Valuing water and wastewater service improvements via impact-weighted numbers of service failures[1]

Ali Chalak[2] and Paul Metcalfe[34]

15 August, 2021

Abstract

Stated preference valuations of water and wastewater service improvements in Great Britain have been found to be highly variable across studies for similar units of service, driven primarily by a substantive dependence on the scope of service change offered to survey participants. The present paper motivates and describes a new approach that is simpler for survey participants and sidesteps the key problems caused by scope insensitivity. It relies on first deriving estimates of the relative impact of different types of service issue, and then using these to apportion a package valuation into valuations of individual service level improvements. The paper presents a case study that implements the new approach in a real business planning context and discusses the impact the new approach has on the valuation results obtained. We contend that the new approach has significant advantages over traditional discrete choice experiment approaches to water and wastewater service valuation and recommend it for future use in similar policy/planning contexts.

Keywords: risk; scope sensitivity; stated preference; valuation; water and wastewater services; willingness-to-pay.

[1] Funding for this research was provided by Dwr Cymru Welsh Water.

[2] Department of Agricultural Sciences, Faculty of Agricultural and Food Sciences, American University of Beirut, P.O. Box 11-0236 Riad el Solh, Beirut 1107-2020, Lebanon

[3] PJM Economics, 1 Walford Close, Wimborne, Dorset, BH21 1PH, UK.

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

1. Introduction

Concerns around the sensitivity to scope of stated preference valuations have been long been voiced by critics of the approach (Diamond & Hausman, 1994; Kahneman & Knetsch, 1992; Hausman, 2012; Desvousges et al., 2012). The issue affects all fields that use stated preference methods for valuation (Ojea & Loureiro, 2011) but is particularly significant where the object of valuation is a reduction in the risk of some physical or environmental harm, such as is the case in studies concerning the value of a statistical life (VSL) or the avoidance of water and wastewater service failures. In such cases, to be consistent with standard economic theory, values must not only be positively correlated with scope – the weak test, they must also be near-proportional to scope – the strong test (Hammitt, 2000). Although there is evidence from many environmental applications that stated preference valuations pass a weak sensitivity to scope test, (Carson et al., 2001; Ojea & Loureiro, 2011), and evidence from some environmental valuation studies of plausible elasticities with respect to scope (Whitehead, 2016), evidence on valuations of risk reductions has rarely shown the required degree of scope sensitivity to be consistent with standard economic theory (Hammit & Graham, 1999; Beattie et al, 1998; Adamowicz et al. , 2011; Lindhjem et al., 2011).

In the context of water and wastewater service valuation, Metcalfe and Sen (2021) present evidence from a meta-analysis of WTP studies undertaken for the 2014/15 periodic price reviews in Great Britain indicating that, consistent with these broader findings, valuations of water and wastewater services also failed the strong sensitivity to scope test. Rather than being constant, as predicted by standard economic theory, values per unit of service change (Unit WTP values) were found to be significantly negatively correlated with the scope of service change offered. Moreover, Unit WTP values varied markedly across studies for the same service measure with the key driver of this variation being differences in the scope of service change offered.

These findings represent a problem for the sector as they suggest that, where valuations are used to determine performance commitment levels via cost-benefit analysis, companies may be prioritising improvements in service areas that were only incrementally improved in the stated preference survey at the expense of service areas where large improvements were tested. Additionally, where valuations are used as the basis of in-period incentives for companies to reward them for exceeding their performance commitments, as has been the case in England and Wales through the

Outcome Delivery Incentive regime since PR14 (Ofwat, 2013; 2017), a high degree of variability in WTP estimates for the same service areas raises questions with respect to the equity of regulatory treatment across companies and their customers. More generally, confidence in the validity of valuations and their implications for business plans is weakened if they are known to be unreliable with respect to reasonable alternative sets of service levels.

The PR14 survey instruments made use of discrete choice experiments offering choices between profiles of service levels, with many of these represented as the risks of a service failure happening. The use of this approach was consistent with industry best practice guidelines (UKWIR, 2011), and was based on similar approaches having been undertaken in previous price reviews in England and Wales (e.g. Willis et al., 2005; Reid et al., 2010) , as well as elsewhere (e.g. Hensher et al., 2005; Snowball et al., 2008). (See Willis and Sheldon, 2021, for a more detailed discussion of the context for water and wastewater service valuation in the UK). Moreover, all of the studies reviewed tended to display good standards of internal validity, including content and construct validity evaluations (Metcalfe and Sen, 2021).

However, notwithstanding the desirable attributes of the methodology used at PR14, the approach was criticised by industry participants as being too complex for participants and resulting in valuations that were too widely variable to be considered reliable for the intended purposes of determining performance commitment levels and outcome delivery incentive rates (UKWIR, 2015).

In this paper, we present an approach that we conceived and implemented for several companies for the 2019 periodic price review in England and Wales (PR19) to address the shortcomings inherent in the typical stated preference instruments fielded at PR14. We contend that the new approach has several advantages over previous approaches. Most importantly, it avoids the problem of relative valuations between service areas being determined by the relative scopes of service change offered, whilst ensuring that absolute valuations are consistent with customers willingness to pay for a realistic full package of service improvements. Additionally, the approach asks simpler and less abstract questions of participants, thereby encouraging a greater validity of response.

The rest of this paper is organized as follows: Section 2 sets out the new valuation approach; Section 3 presents and discusses a case study illustrating the approach based on work we conducted for Welsh Water for PR19 business planning; and Section 4 concludes.

2. New Valuation Approach

The approach proposed is focused on obtaining valuations per service failure that are proportional to the impacts of the service failures if they occur. Formally, for a given set of K service measures, the household's WTP (£/hh/yr) for any given 'package' of improvements across service measures can be written as follows:

$$WTP = c \cdot \sum_{\forall k \in K}^{K} w_k \cdot n_k \quad (1)$$

where w_k measures the impact weight, or disutility, to the customer from a failure in any service measure k in the set K being valued; n_k is the number of properties affected by this service improvement or, equivalently, the change in the risk that an individual household would be affected, on average, multiplied by the number of household (i.e. the scope), and c is a constant.

The value of an improvement (£/hh/yr) in any service measure k can be written as:

$$WTP_k = c \cdot w_k \cdot n_k \quad (2)$$

Finally, the unit WTP per property affected (£/yr/prop) for service measure k, aggregated over the whole customer base, can then be derived as:

$$uWTP_k = c \cdot w_k \cdot N = \left(\frac{WTP}{\sum_{\forall k \in K}^{K} w_k \cdot n_k}\right) \cdot w_k \cdot N \quad (3)$$

where N is the total number of customers.

Note again here that the impact weight w_k is based on the disutility of a single service failure, and is elicited independently from the package *WTP*, in which the scope of improvements is being valued, by means of two separate exercises that will be described in the next section. Note as well how the unit WTP is derived from the package WTP by means of an impact-weighted apportionment of aggregate WTP for a package of improvements and is independent of the scale of the improvement, in keeping with expected utility theory.

A potential complication arises in applying this approach when service measures are not naturally denominated in numbers of properties affected. For example, environmental service measures are often defined in terms of environmental quantities affected, such as length of river. In these cases, the unit of the service measure k in the package valuation, Δq_k, may be different from the number of properties affected. In these cases, Δq_k needs to be mapped to N_k by means of a conversion factor f_k. This ensures that the impact weight, being evaluated based on a single property affected, aligns with the scope of change valued in the package WTP. In such instances, $n_k = f_k \cdot \Delta q_k$, and f_k translates the scope change in the package valuation into number of properties affected. Indeed, when the scope of improvement is itself expressed in the number of properties in the package valuation, then N_k, becomes equivalent to Δq_k and f_k becomes equal to one.

3. Case study – Welsh Water PR19 Business Planning

To illustrate our approach, we present a case study based on work we conducted for Dŵr Cymru Welsh Water (DCWW) in the context of PR19. We conducted a programme of research exploring customers' WTP for a range of possible service level changes. The results from the work ultimately informed the development of the company's 2020-25 business plan and supported its legitimacy to the regulators and other stakeholders. The objectives of the study were to identify, through the use of stated preference surveys or other appropriate methods, which areas of service were most important to DCWW customers, and to estimate the value that customers placed on different levels of service across DCWW's service measure framework.

Survey design and choice formats

The survey questionnaire was designed around two interlinked stated preference exercises: (1) a 'MaxDiff' exercise focussed on which types of service issue would have the most, and least, impact on participants if they were to be affected by them, and as such elicited the disutility or impact of the service issue; and (2) a 'Package' exercise focussed on high level trade-offs between service improvements or deteriorations and changes in the level of the bill, constituting the level at which the scale or risk of service failure would be incorporated and valued.

This questionnaire followed a structure that is typical for stated preference questionnaires, and was consistent with UKWIR (2021) guidelines. It was structured as follows: (1) screening and recruitment; (2) introduction to main survey; (3) Usage, experience and attitude questions; (4) Background information, including service measure definitions; (5) MaxDiff exercise; (6) Follow-up questions on ability to make comparisons between the service measures; (7) Package valuation exercise; (8) Follow-up questions, including reasons for choices, ability to choose, perceived realism of the service levels shown, and understanding of the service measures; and (9) Demographics.

The MaxDiff exercise

The MaxDiff technique, a form of Best-Worst Scaling, is an established, simple and efficient alternative to the use of discrete choice experiments, whilst still being based on the same underlying Random Utility Theory (Louviere et al., 2015). The form of choice question is illustrated in Figure 1 below, alongside a choice card based on guidelines from UKWIR (2011) which illustrates the PR14 valuation approach. The new questions can immediately be seen to be much simpler for respondents to answer as they only require consideration of the impact of each service failure, one at a time, rather than requiring the simultaneous processing of multiple comparisons of small risks of different types of service failure.

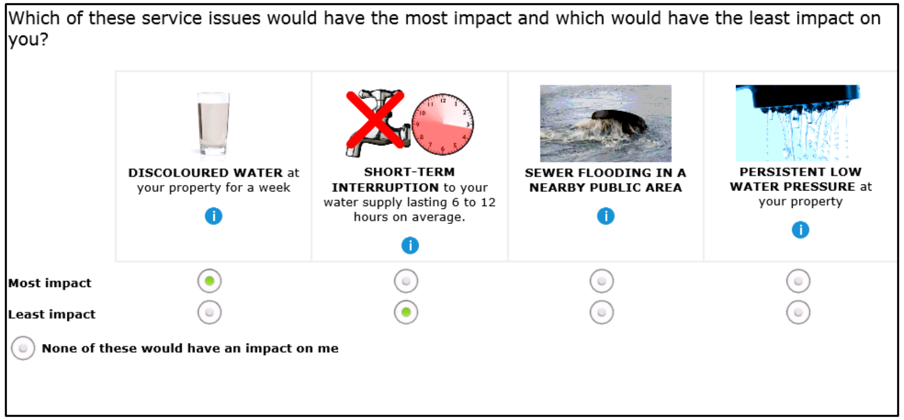
The PR14 approach presented options containing alternate levels of risk (or scope) for each type of service failure in question, plus a change in the water bill, and respondents chose their preferred package. Options varied across the experimental design in terms of the combinations of risk levels, but the service measures themselves were the same in each question within a choice exercise. By contrast, the new approach presented four types of service failure at a time and asked respondents to choose which would have the most impact on them and which would have the least impact. No risks are shown in this case; respondents just had to imagine the service failure itself. In the exercise, the combination of four service measures changed in each question.

Figure 1: Example PR14 and PR19 choice cards

Example Choice Card for PR14[(a)]

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

Example MaxDiff Choice Card for PR19[(b)]



(a) Source: UKWIR (2011), Appendix 10; (b) Source: Accent-PJM Economics (2017)

The service issues addressed in the MaxDiff exercise numbered 19 and are listed in Table 1. Further details on these were provided to participants to ensure they were sufficiently well understood in order to be able to appraise the impacts in relation to one another. The issues themselves were selected in consultation with the company in order to align to its service measure framework.

Table 1: Service issues included in the MaxDiff exercise

Service issue
1. Discoloured water at your property for a week
2. Water taste & smell not ideal at your property for a few days
3. Short-term interruption to your water supply lasting 3 to 6 hours on average.
4. Short-term interruption to your water supply lasting 6 to 12 hours on average.
5. Long-term interruption to your water supply lasting 24-48 hours.
6. Long-term interruption to your water supply lasting 7 days
7. Persistent low water pressure at your property
8. Temporary use ban from May to September.
9. Sewer flooding inside your property- extensive flooding, making it uninhabitable.
10. Sewer flooding inside your property- minor flooding, causing no lasting damage
11. Sewer flooding outside your property
12. Sewer flooding in a nearby public area
13. Odour from sewage works affects your property once or twice per year, for a few days each time
14. Odour from sewage works affects your property several times a year, for a week or two each time
15. Significant pollution incident in your local area caused by welsh water operations
16. Minor pollution incident in your local area caused by welsh water operations
17. River water quality in your local area fails to achieve good status due to the impact of welsh water operations.
18. Coastal bathing water quality in your local area achieves good quality but not excellent due to the impact of Welsh water operations.
19. Coastal bathing water quality in your local area achieves sufficient quality but not good or excellent due to the impact of welsh water operations.

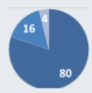
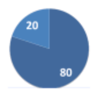
The experimental design for the exercise was generated by means of a D-efficient algorithm aimed at maximising the statistical precision of the estimates, whilst avoiding choice pairs where one option dominated the other one (e.g. two or more identical attributes, or two or more attributes of the same nature but different intensities such as supply interruptions of different durations). A total of 200 choice cards were generated and grouped in 20 blocks of 10 cards each. Thus, each participant was administered choice cards from a randomly selected block answering 10 MaxDiff choice cards each showing 4 service issues.

The MaxDiff exercise provides a means of understanding how impactful each type of service failure would be to customers in relative terms. In order to convert this index to a monetary measure, however, evidence on customers' willingness to trade off money for service level changes is needed at the package level via the package valuation exercise.

Package valuation exercise

An illustrative example of the type of question that was adopted for PR19 to capture WTP at the package level is shown in Figure 2. On the surface, this type of question appears similar to the PR14 valuation approach. However, the questions only need to show a limited range of package alternatives, for example, as in the figure: "Same as now", "Moderate improvement", "Substantial improvement", rather than having each attribute potentially taking any service level. The PR14 approach required variation in service levels in order to estimate values for each service level change in each service measure. By contrast, in the new approach relative values of each of the included elements were measured via the MaxDiff exercise and so the Package exercise only needed to derive one or more package valuations in order to scale these relative values. The PR19 Package questions would therefore be relatively straightforward for respondents to answer.

Figure 2: Illustrative PR19 package valuation choice card

	Option B	Option C
DISCOLOURED WATER lasting for a week at a time The number of properties where this happens in any one year, out of 1.4 million.	5,300 properties	4,000 properties
SHORT-TERM INTERRUPTIONS lasting 3 to 6 hours on average The number of properties where this happens in any one year, out of 1.4 million.	45,000 properties	30,000 properties
LONG-TERM INTERRUPTIONS lasting more than 24 hours The number of properties where this happens in any one year, out of 1.4 million.	1,000 properties	500 properties
TEMPORARY USE BAN from May to September The chance that this happens at your property in any one year	Once every 20 years	Once every 30 years
PERSISTENT LOW WATER PRESSURE The number of properties where this happens in any one year, out of 1.4 million.	100 properties	50 properties
SEWER FLOODING INSIDE CUSTOMERS' PROPERTIES. The number of properties where this happens in any one year, out of 1.4 million.	220 properties	175 properties
SEWER FLOODING OUTSIDE PROPERTIES & IN PUBLIC AREAS. The number of properties where this happens in any one year, out of 1.4 million.	6,500 properties	5,000 properties
ODOUR FROM SEWAGE WORKS. The number of properties where this happens in any one year, out of 1.4 million.	2,150 properties	1,500 properties
SIGNIFICANT POLLUTION Number of significant pollution incidents in Welsh Water's service area each year.	2 incidents	1 incidents
MINOR POLLUTION Number of minor pollution incidents in Welsh Water's service area each year.	110 incidents	80 incidents
RIVER WATER QUALITY. The percentage of river miles that are at "Good" quality.	36%	40%
COASTAL BATHING WATER QUALITY. Percentage of designated beaches in Welsh Water in each water quality category. Excellent Quality Good Quality Sufficient Quality		
THE CHANGE IN YOUR ANNUAL WATER AND SEWERAGE BILL To provide the service quality above	**No change**	**Increase of £11.20 each year, from £372.00 in 2019 to £428.00 in 2024**
	Option B	Option C

The advantage of having multiple profiles in this exercise, rather than just a single package to compare against current service levels, is that customer value is not expected to be linear in service level change. By having multiple package valuations, one can, in principle, derive individual service measure level value estimates that are sensitive to the size/cost of the package within which they are to be carried out. For the purpose of our analysis, four different packages of service levels were defined as follows.

- -1 (A): all service measures in this option deteriorate to '-1' levels. The bill is lower than in SQ.

- SQ (B): all service measures are maintained at their current levels, with the yearly bill either maintained at its current level (in real terms), or slightly decreased or increased.

- +1 (C): all service measures improve to +1 levels, and the bill is higher than in SQ.

- +2 (D): all service measures further improve to +2 levels; the bill is higher than in +1.

These options were presented to participants in the survey in a series of four pairwise package comparisons.

(1) SQ vs. +1: Yields WTP estimate for a status quo (SQ) to +1 improvement

(2) SQ vs. +2: Yields WTP estimate for an SQ to +2 improvement

(3) +1 vs. +2: Yields WTP estimate for a +1 to +2 improvement

(4) SQ vs. -1: Yields WTA estimate for an SQ to -1 deterioration

Table 2 shows the possible bill changes for each of the four options. The bill changes attached to the presented option were randomly drawn as percentages from the levels shown and translated into monetary bill changes for households by multiplying by the current bill. For non-households, percentage changes were shown, in line with industry guidelines (UKWIR, 2011).

Table 2: PR19 Package % bill change levels

		Level		
Package	Definition	1	2	3
-1	% change over and above SQ bill change from 2020 to 2024	-10%	-5%	-1%
SQ	% change of SQ bill from 2020 to 2024	-2.5%	0%	2.5%
+1	% change over SQ bill change from 2020 to 2024	1%	7.5%	15%
+2	% change over and above SQ plus +1 bill changes from 2020 to 2024	1%	5%	10%

For the purpose of illustration, we henceforth restrict our exposition and discussion of the model estimation and WTP results to those based 'SQ to +1' improvement (i.e. Figure 2), knowing that the same principles and algebraic manipulations would apply to any other improvement package.

Survey testing and piloting

The design approach put forward for this study was new to DCWW and the water sector more widely. Therefore, an extensive programme of testing was designed and implemented to refine the design and provide assurance that the instrument was working effectively. This included four phases of pre-testing of the survey instrument with DCWW customers. The first phase of pre-testing consisted of eight pre-tasked focus groups and 24 pre-tasked in-depth telephone interviews. This qualitative work explored customers' response to different sets of show material and ways of presenting the various exercises.

The second phase consisted of 16 cognitive interviews (11 with household customers and 5 with non-household customers), in which participants were encouraged to "think aloud" and give feedback on the questionnaires and showcards as they worked their way through them. The third phase of pre-testing with DCWW customers consisted of a pilot stated preference study of 200 interviews with household customers for each of the PR14 and PR19-style surveys. As for non-household customers, 192 interviews were conducted for the PR14-style survey, and 121 for the PR19-style survey.

Based on the findings from the analysis of the pilot data from the third phase, the design of the PR19-style survey was further refined. The fourth pre-testing phase tested this new survey tool by means of a second round of pilot interviews conducted with 249 household and 68 non-household customers.

Survey administration and data

The overall main stage of the survey comprised a total of 1,550 interviews with DCWW's household and non-household customers. All interviews were conducted using a computer-assisted telephone interviewing (CATI) method, more specifically a phone–post/email–phone approach. All customers were given the option to be interviewed in Welsh. Interviews lasted an average of 36 and 25 minutes for household and non-household interviews, respectively.

For the household sample, Weighting was applied to social grade, age and gender variables to achieve a representative sample; targets were based on Census 2011 data for Wales. A breakdown of all household interviews by these three variables is shown in

Table 3. As for the non-household sample, data were weighted by number of employees. The target profile for Wales was obtained from the Department for Business, Energy & Industrial Strategy's Business Population Estimates, 2016. Table 4 shows the comparison between weighted and unweighted data.

Table 3: Frequency of household interviews by key demographics (N=1,050)

Variable		Unweighted (%)	Weighted (%)
SEG	AB	35%	23%
	C1	26%	27%
	C2	14%	19%
	DE	25%	31%
Age	18-34	6%	17%
	35-54	28%	36%
	55+	66%	47%
Gender	Female	53%	59%
	Male	47%	41%

Table 4: Frequency of non-household interviews by number of employees

Variable		Unweighted (%)	Weighted (%)
Number of employees	Sole trader	12	26
	Less than 4	21	9
	4 to 49	49	22
	50 to 249	10	23
	250 +	9	20
Total interviews		500	500

Econometric anaysis and WTP analysis

Our approach to model estimation and WTP estimation involved the use of econometric analysis. First, impact scores were estimated via a rank-ordered logit (ROL) model. This random utility model was developed by Beggs et al. (1981) and uses the full information contained in stated preference surveys that involve the ranking of alternatives, such as

MaxDiff exercises. The ROL translates rankings into a sequence of 'choices', the probabilities of which are captured by means of sequential conditional logit (CL) probabilities that are functions of the utilities associated with each service issue, until all rankings are exhausted (Hanley et al., 2001). The ROL model will be further formalised below following exposition of the CL model in this same section.

Next, the data from the package exercise are estimated by means of a CL model that estimated the (dis)utility from each of the improvement/deterioration scenarios, in addition to a bill coefficient. In this model, person *n* chooses his/her preferred alternative *i* from a choice set *J*. The utility derived from alternative *i*—U_{ni}—is composed of a deterministic component (V_{ni}), typically specified as a linear index of attributes (X_{ni}) of the alternative and a vector of preference coefficients β to be estimated, and a stochastic error component ε_{ni}:

$$U_{ni} = V_{ni} + \varepsilon_{ni} = \beta' X_{ni} + \varepsilon_{ni} \qquad (4)$$

When ε_{nj} is assumed to be independently and identically distributed (iid) with a Type I Extreme Value distribution, the probability of individual *n* choosing *i*—p_{ni}— assumes a conditional logit form:

$$p_{ni} = Prob\left(U_{ni} > U_{nj}, \forall j \in J\right) = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}} = \frac{e^{\beta' X_{ni}}}{\sum_{j=1}^{J} e^{\beta' X_{nj}}} \qquad (5)$$

The ROL extends the CL model, whereby person *n*'s ranking of *J* alternatives is denoted r_n, a vector whose elements r_n^i represent the rankings of any alternative *i* in *J*. The probability p_n of this ranking can therefore be written as a product of *J-1* CL probabilities. In this product, the first term is the CL probability of the alternative ranked 1st having the highest utility and hence being 'chosen' over the remaining *J-1* alternatives. The second term is the probability of the alternative ranked 2nd having the 2nd highest utility, and therefore being chosen over the remaining *J-2* alternatives after removing the alternative ranked 1st, and so on until reaching the *J-1*th term comprising the two alternatives ranked lowest. Formally, this probability can be written as follows:

$$p_n = Prob\left(U_{r_n^1} > \cdots > U_{r_n^J}\right) = \prod_{j=1}^{J-1} \frac{e^{V_{r_n^i}}}{\sum_{j=i}^{J} e^{V_{r_n^j}}} \qquad (6)$$

In the MaxDiff exercise, β amounts to a vector of utilities β_k associated with each of the K service measures, with one measure selected as a reference measure with β_k=0 for the purpose of identification. The impact scores (i.e. w_k in equations 1-3) were derived as odd ratios in which $w_k = \exp(\beta_k)$. These were adjusted to take account of further research which indicated that the impact scores for three of the attributes were likely to have been overstated due to misinterpreting the service issues shown[5].

In the package exercise, on the other hand, β consists of two coefficients a and b, the former being the CL estimate of the package's alternative-specific constant (ASC)—i.e. the net utility of the change, and the latter the estimate of the bill's marginal utility. The population's *WTP* (see equation 1) for any package can then be derived simply as $-a/b$. Hence we can obtain values for the full range of service change for each package valued: status quo (SQ) to +1, +1 to +2 and the deterioration package SQ to -1.

Finally, unit WTP (aggregated over the whole customer base and expressed in GBP per year per property affected—£/yr/prop) for each of the service measures ($uWTP_k$) can then be derived as per equation 3. Indeed not always is the service change in the package exercise (Δq_k) expressed in number of properties (n_k) affected by the service issue as presented in the MaxDiff exercise. As described earlier, in such cases conversion

---

5 A peer reviewer noted that the values for bathing water quality improvements seemed high in relation to previous research. We judged that this might be due to a potential discrepancy between how participants construed the language defining river and bathing water quality service issues, and how they are primarily to be used in appraisals. A possible reason for overstating these issues was that participants may have interpreted them as deteriorations in quality for which willingness to accept, or WTA, is typically elicited, rather than as states of affairs that could potentially be improved upon, as intended, and for which WTP is typically elicited. This potentially led to typical case of loss/gain or WTA/WTP discrepancy, which according to a meta-analysis by Tunçel and Hammitt (2014), amounts to an average ratio of 6.23. The extent to which they did so was empirically explored via a series of further telephone interviews with customers. Reassuringly, the majority of participants interpreted the wording of the river and coastal water quality service issues in the intended way, i.e. as describing a service level that could potentially be improved (92% and 82%, respectively) and therefore congruent with WTP, with the remaining minorities (8% and 18%) interpreting them as deteriorations subject to WTA. The estimated impact scores for bathing water quality and river water quality were therefore adjusted by dividing them by $0.92+0.08\times6.23$ and $0.82+0.18\times6.23$, respectively. All impact scores were then rescaled such that the revised total added up to 100 again. Both sets of adjusted and unadjusted scaled impact scores are presented in Table 6.

factor (f_k) is to be multiplied with Δq_k to convert it to n_k. These conversion factors, along with the assumptions underlying their derivation, are presented in Table 5.

Table 5: PR19-style MaxDiff units per Package unit

Package service measure	Package unit	MaxDiff service issue	Conversion factor (f_k)
Water service			
Discoloured water	Nr. properties	Discoloured water (a week)	1
Short-term interruptions	Nr. properties	Unexpected interruption (3-6h)	1
Long-term interruptions	Nr. properties	Unexpected interruption (24-48h)	1
Temporary use ban	Chance	Temporary use ban (May-Sep)	$N_{Water,HH}$ (1)
Non-essential use ban	Chance	Non-essential use ban (May-Sep)	$N_{Water,NHH}$ (2)
Persistent low pressure	Nr. properties	Persistent low pressure	1
Wastewater service			
Sewer flooding inside properties	Nr. properties	0.444*(Major sewer flooding inside your property) + 0.556*(Minor sewer flooding inside your property)	1
Sewer flooding outside properties	Nr. properties	0.712*(Sewer flooding outside your property) + 0.288*(Sewer flooding in a nearby public area)	1
Odour from sewage works	Nr. properties	0.406*(Odour from sewage works (frequent)) + 0.594*(Odour from sewage works (infrequent))	1
Significant pollution	Incidents	Significant pollution	N_{Waste}/riverkm(3)
Minor pollution	Incidents	Minor pollution	N_{Waste} /riverkm(3)
River water quality	% Good	River water quality	N_{Waste}/100 (4)
Bathing water quality	% Excellent	Bathing water quality in local area Good but not Excellent	N_{Waste}/100 (5)
	% Sufficient	Bathing water quality in local area Sufficient but not Good	N_{Waste}/100 (5)

(1) $N_{Water,HH}$ represents the number of DCWW household water customers, who are all 'Dual' customers benefiting both from water and wastewater services, and who numbered 1,226,286 in 2017. (2) NWater,NHH represents the number of DCWW non-household water customers, also all 'Dual' customers who numbered 78,492 in 2017. (3) N_{Waste} represents the number of wastewater customers (1,446,538 customers in 2017); 'riverkm' represents the length of river in DCWW's wastewater supply area (7,318 KM). This scaling factor is based on two assumptions: i) that pollution incidents affect 1km of river, and hence that (1/riverkm) represents the proportion of the river network affected per pollution incident; and ii) that 1% of the river is considered local by 1% of customers, or equivalently that an improvement to 1% of the river network benefits 1% of customers. (4) As above, this is based on the assumption that an improvement to 1% of the river network benefits 1% of customers; hence the chance of experiencing the service issue falls by 1%. (5) As with rivers, this is based on the assumption that an improvement to 1% of coastal bathing waters affects 1% of customers, and hence that the chance of experiencing the service issue falls by 1%.

Results and discussion

For ease of illustration, this section will only discuss results from the analysis of the household dual-supply customers (n=1,000). In terms of the package valuation and the unit WTP values, as mentioned above these will be based only on the SQ to +1 improvement scenario.

Impact weights

The rank-ordered logit (ROL) model estimates of impact scores are presented in Table 6. ‘Coastal bathing water quality sufficient but not good’ was omitted as base for model identification. For the remaining attributes, impact scores are presented. As described earlier, the impact score for a given attribute is the perceived disutility of a service failure in ratio to the chosen benchmark; a higher impact score signifies therefore a higher importance of the attribute. Thus, impact scores can therefore be used as weights to apportion package WTPs to individual attributes.

The model results are satisfactory with high a pseudo-R^2 statistic and most coefficients being highly significant. Furthermore, the magnitudes of the coefficients correspond to expectations. Impact scores are found to be increasing for the four interruptions attributes with respect to the duration of the interruption. In addition, extensive internal sewer flooding has a higher impact score than minor internal sewer flooding which in turn has a higher impact score than external sewer flooding. Indeed extensive internal sewer flooding stands out as the clear most-impactful service issue overall, which was consistent with prior expectation. And finally, significant pollution incidents have a higher impact score than minor pollution incidents, as expected. The only result that goes counter expectation is the seeming indifference between frequently and infrequently experienced odour from wastewater treatment works. Overall, the results show a highly plausible pattern of variation.

Table 6: MaxDiff Rank-ordered impact scores

Variable	Impact scores ($\exp(\beta_k)$)	Scaled impact scores (to 100)	Adjusted scaled impact scores (w_k)
Discoloured water (a week)	2.40***	2.04	2.06
Taste & smell not ideal (a few days)	3.22***	2.73	2.76
Short-term interruption (3 to 6 hours)	1.14	0.97	0.98
Short-term interruption (6 to 12 hours)	1.28	1.09	1.10
Long-term interruption (24-48 hours)	6.92***	5.87	5.94
Long-term interruption (7 days)	12.48***	10.59	10.72
Persistent low water pressure	2.35***	2.00	2.02
TUB / NEUB (May to September)	0.72***	0.61	0.62
Sewer flooding inside your property-Severe	40.04***	33.97	34.38
Sewer flooding inside your property-Minor	24.14***	20.48	20.73
Sewer flooding outside your property	5.97***	5.06	5.12
Sewer flooding in a nearby public area	4.03***	3.42	3.46
Odour from sewage works-Minor	1.58***	1.34	1.35
Odour from sewage works-Severe	1.57***	1.33	1.35
Significant pollution incident	4.60***	3.90	3.95
Minor pollution incident	1.79***	1.51	1.53
River water quality less than Good locally	1.83***	1.55	0.81
Bathing water quality less than Excellent locally	0.82***	0.69	0.49
Bathing water quality less than Good locally	*1.00 (base)*	0.85	0.61
Pseudo R-squared	0.239		
Participants	998		

* significant at 10%; ** significant at 5%; *** significant at 1%.

Package valuation

The Package exercise data was analysed by means of conditional logit (CL) models (Table 7). (Dis)utility coefficients for 2 alternative-specific constants (ASC) were estimated: (1) A ‘-1’ ASC representing the deterioration option (A), and (2) a +1 (C) or +2 (D) ASC pooling together the +1 and +2 improvements, assuming there is no additional WTP for the +2 package beyond the WTP for the +1 package (and hence both packages would have the same total WTP relative to the *status quo*). The reason for the restriction in (2) was a recurring problem for the +2 improvement package that was lower than for +1, if not negative altogether. This suggested, against expectations, that the +1 option is preferred to the +2 option. In addition, only a coefficient for bill increase was estimated in the restricted model. This is because the earlier unrestricted model was fitted with a linear spline for bill change coefficients, returned a counterintuitively positive and insignificant coefficient, suggesting that customers prefer lower bill reductions[6].

The model's results conformed to expectations. As anticipated, households showed strong aversion to a potential deterioration of the offered service level signified by the large negative coefficient for "Package –1" in Table 20. Moreover, this model shows that respondents have a strong preference for the improvement scenarios "+1 or +2", as evidence by the positive and significant coefficient. Finally, the significantly negative bill coefficient indicates that respondents are, as expected, averse to bill increases.

Package WTP estimates were then derived as specified earlier. Since the bill change was modelled as a percentage, the initial estimate was multiplied by the average annual household bill in 2017 for dual customers (£439/hh/yr). A WTP point estimate of £50.37/hh/yr for the "SQ to +1" package was derived from this model, and this was used as the basis of the unit WTP values presented and discussed in the next section.

Table 7: Package Exercise Conditional Logit and WTP (£/hh/yr) estimates

Variable	Restricted	*WTP (95% Conf. int)*
Package -1	-1.536***	-
Package SQ (base)	0.000 (omitted)	-
Package +1 or Package +2	0.826***	£50.37 (44.81 ; 55.93)
Bill (%)* Bill increase	-7.195***	-
Observations	*7,992*	
Pseudo-R2	*0.113*	

* significant at 10%; ** significant at 5%; *** significant at 1%

Unit WTP values

The unit WTP values (*$uWTP_k$*) presented in Table 8 were derived for the SQ to +1 service improvements domain as per equation 3. The inputs to the calculations were based on the WTP estimate in Table 7, adjusted impact scores w_k in Table 6, n_k or Δq_k SQ to +1 attribute changes in Figure 2, and f_k conversion factors in Table 5. They represent the

[6] Further analysis was undertaken to derive +1 to +2 utility and WTP values. Results from this analysis and for the original unrestricted model are available upon request.

value of a 1 unit change in the level affecting one property of each service measure; for example, the value of one short-term interruption or one discoloured water incident avoided at one property. Suggested ranges are reported in each case, where these are based on 95% confidence intervals incorporating sampling error in both the main survey and the follow-on survey used to derive weights for the adjustments to impact scores.

Table 8: Unit WTP ($uWTP_k$) values for SQ to +1 service improvements

Service issue	Unit	Central	Range
Water service			
Long-term interruption (7 days)	prop.	£6,743	(£5,691 ; £7,772)
Long-term interruption (24-48 hours)	prop.	£4,009	(£3,348 ; £4,656)
Taste & smell not ideal (few days)	prop.	£1,567	(£1,345 ; £1,784)
Discoloured water (a week)	prop.	£1,162	(£998 ; £1,322)
Persistent low water pressure	prop.	£1,110	(£958 ; £1,258)
Short-term interruption (6-12 hours)	prop.	£702	(£591 ; £810)
Short-term interruption (3-6 hours)	prop.	£636	(£534 ; £735)
Temporary use ban (May to Sep)	prop.	£305	(£268 ; £341)
Non-essential use ban (May to Sep)	prop.	£32	(£23 ; £41)
Wastewater service			
Major sewer flooding inside property	prop.	£22,470	(£17,115 ; £28,512)
Minor sewer flooding inside property	prop.	£13,024	(£10,224 ; £16,094)
Sewer flooding outside property	prop.	£3,090	(£2,457 ; £3,778)
Sewer flooding in a public area	prop.	£1,979	(£1,603 ; £2,382)
Odour from sewage works (freq.)	prop.	£713	(£600 ; £828)
Odour from sewage works (infreq.)	prop.	£712	(£600 ; £825)
Significant pollution incident	local prop.	£2,128	(£1,746 ; £2,537)
	incident	£397,225	(£325,624 ; £473,678)
Minor pollution incident	local prop.	£805	(£674 ; £941)
	incident	£150,459	(£125,822 ; £175,906)
River water quality less than Good to Good	local prop.	£424	(£242 ; £954)
	%	£5,374,271	(£3,095,025 ; £11,991,580)
	km	£73,439	(£42,293 ; £163,864)
Bathing water quality Sufficient to Good	local prop.	£309	(£182 ; £505)
	%	£3,994,362	(£2,373,328 ; £6,497,713)
	site	£3,916,041	(£2,326,792 ; £6,370,306)

* Ranges based on 95% confidence interval of the package WTP estimate.

The figures in Table 8 are to be interpreted as in the following examples:

- The total value to DCWW customers of a reduction by 1 in the number of properties expected to suffer from a long-term interruption lasting 7 days was £6,743 per year for improvements up to the +1 level of service. Similar interpretations apply to all other unit values with ‘prop.’ as the unit of measure.

- In the case of significant pollution incidents, there are two units of measure shown. The first, ‘local prop.’, can be applied when the number of properties local to the pollution incident can be identified. Alternatively, the per incident value of £397,225 (for the SQ to +1 range of improvement) is the expected value to be used which will be valid on average for all significant pollution incidents.

- Multiple units are also shown for river water quality and bathing water quality attributes. Here again, values per local property are shown, which can be used when the number of properties local to the river or beach can be identified. There are also average values per percentage of river/bathing water in the DCWW wastewater supply area, and values per km, in the case of rivers, and per site in the case of bathing waters.

Table 9 shows unit values pertaining to dual-service households for those service measures where comparisons are possible against PR14 reported values, for DCWW and for the industry as a whole. This table suggests that the values are below the PR14 industry range for persistent low pressure and sewer flooding service issues, and above this range for temporary use bans and bathing water quality.

Table 9: Comparison of PR19 and PR14 Unit Values (Dual-service households)

Service issue	Unit	PR19	PR14 (DCWW)	PR14 (Industry range)
Water service				
Discoloured water (for a week)	Prop.	£1,087	£1,525	[£586 ; £3,060]
Short-term interruption (3–6h)	Prop.	£515	£50	[£50 ; £1,670]
Long-term interruption (7 days)	Prop.	£5,646	£13,662	[£277 ; £13,662]
Long-term interruption (24–48h)	Prop.	£3,131		
Temporary use ban (May-Sep)	Prop.	£325	£34	[£0 ; £123]
Persistent low pressure	Prop.	£1,000	£22,345	[£1,408 ; £28,462]
Wastewater service				
Major sewer flooding inside property	Prop.	£20,048	£97,984	[£22,530 ; £367,291]
Minor sewer flooding inside property	Prop.	£11,193		
Sewer flooding outside your property	Prop.	£2,695	£3,397	[£2,869 ; £162,570]
Sewer flooding in a nearby public area	Prop.	£1,755		
Odour from sewage works (frequent)	Prop.	£631	£1,536	[£166 ; £11,820]
Odour from sewage works(infrequent)	Prop.	£631		
Significant pollution	Incident(1)	£366,555	£1,119,639	[£22,200 ; £54,100,000]
Minor pollution	Incident(1)	£135,530	£42,669	
River water quality (Good)	km(2)	£73,071	£4,017	[£4,017 ; £256,889]
Bathing water quality (Suff.)	site(3)	£3,902,383	£202,921	[£79,080 ; £959,575]

Source for PR14 values: Accent-PJM (2014) Comparative review of WTP results. Values in red are those lying above the PR14 industry range; those in blue lie below the PR14 industry range.

In comparison to the DCWW PR14 numbers, the PR19 values show the following key differences:

- PR19 values are higher than at PR14 for short interruptions, Temporary Use Bans, and river and bathing water quality
- PR19 values are lower than at PR14 values for long interruptions, low pressure and sewer flooding

These differences are not unexpected as they can be explained as a consequence of the change in design approach. The PR19 approach is much more sensitive to the scale of service change, by construction, due to the fact that the relative values assigned to different service improvements are now derived as impact-weighted sums of the numbers of customers affected, rather than directly obtained from customers trading them off against one another. As a result, the PR14 approach probably over-valued attributes with small risk reductions (e.g. long interruptions, low pressure and sewer flooding), and under-valued attributes affecting lots of customers (e.g. short interruptions, environmental attributes and temporary use bans). Accordingly, we believe the PR19 numbers to be a more valid and reliable measure of value overall.

4. Conclusion

This paper presents a novel methodology for valuing water and wastewater service improvements via impact-weighted numbers of service failures. Its main advantage is that it obtains WTP estimates per unit of service failure that are proportional to the impact of the service failures themselves without being driven by differences in the scopes of service change offered. The overall size of the values is still dependent on the scope of service change offered overall, but this was selected to be consistent with a realistic package of improvement and so WTP values can be interpreted in this context.

A further advantage of the proposed approach is that it can be implemented using a substantially simpler and less abstract research instrument. This is because it avoids the need for respondents to evaluate lots of different small risk changes, the source of much of the difficulty in the valuation approach widely adopted for PR14.

Notwithstanding these advantages, there are three key potential limitations. Firstly, one practical limitation is that the method is designed to work with a suite of service measures that can be straightforwardly aligned to a (negative) impact experienced by a customer. In most cases this is natural, but with some service measures, particularly those focused on improving environmental outcomes, the construct can become somewhat strained.

A second limitation is that the overall size of the valuations is still likely to be dependent on the scope of service change offered overall. This should be selected to be consistent with a realistic package of improvement, but there could be different sizes of improvement package that are considered realistic and so WTP values may still not be fully reliable with respect to alternative reasonable designs at the package level. This issue is likely to be much less significant than under the previous approach, however, due to the fact it is only the overall scale of valuations that is affected rather than also the relative valuations of each individual service measure.

The final potential limitation is more philosophical in nature. The new approach imposes a rational structure to valuations which is less likely than the previous approach to be consistent with what customers would choose for themselves if given a direct choice between packages of service levels. One could therefore decry the new approach as being less consistent with the principle of consumer sovereignty. The extent to which this is considered to be a concern depends on one's perspective on the role of appraisal – is its purpose to achieve an outcome consumers would choose for themselves, even if that choice is fraught with difficulties, or is its purpose to maximise welfare as experienced by the consumer? Cost-benefit analysis adopts both perspectives at different times, although most often little attention is given practically to such philosophical questions [Smith and Moore, 2010].

Given the various strengths and limitations of alternative approaches, and the different philosophical perspectives that underpin them, a pragmatic recommendation for cases involving appraisal and prioritisation of large-scale business plan programmes would be to approach the task of valuation from multiple angles, subject to resource constraints, incorporating different types of stated preference research, alongside revealed preference and well-being evidence where possible with a view to building a broad and deep evidence base that can be used for triangulation. Within this evidence base, there ought, in our view, to be a significant place for a study utilising the methodology proposed in this paper.

References

Accent and PJM Economics. 2017. ' Dŵr Cymru Welsh Water PR19 Willingness to Pay Research. Final Report, December 2017'.

Adamowicz, Wiktor, Diane Dupont, Alan Krupnick, and Jing Zhang. 2011. 'Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods', *Journal of Environmental Economics and Management*, 61: 213-26.

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.

Beggs, S., S. Cardell, and J. Hausman. 1981. 'Assessing the potential demand for electric cars', *Journal of Econometrics*, 17: 1-19.

Carson, R. T., N. E. Flores, and N. F. Meade. 2001. 'Contingent valuation: Controversies and evidence', *Environmental & Resource Economics*, 19: 173-210.

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.

Hammitt, J. K. 2000. 'Valuing mortality risk: Theory and practice', *Environmental Science & Technology*, 34: 1396-400.

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.

Hanley, N., S. Mourato, and R. E. Wright. 2001. 'Choice modelling approaches: A superior alternative for environmental valuation?' *Journal of Economic Surveys*, 15: 435-462.

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

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

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

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.

Louviere, Jordan J., Terry N. Flynn, and A. A. J. Marley. 2015. *Best-Worst Scaling: Theory, Methods and Applications* (Cambridge University Press: Cambridge).

Metcalfe, P. J., and A. Sen. 2021. "Sensitivity to scope of stated preference water and wastewater service valuations: A meta-analysis of willingness to pay findings from the 2014/15 water price reviews in Great Britain." *Mimeo.*

Ofwat. 2013. "Setting price controls for 2015-20 – final methodology and expectations for companies' business plans. July 2013."

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

Ojea, E., and M. L. Loureiro. 2011. 'Identifying the scope effect on a meta-analysis of biodiversity valuation studies', *Resource and Energy Economics*, 33: 706-24.

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.

Smith, V. K., and E. M. Moore. 2010. 'Behavioral Economics and Benefit Cost Analysis', *Environmental & Resource Economics*, 46: 217-34.

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.

Tunçel, T., and J. K. Hammitt. 2014. 'A new meta-analysis on the WTP/WTA disparity', *Journal of Environmental Economics and Management*, 68: 175-187

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

———. 2015. "Post PR14 Customer Engagement, Communications and Education." In.

Whitehead, J. C. 2016. 'Plausible responsiveness to scope in contingent valuation', *Ecological Economics*, 128: 17-22.

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.