Sensitivity to scope of water and wastewater service valuations: A meta-analysis of findings from water price reviews in Great Britain[1]

P.J. Metcalfe[23] and A. Sen[2]

15 August, 2021

Abstract

At the 2014/15 water price reviews in Great Britain, a notionally similar stated preference methodology was used across multiple customer surveys to derive willingness to pay (WTP) estimates for comparable service improvements. Very different valuations were obtained for the same service measures, however, raising questions regarding their validity and reliability. This paper examines the variation in those WTP estimates via a meta-analysis of household WTP values from 18 water companies for five common service measures. Our main finding is that 59% of the observed variation in WTP estimates could be explained by the scope of service change offered. WTP decreased substantially with the service changes offered for valuation, a finding that is inconsistent with expected utility theory, the standard economic theory of rational choice under uncertainty, but is consistent with predictions from prospect theory, and with empirical evidence from related fields. The study also finds that WTP increased with the number of households supplied, a finding which is considered likely to be due to altruism, and with GDP per capita, and decreased with the number of attributes included in the study design. Significant risk framing effects are also identified.

Keywords: willingness to pay; stated preference; discrete choice experiments; scope sensitivity; water services; meta-analysis

---

1 Funding for this research was provided by Scottish Water.

2 PJM Economics, 1 Walford Close, Wimborne, Dorset, BH21 1PH, UK.

3 Corresponding author. E-mail: paul@pjmeconomics.co.uk; Tel: +44 7786 656 834

1. Introduction

In the past two decades, multi-attribute stated preference surveys, most notably discrete choice experiments, have been extensively used worldwide to value the benefits accruing to customers from water and wastewater service improvements (Price et al., 2019; Kanyoka et al., 2009; Snowball et al., 2008; Vásquez and Adams, 2019; Latinopoulos, 2014; Birol and Das, 2010; Dutta, 2009; Ibrahim et al., 2019; Nur Syuhada et al., 2020; Wahid and Hooi, 2014; Wang et al., 2018), and Australia (MacDonald et al., 2010; Hensher et al., 2005; Hensher et al., 2006; Wilson et al., 2020).

In England and Wales, water and wastewater companies have made extensive use of willingness to pay (WTP) evidence elicited by means of stated preference surveys to derive economically justified performance commitments and outcome delivery incentives for their business plan submissions to the economic regulators of the water industry (Chalak et al., 2013; Hecht et al., 2015; Lanz and Provins, 2015, 2016; Reid et al., 2010; Sayles et al., 2021; Scarpa et al., 2005; Willis et al., 2005).

However, significant concerns have been raised by Ofwat (The Water Services Regulation Authority in England and Wales) in recent price reviews regarding the excessive variability of valuations for similar service measures. At the 2014 England and Wales price review (PR14), for example, household willingness to pay to avoid one property being affected by one internal sewer flooding incident ranged from £22,530 to £367,291, with similarly wide ranges reported for other values per unit of service failure (Accent, 2013). Ofwat cited these differences as a reason to question the stated preference survey methodology in its Water 2020 consultation (Ofwat 2015) and, more recently, has recommended that the industry conducts a national study, rather than having separate studies for each company, in part, because 'a*t PR19, companies' highly divergent estimates of customers' willingness to pay for similar services posed a real challenge in terms of understanding whether these reflected genuine differences in preferences and, if not, which values better reflected customers' views*.' (Ofwat, 2020,p.30.)

An explanation of the divergence in valuations is essential to understand whether the WTP results genuinely reflect customer preferences or whether they are driven by features of the survey design and are, hence, not reliable with respect to alternative reasonable designs. Moreover, an understanding of the drivers of variation can indicate how, if at all, designs might be controlled to ensure economically optimal outcomes for customers and/or an equitable regulatory treatment across companies. Concern in this regard is particularly pertinent in the England and Wales water industry because WTP estimates are used to determine outcome delivery incentive rates – the amount of additional revenue water companies are able to earn through higher customer water charges for achieving higher service levels than the performance commitment level set for them at the price review or, conversely, the penalty paid through lower customer water charges for achieving lower service levels (Ofwat, 2017). Thus, markedly differing WTP estimates across companies for common service measures can give rise to a correspondingly varied treatment across companies in terms of the link between service levels and revenue.

Probably the most significant and sustained challenge to the validity and reliability of stated preference valuations concerns the inadequate sensitivity of valuations to the scope of improvement offered (Diamond and Hausman, 1994; Kahneman and Knetsch, 1992; Hausman 2012; Desvouges et al 2012; Andersson et al, 2016). This challenge is particularly acute with respect to the valuation of small risk reductions such as involved in the valuation of a statistical life (VSL) (Beattie et al., 1998; Hammitt and Graham, 1999; Carson, 2012; Andersson et al, 2016). According to the standard economic theory of rational choice under risk, expected utility theory, WTP for small risk reductions should be near-proportional to the size of the risk reduction offered, and independent of the starting risk (Jones-Lee, 1974; Hammit and Graham, 1999). Thus, for example, a reduction in the risk of death from 3 in 100,000 to 1 in 100,000 should be valued at twice the amount as a reduction in the risk from 3 in 100,000 to 2 in 100,000, (ignoring a negligible income effect), and the same amount as a reduction from 4 in 100,000 to 2 in 100,000. Since VSL is calculated as WTP divided by the risk reduction, this implies a near-constant VSL across studies of the same population with different risk reductions.

In practice, however, while some stated-preference studies find this degree of sensitivity (e.g., Corso et al. 2001; Hammitt and Haninger 2010), VSL estimates tend to be highly sensitive to the scope of the risk reduction offered in the survey (Beattie et al., 1998; Hammit and Graham, 1999; Lindhjem et al, 2011; Gyrd-Hansen et al., 2012; Søgaard et al., 2012; Andersson et al, 2016; Balmford et al, 2019). These findings imply that standard stated preference-derived VSL estimates can be highly dependent on the choice of risk reduction included within the survey design and, if so, should be considered unreliable with respect to alternative reasonable designs.

These findings are directly relevant to the valuation of water and wastewater service levels as these are often measured by the frequency of service failures of various kinds; or, equivalently, by the average risks of a customer experiencing such service failures (e.g. Hensher et al., 2005; Metcalfe and Baker, 2012; Price et al. 2019; Willis et al., 2005; Wang et al,. 2018; Reid et al., 2010; Lanz and Provins, 2015). The sensitivity of valuations to the scope of risk change offered is therefore highly pertinent to the issue of whether water and wastewater service valuations can be considered valid and reliable. A finding that values for avoided water and wastewater service failures vary significantly with respect to the scope of service change offered suggests that these values too might be considered unreliable.

A core focus of the present study is consequently to examine the sensitivity of water and wastewater service valuations to the scope of service change offered. It does so via a meta-analysis of WTP findings from 18 water companies, including all 11 water and sewerage companies in Great Britain. These studies were used to help justify the companies' business plans that they submitted to their economic regulator - to Ofwat for the 2014 water price review (PR14) for all English and Welsh water companies; and to the Water Industry Commissioner for Scotland (WICS) for the 2015 strategic review of charges (SRC15) in the case of Scottish Water.

The study reports are not all publicly available but were obtained directly from the companies for the purposes of this study. They offer a good data source for analysis due to the fact that many aspects of the survey designs were very similar to one another. This was due to their common adherence to an industry set of guidelines for carrying out such surveys (UKWIR, 2011) published shortly prior to the studies commencing. (See Willis and Sheldon, 2021, for a discussion of the context behind the commissioning of these guidelines.) Although some of the studies departed from the guidance, there were a sufficient number of similarly defined service measures, and a sufficiently similar valuation methodology, to enable a worthwhile meta-analysis to explore the drivers of variation in water and wastewater service valuations.

The remainder of the paper proceeds as follows. Section 2 describes the data for the meta-analysis study; Section 3 discusses the variables included in the modelling and theoretical expectations regarding their influence on WTP; Section 4 provides a discussion of our econometric methodology; Section 5 presents the results of our analysis; Section 6 provides a discussion of these results and concludes.

2. Data

For the purposes of this study, permission was sought from all but the very smallest water companies in Great Britain to incorporate their PR14/SRC15 WTP research within the analysis. There were no refusals to cooperate by any water companies. However, two companies had not undertaken comparable stated preference WTP research for PR14 and so could not be included. In total, data for 18 water companies were included in the research, including all 11 water and sewerage companies.

The PR14/SRC15 studies for the 18 contributing water companies all included discrete choice experiments to obtain valuations for improvements in service levels for between 4 and 16 service measures each. Our analysis focusses on household WTP values for five of these service measures: discoloured water, unexpected supply interruptions (3-6 hours and 6-12 hours) and external and internal sewer flooding. These service measures were selected on the basis that they were the ones most commonly included by companies in their studies. WTP values for discoloured water were reported by 16 water companies; WTP values for internal and external sewer flooding were reported by 11 and 10 water companies respectively; and WTP values for unexpected supply interruptions (3-6 hours and 6-12 hours) were reported by 9 and 7 water companies respectively.

The definitions for the service measures differed somewhat across companies. For example, while some companies described service interruptions as an incident that is likely to prevail for a period of 3-6 hours, some other companies specified the exact duration of the incident as say 4 hours. For the purposes of our analysis, we included WTP values in the latter cases within the unexpected supply interruptions (3-6 hours) service measure category. Overall, the service measures were sufficiently similarly defined across studies to be considered comparable within their category, while acknowledging that differences in the definitions and descriptions of service measures introduce an inevitable source of error into comparisons.

The format of the discrete choice experiment questions also differed somewhat across companies, although all adhered to some extent to the guidelines set out in UKWIR (2011). An example choice card from the UKWIR (2011) guidelines is shown in Figure 1. The key differences across studies comprised: the combinations of service measures included; whether a status quo option was included in every choice question; whether visual assists were used to help convey risk sizes; and whether risks were presented as '1 in x' or as 'x in n' ratios, with a common 'n' value across attribute levels. In respect of cases where an 'x in n' ratio was used, some studies consistently used the number of households in the company's supply area as the 'n' value, while others used the smallest multiple of 10 capable of discriminating between the different service levels used.

Figure 1: Example discrete choice experiment question

CHOICE EXPERIMENT EXERCISE A1 (Base Set)			
		OPTION A	OPTION B
1	**AN UNEXPECTED INTERRUPTION to the water supply at your property lasting between 6-12 hours.** The chance that this happens at your property in any one year	**5 in 1,000**	**1 in 1,000**
2	**A BAN ON USING THE HOSE PIPE at your property that would typically last for 5 months beginning in May and ending in September.** The chance that this happens at your property in any one year	**1 in 10**	**1 in 20**
3	**You will be told that FOR UP TO 2 DAYS YOU WILL NEED TO BOIL TAP WATER from your property first before drinking, cooking or preparing food. It may affect your health if you do not boil the water before drinking..** The chance that this happens at your property in any one year	**1 in 5**	**1 in 50**
4	**THE WATER AT YOUR PROPERTY HAS A TASTE AND ODOUR that is not ideal for a week at a time, but it is safe to drink.** The chance that this happens at your property in any one year	**10 in 1,000**	**This does not happen**
5	**THE WATER IS DISCOLOURED AT YOUR PROPERTY for a week at a time, but it is safe to drink. Running the tap for several minutes will not remove the discolouration.** The chance that this happens at your property in any one year	**15 in 1,000**	**This does not happen**
6	**THE CHANGE IN YOUR ANNUAL WATER AND SEWERAGE BILL to provide the service quality above.** The new bill level will also apply in all later years.	**Increase of £5 each year for 5 years, from £350 in 2015 to £375 by 2020**	**Increase of £10 each year for 5 years, from £350 in 2015 to £400 by 2020**

Source: UKWIR (2011), Appendix 10.

Typically, each study included four levels for each attribute, including one deterioration level, a base level, a '+1' medium improvement level, and a '+2' maximum improvement level. From the study reports we directly extracted or derived a 'WTP (Base to +1)' value, equal to the per-household valuation, on a pounds per year bill impact basis of the risk reduction from the Base level to the '+1' level reported, where these risks represented annual chances of experiencing the service issue in question. A 'Unit WTP' value was also directly extracted or derived for each service measure included in the study by dividing WTP (Base to +1) by the size of the risk reduction from the Base level to the '+1' level. For example, if WTP was found to be £5 per household for a risk reduction from '2 in 1000' to '1 in 1000', i.e. a risk reduction of 1/1000, then Unit WTP was calculated by dividing £5 by 1/1000, which equals £5,000. This approach mirrors the calculation of VSL estimates in the context of health and safety valuations (see, for example, Cropper et al., 2011).

Within the econometric analysis we consider a range of variables to explain variations in Unit WTP, in addition to constant terms (dummy variables) to capture underlying values for avoiding each service issue. These factors, along with prior expectations as to their influence on WTP estimates, are detailed as follows.

Base risk level and size of risk reduction

Under expected utility theory, Unit WTP is expected to be perfectly inelastic with respect to base risk level and the size of the risk reduction. To see this, let $V(w, p)$ represent an individual's indirect utility function where w is the person's wealth and p is their risk of experiencing a service failure of some kind, say a short supply interruption. Then, under expected utility theory:

$$V(w, p) = (1 - p)v_0(w) + pv_1(w) \tag{1}$$

i.e. utility is equal to the probability weighted average of utilities in the scenarios with no supply interruption $v_0(w)$ and when a supply interruption occurs, $v_1(w)$. This implies that:

$$Unit\ WTP = \frac{dw}{dp} = \frac{dV}{dp}\Big/\frac{dV}{dw} = \frac{v_1(w) - v_0(w)}{(1-p)v_0'(w) + pv_1'(w)} \tag{2}$$

The denominator is the expected marginal utility of wealth. If this is the same in both scenarios then equation (2) simplifies to $\frac{v_1(w)-v_0(w)}{v'(w)}$, which implies that Unit WTP is proportional to the impact of the supply interruption itself but independent of both the base risk level and the size of the risk reduction.

Although expected utility theory is generally considered to be a good normative theory of decision making under risk in the sense that it corresponds to what a rational person would choose given some basic assumptions about preferences, it is known to fail as a descriptive model of choice in some general and persistent ways. An important one of these ways is the 'certainty effect'. First formally identified by Allais (1953), and henceforth known as the ‘Allais paradox’, researchers found that a desire for certainty meant that the difference between 0.01 and 0, for example, was valued much higher than the difference between 0.11 and 0.1, despite the risk reduction being the same in both cases.

A variety of additional systematic departures from expected utility theory have been observed and modelled in work by psychologists and behavioural economists, which all call into question the expected utility theory prediction of near-constant Unit WTP values. The literature is large and cannot be adequately covered here; the following briefly summarises some of the most relevant findings.

- First, there is substantial evidence that preferences are often constructed in the course of questioning rather than being data retrieved from memory and can, accordingly, be reversed by asking questions in different ways (Lichtenstein and Slovic, 2006). Moreover, when constructing their preferences, people are influenced by a range of factors that should not matter, such as the way in which probabilities and outcomes are framed (Lichtenstein and Slovic, 1971; Tversky and Kahneman, 1981; Slovic, 1995; Grether and Plott, 1979), and by irrelevant ‘anchors’ (Ariely et al, 2003; 2006) but are not influenced sufficiently by factors that should matter, such as the probability of some bad outcome happening (Sunstein, 2003). Because of this, we might expect to see insufficient scope sensitivity in WTP (Base to +1) valuations across studies in relation to expected utility theory predictions.

- When considering probabilities, people often mistakenly consider relative risk changes rather than absolute risk changes. For example, people tend to value a risk reduction from 0.1 to 0.05 more than a reduction from 0.5 to 0.45 even though these would be equivalent under expected utility theory (Baron, 1997). The implication is that Unit WTP values would be negatively correlated with base risk levels because a given-sized risk reduction will appear as a smaller relative risk change from a larger base and will hence command a lower valuation.

- People often tend to make financial decisions within a narrow frame that is more consistent with having multiple ‘mental accounts’ than a single account. For example, people may constrain one kind of purchase when its budget has run out while not constraining another kind of purchase, even though both expenditures draw on the same fungible resource (income). (Thaler, 1985, 1999). If money is 'spent' from a narrow budget rather than via an optimisation constrained only by overall wealth, WTP (Base to +1) valuations will be less than proportionally correlated with the 'Base to +1' scope offered, and hence Unit WTP valuations will be negatively correlated with scope.

Prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) explains the certainty effect, and other commonly observed departures from expected utility theory via two amendments. The first of these is the π function, which transforms probabilities into 'decision weights'. These tend to give greater weight to certainty in comparison with other probabilities. In general, empirical findings suggest that people overweight small probabilities and underweight large probabilities in their decisions. The second amendment is to evaluate outcomes as changes from a reference point rather than from a stable utility function, consistent with the narrow framing of mental accounting (Kahneman, 2003). In general, people value losses greater than gains, and display diminishing sensitivity to both gains and losses (of money). More generally, however, prospect theory no longer necessarily predicts near-perfect sensitivity of WTP to scope.

The number of attributes valued:

Although not expected to be significant under expected utility theory, the number of attributes valued could have a negative impact on WTP for individual service measures for the same reason that we expect to find a less than proportional increase with respect to the scope of service change: where there are more attributes being valued, each attribute is expected to be valued less, in general, due to a less-than-proportional sensitivity to scope. If the number of attributes significantly impacts on Unit WTP, this again means that values are sensitive to design choices and are accordingly unreliable with respect to alternative survey designs.

The number of households in the company area:

If individuals are focussed solely on their own private risk when completing the valuation questions, there should be, under expected utility theory, no significant impact of company size on Unit WTP. Company size simply acts as the denominator in the risk measure and so does not impact valuation in this case. However, there are two potential reasons why WTP may be linked to the number of properties supplied. Firstly, if survey participants are altruistically concerned for the number of people experiencing a service failure rather than simply their own personal risk, then company size acts as a multiplier to the outcome of interest, holding risk change constant, and will hence be positively correlated with Unit WTP. Widespread evidence from a range of fields confirms the existence of altruism as a motivator of human behaviour (Batson, 2011; Fehr and Fischbacher, 2003; Warneken and Tomasello, 2009), including its role in driving valuations of public goods (Viscusi et al., 1988; Arana and Leon, 2002; Zhang et al, 2013).

An alternative reason why the number of households in the company area may drive valuations concerns 'denominator neglect', the widely observed phenomenon whereby people focus on the numerator of a probability and disregard the denominator (Alonso and Fernández-Berrocal, 2003; Garcia-Retamero et al., 2010; Okan et al., 2012; Passerini et al., 2012; Mikuskova, 2015). In the context of our study this would imply that WTP might depend on the absolute number of properties affected by a service failure, controlling for the size of the risk reduction, even if the size of their personal risk reduction was all that mattered to them due to the fact that this represents the numerator of the risk measure in a large number of cases. In these cases, participants may simply ignore the denominator and mistakenly focus on the numerator as the sole measure of risk change. It is consequently necessary to consider the impact of risk framing effects as a potential driver of choices, as discussed below.

Risk framing effects

In order to test for the presence of risk framing effects, we consider three variables:

- The size of the denominator used to represent risk. Where different denominators are used, eg reductions in risk from 1/10 to 1/20, the average of these amounts is used, ie 15.
- The difference in the risk numerator values between Base and +1 levels.
- The difference in the denominator values between Base and +1 levels.

Under expected utility theory, the impact of these two variables on the WTP should be negligible with risk changes controlled for. However, the theory of denominator neglect suggests that the difference in the numerator should have an additional positive effect on WTP after controlling for the size of the risk change.

On the other hand, the size of the denominator used to represent risk has been shown in previous studies to impact on WTP, albeit with mixed findings. Pinto-Prades et al (2006) and Balmford et al. (2019) both find this variable to have a positive impact on WTP, consistent with denominator neglect; however, Zhai and Suzuki (2008) find the opposite effect.

The size the average water bill:

In principle, the size of the bill could have either a positive or negative effect on WTP. On the positive side, a high average bill might make further increases look small in comparison and so appear more acceptable. This would be an instance of narrow framing, consistent with mental accounting. On the other hand, high average bills might mean people are reluctant to pay more per se. Such an outcome could be driven by participants having a wider frame of reference, for example the bills they pay for other utilities, than simply focusing on the relative increase in their water bill.

The effect of package scaling of values:

Differences in study designs and methodologies employed across the customer preference studies could potentially have a systematic effect on WTP findings. One potentially significant difference across studies relates to the use, or otherwise, of a 'package scaling factor'. Where applied, this involves scaling WTP estimates from discrete choice experiments to be consistent with the value measured for a full package of improvements, as obtained from a separate package valuation exercise. The rationale for this scaling is that estimates of the value of a package of improvements can often be higher when obtained by adding up the individual attribute WTP estimates as derived from a set of lower level discrete choice experiments than when obtained directly as a full package valuation (Hanley et al. 1998; Cameron et al. 2002; Bateman et al. 2004; Mogas et al. 2009). The package effect appears to depend on context, however; with some studies finding no significant difference (Jin et al, 2006; Colombo et al, 2006; Tuan and Navrud, 2007) or a mixed set of comparisons (Adamowicz et al, 1998; Foster and Mourato, 2003; Christie and Azevedo, 2009).

In the sample of PR14 studies obtained for the present study, 11 reported results that had been package scaled, with the remainder reporting unscaled results. In order to control for this difference in methodology, we include a package effects dummy variable which equals 1 if a study reports scaled WTP values and equals 0 otherwise. This is expected to enter the model with either a zero or a negative coefficient since scaling would usually be expected to result in similar or lower Unit WTP values all else equal.

GDP per capita of company area:

It is generally to be expected that WTP increases with income, all else equal. We hence include GDP per capita of the area served by the water company as an explanatory factor and expect it to enter the model with a positive coefficient.

In order to obtain this data, an image showing water company borders was georeferenced in ArcGIS and overlaid with a shapefile with the borders of NUTS3 regions. The GDP per capita for each water company was then calculated as the weighted average of the NUTS3 GDP per capita data over all NUTS3 regions inside its border, with weights based on the relative size of each NUTS3 region.

Data summary statistics

Table 1 provides descriptive statistics for WTP (Base to +1), Unit WTP, base risk and risk reduction by service measure. As shown in the table, base risk levels were small for each of the attributes, reaching a maximum of 0.1 (1 in 10) in the case of discoloured water with the mean base risk even for this service measure being below 2 in 100. In the case of the least frequent service issue, internal sewer flooding, base risks were extremely small, ranging from 0.0001 (10 in 100,000) to 0.0004 (40 in 100,000). Consistent with this, the size of the reductions was even smaller, ranging from 0.00001 (1 in 100,000) to 0.00012 (12 in 100,000) in the case of internal sewer flooding, a notable 12-fold difference between minimum and maximum.

Unit WTP values varied much more substantially than WTP (Base to +1) values. For example, mean WTP (Base to +1) ranged narrowly from £2.18 to £4.18 across the five service measures while Unit WTP ranged much more widely from £477 to £126,017 across the same measures. The highest Unit WTP values corresponded to external and internal sewer flooding, while the base risk and risk reduction values for these service measures were also the smallest.

Table 1: Summary statistics for WTP (Base to +1), Unit WTP, base risk and risk reduction by service measures

Service measure	Statistic	N	Mean	Std dev	Min.	Max.
Discoloured water	WTP (Base to +1)	16	£3.26	£2.03	£0.52	£6.12
	Unit WTP	16	£2,502	£3,708	£109	£15,061
	Base risk	16	0.019	0.027	0.0007	0.1
	Risk reduction	16	0.005	0.008	0.0002	0.03
External sewer flooding	WTP (Base to +1)	10	£2.95	£3.88	£0.53	£13.66
	Unit WTP	10	£23,523	£49,164	£2,869	£162,570
	Base risk	10	0.0019	0.0010	0.0005	0.0033
	Risk reduction	10	0.0002	0.0002	0.00002	0.001
Internal sewer flooding	WTP (Base to +1)	11	£4.18	£2.91	£0.99	£9.50
	Unit WTP	11	£126,017	£100,336	£22,530	£367,291
	Base risk	11	0.0002	0.0001	0.0001	0.0004
	Risk reduction	11	0.00005	0.00004	0.00001	0.00012
Unexpected supply interruptions (3-6 hours)	WTP (Base to +1)	9	£2.18	£1.74	£0.55	£6.56
	Unit WTP	9	£477	£601	£50	£1,670
	Base risk	9	0.033	0.0231	0.0049	0.071
	Risk reduction	9	0.009	0.007	0.0012	0.03
Unexpected supply interruptions (6-12 hours)	WTP (Base to +1)	7	£3.37	£2.24	£0.71	£6.34
	Unit WTP	7	£2,678	£4,429	£459	£12,672
	Base risk	7	0.0107	0.0116	0.0007	0.029
	Risk reduction	7	0.0036	0.004	0.0004	0.011

Note: WTP (Base to +1) values are in £ per household per year for an improvement from Base service level to Level +1. Unit WTP values are measured in £ per unit. The mean value of Unit WTP equals the average value of the ratio of WTP (Base to +1) and Risk reduction and not the ratio of the average value of WTP (Base to +1) to the average value of Risk reduction.

Table 2 presents summary statistics for the additional explanatory variables used in the econometric analysis. It shows the average number of households served by the 18 water companies was around 1.7 million, the average household bill was around £352 and the average number of attributes in the company studies was approximately 11. The average GDP per capita of the areas served by the 18 water companies was around £24,634 and around 57% of the WTP estimates reported by water companies were package-scaled. The table shows that the difference in the numerator of the risk variable ranges from 0 to 50,000 while the difference in the denominator of the risk variable ranges from 0 to 1,100. It is pertinent to mention here that that 42 observations in our sample have a zero difference in denominator values while only 12 observations in our sample have a zero difference in numerator values for the risk variable.

Table 2: Summary statistics for additional explanatory variables

Variable	N	Mean	Std dev	Min	Max
Number of attributes	53	10.887	2.025	4	13
Number of households	53	1,771,688	1,304,100	259,129	5,100,507
Average bill (£)	53	£352.89	£113.24	96	499
GDP per capita (£)	53	24,634	3,609	18,571	34,269
Package scaling dummy	53	0.566	0.500	0	1
Risk denominator	53	1,071,211	1,547,158	16	5,100,507
Risk numerator difference	53	-1,698	7,425	-50,000	0
Risk denominator difference	53	1,071,211	1,547,158	16	5,100,507

4. Econometric Methodology

Following previous meta-analysis studies, (e.g. Lindhjem et al. 2011), we focus on a model with a log-log specification. The dependent variable is *ln(Unit WTP)*$_{ij}$, the natural logarithm of Unit WTP for service measure *i* from water company *j*; We include the following explanatory variables, based on our discussions in previous section:

- *D: external sewer flooding*; *D: internal sewer flooding*; *D: supply interruption (3-6 hrs.)* and *D: supply interruption (6-12 hrs.)*: service measure constant terms, equal to 1 for the service measure indicated; equal to 0 otherwise. The omitted service measure is *D*: *discoloured water*, which is captured by the model's constant term.

- *ln(Base risk)*: *discoloured water* to *ln(status quo risk): supply interruption (6-12 hrs)*: the log of the status quo (Base) risk level for each of the service measures.

- *ln(risk reduction): discoloured water* to *ln(risk reduction): supply interruption (6-12 hrs)*: the log of the risk reduction from Base to +1 levels for each of the service measures.

- *ln (no. of attributes*): the log of the total number of attributes included in the company study.

- *ln* (*no. of households*): log of the total number of households served by the water company.

- *ln* (*average bill*): log of the average household bill for the water company in question.

- *ln* (*GDP per capita*): log of GDP per capita for the areas served by the water company

- *Package-scaled*: a dummy variable equal to 1 if the WTP value was package-scaled and equal to 0 otherwise.

- *Risk denominator (millions)*: The average of the Base and '+1' level risk denominators, divided by 1000,000.

- *Risk numerator difference (millions)*: Difference in numerator values between Base and '+1' risk levels, divided by 1,000,000.

- *Risk denominator difference:* Difference in denominator values between Base and '+1' risk levels.

Since there are multiple WTP observations for each company/study, we initially adopted a random effects panel model for estimation. Post estimation, we applied the Breusch-Pagan Lagrange Multiplier test (Breusch and Pagan, 1980) to examine if individual company/study specific variance components were zero. This test indicated that there were no significant panel effects ($p>.05$). Further, a comparison of AIC statistics between the random effect model and the pooled OLS model also showed a preference for the pooled OLS model. Hence, we concluded that a pooled ordinary least squares regression model was to be preferred. A pooled OLS model was therefore estimated with standard errors that were robust to intra-company correlation as well as heteroscedasticity (White, 1980).

5. Results

Table 3 shows the econometric model results. The table presents a 'Full' model, containing all the explanatory variables, and a 'Reduced' model, obtained via a stepwise procedure. The steps followed to obtain the Reduced model specification is as follows:

- Wald tests were performed on the equality of the coefficients of the *ln(risk reduction)* variables; and equality of the coefficients of the *ln(Base risk)* variables. In both cases, we failed to reject the null hypothesis (p=0.88 and 0.83 respectively). Based on these tests, we dropped the separate *ln(risk reduction)* and *ln(Base risk)* variables relating to the five service measures and instead included a single variable in each case that represented the joint impact for all the service measures combined.

- Using a step-down approach, we then removed the least significant explanatory variable at each step until no insignificant variables remained at the 10% level of significance.

The final specification of the Reduced model contains five explanatory variables: D: internal sewer flooding, ln(risk reduction): All service measures, ln(no. of attributes) , ln(no. of households) , ln(GDP per capita) , and Risk denominator, which were all significant at or below the 10% level of significance. All excluded variables were insignificant at the 10% level.

One of the most striking finding from inspection of the models is that the goodness of fit is exceptionally high, with R^2 values of 0.968 and 0.955 in the Full and Reduced models respectively. This indicates that over 95% of the variance in *ln (Unit WTP)* is explained even in the Reduced model. This result is remarkable given the controversy surrounding the large range in WTP estimates amongst water companies in England and Wales over three successive price reviews (Willis and Sheldon, 2021). The decrease in R^2 in moving from the Full to the Reduced model is very small indeed, indicating only a very small loss of explanatory power resulting from exclusion of many of the variables anticipated to have an impact on WTP.

In the following, we focus discussion on the Reduced model, including noting the variables that fail to enter the model due to lack of statistical significance. The latter part of the section examines the explanatory power of each of the variables included in the Reduced model.

Table 3: Meta-regression models for ln (Unit WTP)

Explanatory variables	Full model		Reduced model	
D: external sewer flooding	-4.004*	(2.292)		
D: internal sewer flooding	-1.368	(3.572)	0.530*	(0.271)
D: supply interruption (3-6 hrs.)	-2.426**	(1.005)		
D: supply interruption (6-12 hrs.)	-1.101	(1.472)		
ln(Base risk): discoloured water	-0.090	(0.384)		
ln(Base risk): external sewer flooding	-0.620*	(0.305)		
ln(Base risk): internal sewer flooding	-0.416	(0.594)		
ln(Base risk): supply interruption (3-6 hrs.)	-0.853	(0.906)		
ln(Base risk): supply interruption (6-12 hrs.)	-0.580**	(0.248)		
ln(risk reduction): discoloured water	-0.794*	(0.394)		
ln(risk reduction): external sewer flooding	-0.917***	(0.238)		
ln(risk reduction): internal sewer flooding	-0.764***	(0.217)		
ln(risk reduction): supply interruption (3-6 hrs.)	-0.682	(0.730)		
ln(risk reduction): supply interruption (6-12 hrs.)	-0.536**	(0.253)		
ln(risk reduction): All service measures			-1.011****	(0.051)
ln (no. of attributes)	-1.128**	(0.440)	-1.128***	(0.324)
ln (no. of households)	0.949****	(0.215)	0.873****	(0.137)
ln (average bill)	0.289	(0.450)		
ln (GDP per capita)	2.101***	(0.724)	1.887***	(0.583)
Package-scaled	-0.081	(0.205)		
Risk denominator (millions)	-0.475****	(0.097)	-0.360****	(0.058)
Risk numerator difference (millions)	-6.966	(10.033)		
Risk denominator difference	0.000	(0.001)		
Constant	-31.507***	(8.633)	-27.657****	(5.944)
Observations	53		53	
R^2	0.968		0.955	

Note: The dependent variable is *ln(Unit WTP)*, the natural log of the total willingness to pay per avoided service incident. The dummy variable for discoloured water is the base category against which coefficients for all the other service measure dummies in the Full model should be compared. Significance levels: * p<0.10, ** p<0.05, *** p<0.01, **** p<0.001. Standard errors in parentheses.

First, most of the dummy variables (prefixed 'D:') were statistically insignificant and so did not enter the Reduced model. Only *D: internal sewer flooding* remained, entering the Reduced model with a positive coefficient. This indicates that avoiding this type of service issue was significantly more valued than other types of service issue, a result that is not surprising given the impact that sewer flooding in the home would have on the household in comparison with other service issues.

The finding that no other service issue dummy variables were statistically significant is notable as these variables essentially can be thought of as measuring the constant, underlying, WTP values associated with avoiding each of the service issues, with all other variables fundamentally being design effects. The lack of any statistically significant constant terms associated with any of the other service measures thus contrasts clearly with the anticipated result from expected utility theory whereby we might have instead anticipated the opposite result: that these dummy variables would all be statistically significant, and all others would be insignificant.

Turning to base risk level, the modelling has found that this is not statistically significant (p>.10). This finding is consistent with expected utility theory, which predicts no effect of the base risk level, but is inconsistent with the theory that people evaluate risks in relative, rather than absolute, terms. There is some evidence consistent with relative risk evaluation insofar as all the *ln (Base risk)* variables enter the Full model with the predicted negative coefficient, but the effect is just not strong enough to survive the step-down testing process leading to the Reduced model.

The risk reduction variable, *ln(risk reduction): All service measures*, enters the Reduced model with a negative and a highly statistically significant coefficient of -1.01 (p<.001). This indicates that Unit WTP declined significantly with the overall scope of the risk reduction, a result that emphatically contradicts the expected utility theory hypothesis of zero elasticity. The size of the coefficient, -1.010, indicates that a 1% increase in the size of the risk reduction is consistent with a 1.010% smaller Unit WTP value or, more colloquially, doubling the scope of risk reduction results in an approximate halving of the Unit WTP value.

Our results here are in line with the findings in Lindhjem et. al (2011) who also find a robust, strong, and negative relationship between stated VSL and the size of the risk change valued by the respondents. Lindhjem et.al (2011) show that VSL increases by between 0.29% and 0.83% when the risk reduction valued decreases by 1%.

A further analysis of the significance of the elasticity of the risk reduction measure found by our study with respect to its power to explain the observed differences in Unit WTP across the sector is given below, following our continued exposition of the model findings.

The (log of the) number of attributes shown in the survey is negative and statistically significant in the Reduced model, a finding which again contradicts expected utility theory predictions but is consistent with findings on the sensitivity to the scope of risk change offered. The coefficient of -1.128 indicates that a 1% increase in the number of attributes valued is consistent with a 1.128% smaller Unit WTP value. This result again implies that the WTP values are sensitive to design choices.

The coefficient with respect to the log of the number of households in the company area is positive and significant in the Reduced model. This indicates that larger companies tended to generate higher WTP measures on average, all else equal. Thus, the coefficient of 0.873 indicates that a 1% increase in the number of households in the company area is associated with a 0.873% larger Unit WTP value. This finding might be considered indicative of the presence of altruism as a component of customers' preferences, consistent with findings elsewhere (Viscusi et al., 1988; Arana and Leon, 2002; Zhang et al, 2013).

To further examine the presence of altruistic effects, we noted that one study had defined the internal and external sewer flooding attributes in terms of the number of "other peoples' properties" that would be affected, rather than the number of properties overall. This definition implies that the personal risk reduction in this case was zero, but that there would be other people affected by the overall service improvement.[4] Somewhat counter-intuitively, the Unit WTP values of the company in question were higher than average, particularly for internal sewer flooding, suggesting that customers were altruistic with regard to sewer flooding to the extent that they may prioritise other people more than themselves. As a counter to this conclusion, however, it is possible that participants did not take account of the difference in the form of presentation between the sewer flooding attributes and other attributes and saw all of them as representing a certain level of service. Certainly, a more detailed exploration would be needed to confirm the extent to which altruism drives choices than is possible here.

An alternative explanation in principle for the positive impact of the number of properties supplied on Unit WTP is the phenomenon of denominator neglect. However, this is controlled for in our analysis via the inclusion of three risk framing variables, *Risk denominator (millions)*, *Risk numerator difference (millions)* and *Risk denominator difference*. Of these three variables, only *Risk denominator (millions)* enters the Reduced

[4]. In the econometric models and descriptive statistics reported, the risk reduction used was the non-zero ratio of other people's properties to total properties. Replacing these values with zero made no substantive difference to the model findings.

model, with the other two being dropped due to statistical insignificance ($p>.10$). The strong negative coefficient on this variable runs counter to our prior expectation based on the principle of denominator neglect, which would predict that larger denominators would be associated with larger valuations when controlling for the size of the risk reduction and the number of number of households affected in total.

Although not unique – Zhai and Suzuki (2008) find a similar result - it is difficult to understand the reasons for this pattern of valuations. Whilst it seems likely that the result reflects a genuine sensitivity to risk framing, it is also likely that we have been unable to capture fully how framing affects valuations within the relatively small dataset available for analysis. More data would be needed to adequately explore alternative mechanisms potentially driving this unexpected effect.

In other results from the model, neither the size of the average bill in the company area nor the package scaling effect were found to have a statistically significant impact on WTP values. Neither finding runs counter to expected utility theory.

The elasticity with respect to GDP per capita is positive and statistically significant in the Reduced models indicating that income has a significant positive influence on WTP values, as expected. This result indicates that larger incomes tended to generate higher WTP measures on average, all else equal. Thus, the coefficient of 1.887 in the Reduced model indicates that a 1% increase in the GDP per capita value is consistent with a 1.887% larger Unit WTP value. This result is somewhat higher than findings from the related VSL and environmental valuation literature. For example, Lindhjem et.al (2011) found that the elasticity of GDP per capita to VSL was between 1.1 and 1.3 in a global meta-analysis of stated preference surveys of mortality risk valuation in environmental, health, and transport risk contexts and. Similarly, Tyllianakis and Skuras (2016) conducted a meta-analysis of WTP values to restore Good Ecological Status (GES) under the EU Water Framework Directive (WFD). The estimated meta-regression models showed that income had a positive and statistically significant impact on WTP values and its elasticity was considerable in magnitude ranging from 0.6 to almost 1.7. The Reduced model estimate is thus above the upper range of estimates in the literature, but does not appear unreasonably high.

Shapley values

In order to quantify the relative contribution of each of the explanatory variables to the overall goodness of fit (i.e. R^2) of the Reduced model, we use the Shorrocks-Shapley decomposition methodology (Shorrocks 1982). model. This approach begins by estimating a full model which produces an R^2. Next, each of the regressor variables are successively removed one by one, based on a particular ordering of the variables. The difference in R^2 associated with the removal of a variable is regarded as that variable's marginal contribution in that particular ordering of the variables. Based on the assumption that all such orderings occur with equal probability, the Shapley value of a variable is defined as the variable's average estimated marginal contribution over all possible orderings. For example, if the full model contains *k* regressor variables, then we consider 2^k sub-models, estimate the average marginal contribution of each regressor *j* over all possible permutations and compute the Shapley value estimate for each regressor *j* as follows:

$$\mathrm{Sh}(x_j) = \frac{1}{k\,!} \sum_{all\ permutations} [Marginal\ contribution] \tag{3}$$

Table 4 below computes the additive decomposition of the R^2 value based on the estimated Reduced model[5].

Table 4: Shapley decomposition of R^2: Reduced model

Regression component	Shapley value estimates[(1)]
ln(risk reduction): All service measures	58.7%
D: internal sewer flooding	24.5%
ln (no. of households)	5.3%
Risk denominator (millions)	4.3%
ln (no. of attributes)	1.7%
ln(GDP per capita)	0.8%
Residual	4.5%
Total	**100%**

(1) The Shapley value estimates for the explanatory variables in the first column add up to the overall R^2 value. The residual part is indicated, which corresponds to the part of the variation that could not be attributed to any of the included explanatory variables in the model.

The results of the Shorrocks-Shapley decomposition method show that the risk reduction variable accounts for 58.7% of the observed variance in *ln (Unit WTP)*. The explanatory

[5] The Shapley value estimates were obtained using the shapley2 Stata package developed by Chávez Juárez (2012).

power of this variable is quite striking, particularly since, under expected utility theory, the variable should account for precisely none of the variance.

The variable with the second-most explanatory power is D: internal sewer flooding. Thus, despite this variable having the weakest coefficient in the Reduced model, in terms of its significance level, it is responsible for explaining 24.5% of the variance of *ln (Unit WTP)*.

The remaining variables collectively explain a fairly low proportion of the variance in Unit WTP. The most significant are *ln (no. of households)*, which explains 5.3% of the variance and *Risk denominator (millions)*, which explains 4.3%. Much less significant are ln (no. of attributes) and ln(GDP per capita) which explain 1.7% and 0.8% respectively. In part, this reflects the fact that these variables did not vary so substantially across companies (see Table 2).

6. Discussion and Conclusions

The results presented here emphatically contradict the expected utility theory prediction that elasticities of Unit WTP with respect to the scope of the risk reduction are jointly equal to zero ($p<.001$). Moreover, the size of the estimated scope elasticity, -1.0, and the fact that it explains 59% of the observed variance in Unit WTP, are both striking in the context of expected utility theory predictions.

The findings were not wholly unexpected. They are consistent with the VSL literature, which has often found VSL estimates to be highly sensitive to the scope of the risk reduction offered in the survey (Beattie et al; 1998; Hammitt and Graham, 1999; Lindhjem et al., 2011; Andersson et al., 2016), and with a range of theories originating from the cognitive psychology literature most notably including prospect theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992).

However, they have important implications for the interpretation of WTP evidence in the water sector and for the future measurement of WTP for water and wastewater service enhancements. Firstly, WTP estimates obtained from these studies, and from the methodology used to generate them, cannot be considered reliable in a broad sense. Whilst reported confidence intervals within the studies themselves generally suggest a good degree of reliability of the estimates, these are misleading with respect to a broader conception of reliability which considers the possibility of alternative reasonable designs, each with smaller or larger amounts of risk reduction offered.

Nor, as such, can the estimates be considered valid measures of customer preference in the usual sense. Whilst customers themselves may have considered them to be valid when answering, and whilst results may have been internally coherent within studies, it is only via the more powerful test afforded by external comparison, either via split sample treatments or meta-analyses such as the present study, that a complete picture of validity emerges. And unfortunately, in the present case, the picture is discouraging.

In the broader literature on valuing risk reductions, particularly in the VSL literature, emphasis has been placed on the need to spend time and effort within the survey, including appropriate visual aids and verbal analogies, to help participants understand and evaluate risk changes objectively, and to implement an external sensitivity to scope test to demonstrate validity (Hammitt and Graham, 1999; Corso et al., 2001; Krupnick et al,. 2002). Some studies have successfully demonstrated an appropriate sensitivity to scope via this approach (Corso et al., 2001; Hammitt and Haninger, 2010; Alolayan et al., 2017), but most discrete choice experiment studies including risk as an attribute have not included visual aids, and strong external tests of sensitivity to scope are extremely rare (Harrison et al., 2014; Andersson et al., 2016).

With regard to ensuring the validity of future estimates, whilst there may be some benefit in re-examining the use of visual aids and risk analogies to encourage a more objective basis to participants' choices, we are not optimistic that this approach would generate any substantial improvement. This is because many of the studies included in a major VSL meta-analysis (Lindhjem et al., 2011) had included visual aids, as recommended, and yet the external comparison across studies showed that VSL estimates were still highly sensitive to the scope of risk change offered. Furthermore, for the UKWIR (2011) guidelines, a set of depth interviews was undertaken with water customers which included testing of a number of visual aids to represent risks. The results suggested that the use of visual aids might actually hinder some respondents' understanding. When presented with filled-in grids, some participants interpreted the visualisation as representing the underlying distribution (i.e., clustered vs. random) as opposed to using the visual simply as a means of understanding the associated probability. As a consequence, the UKWIR (2011) guidelines explicitly stated that visual aids (e.g., filled-in grids) were not required and, accordingly, only a few of the studies in our meta-analysis contained visual aids to represent risk.

An alternative approach, which is currently under active consideration in the England and Wales water sector, would be to restrict the scope, and other methodological features, to be the same across companies via the undertaking of a national study. This approach would have the advantage of maximising the comparability of valuations across companies, but this would come with the disadvantage of imposing restrictions on the content of each company's research. It is not necessarily desirable to include common attributes for each company, nor to value common risk changes when each company starts at a different service level and has different opportunities for service improvement. Most significantly, however, a single national study could not be expected to result in valid and reliable values for any company unless a better methodology can be found than that underlying the results examined here.

In light of the issues raised here, subsequent work by one of this study's authors has focussed on developing a new methodology for water and wastewater service valuation which involves combining contingent valuation estimation of the value of a package of service improvements with a MaxDiff exercise focused on measuring the relative disutility of certain experience of different types of service failure. This methodology is set out in Chalak and Metcalfe (2021) along with a case study demonstrating its implementation. Although this new approach has some limitations, as described therein, we believe it offers much potential to resolve the problems caused by scope insensitivity documented here and, as such, offers some hope that valid and reliable WTP estimates for water and wastewater service improvements are achievable.

References

Accent. 2013. 'Comparative Review of Willingness to Pay Results. Final Report prepared for a club of UK water companies.'.

Adamowicz, W., P. Boxall, M. Williams, and J. Louviere. 1998. 'Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation', *American Journal of Agricultural Economics*, 80: 64-75.

Allais, M. 1953. 'Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole Americaine', *Econometrica*, 21: 503-46.

Alolayan, M. A., J. S. Evans, and J. K. Hammitt. 2017. 'Valuing Mortality Risk in Kuwait: Stated-Preference with a New Consistency Test', *Environmental and Resource Economics*, 66: 629-46.

Alonso, D., and P. Fernández-Berrocal. 2003. 'Irrational decisions: Attending to numbers rather than ratios', *Personality and Individual Differences*, 35: 1537-47.

Andersson, H., A. R. Hole, and M. Svensson. 2016. 'Valuation of small and multiple health risks: A critical analysis of SP data applied to food and water safety', *Journal of Environmental Economics and Management*, 75: 41-53.

Araña, J. E., and C. J. León. 2002. 'Willingness to pay for health risk reduction in the context of altruism', *Health Economics*, 11: 623-35.

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

———. 2006. 'Tom Sawyer and the construction of value', *Journal of Economic Behavior and Organization*, 60: 1-10.

Balmford, B., I. J. Bateman, K. Bolt, B. Day, and S. Ferrini. 2019. 'The value of statistical life for adults and children: Comparisons of the contingent valuation and chained approaches', *Resource and Energy Economics*, 57: 68-84.

Baron, J. 1997. 'Confusion of relative and absolute risk in valuation', *Journal of Risk and Uncertainty*, 14: 301-09.

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

Batson, C. D. 2011. *Altruism in Humans*.

Beattie, J., J. Covey, P. Dolan, L. Hopkins, M. Jones-Lee, G. Loomes, N. Pidgeon, A. Robinson, and A. Spencer. 1998. 'On the Contingent Valuation of Safety and the Safety of Contingent Valuation: Part 1-Caveat Investigator', *Journal of Risk and Uncertainty*, 17: 5-26.

Birol, E., and S. Das. 2010. 'Estimating the value of improved wastewater treatment: The case of River Ganga, India', *Journal of Environmental Management*, 91: 2163-71.

Breusch, T. S., and A. R. Pagan. 1980. 'The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics', *The Review of Economic Studies*, 47: 239-53.

Cameron, T. A., G. L. Poe, R. G. Ethier, and W. D. Schulze. 2002. 'Alternative non-market value-elicitation methods: Are the underlying preferences the same?', *Journal of Environmental Economics and Management*, 44: 391-425.

Carson, R. T. 2012. 'Contingent valuation: A practical alternative when prices aren't available', *Journal of Economic Perspectives*, 26: 27-42.

Chalak, A., and P. J. Metcalfe. 2021. "Valuing water and wastewater service improvements via impact-weighted numbers of service failures."

Chalak, A., S. Reid, M. Silcock, and M. G. Abiad. 2013. 'Customer perspectives on water quality, safety, and service failure risks', *Journal - American Water Works Association*, 105: E449-E59.

Chavez Juarez, F. 2012. 'SHAPLEY2: Stata Module to Compute Additive Decomposition of Estimation Statistics by Regressors or Groups of Regressors', *Statistical Software Components*, S457543.

Christie, M., and C. Azevedo. 2009. 'Testing the consistency between standard contingent valuation, repeated contingent valuation and choice experiments', *Journal of Agricultural Economics*, 60: 154-70.

Colombo, S., J. Calatrava-Requena, and N. Hanley. 2006. 'Analysing the social benefits of soil conservation measures using stated preference methods', *Ecological Economics*, 58: 850-61.

Corso, P. S., J. K. Hammitt, and J. D. Graham. 2001. 'Valuing mortality-risk reduction: Using visual aids to improve the validity of contingent valuation', *Journal of Risk and Uncertainty*, 23: 165-84.

Cropper, M., J. K. Hammitt, and L. A. Robinson. 2011. 'Valuing mortality risk reductions: Progress and challenges', *Annual Review of Resource Economics*, 3: 313-36.

Desvousges, W., K. Mathews, and K. Train. 2012. 'Adequate responsiveness to scope in contingent valuation', *Ecological Economics*, 84: 121-28.

Diamond, Peter A., and Jerry A. Hausman. 1994. 'Contingent Valuation: Is Some Number Better than No Number?', *Journal of Economic Perspectives*, 8: 45-64.

Dutta, V. 2009. "Economic assessment of differential-quality water demands from a metropolis and its peri-urban environments." In *IAHS-AISH Publication*, 347-53.

Fehr, E., and U. Fischbacher. 2003. 'The nature of human altruism', *Nature*, 425: 785-91.

Foster, V., and S. Mourato. 2003. 'Elicitation Format and Sensitivity to Scope: Do Contingent Valuation and Choice Experiments Give the Same Results? ', *Environmental and Resource Economics*, 24 141-60.

Garcia-Retamero, R., M. Galesic, and G. Gigerenzer. 2010. 'Do icon arrays help reduce denominator neglect?', *Medical Decision Making*, 30: 672-84.

Grether, D. M., and C. R. Plott. 1979. 'Economic theory of choice and the preference reversal phenomenon', *American Economic Review*, 69: 623-38.

Gyrd-Hansen, D., T. Kjaer, and J. S. Nielsen. 2012. 'Scope insensitivity in contingent valuation studies of health care services: should we ask twice?', *Health Economics*, 21: 101-12.

Hammitt, J. K., and J. D. Graham. 1999. 'Willingness to pay for health protection: Inadequate sensitivity to probability?', *Journal of Risk and Uncertainty*, 18: 33-62.

Hammitt, J. K., and K. Haninger. 2010. 'Valuing fatal risks to children and adults: Effects of disease, latency, and risk aversion', *Journal of Risk and Uncertainty*, 40: 57-83.

Hanley, N., D. MacMillan, R. E. Wright, C. Bullock, I. Simpson, D. Parsisson, and B. Crabtree. 1998. 'Contingent valuation versus choice experiments: Estimating the benefits of environmentally sensitive areas in Scotland', *Journal of Agricultural Economics*, 49: 1-15.

Harrison, M., D. Rigby, C. Vass, T. Flynn, J. Louviere, and K. Payne. 2014. 'Risk as an Attribute in Discrete Choice Experiments: A Systematic Review of the Literature', *Patient-Patient Centered Outcomes Research*, 7: 151-70.

Hausman, Jerry. 2012. 'Contingent Valuation: From Dubious to Hopeless', *Journal of Economic Perspectives*, 26: 43-56.

Hecht, J., S. Reid, and A. Chalak. 2015. 'A marketing analysis of the equitability of water/sewage service improvements', *Journal - American Water Works Association*, 107: E224-E33.

Hensher, D., N. Shore, and K. Train. 2005. 'Households' willingness to pay for water service attributes', *Environmental & Resource Economics*, 32: 509-31.

———. 2006. 'Water supply security and willingness to pay to avoid drought restrictions', *Economic Record*, 82: 56-66.

Jones-Lee, Michael. 1974. 'The Value of Changes in the Probability of Death or Injury', *Journal of Political Economy*, 82: 835-49.

Kahneman, D. 2003. 'A Perspective on Judgment and Choice: Mapping Bounded Rationality', *American Psychologist*, 58: 697-720.

Kahneman, Daniel, and Jack L. Knetsch. 1992. 'Valuing public goods: The purchase of moral satisfaction', *Journal of Environmental Economics and Management*, 22: 57-70.

Kahneman, D., and A. Tversky. 1979. 'Prospect theory: An analysis of decision under risk', *Econometrica*, 47: 263-91.

Kanyoka, P., S. Farolfi, and S. Morardet. 2009. 'Households' preferences and willingness to pay for multiple use water services in rural areas of South Africa: An analysis based on choice modelling', *Water SA*, 34: 715-24.

Krupnick, A., A. Alberini, M. Cropper, N. Simon, B. O'Brien, R. Goeree, and M. Heintzelman. 2002. 'Age, health and the willingness to pay for mortality risk reductions: A contingent valuation survey of Ontario residents', *Journal of Risk and Uncertainty*, 24: 161-86.

Lanz, B., and A. Provins. 2015. 'Using discrete choice experiments to regulate the provision of water services: do status quo choices reflect preferences?', *Journal of Regulatory Economics*, 47: 300-24.

Latinopoulos, D. 2014. 'Using a choice experiment to estimate the social benefits from improved water supply services', *Journal of Integrative Environmental Sciences*, 11: 187-204.

Lichtenstein, S., and P. Slovic. 1971. 'Reversals of preference between bids and choices in gambling decisions', *Journal of Experimental Psychology*, 89: 46-55.

Lichtenstein, Sarah, and Paul Slovic. 2006. 'The Construction of Preference: An Overview.' in Paul Slovic and Sarah Lichtenstein (eds.), *The Construction of Preference* (Cambridge University Press: Cambridge).

Lindhjem, H., S. Navrud, N. A. Braathen, and V. Biausque. 2011. 'Valuing Mortality Risk Reductions from Environmental, Transport, and Health Policies: A Global Meta-Analysis of Stated Preference Studies', *Risk Analysis*, 31: 1381-407.

MacDonald, D. H., M. D. Morrison, and M. B. Barnes. 2010. 'Willingness to Pay and Willingness to Accept Compensation for Changes in Urban Water Customer Service Standards', *Water Resources Management*, 24: 3145-58.

Metcalfe, P. J., and W. Baker. 2012. "Willingness to Pay to Avoid Drought Water Use Restrictions."

Mikušková, Eva. 2015. 'The denominator neglect in decision-making', *Studia Psychologica*, 57: 255-69.

Mogas, Joan, Pere Riera, and Raul Brey. 2009. 'Combining Contingent Valuation and Choice Experiments. A Forestry Application in Spain', *Environmental and Resource Economics*, 43: 535-51.

Nur Syuhada, C. I., K. Mahirah, and M. A. Roseliza. 2020. 'Dealing with attributes in a discrete choice experiment on valuation of water services in East Peninsular Malaysia', *Utilities Policy*, 64.

Ofwat. 2015. "Water 2020: Regulatory framework for wholesale markets and the 2019 price review. Final Report."

———. 2017. "Delivering Water 2020: Our final methodology for the 2019 price review. December 2017."

———. 2020. "PR24 and beyond: Reflecting customer preferences in future price reviews – a discussion paper

Okan, Y., R. Garcia-Retamero, E. T. Cokely, and A. Maldonado. 2012. 'Individual Differences in Graph Literacy: Overcoming Denominator Neglect in Risk Comprehension', *Journal of Behavioral Decision Making*, 25: 390-401.

Passerini, Gabriella, Laura Macchi, and Maria Bagassi. 2012. 'A methodological approach to ratio bias', *Judgment and Decision Making*, 7: 602-17.

Pinto-Prades, José-Luis, Jorge-Eduardo Martinez-Perez, and JoséMaría Abellán-Perpiñán. 2006. 'The influence of the ratio bias phenomenon on the elicitation of health states utilities', *Judgment and Decision Making*, 1: 118-33.

Price, J. I., P. R. Lloyd-Smith, D. P. Dupont, and W. L. Adamowicz. 2019. 'Floods and Water Service Disruptions: Eliciting Willingness-to-Pay for Public Utility Pricing and Infrastructure Decisions', *Water Economics and Policy*, 5.

Reid, S., A. Chalak, and J. Hecht. 2010. 'Determining the optimal investment plan for water utilities: The case of Veolia Water Central', *Water Science and Technology: Water Supply*, 10: 367-75.

Sayles, R., H. M. Smith, and P. Jeffrey. 2021. 'Customer priorities for water and wastewater services: a comparative evaluation of three elicitation methods', *Water and Environment Journal*, 35: 55-66.

Scarpa, Riccardo, Silvia Ferrini, and Kenneth Willis. 2005. 'Performance of Error Component Models for Status-Quo Effects in Choice Experiments.' in Riccardo Scarpa and Anna Alberini (eds.), *Applications of Simulation Methods in Environmental and Resource Economics* (Springer Netherlands: Dordrecht).

Shorrocks, A. F. 1982. 'Inequality Decomposition by Factor Components', *Econometrica*, 50: 193-211.

Slovic, P. 1995. 'The Construction of Preference', *American Psychologist*, 50: 364-71.

Snowball, J. D., K. G. Willis, and C. Jeurissen. 2008. 'Willingness to pay for water service improvements in middle-income urban households in South Africa: A stated choice analysis', *South African Journal of Economics*, 76: 705-20.

Søgaard, R., J. Lindholt, and D. Gyrd-Hansen. 2012. 'Insensitivity to scope in contingent valuation studies: Reason for dismissal of valuations?', *Applied health economics and health policy*, 10: 397-405.

Sunstein, C. R. 2003. 'Terrorism and Probability Neglect', *Journal of Risk and Uncertainty*, 26: 121-36.

Thaler, R. 1985. 'Mental accounting and consumer choice', *Marketing Science*, 4: 199-214.

Thaler, R. H. 1999. 'Mental accounting matters', *Journal of Behavioral Decision Making*, 12: 183-206.

Tuan, Tran Huu, and Stale Navrud. 2007. 'Valuing cultural heritage in developing countries: comparing and pooling contingent valuation and choice modelling estimates', *Environmental and Resource Economics*, 38: 51-69.

Tversky, A., and D. Kahneman. 1981. 'The framing of decisions and the psychology of choice', *Science*, 211: 453-58.

———. 1992. 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty*, 5: 297-323.

Tyllianakis, E., and D. Skuras. 2016. 'The income elasticity of Willingness-To-Pay (WTP) revisited: A meta-analysis of studies for restoring Good Ecological Status (GES) of water bodies under the Water Framework Directive (WFD)', *J Environ Manage*, 182: 531-41.

UKWIR. 2011. "Carrying Out Willingness to Pay Surveys "

Vásquez, W. F., and E. A. Adams. 2019. 'Climbing the water ladder in poor urban areas: Preferences for 'limited' and 'basic' water services in Accra, Ghana', *Science of the Total Environment*, 673: 605-12.

Viscusi, W. K., W. A. Magat, and A. Forrest. 1988. 'Altruistic and private valuations of risk reduction', *Journal of Policy Analysis and Management*, 7: 227-45.

Wahid, Nabsiah, and Chew Hooi. 2014. 'Factors Determining Household Consumer's Willingness to Pay for Water Consumption in Malaysia', *Asian Social Science*, 11.

Wang, Jia, Jiaoju Ge, and Zhifeng Gao. 2018. 'Consumers' Preferences and Derived Willingness-to-Pay for Water Supply Safety Improvement: The Analysis of Pricing and Incentive Strategies', *Sustainability*, 10: 1704.

Warneken, F., and M. Tomasello. 2009. 'The roots of human altruism', *British Journal of Psychology*, 100: 455-71.

White, H. 1980. 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity', *Econometrica*, 48: 817-38.

Willis, Ken, and Rob Sheldon. 2021. 'Research on customers' willingness-to-pay for service changes in UK water company price reviews 1994–2019', *Journal of Environmental Economics and Policy*: 1-17.

Willis, K. G., R. Scarpa, and M. Acutt. 2005. 'Assessing water company customer preferences and willingness to pay for service improvements: A stated choice analysis', *Water Resources Research*, 41.

Wilson, C., W. Athukorala, B. Torgler, R. Gifford, M. A. Garcia-Valiñas, and S. Managi. 2020. 'Willingness to pay to ensure a continuous water supply with minimum restrictions', *Empirical Economics*.

Zhai, G. F., and T. Suzuki. 2008. 'Effects of risk representation and scope on willingness to pay for reduced risks: Evidence from Tokyo Bay, Japan', *Risk Analysis*, 28: 513-22.

Zhang, J., W. Adamowicz, D. P. Dupont, and A. Krupnick. 2013. 'Assessing the extent of altruism in the valuation of community drinking water quality improvements', *Water Resources Research*, 49: 6286-97.