management 33, 71-86.

Liechty, J., Ramaswamy, V., Cohen, S H. (2001) Choice menus for mass customization: an experimental approach for analyzing customer demand with an application to a web-based information service. Journal of Marketing Research 38, 183-196.

Lienhoop, N., MacMillan, D C. (2007) Contingent valuation: comparing participant performance in group-based approaches and personal interviews. Environmental Values 16, 209-232.

Little, J., Berrens, R. (2004) Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis. Economics Bulletin 3, 1-13.

Luechinger, S. (2009) Valuing air quality using the life satisfaction approach. Economic Journal Conference Papers 119, 482-515.

Luechinger, S., Raschky, P A. (2009) Valuing flood disasters using the life satisfaction approach. Journal of Public Economics 93, 620-633.

Machado, F S., Mourato, S. (2002) Evaluating the multiple benefits of marine water quality improvements: how important are health risk reductions? Journal of Environmental Management 65, 239-250

Macmillan D C., Philip, L., Hanley, N., Alvarez-Farizo, B. (2002) Valuing the non-market benefits of wild goose conservation: a comparison of interview and group based approaches. Ecological Economics 43, 49-59.

Magat, W A., Huber, J., Viscusi, W K. (2000) An iterative choice approach to valuing clean lakes, rivers, and streams. Journal of Risk and Uncertainty 21, 7-43.

Mariel, P., Hoyos, D., Meyerhoff, J., Czajkowski, M., Dekker, T., Glenk, K., Jacobsen, J B., Liebe, U., Olsen, S B., Sagebiel, J., Thiene, M. (2021) Environmental Valuation with Discrete Choice Experiments. Springer, Cham.

McFadden, D., Train, K. (eds.) (2017) Contingent valuation of environmental goods: A comprehensive critique. Elgar, Cheltenham, UK.

McNamee, P., Mendolia, S. (2019) Changes in health-related quality of life: a compensating income variation approach. Applied Economics 51, 639-650.

Mei, Y., Gao, L., Zhang, J., Wang, J. (2020) Valuing urban air quality: a hedonic price analysis in Beijing, China. Environmental Science and Pollution Research 27, 1373-1385.

Melstrom, R. T., Lupi, F., Esselman, P C., Stevenson, R J. (2015) Valuing recreational fishing quality at rivers and streams. Water Resources Research 51, 140-150.

Mesa-Jurado, M A., Martin-Ortega, J., Ruto, E., Berbel, J. (2012) The economic value of guaranteed water supply for irrigation under scarcity conditions. Agricultural Water Management 113, 10-18.

Metcalfe, P J., Sen, A. (2021) Sensitivity to scope of water and wastewater service valuations: a meta-analysis of findings from water price reviews in Great Britain, Journal of Environmental Economics and Policy, DOI: 10.1080/21606544.2021.1984314

Meyerhoff, J., Dehnhardt, A., Hartje, V. (2010) Take your swim suit along … – the value of improving urban bathing sites in the Metropolitan Area of Berlin. Journal of Environmental Planning and Management 53, 107-124.

Moore, M R., Doubek, J P., Xu, H., Cardinale, B J. (2020) Hedonic price estimates of lake water quality: valued attribute, instrumental variables, and ecological-economic benefits. Ecological Economics 176: 106692

Morrisson, M., Bennett, J. (2004) Valuing New South Wales Rivers for use in Benefit Transfer. The Australian Journal of Agricultural and Resource Economics 48, 591-611.

Murphy, J J., Allen, P G., Stevens, T H., Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation. Environmental and Resource Economics 30, 313-25.
Navrud, S., Ready, R. (2007) Environmental Value Transfer: Issues and Methods. Springer, Dordrecht.

OECD (2013) OECD Guidelines on Measuring Subjective Well-being, OECD Publishing, Paris

OECD (2018) Cost-benefit Analysis and the Environment – Further Developments and Policy Use. OECD, Paris., https://www.oecd.org/env/cost-benefit-analysis-and-the-environment-9789264085169-en.htm

Orgill-Meyer, J., Jeuland, M., Albert, J., Cutler, N. (2018) Comparing contingent valuation and averting expenditure estimates of the costs of irregular water supply. Ecological Economics 146, 250-264.

Pakalniete, K., Aigars, J., Czajkowski, M., Strake, S., Zawojska, E., Hanley, N. (2017) Understanding the distribution of economic benefits from improving coastal and marine ecosystems. Science of the Total Environment 584-585, 29-40.

Penn, J., Hu, W., Cox, L., Kozloff, L. (2016) Values for recreational beach quality in Oahu, Hawaii. Marine Resource Economics 31, 47-62.

Perni, A., J. Martinez-Paz, Martínez-Carrasco, D. (2012) Social preferences and economic valuation for water quality and river restoration: the Segura River, Spain. Water and Environment Journal 26, 274-284.

Poor, P J., Pessagno, K L., Paul, R W. (2007) Exploring the hedonic value of ambient water quality: a local watershed-based study. Ecological Economics 60, 797-806.

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

Rosado, M., Cunha-e-Sa, M., Ducla-Soares, M M., Nunes, L C. (2006) Combining averting behavior and contingent valuation data: an application to drinking water treatment in Brazil. Environment and Development Economics 11, 729-746.

Rozan, A. (2004) Benefit transfer: a comparison of WTP for air quality between France and Germany. Environmental and Resource Economics 29, 295-306.

Seo, K., Salon, D., Kuby, M., Golub, A. (2018) Hedonic modeling of commercial property values: distance decay from the links and nodes of rail and highway infrastructure. Transportation 46, 859–882

Sinclair, M., Mayer, M., Woltering, M., Ghermandi, A. (2020) Valuing nature-based recreation using a crowdsourced travel cost method: A comparison to onsite survey data and value transfer. Ecosystem Services 45: 101165.

SITF (Social Impacts Task Force) (2021) Wellbeing Guidance for Appraisal: Supplementary Green Book Guidance., https://www.gov.uk/government/publications/green-book-supplementary-guidance-wellbeing

Soekhai, V., Donkers, B., Levitan, B., D Bekker-Grob, E W. (2021) Case 2 best-worst scaling: for good or for bad but not for both. Journal of Choice Modelling 41: 100325.

Spash, C L. (2007) Deliberative monetary valuation (DMV): issues in combining economic and political processes to value environmental change. Ecological Economics 63, 690-699.

Street, J., Duszynski, K., Krawczyk, S., Braunack-Mayer, A. (2014) The use of citizens' juries in health policy decision-making: a systematic review. Social Science and Medicine 109, 1-9.

Szabó, Z. (2011) Reducing protest responses by deliberative monetary valuation: Improving the validity of biodiversity valuation. Ecological Economics 72, 37-44.

Thangavelu, T., Paulrud, A., Stage, J. (2017). Understanding heterogeneous preferences for angling site attributes: application of a choice experiment. Journal of Environmental Economics and Policy 6, 324-340.

Toivonen, A. L., Roth, E., Navrud, S., Gudbergsson, G., Appelblad, H., Bengtsson, B., Tuunainen, P. (2004) The economic value of recreational fisheries in Nordic Countries. Fisheries Management and Ecology 11, 1-14.

Tsurumi, T., Managi, S. (2015) Environmental value of green spaces in Japan: an application of the life satisfaction approach. Ecological Economics 120, 1-12.

UKWIR (UK Water Industry Research) (2011) Carrying Out Willingness to Pay Surveys. UKWIR, London.

Um, M-J., Kwak, S-J., Kim, T-Y. (2002) Estimating willingness to pay for improved drinking water quality using averting behavior method with perception measure. Environmental and Resource Economics 21: 287–302, 2002.

Van Praag, B., Baarsma, B. (2005) Using happiness surveys to value intangibles: the case of airport noise. Economic Journal 115, 224-246.

Vargas, A., Lo, A Y., Rohde, N. Howes, M. (2017) Social influences on expressed willingness to pay: results of a deliberative monetary valuation study in Colombia. Journal of Environmental Planning and Management 60, 1511-1528.

Völker, M., Lienhoop, N. (2007) Exploring group dynamics in deliberative choice experiments. Ecological Economics 123, 57-67.

Wallentin, E. (2016) Choice of the angler: estimating single-site recreation demand using revealed preference data. Tourism Economics 22, 1338-1351.

Walsh, P J., Milon, J W., Scrogin, D O. (2011) The spatial extent of water quality benefits in urban housing markets. Land Economics 87, 628-644.

Welsch, H. (2007) Environmental welfare analysis: a life satisfaction approach. Ecological Economics 62, 544–551.

Whitehead, J C., Lee, D K. (2020) Estimating recreation benefits through joint estimation of revealed and stated preference discrete choice data. Empirical Economics 58, 2009-2029.

Whitehead, J C., Lew, D K. (2020) Estimating recreation benefits through joint estimation of revealed and stated preference discrete choice data. Empirical Economics 58, 2009-2029.

Willis, K G., Scarpa, R., Acutt, M. (2005) Assessing water company customer preferences and willingness to pay for service improvements: a stated choice analysis. Water Resources Research 41, DOI:10.1029/2004WR003277

Willis, K., Sheldon, R. (2021) Research on customers' willingness-to-pay for service changes in UK water company price reviews 1994-2019, Journal of Environmental Economics and Policy, DOI: 10.1080/21606544.2021.1927850.

Winke, T. (2017) The impact of aircraft noise on apartment prices: a differences-in-differences hedonic approach for Frankfurt, Germany, Journal of Economic Geography 17, 1283-1300.

Wu, P I., Huang, C L. (2001) Actual averting expenditure versus stated willingness to pay. Applied Economics 33, 277-283.

Yang, L., Chau, K W., Szeto, W Y., Cui, X., Wang, X. (2020) Accessibility to transit, by transit, and property prices: Spatially varying relationships. Transportation Research Part D: Transport and Environment 85: 102387.