IMPACT
FROM INSIGHT TO INFLUENCE

PR24 Triangulation

(Phase 2 Results)

Final Report

Prepared for South Staffordshire Water (SSC)

Prepared by Impact Research

22 June 2023

Project No. 1345

South Staffs Water   Cambridge Water

Table of contents

GLOSSARY

Industry Terms

AMP7	Seventh Asset Management Period
BAU	Business As Usual
CAC	Cost Adjustment Claim
CAPEX	Capital expenditure
CBA	Cost Benefit Analysis
ICG	Independent Challenge Group
CMEX	Customer measure of experience
CRI	Compliance Risk Index
CTS	Cost to Serve
DMEX	Developer services measure of experience
ERI	Event Risk Index
ESG	Environmental, Social and Governance
FCR	First Contact Resolution
IIP	Investors in People
LTDS	Long Term Delivery Strategies
LTF	Looking to the Future
MCDA	Multi-Criteria Decision Analysis
ODI	Outcome delivery incentive
PC	Performance Commitment
PCC	Per Capita Consumption
PCD	Price Control Deliverables
PIC	Public Interest Commitment
PR19	Price Review 2019
PR24	Price Review 2024
WASC	Water and Sewerage Company
WOC	Water Only Company
WRMP	Water Resource Management Plan

Organisations

CAM	Cambridge Water supply region
CCW	The Consumer Council for Water
Ofwat	The Water Services Regulation Authority
SSC	South Staffs Water & Cambridge Water
SSW	South Staffs Water supply region

Customer Engagement

DCE	Discrete Choice Experiment
CVM	Contingency Value Method
HH	Household
H2Online	SSC customer community – household
NHH	Non-household
RP	Revealed Preference
SP	Stated Preference
WTP	Willingness to Pay

1. BACKGROUND

Aims

Impact Research was commissioned in June 2022 by SSC to provide the following:

- To deliver a robust triangulation of customers' and stakeholders' priorities that underpins the narrative of SSC's Price Review 2024 (PR24) plans.
  - o Robustly triangulate evidence relating to WRMP to support all key decisions.
  - o Support the development of SSC's Performance Commitment (PC) package.
  - o Triangulate WTP values to set central, upper and lower values for use in Cost Benefit Analysis (CBA).
- To support the development of SSC plans with triangulated valuations and insights to best deliver 'public value'.
- Create an insight matrix from SSC's strategic trackers to assist in the delivery of the PR19 plan and guide PR24.
- Enable both SSC's Stakeholder Challenge Panel (or ICG) and board to effectively challenge the approach, plus independent review by a third-party expert.

The key focus of this report is to assess previously available **Willingness to Pay (WTP)** information and the latest PR24 WTP results commissioned by SSC, together with the Willingness to Accept (WTA) values from the Outcome Delivery Incentive Research developed centrally on behalf of Ofwat[1].

Executive Summary

This triangulation was developed from the approach used for SSC in PR19, which was extensively peer reviewed and commented on by Ofwat as showing good evidence of triangulation. The main developments were the extension of the criteria by which sources were evaluated and weighted (the 'RAG' ratings) and the inclusion of an external 'expert panel' (Delphi method). Sources older than six years from PR14, which had featured in PR19, were removed this time around.

Monetary-equivalent values were derived from the weighted averages of relevant values from each of the sources, with the most focus placed on studies specifically commissioned for PR24. These were provided to SSC as low-central-high values for inputs to the 'Copperleaf' investment modelling process, allowing sensitivity analyses over a range of outcomes.

One of the main criticisms of the approaches taken by water companies in PR19 was the variety of survey methods employed and the resulting wide variations in WTP values. However, this was accompanied by a call for more innovation in the way WTP values are measured. For PR24, this appears to have resulted in an even greater range of values, driven by the use of new approaches that departed from more established methods such as DCE and CVM and were largely untested prior to this round of Price Review research and required more application to build confidence in their use.

For the PR29 review, it is recommended that OfWat spend the intervening time evaluating these new methods and to proscribe the most acceptable approach to WTP measurement. This can then become the standard method which all water companies can use to evaluate their schemes on a consistent basis.

---

[1] https://www.ofwat.gov.uk/wp-content/uploads/2022/12/PR24_final_methodology_Appendix_8_Outcome_delivery_incentives.pdf

2. TECHNICAL TRIANGULATION DETAILS

The details of the triangulation methodology for PR24, an enhancement of the PR19 SMARTS approach, are outlined in '**SSC08 PR24 Technical triangulation - Phase 1 Methodology**' and can be summarised as:

- SCREEN data sources to identify those with potentially comparable measures
- MAP non-core evidence to core measures where possible to enable comparison
- ASSESS theoretical and statistical validity of the resulting measures
- RATE measures as Red/Amber/Green (RAG) depending on how well they perform with respect to the validity measures
- TRIANGULATE to conclude on the values to take forward based on applying RAG weights to obtain central values and ranges.
- SENSITIVITY TEST the results based on amending the weights to conform with alternative reasonable perspectives.

A detailed account of the original approach taken for PR19 is given in '**SSC PR19 Customer Data Triangulation - Final Report**'[2].

Here we outline the specific steps taken to establish the set of WTP values suitable for use in SSC's Copperleaf investment modelling. These figures were assessed by a 'Delphi panel,' a small group of external evaluators convened specifically for the purpose of assessing the triangulation outputs, and by the external peer reviewer, Professor Iain Fraser. Their observations and inputs are incorporated in this report.

The PR19 Approach

Table 2.1 lists the sources that were included in the PR19 analysis, together with the Red-Amber-Green ratings that were assigned to them at the time (based on an assessment of theoretical and statistical validity), together with the corresponding weights that determined the contribution of each set of results to the final PR19 WTP central values. For example, over half (52%) of the final WTP mean average values derived from this process were driven by the results of the Discrete Choice surveys conducted with household customers.

Table 2.1: Sources used in the PR19 SMART analysis

Source		RAG Rating	Weight	Contribution
WTP core_DCE	Willingness-to-pay research to support PR19: Technical Report (Jan / Feb 2018)	**Green**	100%	26%
WTPCore_DCE2		**Green**	100%	26%
WTP core_Maxdiff		Green / Amber	50%	13%
PC Slider	SSW PC slider data main results 2018	Green / Amber	50%	13%
Priorities	SSC Customer priorities online survey results 2018	**Amber**	25%	6%
Contacts	Triangulation approach using SSC BAU Data	**Amber**	25%	6%
Satisfaction	SS HH Tracker (2017-2018) Regional Annual summary FINDINGS	**Amber**	25%	6%
WTPPR14	Household customers only - Accent and PJM Economics, July 2018, PR19 Data Triangulation, SSC	**Amber**	25%	6%
WRMP online	Household customers only - SSC WRMP Quant Survey 2018	**Amber**	25%	6%

---

[2] Accent and PJM Economics, July 2018, PR19 Data Triangulation, Report for South Staffordshire Water

Source		RAG Rating	Weight	Contribution
WRMP workshops	WRMP & Long Term Resilience Customer Engagement Insight: Full report (September 2017)	**Amber**	25%	6%
ExternalWTP19	Comparative Review of PR19 WTP Results (Accent and PJM)-May 2018	**Amber**	25%	6%
ExternalWTP14	Comparative Review of Willingness to Pay Results (Accent and PJM)-June 2014	Amber / Red	10%	3%
WTPCore_DCE2a	Non-household (DCE2a), Household customers (tests of sensitivity to Low bills) and both types of customers ('Private' priorities) - Willingness-to-pay research to support PR19: Technical Report (Jan / Feb 2018)	**Red**	0%	0%
WTPCore_DCE2_LowBill		**Red**	0%	0%
WTP core_DCE_Private		**Red**	0%	0%

SCREEN - Source documents considered and screened for triangulation at PR24

Studies that were considered as potential evidence for triangulation for PR24 included all those used in PR19. Table 2.2 below lists a summary of the additional sources of information gathered since then and screened into the process on the basis that they provide WTP values suitable for inclusion.

Table 2.2: New Sources screened in for use in the PR24 SMART analysis

Source		Comment
Priorities	SSC Customer priorities online survey results 2022	Updated information supplied by SSC drawn from recent research on household customers' priorities
Contacts	Triangulation approach using SSC BAU Data – 2021-2022	Updated customer contact information supplied by SSC
Satisfaction	Customer satisfaction metrics 2020-2022	Updated information supplied by SSC drawn from Qualtrics point of contact customer satisfaction survey tracking research. An 'FCR' score was derived from the correlations for each measure against overall satisfaction, to represent the relative importance of each service attributes to customers.
WRMP24 MCDA	Accent Quant MCDA Study - Feb 2022 - FINAL REPORT.pptx	Of the range of items tested, only 'Reduce leakage' and 'Habitats for native wildlife and plants' were relevant for this study. Households only.
Strategic Research Options (SRO)	3543rep02_master_SRO_Final_v7.pdf	Of the range of items tested, the only item of relevance to this PR24 work was 'Specialist habitats created for wildlife'
ODI rates	3524m_ODIResultsGuidance.pdf	New research undertaken for Ofwat that covered half of the service attributes tested for SSC.
		Results available for both SSC and nationally
NERA WTP PR24	221215 SSCW Report Draft - Revised.pdf	New research undertaken for SSC to evaluate 12 areas of improvement
	Wessex and Yorkshire WTP PR24 study results.pdf	New research undertaken for Wessex Water included 4 relevant areas of improvement. New research undertaken for Yorkshire Water included 3 relevant but different areas of improvement

The information from these sources, combined with the key calculations of the SMART process (specifically the RATE element discussed later in this chapter) are all contained in **'SSC09 PR24 Technical triangulation – Phase 2 dataset**. This resource contains a detailed overview of each source of insights including project objectives and sample sizes. Other sources which contained potentially relevant WTP values but were rejected are listed in Table 2.3a.

Table 2.3a: Sources screened out of the PR24 SMART analysis

Source		Reasons for omission
WTPPR14	Accent and PJM Economics, July 2018, PR19 Data Triangulation, SSC	Values approximately 10 years old
ExternalWTP14	Comparative Review of Willingness to Pay Results (Accent and PJM)-June 2014	Values approximately 10 years old
External WRW 2021	WRW valuations March 2021.docx	Used unaltered PR19 WTP results – i.e., duplication of PR19 values

Finally, a wide range of other sources were considered, all of which we rejected on the basis that they did not provide specific WTP values and / or were not quantitative in nature. This is in line with the PR19 approach, where triangulation focussed on those sources that offered quantitative values for relevant service attributes. The summary of studies rejected at PR19 (and now also for PR24) are listed in Table 2.3b.

Table 2.3b: Other Data Sources Screened Out[3]

Source	Reasons for screening out
Foundation priorities	Customers priorities for service delivery both at present and over the longer term (prompted and spontaneous). Purely qualitative and discursive hence no useable measure for this application.
Metering study	Focussed on customers' reasons for not switching to a meter, hence not comparable.
Bright (point of contact satisfaction surveys and Ofwat SIM survey	Satisfaction measurement of key interaction points – billing and operational hence cannot be used directly as we will need to relate satisfaction to experience of service failure. Satisfaction is not itself a good measure of WTP since satisfaction with different service elements contributes differently to overall satisfaction depending on how important the service element is to the customer.
SSC Web survey	Satisfaction measures relating several service measures. No comparable measures.
CCWater studies	Qualitative research about attitudes to water use/saving, behaviours and messaging. No comparable measures.
External WTP evidence	*VoLL (energy)*: This was screened out for two reasons: firstly, our approach requires at least two measures since we package scale everything; secondly, VoLL is insufficiently closely related to supply interruptions, particularly given that there are many direct valuations obtained for interruptions in the water sector.

For PR24, the ambition was to broaden the scope of the work to make more explicit use of qualitative and more generic studies. This took the form of inputs to summary materials developed for use in the Delphi approach conducted with a panel of four experts covering a range of specialisms of customer engagement, utility policy development and valuations research. They would consider this information alongside the central, lower and upper WTP values derived from the studies reported in this document and inform their view as to whether the values given from the main quantitative sources are low, high or representative of customer priorities as expressed in summary material and supporting documents.

---

[3] Accent and PJM Economics, July 2018, PR19 Data Triangulation, Report for South Staffordshire Water, p32

MAP - Areas tested and Relevant Attributes

The evidence was assessed in relation to the specific service attributes under consideration by SSC for use in its Copperleaf investment tool. Figure 2.1 summarises the process by which these 12 service attributes were selected. The final set of service attributes to be triangulated for WTP purposes are summarised in Table 2.4.

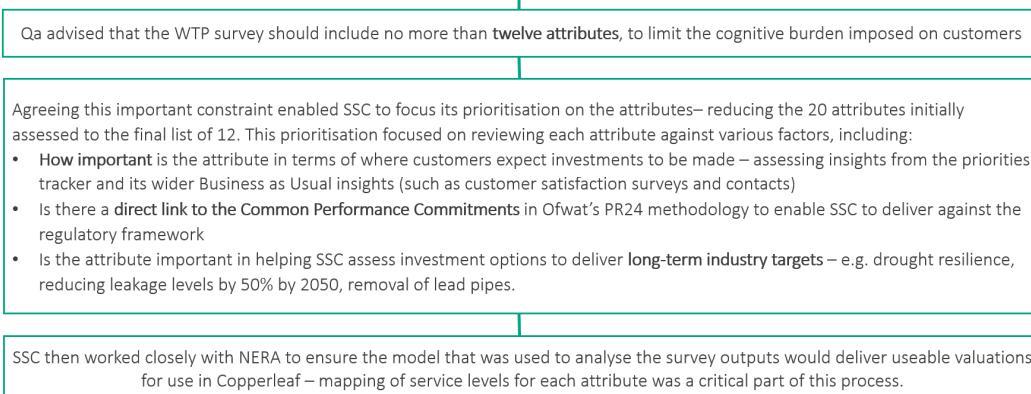*Figure 2.1: Selection of service attributes*
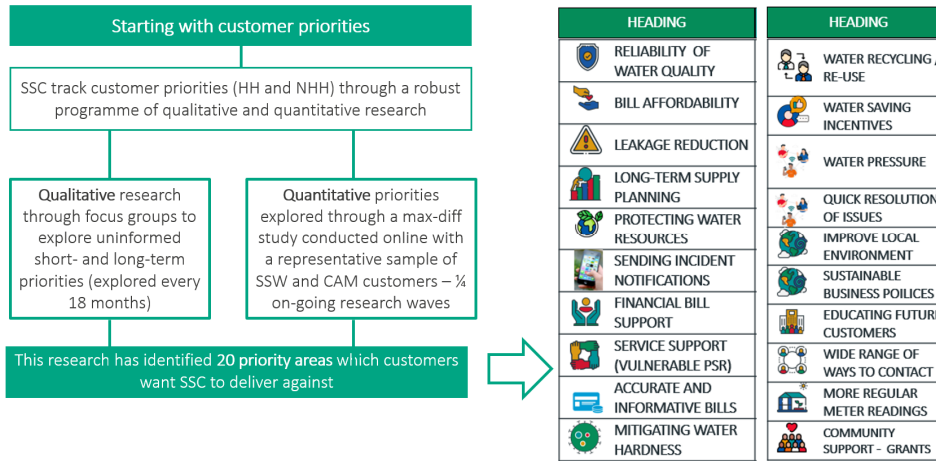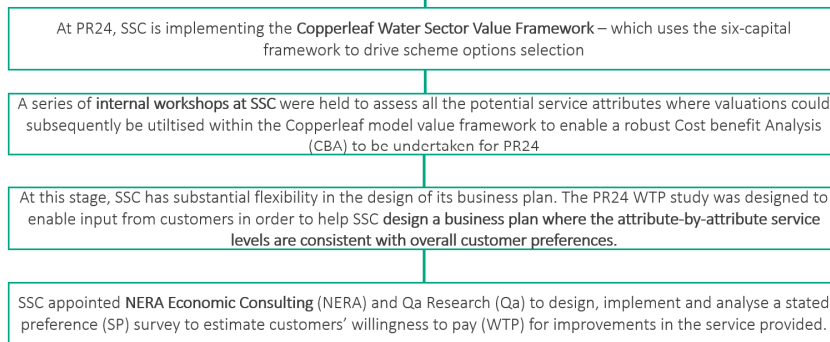
Table 2.4: Areas tested for PR24 in the NERA SSC Study

SSC outcome	SSC's Technical Description of Service (the issue)
Customer Service	To provide excellent levels of service when customers get in touch with queries – by phone, email, online, letter, or face-to-face. In 2021/22 (TEXT SUB: South Staffs Water / Cambridge Water) customer satisfaction was rated 3rd out of all 17 water & sewerage companies in England and Wales.
Risk of a temporary "do not drink" notice	Occasionally, water companies have to send customers a notice saying not to drink the tap water because of an issue with the water quality. Usually this would last about 2-3 days, and (TEXT SUB: South Staffs Water / Cambridge Water) would provide safe drinking water near your property at temporary water stations and would deliver bottled water directly to vulnerable households.
Installing 'smart' water meters	(TEXT SUB: South Staffs Water / Cambridge Water) needs to carefully manage demand for water to ensure there is enough for the future. 'Smart' water meters automatically send regular readings. Having more information helps the water company and customers to understand where and when water is being used, or lost to leaks.
Hard water supply	(TEXT SUB: South Staffs Water / Cambridge Water) has a hard water supply. Hard water is not harmful to human health, but it can lead to limescale damage on taps, showerheads and appliances (e.g., washing machines).
Lead pipes	Some properties in your area are served by a lead supply pipe. Most of these pipes are owned by the customer and not your water company. (TEXT SUB: South Staffs Water / Cambridge Water) treats the water supply to ensure lead levels in the water are safe, but there are some circumstances where it can become unsafe (e.g., if lead pipes are badly damaged). Over time, lead exposure can be damaging to health.
Water lost to leakage from pipes	Every day, treated water is lost to leakage from the (TEXT SUB: South Staffs Water / Cambridge Water) pipe network as pipes age or are damaged. The majority of the water lost to leaks is from the water company's pipes (70%) and the rest is from customer pipes. The company aims to fix the largest and most disruptive leaks first.
Issues with tap water colour, taste, or smell	Every year, some (TEXT SUB: South Staffs Water / Cambridge Water) customers suddenly experience a temporary issue with the look, taste or smell of their tap water. The water is still safe to drink. The most common issues are the water turning a light brown colour or a chlorine smell, typically lasting up to 24 hours.
Chance of property flooding from a burst pipe	Sometimes the main water supply pipe owned by the water company can burst and flood the ground floor of a customer's home or business. When this happens, (TEXT SUB: South Staffs Water / Cambridge Water) covers the cost of the repair through its insurance to get the property put back as it was.
Low water pressure	Every year some properties experience temporary periods of low water pressure, normally lasting less than 6 hours. These periods of low pressure are usually caused by problems with the pipe network.
Supporting nature and wildlife	(TEXT SUB: South Staffs Water / Cambridge Water) has a legal duty to protect and enhance nature and wildlife and ensure there is no permanent damage to the areas where it operates. The company aims to ensure rivers, (TEXT SUB IF CAM: chalk) streams, reservoirs and underground water stores are healthy.
Unplanned short interruptions to water supply	Every year some customers will experience a short interruption to their property's water supply, where it suddenly stops working without warning for 3-6 hours. During this type of interruption, (TEXT SUB: South Staffs Water / Cambridge Water) would deliver bottled water directly to the homes of vulnerable people.
Risk of temporary use ban, including hosepipes	To protect essential water supplies during extended periods of dry weather, (TEXT SUB: South Staffs Water / Cambridge Water) may send you a notice saying you must not use a hosepipe or sprinkler or use water for other non-essential uses. The length of temporary use bans can vary, but are usually issued for five months, between May and September.

Table 2.5 lists the attributes from the chosen triangulation sources that are applicable to the attributes selected by SSC for PR24 investment analysis. It should be stressed that each study had different definitions of each attribute and the extent to which these differ from those listed in Table 2.4 contributed to the RAG theoretical ratings discussed in the RATE section later in this chapter.

Table 2.5: Sources used in the PR24 SMART analysis

Source		Relevant attributes
WTP core_DCE	Willingness-to-pay research to support PR19: Technical Report (Jan / Feb 2018)	Water not safe to drink
WTPCore_DCE2		Taste and smell of water
WTP core_Maxdiff		Water hardness
		Lead pipes
		Unexpected temporary loss of water supply
		Flooding from a burst pipe
		Low water pressure
		Leakage
		Protecting wildlife habitats
		Temporary use ban
		Water metering
PC Slider	SSW PC slider data main results 2018	Protecting wildlife
		Leakage levels
		Interruptions to water supply
Priorities	Accent/PJM, 2020 - 2022, Priorities Research Qualitative and Quantitative Insight	Hardness - Not impacted by the effects of hard water
		Leakage - Quickly repair and replace pipes
		Pressure - The water always comes out of the taps at pressure
		Reliability - High quality and always safe to drink
		Local Environment - grants that improve local habitats
		Water Resources - Actively protect the environment
		Quick Resolution - They are easy to deal with
Contacts	Triangulation approach using SSC BAU Data	Water not safe to drink
		Discoloured water
		Taste and smell of water
		Unexpected temporary loss of water supply
		Low water pressure
		(Hardness not included due to too few contacts)
Satisfaction	Customer journey satisfaction 2020-2022 (Qualtrics survey)	Discoloured water
		Taste and smell of water
		Unexpected temporary loss of water supply
		Low water pressure
Strategic Research Options (SRO)	Accent SRO Schemes Research, July 2022	Specialist habitats created for wildlife
WRMP24 MCDA	Accent Quant MCDA Study - Feb 2022 - FINAL REPORT.pptx	Habitats for native wildlife and plants
WRMP online	SSC WRMP Quant Survey 2018	Reducing leakage
		Water metering/Installing smart meters
WRMP workshops	WRMP & Long Term Resilience Customer Engagement Insight: Full report (September 2017)	Leakage
		Water metering/Installing smart meters
External WTP19	Comparative Review of PR19 WTP Results (Accent and PJM)- May 2018	Water not safe to drink
		Discoloured water
		Taste and smell of water
		Unexpected temporary loss of water supply
		Low water pressure
		Temporary use ban
		Leakage
		Water metering
		Protecting wildlife habitats
ODI rates	3524m_ODIResultsGuidance.pdf	Water not safe to drink
		Unexpected temporary loss of water supply

Source		Relevant attributes
		Taste and smell of water Discoloured water Low water pressure Temporary use ban
NERA WTP PR24	221215 SSCW Report Draft - Revised.pdf	New research undertaken for SSC to evaluate the 12 areas of service improvement
	Wessex and Yorkshire WTP PR24 study results.pdf	Supply interruptions Water quality Customer service Supporting nature and wildlife
		Drinking water colour, taste and smell Water lost through leaks Low water pressure

ASSESS – the Validity Criteria

The PR19 approach to assessing evidence began with defining two types of validity: 'theoretical' and 'statistical'[4]. To make the approach more inclusive of non-numeric outputs to the process, a third dimension of validity was added for PR24: 'depth'. This relates to the use of qualitative research in the development of the quantitative research instruments and quality and detail of information given to survey participants. The intention was to encourage greater consideration of qualitative inputs to the measurement of WTP. These in themselves would not provide numeric values comparable to those provided by quantitative sources, but there use in the development of those quantitative sources would give more confidence that issues have been covered in sufficient depth for customers to express an informed opinion.

Table 2.6 lists the criteria against which each of the screened-in source material was assessed for theoretical and statistical validity (definitions taken from the PR19 approach) and depth validity (new to PR24).

Table 2.6: Validity Criteria

Theoretical validity	Statistical validity	Depth validity
Are definitions of candidate and target measure the same?	How large is the sample?	What was the extent of explorative and developmental work behind the design of the study?
Are contextual conditions the same between candidate and target measures?	How representative is the sample? • Are the populations the same and, if not, how different are they? • How old is the data? • Are there any biases due to the timing of the research?	What was the quality and detail of information given to survey participants?
If no to either of these, what issues do the differences give rise to?	How wide are the confidence intervals?	
	Have the results been derived using best practice techniques?	

---

[4] Accent and PJM Economics, July 2018, PR19 Data Triangulation, Report for South Staffordshire Water

RATE – Red/Amber/Green (RAG) Ratings

For each of the dimensions of validity (Theoretical, Statistical and Depth), Red/Amber/Green ('RAG') ratings were applied. This was a five-point scale where Green represents strong and Red represents weak. Weights would then be applied according to these classifications to derive final mean average WTP values, as discussed in the next chapter on Triangulation ('Weighted WTP (per unit) Calculations'). A number of sources provided customer values that were then subsequently converted to approximate WTP values using data from the Core WTP research and analytical methods developed by Accent and PJM. For detailed accounts of these methods, the reader is referred to **'SSC PR19 Customer Data Triangulation - Final Report**'.

Tables 2.7a to 2.7m summarise the assessments of validity for each of the screened-in sources, viewed in relation to the attributes now being tested in the new WTP work for SSC in preparation for PR24.

Table 2.7a: WTP core_DCE / WTPCore_DCE2 (2017 / 2018)

Description	Validity	Criteria	Comment	RAG Rating
The 'core' survey research used in PR19 that derived WTP from DCE surveys with attributes that were closely aligned to the planned investment options	Theoretical	Are definitions of candidate and target measure the same?	Covered almost all the areas of interest to PR24 except customer service. Most were defined in terms of the likelihood of an event occurring or specific units of impact (volume of leakage, % households receiving a smart meter, hectares of rives/habitats affected). This is broadly in line with the way attribute levels are expressed in the new PR24 research.	**Green**
		Are contextual conditions the same between candidate and target measures?	In the second wave of this research, respondents were informed as to which level represented the current level of service, but the effect the this versus the first wave was not found to be significant. The study also tested for differences in presenting attributes in terms on 'private' (likely individual customer impacts) and 'public' (impacts on customers as a whole). The latter is more applicable to the current research definitions, but while 'private' values were generally higher than 'public' values, the differences were not considered large.	
		If no to either of these, what issues do the differences give rise to?	Comparisons of results from the different waves indicated that unit WTP values were very sensitive to the scope of change offered, so that differences between the results of these studies and the new WTP research are therefore likely to reflect differences due to the scope definitions	
	Statistical	How large is the sample?	Statistically robust samples were covered for households in both waves of research, but only the second wave obtained a sufficiently large sample of non-household customers for these values to be considered robust.	Green / Amber
		How representative are the sample / timings?	Quotas and the subsequent weighting of data to Census profiles ensured a representative profile of customers. **The data itself is now over 5 years old, since when significant events have occurred** (Covid-19, Brexit, the cost-of-living crisis and increased concern over climate change)	
		How wide are the confidence intervals?	Confidence intervals of up to ±20% of the mean values are fairly common across the attributes tested	
		Have the results been derived using best practice techniques?	The surveys used conventional DCE, rigorously tested with a series of large scale pilot surveys and independent peer review. The study was considered by the latter to have exercised best practice methods.	
	Depth	Extent of explorative and developmental work?	The development of the surveys involved customers extensively in the design of the survey and service measures wordings and supporting stimulus materials.	**Green**
		Quality and detail of information given to survey participants?	Qualitative work was carried out with the aim of producing informative and digestible introductions to the investment areas that would be presented to respondents in the WTP surveys. As a result, comprehension among respondents was considered high, but **descriptions were necessarily brief and simplified**.	

Table 2.7b: WTP core_Maxdiff (2017)

Description	Validity	Criteria	Comment	RAG Rating
Conducted alongside the 'core' DCE survey research conducted for PR19, Max Diff exercises were included to test customers' general priorities, independently of the scope of service level change or bill impacts	Theoretical	Are definitions of candidate and target measure the same?	The Max Diff exercises covered almost all the areas of interest to PR24 except customer service. They were **only presented in terms of improving each area from the current experience** and were not defined in terms of detailed service levels.	Amber
		Are contextual conditions the same between candidate and target measures?	Because **specific levels of service were not specified**, the attributes are relevant at an overall level, though current customers may potentially perceive priorities differently in relation to a changed perception of current service experience	
		If no to either of these, what issues do the differences give rise to?	These results were used in conjunction with the DCE values in the triangulation approach developed by Accent/PJM, so these in turn are ultimately **subject to any scope effects**. By scope effects we refer to the 'span' of service attribute levels – eg one study may present the number of houses experiencing a 'do not drink' order ranging from 1 to 8, another may talk of 100 to 2,000.	
	Statistical	How large is the sample?	Statistically robust samples were covered for households in both waves of research, but only the second wave obtained sufficiently large sample of non-household customers for these values to be considered robust for the latter.	Amber
		How representative are the sample / timings?	Quotas and the subsequent weighting of data to Census profiles ensured a representative profile of customers. **The data is now over 5 years old, since when significant events have occurred**	
		How wide are the confidence intervals?	Confidence intervals of up to ±20% of the mean values are common across the attributes tested, but **when combined with the core DCE values, the resulting intervals become larger.**	
		Have the results been derived using best practice techniques?	The surveys used conventional Max Diff designs, tested with a series of large scale pilot surveys and independent peer review. The study was considered by the latter to have exercised best practice methods.	
	Depth	Extent of explorative and developmental work?	The development of the surveys involved customers extensively in the design of the survey and service measures.	Amber
		Quality and detail of information given to survey participants?	As a result of extensive qualitative work and the simplified presentation implicit to Max Diff, comprehension among respondents was considered high, but **descriptions were necessarily brief and simplified. They were limited to very simple summary descriptions for the purpose of the Max Diff format**.	

Table 2.7c: PC Slider (2018)

Description	Validity	Criteria	Comment	RAG Rating
Performance Commitments (PC) Slider research - Explain Research in 2018. Household customers sked to move the sliders up and down for 11 attributes and seeing the dynamic impact on a typical bill.	Theoretical	Are definitions of candidate and target measure the same?	The research covered three areas of interest to PR24 and 8 others. The service attributes covered were **only presented in terms of movements from the current service level** and were not defined in terms of detailed service levels.	Amber
		Are contextual conditions the same between candidate and target measures?	Because **specific levels of service were not specified**, the attributes are relevant at an overall level, though current customers may potentially perceive priorities differently in relation to a changed perception of current service experience.	
		If no to either of these, what issues do the differences give rise to?	These results were used in conjunction with the DCE values in the triangulation approach developed by Accent/PJM, so these in turn are ultimately **subject to any scope effects**.	
	Statistical	How large is the sample?	Modest household sample sizes (319 SSW and 139 CAM)	Amber / Red
		How representative are the sample / timings?	Online survey only. **The data is now over 5 years old, since when significant events have occurred.**	
		How wide are the confidence intervals?	Confidence intervals of up to ±10% of the mean values are observed, but **when combined with the core DCE values, the resulting intervals become larger.**	
		Have the results been derived using best practice techniques?	The slider method encouraged trade-offs but did not specify service levels. In this respect the data is similar in detail to the Max Diff data, but did not use statistical choice modelling to derive the importance measures.	
	Depth	Extent of explorative and developmental work?	Used feedback from qualitative workshops to help shape the descriptions.	Amber

		Quality and detail of information given to survey participants?	Descriptions were limited to simple summary descriptions for the purpose of the slider format.	

Table 2.7d: Priorities (2022)

Description	Validity	Criteria	Comment	RAG Rating
A regular annual survey, fieldwork undertaken quarterly, partly designed to deliver an index of priority for a range of service measures	Theoretical	Are definitions of candidate and target measure the same?	The research covered seven areas of interest to PR24. They were **only presented in terms of priorities for improvement** and were not defined in terms of detailed service levels or costs of improvement with potential impacts on bills.	**Amber**
		Are contextual conditions the same between candidate and target measures?	Because **specific levels of service were not presented**, the attributes are relevant at an overall level, though current customers may potentially perceive priorities differently in relation to a changed perception of current service experience.	
		If no to either of these, what issues do the differences give rise to?	These results must be used in conjunction with the DCE values in the triangulation approach developed by Accent/PJM, so these in turn are ultimately **subject to any scope effects**.	
	Statistical	How large is the sample?	Statistically robust samples of HH customers were obtained. (n=1073 in 2023 – 746 in SSW and 327 in CAM)	**Amber**
		How representative are the sample / timings?	Online self-completion surveys with appropriate quotas set. Data is very recent	
		How wide are the confidence intervals?	The results have to be **combined with the core DCE values; the resulting intervals become larger.**	
		Have the results been derived using best practice techniques?	Priority scores derived from an established modelling approach (ordered logit)	
	Depth	Extent of explorative and developmental work?	Regular tracking research using short statements to describe broad service areas / issues. Two waves of qualitative research conducted in 2020 and 2022 ensured that SSC is tracking the right priority areas using attribute descriptions that are customer friendly.	**Amber**
		Quality and detail of information given to survey participants?	**Although the main descriptions were limited to simple summaries,** 'pop up' text for each attribute contained comparative and other details to help customers make their choice descriptions.	

Table 2.7e: Contacts (2022)

Description	Validity	Criteria	Comment	RAG Rating
SSC records of customer contacts	Theoretical	Are definitions of candidate and target measure the same?	The information, based on contacts received from customers by SSC, cover four areas of interest to PR24. **They relate to any contacts classified as being related to these attributes.**	**Amber**
		Are contextual conditions the same between candidate and target measures?	Because **a variety of issues are covered within each contact type**, the attributes are relevant at an overall level	
		If no to either of these, what issues do the differences give rise to?	These results must be used in conjunction with the DCE values in the triangulation approach developed by Accent/PJM, so these in turn are ultimately **subject to any scope effects**.	
	Statistical	How large is the sample?	Collected as a mix of unsolicited contacts from customers and contacts that are responses to an initial prompt from SSC (eg at text requesting a meter read). For the four areas relevant to the study (Water not safe to drink, Discoloured water/ Taste and smell of water, Unexpected temporary loss of water supply) the number of contacts in the 12 months period to Sep. 2022 were 108, 1428, 7176 respectively across both SSC supply areas.	Amber / Red
		How representative are the sample / timings?	Data is very recent or recent.	
		How wide are the confidence intervals?	The results have to be **combined with the core DCE values; the resulting intervals become larger.**	
		Have the results been derived using best practice techniques?	Counts of number of contacts for each service area as a percentage of all properties known to be affected by these issues.	
	Depth	Extent of explorative and developmental work?	N/a	**Red**

| | | Quality and detail of information given to survey participants? | The reason for the contact is a classification by SSC staff receiving the calls and covers a range of diverse issues within that classification. | |

Table 2.7f: Satisfaction (2022)

Description	Validity	Criteria	Comment	RAG Rating
Qualtrics point of contact tracking survey. Short-snap satisfaction surveys sent to customer following a contact with the company	Theoretical	Are definitions of candidate and target measure the same?	The customer satisfaction surveys covered four attributes of interest to PR24. They are **only presented in terms of the correlation of service area satisfaction ratings** with overall satisfaction and are not defined in terms of detailed service levels.	**Amber**
		Are contextual conditions the same between candidate and target measures?	Because **specific levels of service are not specified**, the attributes are only relevant at an overall level.	
		If no to either of these, what issues do the differences give rise to?	These results must be used in conjunction with the DCE values in the triangulation approach developed by Accent/PJM, so these in turn are ultimately **subject to any scope effects**.	
	Statistical	How large is the sample?	The information is gathered in relation to customers who feedback regarding a specific event that SCC have responded to or a contact a customer has made about their water services. For some of the attributes relevant here (Discoloured water/Taste and smell of water and low water pressure), the **sample sizes are low** (n= 14 and 73 respectively).	Amber / Red
		How representative are the sample / timings?	Only of customers who have experienced an issue and/or raised a query and agreed to rate their experience of the service. Data is recent.	
		How wide are the confidence intervals?	The results have to be **combined with the core DCE values; the resulting intervals become larger.**	
		Have the results been derived using best practice techniques?	Regression values representing the derived relative impact on overall satisfaction.	
	Depth	Extent of explorative and developmental work?	Regular tracking research using short statements to describe broad service areas / issues. Although the questions were developed using best practice from customer experience communities and tested and refined with customers over time, they are not designed to give the detail required for WTP	**Red**
		Quality and detail of information given to survey participants?	**No attribute descriptions** – just single statements of topics to be rated (e.g., 'satisfaction', 'effort', First Contact Resolution (FCR), agent performance' etc).	

Table 2.7g: Strategic Research Options (SRO) (2022)

Description	Validity	Criteria	Comment	RAG Rating
Commissioned by a club of water companies to obtain primary evidence on customer preferences for 'added value' elements to inform the development of strategic resource options (SROs).	Theoretical	Are definitions of candidate and target measure the same?	The research covered one area of interest to PR24 'Specialist habitats created for wildlife' among a range of topics relating to longer term social, economic, and environmental issues. The attribute wase **only presented in terms of general changes the current experience** (e.g., 'Moderate positive impact', 'Major positive impact') and not in terms of specific service levels.	Green/Amber
		Are contextual conditions the same between candidate and target measures?	Because **specific levels of service were not specified**, the attributes are relevant at an overall level	
		If no to either of these, what issues do the differences give rise to?	The direct WTP result derived from this work was adjusted to reflect assumed number of hectares that would be affected.	
	Statistical	How large is the sample?	Statistically robust samples were obtained (5,902 HH and 553 NHH), though this covered several regions, of which CAM customers was represented by only 5% of the sample.	Green/Amber
		How representative are the sample / timings?	Recruited to representative quotas. Data is very recent.	
		How wide are the confidence intervals?	Values derived from a pair-wise Stated Preference and CVM approach using established design and analysis procedures.	
		Have the results been derived using best practice techniques?	The SP approach encouraged trade-offs but did not specify service levels, similar in detail to the Max Diff data.	
	Depth	Extent of explorative and developmental work?	Careful development work and piloting was carried out to test the materials with customers.	Amber / Red

Description	Validity	Criteria	Comment	RAG Rating
		Quality and detail of information given to survey participants?	Service area contained detailed descriptions.	

Table 2.7h: WRMP MCDA (2022)

Description	Validity	Criteria	Comment	RAG Rating
Study drawing on initial qualitative research, explored through stated preference choice exercises conducted with a representative sample of SSW and CAM customers	Theoretical	Are definitions of candidate and target measure the same?	The research covered one area of interest to PR24 'Habitats for native wildlife and plants' among a range of topics relating to longer term social, economic, and environmental issues. They were **only presented in terms of movements from the current experience** and were not defined in terms of detailed service levels.	Amber
		Are contextual conditions the same between candidate and target measures?	Because **specific levels of service were not specified** (impact levels were defined as negative/positive, moderate major) the attributes are relevant at an overall level.	
		If no to either of these, what issues do the differences give rise to?	The direct WTP result derived from this work was adjusted to reflect assumed number of hectares that would be affected.	
	Statistical	How large is the sample?	Statistically robust samples were obtained (1,015 HH), covering both the SSW and CAM areas.	Amber
		How representative are the sample / timings?	Recruited to representative quotas. Data is very recent.	
		How wide are the confidence intervals?	Values derived from a pair-wise Stated Preference approach using established design and analysis procedures.	
		Have the results been derived using best practice techniques?	The SP approach encouraged trade-offs but did not specify service levels. In this respect the data is similar in detail to the Max Diff data.	
	Depth	Extent of explorative and developmental work?	Piloting was carried out ahead of fieldwork launch.	Amber / Red
		Quality and detail of information given to survey participants?	Descriptions limited to short summary descriptions.	

Table 2.7i: WRMP online (2018)

Description	Validity	Criteria	Comment	RAG Rating
Part of a multi-stage project covering priorities / preferences around WRMP and also level of support for various demand and supply side options	Theoretical	Are definitions of candidate and target measure the same?	Covered the relevant service areas of 'Reducing leakage' and 'Water mering / Installing smart meters'. They were **only presented in terms of rank order of priorities** and were not defined in terms of detailed service levels.	Amber
		Are contextual conditions the same between candidate and target measures?	Because **specific levels of service were not specified**, the attributes are relevant at an overall level, though current customers may potentially perceive priorities differently in relation to a changed perception of current service experience.	
		If no to either of these, what issues do the differences give rise to?	These results must be used in conjunction with the DCE values in the triangulation approach developed by Accent/PJM, so these in turn are ultimately **subject to any scope effects.**	
	Statistical	How large is the sample?	Statistically robust samples were obtained (512).	Amber / Red
		How representative are the sample / timings?	Online surveys using representative quotas	
		How wide are the confidence intervals?	Confidence intervals around ±10% of the reported values	
		Have the results been derived using best practice techniques?	Frequency counts of preferences for different investment priorities.	
	Depth	Extent of explorative and developmental work?	Learnings from the workshops (see below) informed the design of the survey.	Amber / Red
		Quality and detail of information given to survey participants?	**Specific levels of service were not defined.**	

Table 2.7j: WRMP Workshops (2017)

Description	Validity	Criteria	Comment	RAG Rating
Part of a multi-stage project covering priorities / preferences around WRMP and also level of support for various demand and supply side options	Theoretical	Are definitions of candidate and target measure the same?	Covered the relevant service areas of 'Reducing leakage' and 'Water mering / Installing smart meters'	**Amber**
		Are contextual conditions the same between candidate and target measures?	The qualitative format encouraged participants to focus on the things that mattered most to them, but this also suggests greater variation in how the different service areas would be considered by customers.	
		If no to either of these, what issues do the differences give rise to?	These results must be used in conjunction with the DCE values in the triangulation approach developed by Accent, so these in turn are ultimately **subject to any scope effects**.	
	Statistical	How large is the sample?	Small: a series of workshops with a total of 62 people divided across the two regions of SSE and CAM.	Amber / Red
		How representative are the sample / timings?	Participants were recruited from a range of demographics and other key characteristics	
		How wide are the confidence intervals?	n/a	
		Have the results been derived using best practice techniques?	Frequency counts of preferences for different investment priorities	
	Depth	Extent of explorative and developmental work?	Issues were described in detail and discussed at length	**Amber**
		Quality and detail of information given to survey participants?	**Specific levels of service were not defined**	

Table 2.7k: External WTP19 (2017)

Description	Validity	Criteria	Comment	RAG Rating
A comparative anonymised review of stated preference (SP) willingness to pay (WTP) results for 13 water companies from England and Wales	Theoretical	Are definitions of candidate and target measure the same?	Covered almost all the areas of interest to PR24 except customer service and water hardness. As this was an analysis of a wide range of research programs conducted for PR19 across 13 Water and sewerage companies, the definitions of the service areas and their levels of service varied widely.	**Amber**
		Are contextual conditions the same between candidate and target measures?	The wide variation in definitions and regional contexts suggests that only 'top level' comparisons (i.e., comparative WTP/unit) are meaningful.	
		If no to either of these, what issues do the differences give rise to?	Potentially large differences.	
	Statistical	How large is the sample?	Statistically robust samples were covered across the studies, each of which aimed to meet Ofwat guidelines as to what was required for statistical robustness.	Green/Amber
		How representative are the sample / timings?	Quotas and the subsequent weighting of data to Census profiles was common. However, **the data itself is now over 5 years old, since when significant events have occurred** (Covid-19, Brexit, the cost-of-living crisis and increased concern over climate change).	
		How wide are the confidence intervals?	Confidence intervals of up to ±20% of the mean values are fairly common across the attributes tested.	
		Have the results been derived using best practice techniques?	The surveys used a range of methods, though a number followed the route of conventional DCE, tested with a series of large scale pilot surveys and independent peer review.	
	Depth	Extent of explorative and developmental work?	The development of many of contributing the surveys involved customers extensively in the design of the survey and service measures.	**Red**
		Quality and detail of information given to survey participants?	Qualitative work was frequently carried out with the aim of producing informative and digestible introductions to the investment areas that would be presented to respondents in the WTP surveys. A key limitation is that each survey was designed specifically for the needs of the commissioning Water Companies, so **the scope for accurate benefit transfer is limited.**	

Table 2.7l: ODI Rates (Accent / PJM Research, Dec 2022)

Description	Validity	Criteria	Comment	RAG Rating SSC	RAG Rating National
A national study on behalf of CCW and Ofwat, designed to establish a single consistent approach to obtaining WTA values attached to single service events	Theoretical	Are definitions of candidate and target measure the same?	**Only six relevant attributes covered.** Advice from PJM was to estimate pivoted values for the remaining attributes, using information from other research (eg the NERA PR24 study). Ratios of the values of these other attributes against an attribute common to the ODI study (eg 'water not safe to drink').	**Amber**	**Amber**
		Are contextual conditions the same between candidate and target measures?	The ODI approach is **presented in terms of WTA events** experienced by the individual customers, so that context is quite different from the attributes as covered in the NERA study. Also, the **attributes are defined at a general level**, not specific to degrees of impact[5] or presented in the context of the SSC region.		
		If no to either of these, what issues do the differences give rise to?	The magnitude of some of the monetary values appear relatively small when considered over a region-wide impact. For example, the impact of a 'do not drink' notice is represented by a single customer only experiencing that issue.		
	Statistical	How large is the sample?	**Statistically robust samples** were covered for households (n= 12,567) and non-households (n=3,669). Samples for SSC were n=601 and n=200 respectively, so that by SSW/CAM these numbers begin to be relatively small.	**Green**	Amber / Red

---

[5] Correspondence with NERA (23/03/23): 'The impact rankings are not truly cardinal – individuals are not asked how much more of an impact one scenario will have than another. If even a subset of customers are extremely concerned about a "do not drink notice", e.g. because of the health risks, Ofwat's study will not capture this degree of concern at the individual level and so may underestimate WTP.'

		How representative are the sample / timings?	Profile matched to Census profile for the SSC region and nationally.		
		How wide are the confidence intervals?	Confidence intervals around the usable WTA values lie in the region of ±20-40% at the regional level.		19
		Have the results been derived using best practice techniques?	The study used **a new trade-off method** designed to address concerns raised by Ofwat and others in relation to SP methods used in PR19: • Positively, the approach appears to be an improvement over conventional SP methods, insofar as it is **easier for customers to comprehend** and complete when considering the impact of specific events on them personally. This seems likely to result in better quality / more plausible survey responses. • Less positively, the approach is **inconsistent with all other studies** insofar as it measures Willingness to Accept (WTA), the values from which are usually much higher than WTP. Also, they apply to personal experience rather than customers views on what will benefit the 'general good' in their region. This is not in itself a statistical weakness, more of a theoretical issue (hence the amber/red rating for theoretical validity).		
	Depth	Extent of explorative and developmental work?	The development of the surveys involved customers extensively in the design of the survey and service measures.	**Green**	**Green**
		Quality and detail of information given to survey participants?	Qualitative work was carried out with the aim of producing informative and digestible introductions to the investment areas that would be presented to respondents in the WTA surveys. As a result, comprehension among respondents was considered high.		

Table 2.7m: Willingness to pay for water services at PR24 (NERA, Nov 2022)

Description	Validity	Criteria	Comment	RAG Rating SSC	RAG Rating WW/YW
A bespoke piece of research commissioned by SSC to address the specific requirements of PR24 in their region. Similar studies were commissioned by Wessex Water (WW) and Yorkshire Water (YW)	Theoretical	Are definitions of candidate and target measure the same?	Specifically **designed to cover the 12 principal areas of interest** to PR24 for SSC.	**Green**	**Amber**
		Are contextual conditions the same between candidate and target measures?	WW/YW attributes couched in terms appropriate to those regions, and therefore of less direct relevance to SSC.		
		If no to either of these, what issues do the differences give rise to?	**Values represent an average per household**, so that when some of these values are multiplied through by the total number of customers, the resulting aggregate values can appear very large when compared with other studies.		
	Statistical	How large is the sample?	**Statistically robust samples** were covered for households (n=1,690) and non-households (n=247). Samples become small for NHH when split by region.	**Amber**	Amber / Red
		How representative are the sample / timings?	Profile matched to Census profile for the SST and CAM regions.		
		How wide are the confidence intervals?	Confidence intervals around the usable WTP values lie in the region of ±25-50%.		
		Have the results been derived using best practice techniques?	The study used **a new trade-off method** designed to address concerns raised by Ofwat and others in relation to SP methods used in PR19: • Positively, the approach appears to be an improvement over conventional SP methods, insofar as it is **easier for customers to comprehend** and complete when considering water investments. This seems likely to result in better quality / more plausible survey responses. • Less positively, the approach lacks the body of published material associated with other established SP methods. The agency responded to detailed criticism from the peer reviewer and the analytical approach was developed extensively. The main conclusion from the peer reviewer was the results needed to be carefully triangulated (as is the case in this report) [6]. Some of the improvements are relatively small (e.g., moving from 2 properties to 1 property) and yet the value given is per customer. Where are value is more than £0, this results in some **very large values for some quite small service improvements**.		
	Depth	Extent of explorative and developmental work?	The development of the surveys involved customers extensively in the design of the survey and service measures. In the modelling, certain respondents were identified as having 'protest attitudes', concerns over ability-to-pay and vulnerability. These were all used to test the sensitivity of the results.	**Green**	**Green**
		Quality and detail of information given to survey participants?	Extensive qualitative work, including cognitive testing, was carried out with the aim of producing informative and digestible introductions to the investment areas that would be presented to respondents in the WTP surveys. As a result, comprehension among respondents was considered high.		

---

[6] 'A triangulation of findings is surely a necessary step before using the results to inform developments of the companies' business plan. Given the innovative design and estimation strategy and the lack of external validity, I express a circumspect opinion on the overall validity of results. This is not a criticism of what is done here but a generic problem introduced by Ofwat to welcome innovation but without specifying how to conduct a sensible validation.' Dr. Silvia Ferrini , 15 December 2022, Review of the report "Willingness to Pay for Water Services at PR24. Prepared for South Staffs and Cambridge Water"

TRIANGULATE – application of the RAG Ratings

For each source, we took the ratings given for each aspect of validity and created an overall single value to represent them. This was the approach developed for the PR19 triangulation work. The rule we applied was to take the highest rating for each type of validity to represent each source. The hierarchy of importance was: Theory, Statistical and Depth validity. However, where there were two cases that were highest, this was taken as the highest value; for example, if Statistical and Depth validity both scored green while Theory scored amber, the high score would be green. The full application of this approach to triangulation is covered in the next chapter.

SENSITIVITY

To allow sensitivity testing, we then identified the lowest rating (from any of the three types of validity). These final overall classifications for 'highest' and 'lowest' ratings are shown in Table 2.8.

Table 2.8: RAG Ratings (high and low definitions)

Source		Overall RAG Rating (highest)	Overall RAG Rating (lowest)
WTP core_DCE		Green	Green / Amber
WTPCore_DCE2		Green	Green / Amber
WTP core_Maxdiff		Amber	Amber
PC Slider		Amber	Amber / Red
Priorities		Amber	Amber
Contacts		Amber	Red
Satisfaction		Amber	Red
Strategic Research Options (SRO)		Green / Amber	Amber / Red
WRMP MCDA		Amber	Amber / Red
WRMP online		Amber	Amber / Red
WRMP workshops		Amber	Amber / Red
External WTP19		Amber	Red
NERA WTP PR24	SSC	Green	Amber
	WW & YW	Amber	Amber / Red
ODI rates PR24	SSC	Green	Amber
	National	Amber	Amber / Red

Weights were then associated with the RAG ratings to represent the relative strength of overall validity. We used the same weights that were used in the PR19 research to estimate the core values, plus two alternative weighing approaches as sensitivity tests. These are shown in Table 2.9:

Table 2.9: RAG value weightings

Overall RAG rating	Weight used in PR19	Alternative (sensitivity tests)	
		Test 1	Test 2
Green	100%	100%	100%
Green / Amber	50%	75%	50%
Amber	25%	50%	0%
Amber / Red	10%	25%	0%
Red	0%	0%	0%

These weights attached to the RAG ratings were used to derive a mean average value for each service attribute across the different sources.

3. TRIANGULATION

Weighted WTP (per unit) Calculations

In the same manner as for PR19, the final WTP triangulation results were derived from a process of applying weights to each of the data sources based on their overall RAG ratings and then combining these measures to derive central values and ranges for the core WTP and customer preference service measures to be used subsequently in investment modelling. The specific approach is to calculate a weighted average WTP across the available values, each WTP value weighted by the figures shown in Table 2.8. An example of how these weights are applied is given in Table 3.1.

Table 3.1: RAG value weightings applied to 'Water not safe to drink' attribute

| Water not safe to drink / Risk of a temporary "do not drink" notice (per property affected) HH in SSW[7]						
Source		**WTP per Property affected**	**High RAG rating**	**Weight**	**Low RAG rating**	**Weight**
WTP core_DCE		£759	Green	100%	Amber	25%
WTP core_Maxdiff		£3,396	Amber	25%	Red	0%
WTPCore_DCE2		£253	Green	100%	Amber	25%
WTPCore_DCE2a		-	Red	0%	Amber	25%
WTPCore_DCE2_LowBill		-	Red	0%	Red	0%
WTP core_DCE_Private		-	Red	0%	Amber	25%
Priorities		£970	Amber	25%	Red	0%
Contacts		£12	Amber	25%	Red	0%
Satisfaction		-	Amber	25%	Red	0%
WTPPR14		-	Red	0%	Amber / Red	10%
WRMP online		-	Amber	25%	Amber / Red	10%
WRMP workshops		-	Amber	25%	Red	0%
ExternalWTP14		-	Red	0%	Red	0%
ExternalWTP19		£475	Amber	25%	Amber / Red	10%
PC Slider		-	Amber	25%	Amber / Red	10%
WRMP MCDA		-	Amber	25%	Amber / Red	10%
SRO		-	Green / Amber	50%	Amber / Red	10%
NERA WTP PR24	SSC	£398,655	Green	100%	Amber	25%
	WW/YW	-	Amber	25%	Amber / Red	10%
ODI rates	SSC	£211	Green	100%	Amber	25%
	National	£184	Amber	25%	Amber / Red	10%
WEIGHTED AVERAGE	Exc. WTP PR24			£946		£1,065
	Inc. WTP PR24			£76,407		£48,254
	Inc. WTP PR24 ODI rates only			£211		£203

---

[7] Although some sources have a rating above red, they will not always include a value relevant to the chosen attribute.

The example also illustrates an important issue related to the values obtained in the work by NERA, where there is a very large difference in magnitude both for this attribute and 'flooding from a burst pipe', when compared to other sources. Both areas are associated with very small changes in service improvement (a single property for 'water not safe to drink' and 5 properties for 'flooding from a burst pipe'), but the value is an average WTP per customer, whether they are personally affected or not. This contrasts with other studies, such as the ODI rates research, where the value only related to the customer experiencing the impact, or the PR19 DCE studies, where the number of properties affected were in '000s. This issue is discussed further when the results from the NERA and ODI are shown separately later in this chapter.

All Household customers

Figure 3.1a shows the central WTP values for all household customers (SSW and CAM combined, weighted by population), using the two sets of RAG ratings listed in Table 2.7. Units for each service area are shown below each one. Note that the last three items are not directly comparable to the others, because they are represented by area-specific units. See the next section ('Combined WTP calculations') where the application of these values to the total number of customers aims to allow for more direct comparison.

Figure 3.1a: Household Central WTP Values (Highest and Lowest overall RAG Ratings)
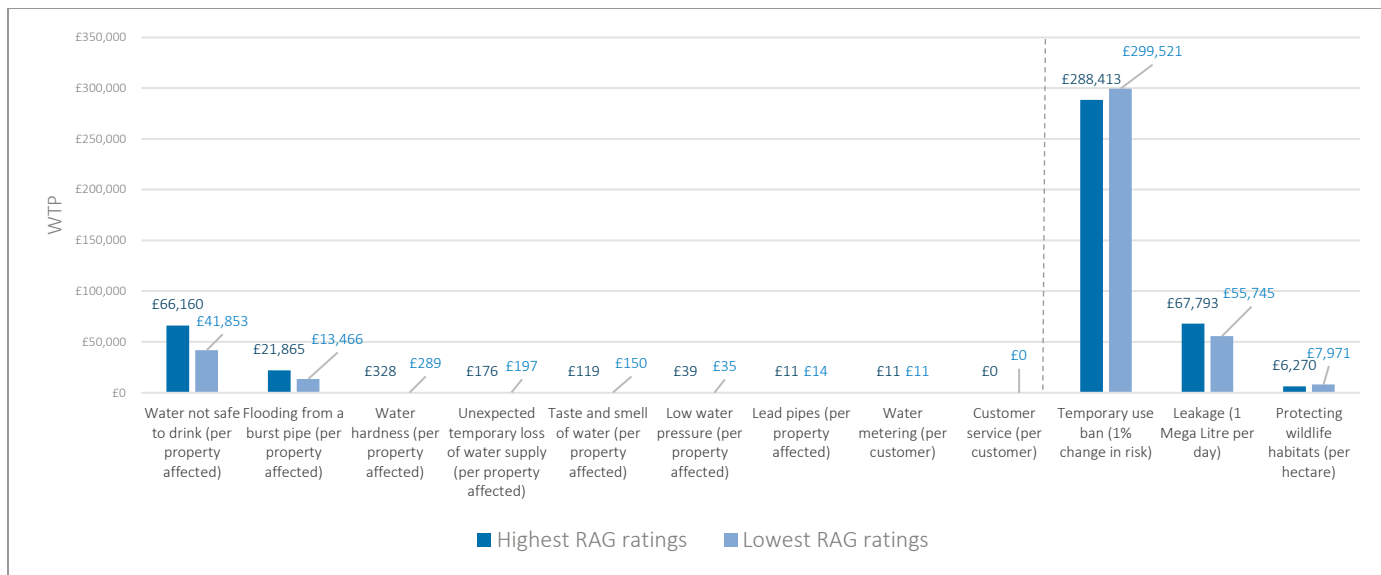
The effect of using the lowest overall RAG Ratings is to shift the values more towards the 'WTP core_DCE/DCE2' PR19 WTP values. The single attribute most impacted is for 'Water not safe to drink', where the difference in values is about one third. This suggests that it will be important to test a wide range of values for this attribute in the CopperLeaf calculations. Other attributes are affected less strongly, so that the order of priority remains fairly consistent.

Figure 3.1b shows the largest 'upper' and smallest 'lower' values across the two sets of RAG Ratings with a weight of more than zero, together with the central values from the previous Figure 3.1a. The high variability around 'Water not safe to drink' and 'Flooding from a burst pipe' is shown very clearly here driven by the large values derived from the NERA PR24 study .

Figure 3.1b: Household lower and upper WTP Values (Highest and Lowest overall RAG Ratings)
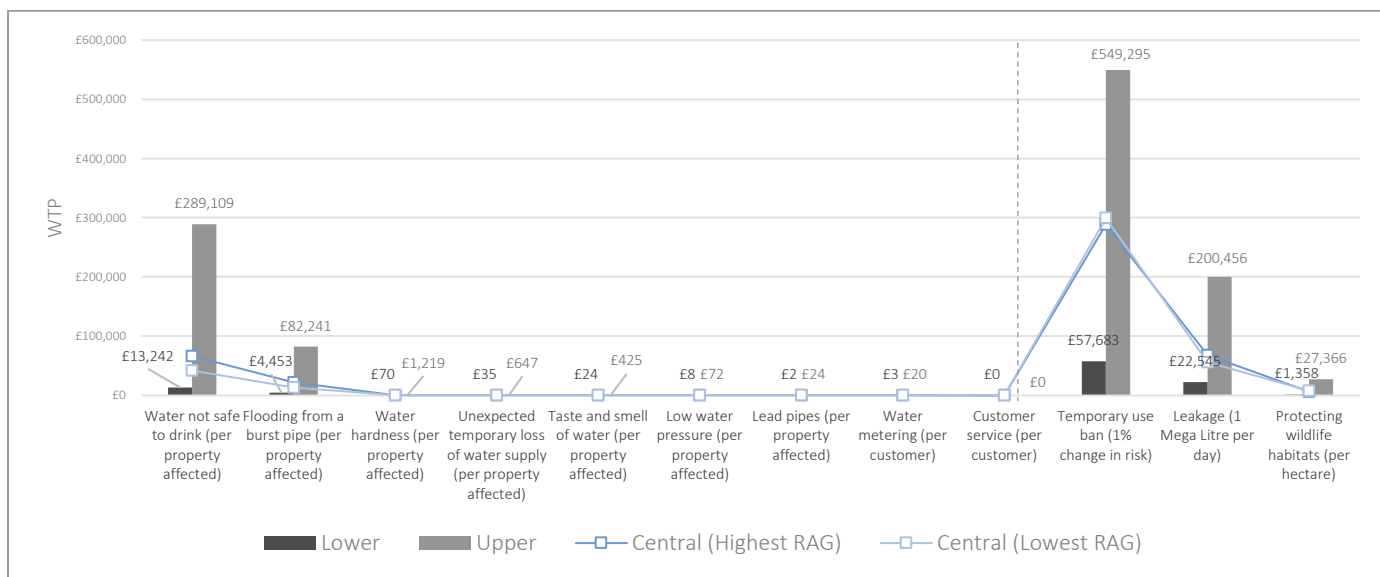
A further test of sensitivity to the values drawn from pre PR24 studies is to use different RAG weights as indicated in Table 2.9. Figure 3.1c compares central values derived from the highest RAG ratings by the three sets of weights reported in Table 2.9.

Figure 3.1c: Household Central WTP Values (Highest RAG Ratings) by alternative weights
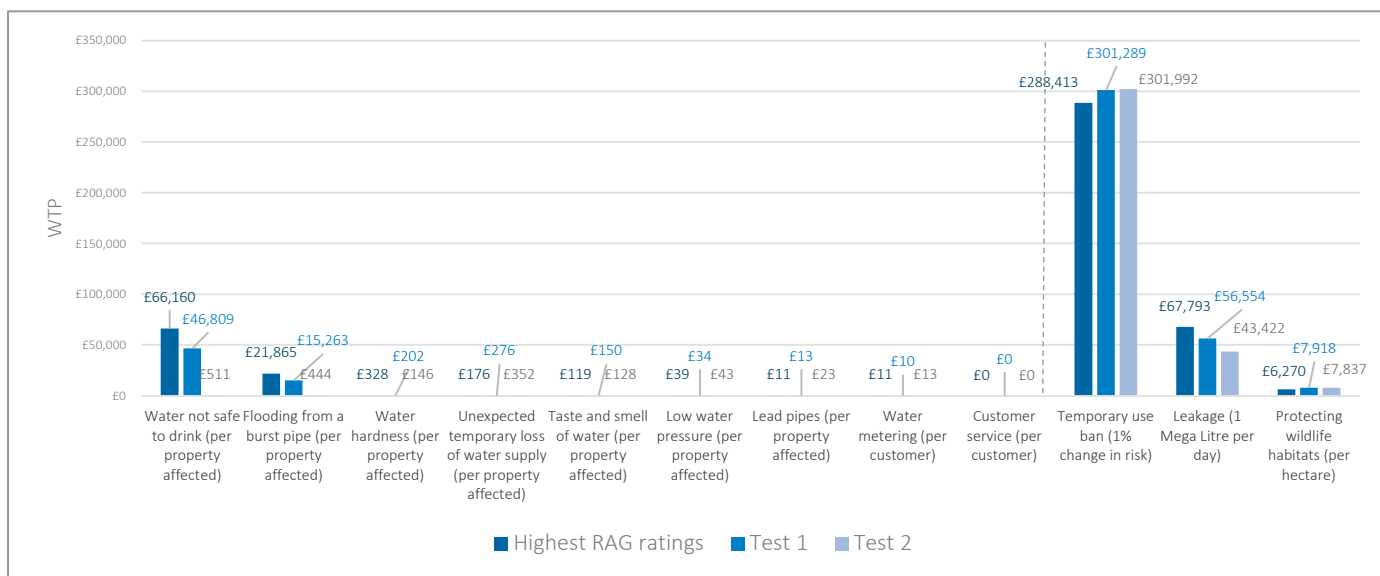
Test 1 placed greater weight on all lower measures (Green/amber through to Red/amber) when compared to the top measure (Green), while Test 2 only place weights on Green and Green/amber. The attributes most affected proportionally by the alternative weights are Water not safe to drink' and 'Flooding from a burst water pipe'. Where test 2 is applied, the influence of the NERA results is greatly reduced, so that these two attributes come closer in value to the other 'per property affected' attributes.

However, the influence of the weights is relatively small compared to the lower and upper WTP values from across all the sources, indicating that it is the way the sources are assessed, rather than the method for calculating the weighted averages, that will have the biggest influence on the final values. The implication for the Copperleaf calculations is that where resources allow, these sensitivity tests should be included to fully explore the variability of the inputs and their impact on the final investment returns.

All Non-Household customers

Figure 3.2a shows the central WTP values for non-household customers. The effect of using the lowest overall RAG Ratings is generally smaller when compared to what was observed for household customers, with the exception of temporary use bans which increases in value as the lower RAG weights shift the result towards the per-PR24 values.

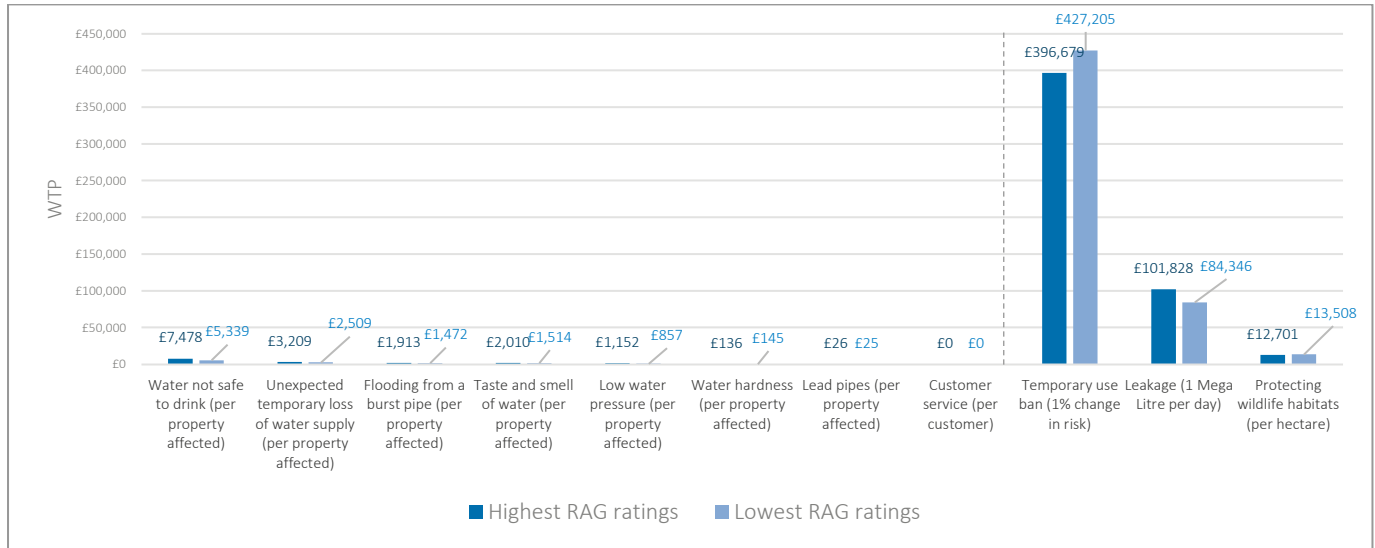*Figure 3.2a: Non-Household Central WTP Values (Highest and Lowest overall RAG Ratings)*

Figure 3.2b shows the largest 'upper' and smallest 'lower' values across the two sets of RAG Ratings, together with the central values. In these results, the range for 'Temporary Use Ban', is the largest, followed by 'Leakage'.

Figure 3.2b: Non-Household lower and upper WTP Values (Highest and Lowest overall RAG Ratings)
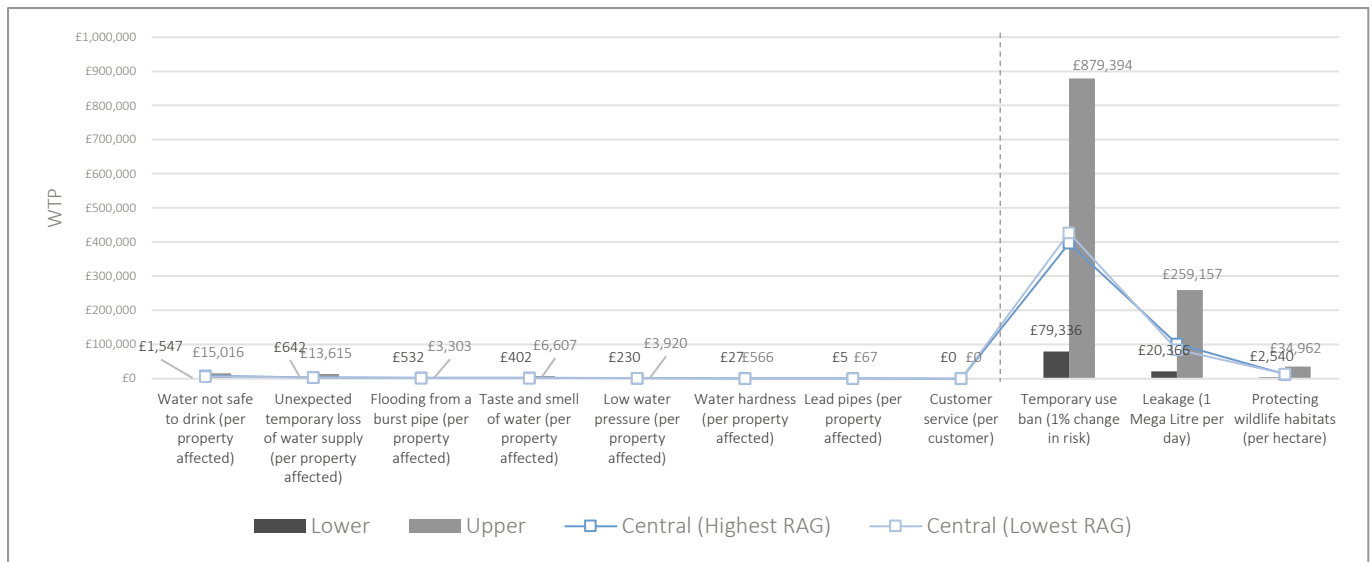
Finally, the sensitivity test related to different RAG weights as per Table 2.9 are shown in Figure 3.2c. The pattern is broadly similar to that observed for Households, with the feature that 'temporary use ban' is more variable. As indicated for households, there will be merit in including these values as additional sensitivity tests in Copperleaf.

Figure 3.2c: Non-Household Central WTP Values (Highest RAG Ratings) by alternative weights
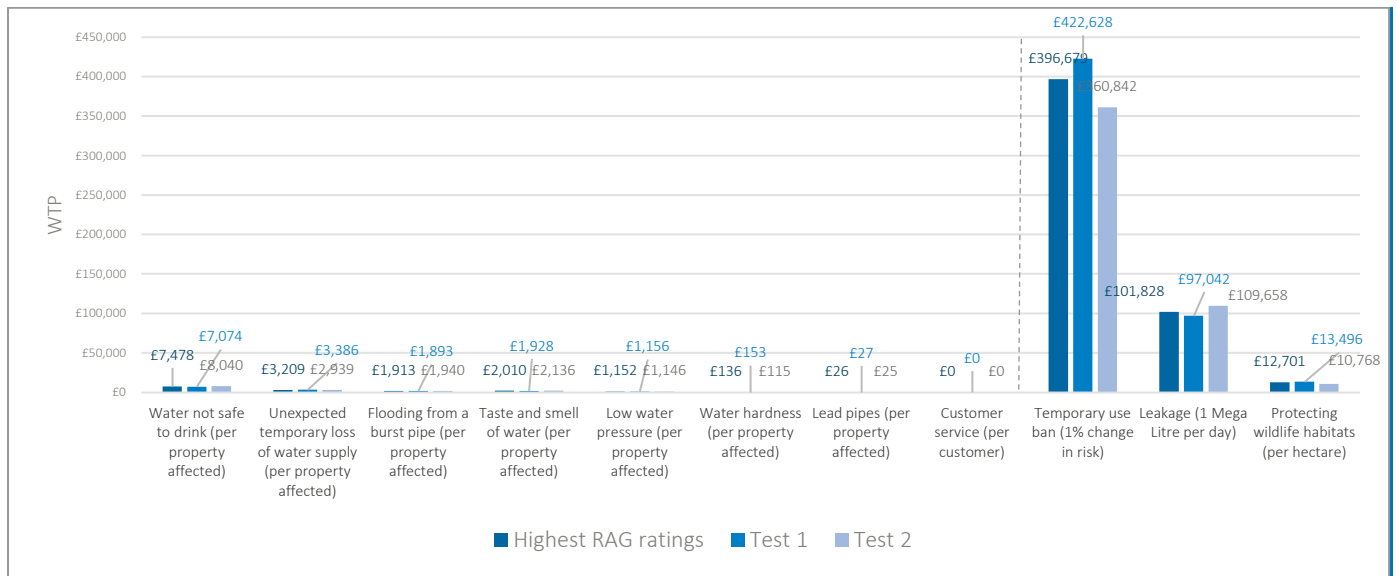
Application of the Triangulated WTP Calculations to Specific Improvements

The unit values reported above give an indication of the average WTP per customer, but to establish how these may potentially impact investment appraisals, it is necessary to calculate the collective value across all customers together.

To illustrate this, we can summarise the total value by applying the WTP unit values to the first level of improvements described in the latest WTP surveys for PR24. These are summarised for SSW in Table 3.2. Specific levels differed for some areas in the CAM region, but the list and definition of areas was the same. We have used this as a consistent basis for applying the overall value to customers of each improvement, using the triangulated figures reported above. For example, the improvement of 'Do not drink' from 2 to 1 property.

Table 3.2a: Example of Current and first improvement attribute levels tested in PR24 (NERA, 2022, WTP for PR24) – SSW Region[8]

	Current Level	Small Improvement	Current	Improvement	Units
Customer Service	1 in 3 customers (30%) waits more than 10 minutes	1 in 6 customers (20%) waits more than 10 minutes	30%	20%	% of customers
Risk of a temporary "do not drink" notice	2 properties per year receive "do not drink" notice	1 property per year receives "do not drink" notice	2	1	No. of properties
Installing 'smart' water meters	24% of properties have an operational 'smart' meter by 2030	42% of properties have an operational 'smart' meter by 2030	24%	42%	% of properties
Hard water supply	South Staffs Water does not invest in water softening	South Staffs Water contributes to the cost of installing water softening devices in 5,000 properties	0	5000	No. of properties
Lead pipes	2 in 8 properties will still have a lead supply pipe by 2030	2 in 9 properties will still have a lead supply pipe by 2030	25%	22%	% of properties
Water lost to leakage from pipes	20% of treated water lost to leakage	18% of treated water lost to leakage	20%	18%	% of properties
Issues with tap water colour, taste, or smell	1-in-26 properties per year experience issues with tap water	1-in-29 properties per year experience issues with tap water	3.8%	3.4%	% of properties
Chance of property flooding from a burst pipe	51 flooding incidents per year	46 flooding incidents per year	51	46	No. of properties
Low water pressure	2-in-26 properties experiences low pressure per year	2-in-29 properties experiences low pressure per year	7.7%	6.9%	% of properties
Supporting nature and wildlife	1280 acres (720 football pitches) protected and enhanced	2030 acres (1150 football pitches) protected and enhanced	518	822	No. of Hectares
Unplanned short interruptions to water supply	1 in 130 properties experience a short interruption per year	1 in 140 properties experience a short interruption per year	0.8%	0.7%	% of properties
Risk of temporary use ban, including hosepipes	Temporary use ban occurs once in 40 years	Temporary use ban occurs once in 45 years	2.5%	2.2%	Probability

---

[8] For the full list of levels in SSW and CAM, for HH and NHH, see NERA, 2022, Willingness to pay for water services at PR24, SSC, Tables 2.3 and 2.4

Table 3.2b: Example of Current and first improvement attribute levels tested in PR24 (NERA, 2022, WTP for PR24) – CAM Region

	Current Level	Small Improvement	Current	Improvement	Units
Customer Service	1 in 3 customers (30%) waits more than 10 minutes	1 in 6 customers (20%) waits more than 10 minutes	30%	20%	% of customers
Risk of a temporary "do not drink" notice	2 properties per year receive "do not drink" notice	1 property per year receives "do not drink" notice	2	1	No. of properties
Installing 'smart' water meters	66% of properties have an operational 'smart' meter by 2030	74% of properties have an operational 'smart' meter by 2030	66%	74%	% of properties
Hard water supply	Cambridge Water does not invest in water softening	Cambridge Water contributes to the cost of installing water softening devices in 2,600 properties	0	2600	No. of properties
Lead pipes	2 in 8 properties will still have a lead supply pipe by 2030	2 in 9 properties will still have a lead supply pipe by 2030	25%	22%	% of properties
Water lost to leakage from pipes	15% of treated water lost to leakage	13% of treated water lost to leakage	15%	13%	% of properties
Issues with tap water colour, taste, or smell	1-in-47 properties per year experience issues with tap water	1-in-52 properties per year experience issues with tap water	2.1%	1.9%	% of properties
Chance of property flooding from a burst pipe	12 flooding incidents per year	11 flooding incidents per year	12	11	No. of properties
Low water pressure	2-in-26 properties experiences low pressure per year	2-in-29 properties experiences low pressure per year	7.7%	6.9%	% of properties
Supporting nature and wildlife	60 acres (40 football pitches) protected and enhanced	200 acres (110 football pitches) protected and enhanced	24	81	No. of Hectares
Unplanned short interruptions to water supply	1 in 130 properties experience a short interruption per year	1 in 140 properties experience a short interruption per year	0.8%	0.7%	% of properties
Risk of temporary use ban, including hosepipes	Temporary use ban occurs once in 20 years	Temporary use ban occurs once in 25 years	5.0%	4.0%	Probability

The size of these movements from current to the first level of improvement will greatly affect the results that follow the application of the WTP values. For example, the 'risk of a temporary use ban' first level improvement applies to a very small number of properties (the goal is to move from 2 properties per year to 1 property per year

experiencing this), while 'supporting nature and wildlife' (increasing protected areas by some 300 hectares) in principle benefits all customers. Where attributes had more than one level of improvement in the NERA study, the impact was combined, to give a single average figure (£/unit) for each attribute. For the purpose of illustrating the application of these averaged WTP values, we have calculated values just to the first level movements, as shown in Table 3.2 are summarised in Figures 3.3 and 3.4.

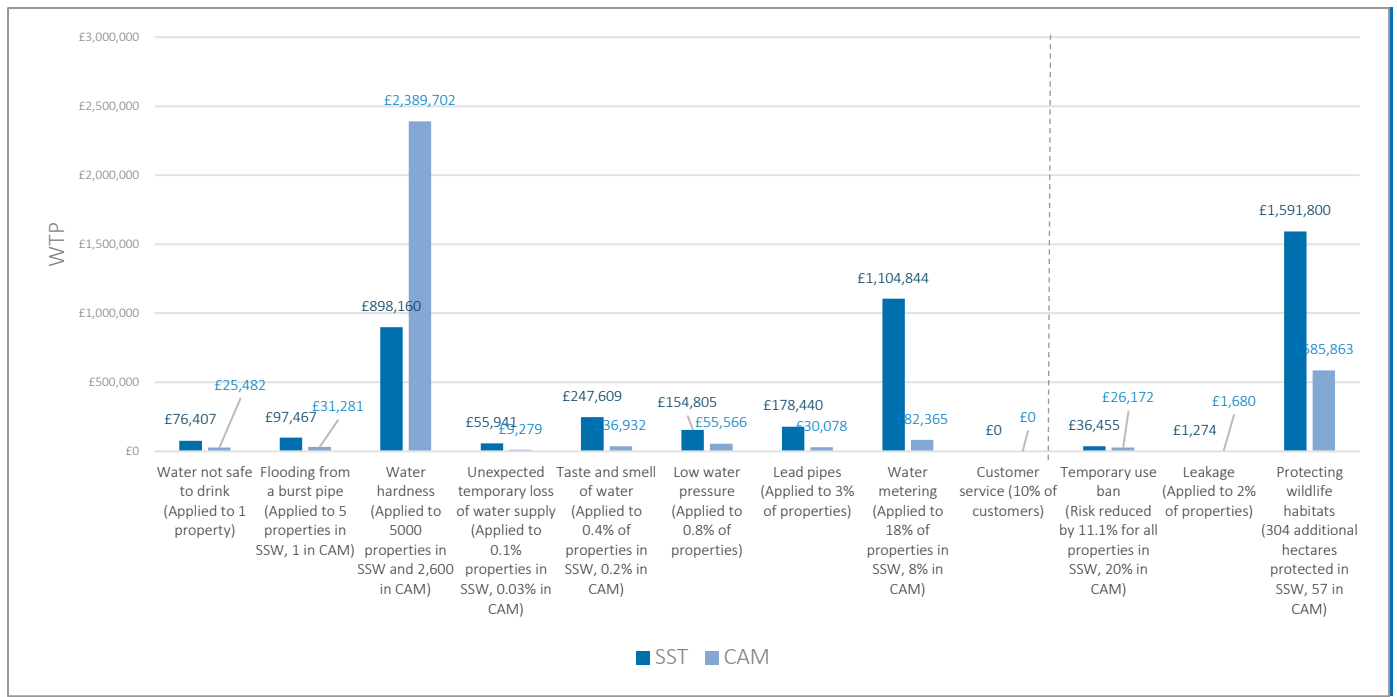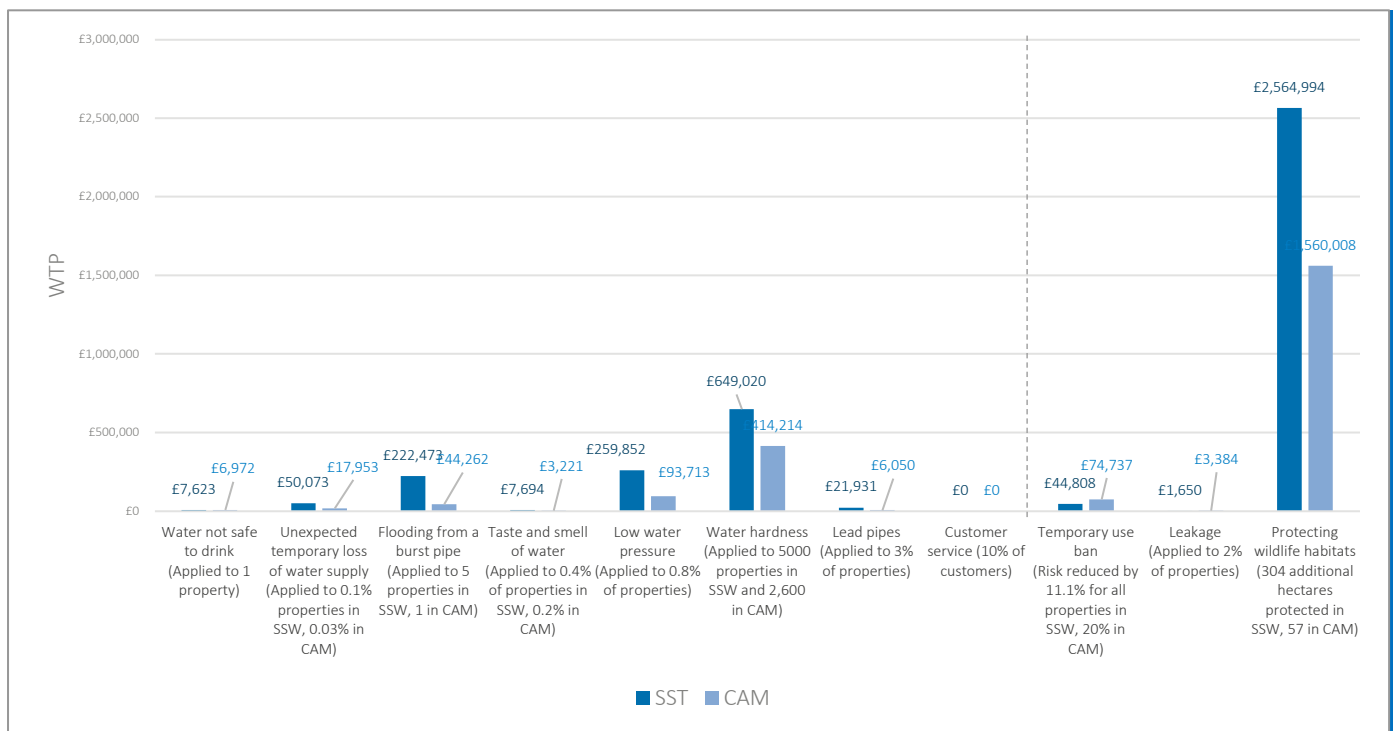*Figure 3.3: HH WTP Values (Highest overall RAG Ratings) applied to improvements across the regions*

Figure 3.4: NHH WTP Values (Highest overall RAG Ratings) applied to improvements across the regions

The application of the WTP figure in this way raises important issues as to what the WTP values ultimately represent in terms of benefits to customers. 'Use' values, where customers directly experience the impact of a service area on their household or property, can relate to a relatively small number of properties/customers (e.g., even water hardness affects a few thousands), whereas 'Non-use' values, where the benefits are in theory applicable to all customers but are directly experienced by only a few, relate to a very large number of properties/customers.

When compared to the average WTP figures presented earlier in Figures 3.1a and 3.2a, we see a greater focus on some items that have a relatively low average WTP figure (hardness, water metering) and what would appear to be a major over-statement of the value of environmental benefits. One way to address this latter concern might be to agree a basis for converting this to a 'use value' – for example to adjust down by the number of households that directly engage with natural habitats (e.g. regular go for walks in the country or visit nature reserves). However, this would be an external assumption, because in each of the studies, the areas were presented to respondents in the broadest terms, and not local in terms of local amenities. It has therefore not been applied here.

Comparison of PR24 and Pre-PR24 results

When the outputs from the all the studies are converted to represent the same impact per unit, the contrast in magnitude for some of the NERA results is striking. They are compared to the values from pre-PR24 research and the latest ODI results in Tables 3.5 and 3.6. There are no instances where the new values from the SSC PR24 study fall within the range of values derived from pre-PR24 research or ODI research.

Table 3.5: HH PR24 and PR19 SSC £/unit values compared (Highest RAG Ratings, total SSC region)

	Attribute	Unit	NERA SSC PR24	Accent/PJM ODI Ratings	Pre-PR24
A	Customer service	Per customer	£0	£0	£0
B	Risk of temporary "do not drink" notice	Per property affected	£344,846	£211	£952
C	Installing "smart" water meters	Per household	£0	£3	£10
D	Hard water supply	Per property affected	£802	£73	£256
E	Lead pipes	Per property affected	£0	£5	£21
F	Water lost to leakage from pipes	1 Mega Litre per day	£224,112	£12,207	£44,366
G	Issues with tap water colour, taste, or smell	Per property affected	£0	£118	£192
H	Chance of property flooding from a burst pipe	Per property affected	£97,331	£110	£451
I	Low water pressure	Per property affected	£0	£70	£41
J	Supporting nature and wildlife	Per hectare	£1,862	£2,673	£10,384
K	Unplanned short interruptions to water supply	Per property affected	£0	£172	£270
L	Risk of temporary use ban, including hosepipes	Per 1% change in risk	£0	£536,523	£268,955

Table 3.6: NHH PR24 and PR19 SSC £/unit values compared (Highest RAG Ratings, total SSC region)

	Attribute	Unit	NERA SSC PR24	Accent/PJM ODI Ratings	Pre-PR24
A	Customer service	Per customer	£0	£0	£0
B	Risk of temporary "do not drink" notice	Per property affected	£15,144	£14,669	£576
C	Installing "smart" water meters	Per customer	£0	£0	£0
D	Hard water supply	Per property affected	£0	£308	£128
E	Lead pipes	Per property affected	£0	£60	£21
F	Water lost to leakage from pipes	1 Mega Litre per day	£117,010	£190,511	£64,316
G	Issues with tap water colour, taste, or smell	Per property affected	£0	£7,756	£345
H	Chance of property flooding from a burst pipe	Per property affected	£2,364	£3,651	£643
I	Low water pressure	Per property affected	£0	£4,238	£33
J	Supporting nature and wildlife	Per hectare	£47	£31,830	£11,867
K	Unplanned short interruptions to water supply	Per property affected	£0	£10,709	£240
L	Risk of temporary use ban, including hosepipes	Per 1% change in risk	£0	£820,819	£388,092

We consider that this extreme variation across the three sets of values is likely to reflect the differences in survey design and statistical analysis between the latest study and more conventional SP approaches used in the past, rather than any fundamental shift in customers' priorities when compared to previous years, or to the country. The only exception to this may be 'risk of temporary use ban', as the PR24 work followed a summer in which these were threatened, and in some areas implemented. However, this would not explain the large value observed for the ODI ratings, compared to the zero rating from the NERA study.

To illustrate the potential impacts on investment modelling, the separate figures for NERA, ODI and Pre-PR24 were all applied to the specific examples of improvements listed previously in in Table 3.2. As an example, the calculations for Pre-PR24 WTP and PR24 ODI WTA were per property x number of properties affected; for SSC PR24: WTP per customer x number of customers). The results are summarised in Figures 3.5 and 3.6 for households and non-households respectively. These results demonstrate some particularly high values for some attributes for HH customers drawn from the SSC PR24 study, notably water hardness, but also flooding and water not safe to drink.

In the same way as for Figures 3.3 and 3.4, The £/unit are taken for each attribute and multiplied through by the size of the improvement. This is done at the regional level (SSW/CAM) and the results combined according to the population of each region. For example, for the risk of receiving a 'do not drink notice', the NERA study gave values of £0.74 *per customer* per property affected in the SSW region and £0.97 in the CAM region. The domestic customer numbers of the regions are 539,437 and 135,882 households respectively. Therefore, the total value of reducing the number of properties affected by a 'do not drink' notice from 2 to 1 *in each region* (i.e. two properties in total) is: £0.74 x 539,437 + £0.97 x 135,882 = £529,886[9]. The large value from the NERA study associated with properties receiving hard water treatment is entirely driven by the CAM region, where a value of £0.029 per property affected per customer was derived, and the number of properties affected was 2,600 (see Table 3.2). This a total value of (£0 x 5,000 properties x 539,437 households) + (£0.029 x 2,600 properties x 135,882 households) = £10,364,385.

---

[9] This final result based on £/HH at 15 decimal places, which are shown here at 2 decimal places

Figure 3.5: HH WTP Values (Highest overall RAG Ratings) applied to improvements across the entire SSC region
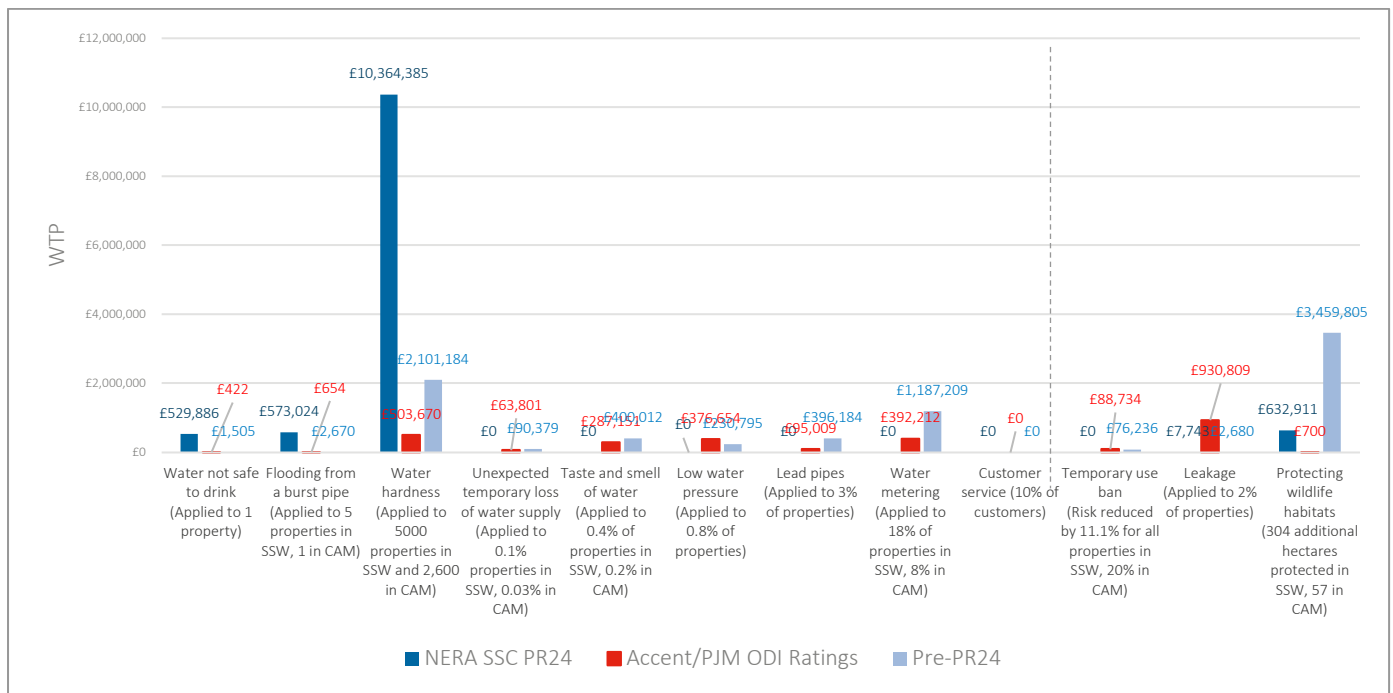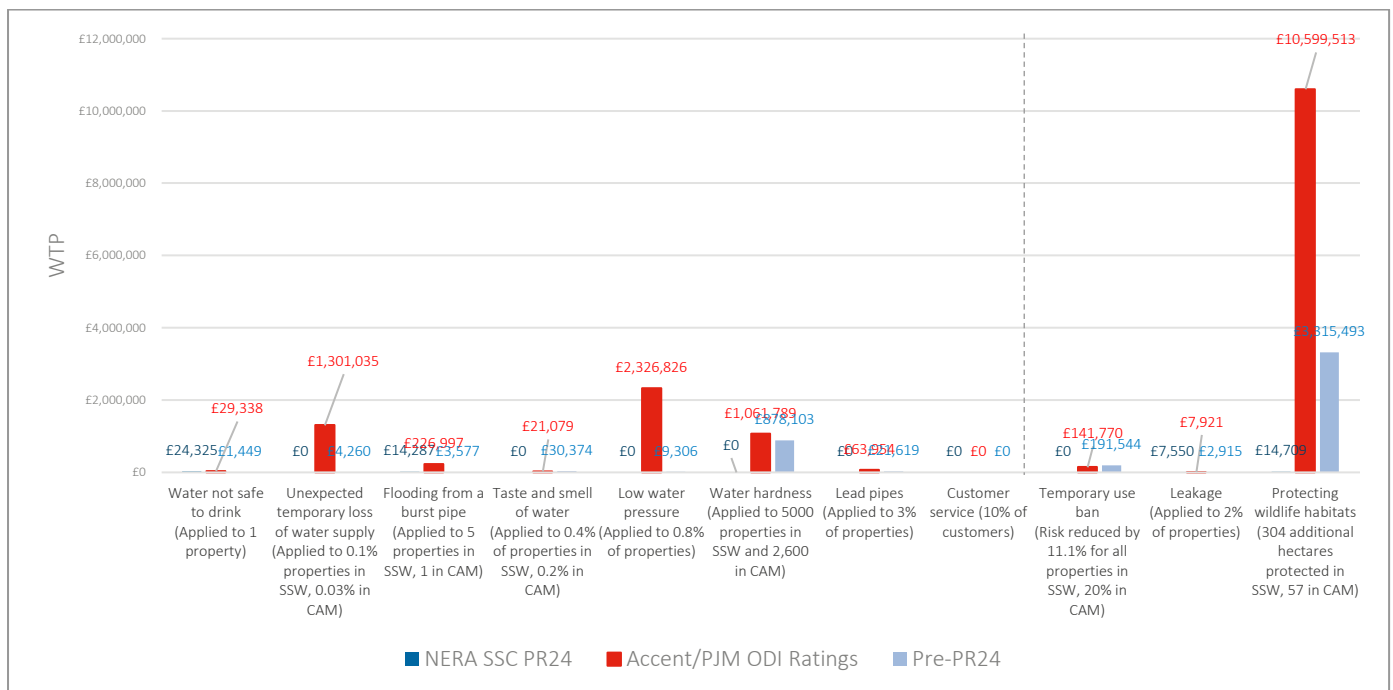

Figure 3.6: NHH WTP Values (Highest overall RAG Ratings) applied to improvements across the entire SSC region

4. DELPHI VALIDATION

External Validation

The Delphi Method

A key innovation compared to PR19 was the expansion of the number of people who will be involved in assessing the WTP evidence, using the Delphi method to encourage movement to a consensus view (or, in cases where consensus cannot be reached, a clear set of arguments for different outcomes which SSC could use to choose a final set of low/central/upper values) for use in Copperleaf. See '**SSC08 PR24 Technical triangulation - Phase 1 Methodology**' on the development of the Delphi approach.

The method was applied to two phases:

- Phase 1 (December 2022) – four participants were given an extensive summary of all available information on the 12 service areas to be covered SSC's PR24 WTP research. From this they were asked to identify what they considered to be the rank order of customer priorities. They were also introduced to the results from the PR19 triangulation in preparation for Phase 2.

- Phase 2 (February 2023) – the same four participants were given an information pack with feedback on their comments from Phase 2 and were asked to reconsider their rank ordering of customer priorities. They were also presented with summary reports of the WTP/WTA results for PR19 (triangulated), the NERA SSC PR24 study and the Accent/PJM ODI PR24 study. As well as giving their views on the credibility of these different information sources, they were asked to make one final reassessment of the rank ordering of customer priorities in the light of these results.

A full report on the results of the Delph approach is reported in '**SSC10 PR24 Technical Triangulation – Application of the Delphi Method**'. Below, we draw out the main findings that are pertinent to the triangulation results.

Customer Priority Rankings (Phase 1)

Table 4.1 shows the priority rankings for Household (HH) customers as determined by the Delphi panelists, based on their assessment of the summary material given to them and in advance of being presented with any WTP / WTA values. 'Water lost to leakage from pipes' receives a consistently high ordering, followed by 'issues with tap water colour, taste or smell'. 'Lead pipes' is more diverse, though highly ranked by two of the three panelists who were considering all HH customers in their assessment.

Table 4.2 shows the priority rankings for Non-Household (NHH) customers as determined by the Delphi panelists. 'Unplanned interruptions of water supply' is consistently highest in rank, with 'issues with tap water colour, taste or smell' then taking second place in a similar way as for HH customers. Beyond that, opinions are more diverse. It was observed by one participant that the needs of NHH customers vary widely according to the types of business and this could be behind the diversity of opinion.

Reactions to Triangulation PR19 values (Phase 1)

In phase 1, Delphi participants were challenged when they attempted to interpret the values, regardless of whether they were presented as 'per property affected/unit' or as 'total values (all customers x the total number of properties / units affected). The biggest concern was the diverse range of some of the values, both across and within the service areas. The main learning that were taken forward into Phase 2, where the results of the three different sources would be compared (PR19 triangulation, SSC PR24, ODI PR24), was to clarify that the results were for one year (v annual bill) and to separate out 'per property' attributes from 'unit change'. Also, it would be instructive to summarise the triangulated results in terms of the implied *rank order of priority* so that this could be compared with the 'expert' rankings from this first phase.

Table 4.1: Ranking of HH Customer Priorities by Delphi Participants

The priorities expressed by panelist 1 were based upon a vulnerable customer subset. The general pattern of priorities across the group are therefore determined by panelists 2 to 4.

	Delphi panelist 1 ranking	Delphi panelist 2 ranking	Delphi panelist 3 ranking	Delphi panelist 4 ranking
Water lost to leakage from pipes	2	2	1	3
Issues with tap water colour, taste or smell	8	1	4	1
Lead pipes	10	9	2	2
Unplanned interruptions to water supply	4	3	8	4
Customer Service	5	4	6	7
Supporting nature and wildlife	7	5	5	11
Risk of temporary 'do not drink' notice	6	12	3	9
Installing 'smart' water meters	1	6	9	12
Chance of property flooding from a burst pipe	3	11	7	10
Hard water supply	12	7	10	6
Low water pressure	11	10	11	5
Risk of temporary usage bans, including hosepipes	9	8	12	8

Table 4.2: Ranking of NHH Customer Priorities by Delphi Participants

	Delphi panelist 1 ranking	Delphi panelist 2 ranking	Delphi panelist 3 ranking	Delphi panelist 4 ranking
Unplanned interruptions to water supply	3	1	2	1
Risk of temporary usage bans, including hosepipes	7	2	5	8
Installing 'smart' water meters	1	9	12	2
Customer Service	2	7	9	3
Water lost to leakage from pipes	6	10	1	9
Low water pressure	8	6	3	5
Issues with tap water colour, taste or smell	11	3	4	4
Supporting nature and wildlife	4	5	8	11
Risk of temporary 'do not drink' notice	12	4	6	7
Chance of property flooding from a burst pipe	5	8	10	12
Hard water supply	9	12	11	6
Lead pipes	10	11	7	10

When shown this summary of rankings in phase 2, participants did not feel the need to alter their personal rankings. However, when asked to consider potential regional differences and industry comparisons, some changes in priority were identified. In Tables 4.3 and Tables 4.4 , we summarise the most relevant points relating to specific services attributes, for HH and NHH customers respectively.

Table 4.3: Key Observations on Service Attributes Tested for HH Customers

Attribute	Delphi Feedback (Summary)	Comment
Risk of temporary 'do not drink' notice	Water safety is an essential element of the service, particularly for HH customers. To guide customers, more could be said about the impact, for example whether one can still use the water to cook with. One panelist was surprised by the relatively high WTP value on risk of temporary 'do not drink' notice: "I would have thought customers would expect clean safe water as core service and therefore less WTP for it'."	*The panelist's comment highlights one of the ambiguities of the values derived for PR24: are they a measure of what customers say they will pay in terms of bill increases in order to reduce the risk (as in the NERA and PR19 DCE studies), or are they the amount they expect to be compensated for the loss of this essential service (WTA, as in the ODI study)?*
Water quality: 'Issues with tap water colour, taste or smell'	Customers expect clean, safe water from their water company. It is therefore an important attribute, but in practice rarely seems to be an issue for customers. Information and transparency are key to reassuring customers.	*When panelists understood that performance from SSC was generally high compared to the industry average, it was suggested that this could be given lower priority for improvement (ie sufficient to maintain current levels of service).*
Water quality: 'Hard water supply'	Not felt to be a major issue, though expected to be more relevant to CAM, where the water is harder. However, customers may take note of impacts beyond the effect on appliances, such as health-related concerns: 'Many people care about the impact on appliances and skin and hair but taste wise it's a personal preference'.	*These comments, together with the relatively low ranking indicate that this should be considered a low priority.*
Water quality: 'Lead pipes'	There was felt to be some potential confusion about this topic: who was responsible for which pipes and most importantly, the true nature of the heath risks, which cannot be understood simply in terms of the amount of piping replaced. "Lead pipes I put as unsure, as I don't know how much of an issue that is in the company's area, health risks or levels of awareness."	*This potential for confusion may show itself in the customer valuations, where the specific benefits of investment to individual customers may not be readily understood.*
Unplanned interruptions to water supply	"Water reliability is considered a key determinant of overall customer satisfaction and the expectation is that water must be reliable". However, customers also need to know that it is sometimes unavoidable, so communication is important.	*Attention by the national media may keep this high in customers' minds. There was some question as to whether customers understood fully the impact on vulnerable customers and whether the impact on themselves may be greater if part of a bigger outage of services.*
Chance of property flooding from a burst pipe	It was felt that customers may not fully comprehend the impact of flooding, for example the negative experience and the cost of drying it out.	*It was suggested that question framing could be influential here. For example, the 51 properties mentioned in the NERA approach could make it sound more of an issue than it really is (less than 0.01% of properties will be affected). On the other hand, the ODI approach appeared to understate the impact, suggesting 1 month's recovery when 3 months might be more realistic).*

Attribute	Delphi Feedback (Summary)	Comment
Low water pressure	Thought to be of relatively low importance since the expectation is that water is or must be reliable and relatively few people are impacted for any length of time. Customers seem prepared to occasionally experience low pressure as long as notice is clear and in advance. Communication is therefore key.	*This could be a longer-term issue – i.e. reflective of sufficient investment in infrastructure to ensure adequate pressure. Therefore, not likely to be an immediate priority for customers.*
Water lost to leakage from pipes	"Once informed [customers] are vocal in raising the importance to tackle this problem." "The urgency of this service is probably also triggered by the need to save/preserve water". There was a suggestion that local issues might influence regional variations, for example the aquifer in CAM.	*Primarily, leakage could be a prominent indicator of the efficiency of the service: "Leakage was put highest, as is prominent in the media, people hate the waste, it has benefit for a number of the other areas e.g. environment, supply interruptions, water pressure. This, and the generally high coverage of the issue from time to time in the media, may be evident in the fact that all studies derive some WTP value for this attribute.*
Supporting nature and wildlife	Generally seen to be an emotive issue, for which water companies have a poor public image, linked to reports of sewage discharges into rivers. "Less than half of customers think that water companies are doing a good job". "No mention of lived experience of nature and wildlife, e.g. dog walks or water sports, which are important for customers to understand the impact."	*The reference to 'lived experience' infers that the descriptions of environmental improvements in the WTP studies are rather generalised and touches on the issue of 'use' value. It is not indicated what the personal benefit would be to a customer. This was felt to potentially lower the importance attached to this attribute. There was also discussion about the extent this attribute is covered by other attributes, such as leakage and general water management, and to what extent the Cost of Living / COVID crises may have impacted on WTP, both negatively for the former and positively for the latter.*
Risk of temporary usage bans, including hosepipes	This was another area where public perceptions would reflect the information available to them, for example the reasons why a TUB is triggered. Experience would also be an influence: "TUBs [occurred] recently and for many not as bad as thought. Though there may be evidence suggesting this caused real problems for some."	*It was thought that customers may struggle to interpret probabilities (e.g. risk of 1 in 40) but there seemed an expectation that TUBS would become more frequent in the future, so it needed to be planned for. Mitigation to reduce the impact of TUBs, such as use of brown water, soakaways instead of drains, might change the way customers value this attribute.*
Installing smart water meters	Guidance for customers is important in this area – for example, the potential impact on bills and whether experience of energy smart meters influences opinion. The benefits of smart meters to customers are not always apparent. Not seen as a short-medium term priority. Also, potential issues of fairness (e.g. some vulnerable people whose consumption may be higher due to being located mostly at home).	*One panelist elaborated on the HH priorities ranking: "Considered putting 3rd but in end put 9th - lower than PR19 as half of energy smart meters don't work, confidence in smart meters has declined. Not seen as value for money. Though people do want the data and control and properly smart meters could help with managing water demand and resilience." This uncertainty about the benefits and effectiveness of smart meters may be behind this attribute being a low priority for customers*
Customer service	This was seen to be a rather broad term, as it could cover a lot of service aspects: self-serve, first call resolution, notice of interruptions, multi-channel contact, supply issue reporting and update, etc, 24/7 contact etc, and how SSC perform and costs/impacts of variation across the service. Regarding the PR24 research "the question is about call waiting times, not customer service more generally. Call waiting times are important, but not a proxy for wider service or consumer satisfaction." Customers in CAM were thought to be potentially more demanding than in SSW.	*"10 mins seems a long waiting time compared to other service providers, e.g. DNOs 98% of calls answered within 60 seconds." One panelist "considered putting customer service as a lower priority, especially for SSW, as its performance is relatively good. But kept it at 4, as most people won't have contacted SSC, but would want to know its good service when they need it" "People expect excellent customer service. Expectations are rising." As this attribute does not have any WTP value from the latest research, it may reflect the fact that only waiting times were represented in the attribute descriptions*

Attribute	Delphi Feedback (Summary)	Comment
Risk of temporary 'do not drink' notice	"Essential service for HH and NHH respondents"	*Despite this comment, this attribute was ranked relatively low for NHH and HH customers.*
Unplanned interruptions to water supply	"No surprise really that continued, uninterrupted supply is key. This is an enabler for most businesses, or a simple hygiene factor." Reactions to ODI NHH values: 'I had expected that an interruption to supply would over index given that water supply is critical for most NHHs - either being used for the business' core business, or as an enabler for staff or customers.	*This attribute was most consistently ranked as important and the expectation was that this would be reflected in higher WTP / WTA figures.*
Chance of property flooding from a burst pipe	One panelist asked "does SSC cover the loss of income and the income of employees if a business can't operate?"	*The WTP values may reflect assumptions by NHH customers about how the impact can be mitigated through insurance and/or compensation.*
Customer service	"I would put customer service higher up for water dependent businesses in particular, as when things go wrong they would want quick resolution of problems given impact on daily operations."	*This, together with comments for other attributes, reflects the diversity of NHH needs – while the overall value of some attributes may not be high, they will be particularly critical to some businesses.*

Delphi rankings v measured priorities.

In Tables 4.5 and 4.6 we have categorised the values from the NERA and WTP studies to be able to compare them with one another and with the broad Delphi rankings from Tables 4.1 and 4.2. The aim is to establish where the research results appear to broadly align with the thinking of Delphi panelists and where they differ.

The most prominent differences for households are:

- Low WTP value rankings: leakage (ODI), colour, taste and smell of water (NERA) and lead pipes (both studies), which were all anticipated to be high priority by Delphi panelists but received relatively low WTP values.
- High WTP rankings: Low pressure and TUBs (ODI), flooding (NERA) and hard water (both studies), which were considered lower priority but received higher values.

The most prominent differences for non-households are:

- Low WTP value rankings: TUBs (NERA) and smart meters (both studies), which were all anticipated to be high priority by Delphi panelists but received relatively low WTP values.
- High WTP rankings: Flooding (NERA), Nature & Wildlife (both studies) and risk of 'do not drink' notice (both studies), which were considered lower priority but received higher values.

Table 4.5: Comparison of Delphi and NERA/ODI HH Priorities

NERA and ODI priorities (Both coded in purple)						
		Very High	High	Middling	Low	Very Low
Delphi rankings	Very high		• Leakage		• Leakage	
	High	• Colour, taste or smell	• Unplanned interruptions	• Unplanned interruptions	• Colour, taste or smell	• Lead pipes
	Middling	• Nature & wildlife • 'Do not drink'		• Customer Service • Nature & wildlife	• 'Do not drink'	• Customer Service
	Low	• Flooding		• Flooding	• 'Smart' meters	
	Very Low	• Low pressure • TUBs	• Hard water			• Low pressure • TUBs

Table 4.6: Comparison of Delphi and NERA/ODI NHH Priorities

NERA and ODI priorities (Both coded in purple)						
		Very High	High	Middling	Low	Very Low
Delphi rankings	Very high	• Unplanned interruptions	• TUBs	• Unplanned interruptions		• TUBs
	High				• 'smart' meters	
	Middling	• Low pressure • Colour, taste or smell	• Leakage	• Low pressure	• Leakage • Colour, taste or smell	
	Low	• Nature & wildlife • 'Do not drink' • Flooding	• Nature & wildlife • 'Do not drink'	• Flooding		
	Very Low			• Hard water		• Hard water • Lead pipes

Delphi Assessment of the NERA / WTP approaches

NERA 2022 SSC WTP Study

All 4 panelists said they understood the rationale and objectives behind the NERA WTP study:

- "The approach is clear and allows the company to gauge what is prioritised (or not) by both HH and NHH customers."
- "Positive that attributes are assessed one at a time in addition to the package test."
- "The approach is simple enough and participants are offered symmetrical WTP options e.g., Increase or reducibly similar values."
- "I found the information was presented clearly and consistently and provided a range of scenarios that felt meaningful and easy to understand."

When asked about the robustness of the approach, there were mixed responses. One panelist did not find it to be an obvious alternative method (to previous SP approaches) and believed it is important to triangulate with wider research. Another panelist considered it "another set of data to test for sensitivity", again suggesting the value of triangulation on not relying on one source of values.

When considering the WTP values from this study, some panelists were surprised by a few of the differences between the regions, with some also showing scepticism for the NERA method:

- "Surprised about lower water lost to leakage figure for SSW compared to Cambridge."
- "Supporting wildlife and nature was not surprising in Cambridge given the local focus, but I would have imagined higher focus."
- "Not surprised by low value placed on supporting nature and wildlife – I think the way the question is asked means true value is not understood."
- "Higher value for 'risk of temporary do not drink notices' than I expected."
- "High value for 'chance of property flooding from a burst pipe' in CAM seems out of proportion to SSW.
- "[WTP values only] represent between 0.0004% to 0.001% changes in current water bill (0.0008% to 0.01% of average company bill) - but then when aggregated for CBA might not represent reality."
- "Polarised results, with customers focusing on adequacy and quality of supply."

In the final assessment, the panelists appreciated the merits of the exercise in terms of respondents' ability to assess the attributes and express their priorities, but for the outputs there was caution, with emphasis on the need to triangulate the results with other studies:

- "the importance of attributes is very different if compared to other studies and differs from experts' assessments. The most important attribute according to NERA was not marked as important by any of the experts. Would use these results with caution, especially considering their potential lower bounds compared to others."
- "WTP is an imperfect tool, need to triangulate between impact compensation exercise [from ODI study] and WTP. Some sample sizes, e.g., future customers, are small."
- "Each approach has inherent defects, but in combination should provide a fairly robust view."

Accent / PJM ODI Ratings 2022 WTA Study

All 4 panellists seemed to generally understand the method, with some questions remaining.

- "Straightforward, since it tests and probes perceptions around preferences between scenarios to identify which is the least or most preferable."
- "This is effectively asking how much would you need to be paid to put up with something."
- "Simple enough background and approach."

- "Detail was clear, respondent would be able to understand how the issue may impact them."

The majority of the panelists questioned the robustness of the approach[10] and the decision to measure compensation rather than WTP raised some issues:

- "Appreciates innovation to overcome the limits of choice modelling and benefits of ranking services, but not sure that the compensation is a neutral approach to value benefits."
- "Useful context and some sure foundations, but unclear whether all options will have been compared against each other."
- "Provides a single value of compensation against a service issue rather than determining the value of compensation customers would accept."
- "This is again a novel approach that needs testing".
- "Of the 3 [approaches: PR19, NERA and ODI), ODI seems the weakest."

When considering the outputs from the ODI research, some of the rank ordering was queried, for example "I thought there would be higher levels of compensation for hosepipe bans." However, as one panelist put it, the values seemed generally in line with expectations, despite concerns about the method:

"These values are WTA which is in principle an infinite economic measure and I wonder if these reflect benefits. But seems to reflect well the level of importance expressed by customers and reviewed in Delphi phase 1."

Implications for Copperleaf Input Values

The NERA approach was generally well regarded as an engaging and accessible exercise for respondents, but there were concerns around the outputs, where some attribute improvements appeared very highly valued and others undervalued (often zero value). There was less support for the ODI approach, particularly in relation to the measurement of WTA and the open-ended nature of compensation figures. However, when it came to outputs, there seemed less controversy and closer alignment with general expectations.

The marked variation in how the methods have been assessed by the panel, and the variability of the outputs, echoes an observation from the Peer Reviewer for our triangulation process: "What is very striking about all of the research being done, is that all researchers can claim they are being consistent in terms of the HMT Green Book (i.e. undertaking stated preference research in one form or another), but this doesn't rule out methods being employed in non-standard ways."[11]

Our assessment is that the two PR24 studies are perceived by the Delphi panel to be successful in the way they have presented the service attributes and asked respondents to express their priorities. The exercises in both studies are more intuitively appealing to respondents and easier to complete when compared to previous (DCE) approaches that characterised PR19. However, this departure from more established approaches and the attendant lack of a body of supporting technical knowledge raises a range of questions when it comes to interpreting the outcomes.

All of this serves to support the importance of using triangulated values in the Copperleaf investment modelling and testing the full range of variation in values that result from these very different studies. This includes running scenarios where the contribution of the new studies is down weighted, excluded altogether and/or used exclusively. That way SSC will have a full understanding of how the variation in inputs will ultimately affect the investment calculations in Copperleaf.

---

[10] See the appendix (Peer Review: Final Observations, 'WTP v WTA' and 'Delphi Validation') as to how this may reflect a lack of understanding of the appropriateness of WTA
[11] Correspondence with Iain Fraser, 30/3/2023

5. Copperleaf Inputs

Total SSC Scenarios

A series of alternative scenario sets were constructed using the RAG approach, summarised in Tables 5.1a and 5.1b and divided into a 'core' set of six plus five additional 'sensitivity tests'. The key terms in the titles are:

- LOWEST / HIGHEST RAG – Results based on the RAG ratings being all at their lowest or (in most cases) at their highest (see Chapter 2 for the detail on RAG ratings)

- Lower – Central – Upper value – Results relating to a range, where the central vale equates to a mean average WTP value and lower / upper values represent a confidence range, partly based in statistical measures (standard errors) but also on the RAG rating used.

Table 5.1a: Scenarios prepared for Copperleaf inputs

	Description	Notes
1. ALL – HIGHEST RAG Central value	All sources have been used to calculate these central values, with each source weighted by the highest RAG value across ratings of theoretical, statistical and depth validity.	*An example of using the highest RAG rating: for NERA PR24, we rated Theoretical=Green, Statistical=Amber, Depth=Green. The 'highest rating' used for this study was therefore Green.*
2. LOW NERA AND ODI RAG - ALL OTHERS HIGHEST RAG Central value	As input 1 above, but with NERA and ODI (national and SSC) given their lowest RAG values	*An example of using the highest RAG rating: for NERA PR24, we rated Theoretical=Green, Statistical=Amber, Depth=Green. The 'lowest rating' used for this study was therefore Amber.*
3. NO NERA – HIGHEST RAG Central value	As input 1 above, but with NERA removed. ODI pivot values (estimated for attributes that were not covered by ODI) now based on pre-PR24 values	*The impact of using pre-PR24 values to pivot attributes not included in the ODI study is to produce higher values, particularly for leakage and protecting wildlife habitats)*
4. PRE PR24 – HIGHEST RAG Central value	As input 1 above, but with NERA and ODI removed. Effectively PR19 values with some updates	
5. ALL – HIGHEST RAG Lower value	As input 1 above, but all 'lower' values	*All input values come with a low, central and high value that reflect the range of sources and the weightings applied to them*
6. ALL – HIGHEST RAG Higher value	As input 1 above, but all 'higher' values	*See above*

In addition to the inputs based on all sources (input 1), the purpose of the other value sets was to reflect the wide variation due to the inclusion or exclusion of different sources (inputs 2 to 4) and the range of variation around the central values calculated for each triangulated set of inputs (inputs 5 and 6). The sensitivity tests in Table 5.1b covered the impact of RAG ratings (second, third and fourth tests) and the role of using ODI values alone (first test) or excluding them completely (fifth test).

Table 5.1b: Sensitivity Tests prepared for Copperleaf inputs

	Description	Notes
ODI only – HIGHEST RAG Central value	ODI results for SSC only	
ALL – LOWEST RAG Central value	All sources have been used to calculate these central values, with each source weighted by the lowest RAG value across ratings of theoretical, statistical and depth validity	*An example of using the highest RAG rating: for NERA PR24, we rated Theoretical=Green, Statistical=Amber, Depth=Green. The 'lowest rating' for this study was therefore Amber.*
ALL - HIGHEST test 1 Central value	As 1 above, but with alternative RAG weightings (Green=100, Green/Amber=75, Amber=50, Amber/Red=25, Red=0)	
ALL - HIGHEST test 2 Central value	As 1 above, but with alternative RAG weightings (Green=100, Green/Amber=50, Amber=0, Amber/Red=0, Red=0)	
NO ODI – HIGHEST RAG Central value	As input 1 above, but with ODI removed.	

Table 5.2 summarises the values that could be used in Copperleaf for HH and NHH combined, corresponding to the six sets of inputs listed in Table 5.1a.

Table 5.2: Values (per unit) to be tested in Copperleaf (High RAG ratings, HH and NHH combined, total SSC)

COMBINED SSC	ALL - HIGHEST Central value	LOW NERA AND ODI - ALL OTHERS HIGHEST Central value	NO NERA - HIGHEST Central value	PRE PR24 Central value	ALL - HIGHEST Lower value	ALL - HIGHEST Higher value
Water not safe to drink (per property affected)	£73,592	£27,985	£5,983	£1,510	£14,779	£303,914
Flooding from a burst pipe (per property affected)	£23,775	£10,102	£2,090	£1,064	£4,983	£85,550
Unexpected temporary loss of water supply (per property affected)	£3,369	£1,832	£4,573	£506	£674	£14,259
Water hardness (per property affected)	£484	£437	£404	£381	£98	£1,762
Taste and smell of water (per property affected)	£2,116	£1,166	£2,876	£520	£423	£7,030
Low water pressure (per property affected)	£1,185	£582	£1,612	£74	£237	£3,991
Lead pipes (per property affected)	£39	£40	£50	£42	£8	£89
Water metering (per customer)	£8	£10	£8	£10	£3	£20
Customer service (per customer)	£0	£0	£0	£0	£0	£0
Temporary use ban (1% change in risk)	£685,465	£671,469	£875,589	£646,860	£137,093	£1,429,639
Leakage (1 Mega Litre per day)	£170,328	£131,704	£140,076	£107,278	£42,578	£450,886
Protecting wildlife habitats (per hectare)	£19,109	£19,666	£24,285	£21,779	£3,822	£61,063

Table 5.3 summarises the values for the sensitivity tests also listed in Table 5.1b.

Table 5.3: Values (per unit) to be tested in Copperleaf (Sensitivity tests)

COMBINED SSC	ODI ONLY- SSC ONLY Central value	ALL - LOWEST Central value	ALL - HIGHEST - RAG TEST 1 Central value	ALL - HIGHEST - RAG TEST 2 Central value	NO ODI - HIGHEST Central value
Water not safe to drink (per property affected)	£14,880	£47,149	£60,668	£94,499	£91,582
Flooding from a burst pipe (per property affected)	£3,711	£14,924	£21,595	£26,502	£31,464
Unexpected temporary loss of water supply (per property affected)	£10,880	£2,690	£3,577	£3,038	£304
Water hardness (per property affected)	£381	£444	£465	£511	£516
Taste and smell of water (per property affected)	£7,875	£1,649	£2,047	£2,216	£280
Low water pressure (per property affected)	£4,308	£886	£1,190	£1,178	£49
Lead pipes (per property affected)	£65	£39	£39	£38	£28
Water metering (per customer)	£3	£9	£8	£10	£11
Customer service (per customer)	£0	£0	£0	£0	£0
Temporary use ban (1% change in risk)	£1,374,351	£723,646	£714,318	£645,648	£453,509
Leakage (1 Mega Litre per day)	£202,189	£140,525	£160,296	£188,605	£157,401
Protecting wildlife habitats (per hectare)	£33,759	£21,390	£20,654	£15,854	£14,726

A breakdown of these figures by region and customer type are available in file '**Summary WTP Tables 210523.xlsm**'. The main differences between regions and customer types that emerge when comparing the Highest RAG rating scenario 1 (central values) are:

- 'Water not safe to drink' and 'Temporary use ban' are more highly valued in SST than in CAM
- 'Flooding from a burst pipe', 'water hardness', 'Taste and smell of water', 'leakage' and 'protecting wildlife habitats' are more highly valued in CAM than in SST
- 'Water not safe to drink', 'Flooding from a burst pipe' and 'Water hardness' are more highly valued by HH customers than NHH customers
- 'Unexpected temporary loss of water supply', 'Taste and smell of water', 'low water pressure', 'temporary use ban', 'leakage' and 'protecting wildlife habitats' are more highly valued by NHH customers than HH customers

Scenarios for Sub-Groups

Further work was commissioned by SSC from Accent/PJM to derive values for vulnerable customers and values were also available from the NERA research. Table 5.4 shows the values derived for a range of vulnerable groups: Medical = Disabled or suffers from a debilitating illness, has a learning difficulty, relies on water for medical reasons, visually impaired, deaf or hard of hearing; Communication = Speaks English as a second language; Life stage = Over the age of 75 years old or a new parent. Financial vulnerability was defined as 'I usually / always struggle to pay my bills'.

Table 5.4a: Values (per unit) for Vulnerable Customers (ODI)

COMBINED SSC	MEDICAL	COMMUNICATION	LIFE STAGE	FINANCIALLY
Water not safe to drink (per property affected)	£314	£329	£273	£46
Flooding from a burst pipe (per property affected)	£163	£171	£142	£24
Unexpected temporary loss of water supply (per property affected)	£238	£287	£225	£40
Water hardness (per property affected)	£108	£114	£94	£16
Taste and smell of water (per property affected)	£165	£194	£157	£29
Low water pressure (per property affected)	£95	£112	£88	£18
Lead pipes (per property affected)	£8	£8	£7	£1
Water metering (per customer)	£5	£5	£5	£1
Customer service (per customer)	£0	£0	£0	£0
Temporary use ban (1% change in risk)	£766,671	£907,754	£724,012	£123,431
Leakage (1 Mega Litre per day)	£18,146	£19,048	£15,783	£2,689
Protecting wildlife habitats (per hectare)	£3,973	£4,171	£3,456	£589

Table 5.4b: Values (per unit) for Vulnerable Customers (ODI)

COMBINED SSC	FINANCIALLY VULNERABLE	SOCIALLY VULNERABLE	ALL VULNERABLE
Water not safe to drink (per property affected)	£250,248	£278,562	£244,364
Flooding from a burst pipe (per property affected)	£0	£119,262	£83,904
Unexpected temporary loss of water supply (per property affected)	£0	£0	£0
Water hardness (per property affected)	£0	£1,367	£1,094
Taste and smell of water (per property affected)	£0	£0	£0
Low water pressure (per property affected)	£0	£0	£0
Lead pipes (per property affected)	£0	£0	£0
Water metering (per customer)	£0	£0	£0
Customer service (per customer)	£0	£0	£0
Temporary use ban (1% change in risk)	£0	£0	£0
Leakage (1 Mega Litre per day)	£8,346	£164,509	£136,415
Protecting wildlife habitats (per hectare)	£1,436	£2,086	£1,573

Reflecting the Delphi Validation

In his final observations, the Peer Reviewer reflected that "given the difference in ranks between the methods [NERA/ODI v Delphi], can these differences be used to make any changes to the WTP/WTA estimates"? Table 5.5 summarises the most distinctive variations in Tables 4.5 and 4.6:

Table 5.5: Delphi Rankings v NERA/ODI values

Attribute	HH Customers	NHH Customers
Leakage	ODI is low	-
Lead pipes	NERA and ODI are low	-
Protecting wildlife habitats	NERA is high	NERA is high
Do not drink notice	NERA is high	NERA is high
Flooding from a burst pipe	NERA is high	-
Low pressure	ODI is high	-
Hard water	NERA and ODI are low	-
TUBS	ODI is high	NEAR is low

On this basis, we have selected values from Table 5.2 that omit ODI, NERA or both to reflect this variation. In some cases, an attribute might be seen to be ranked higher or lower in a study, but the actual value in £ per unit is still higher than in the main scenario ('ALL – HIGHEST Central value'). In such cases we have kept with the main scenario value. The set of values based on this selection is presented in Table 5.6. This forms one additional set of values for a sensitivity test that could be used as Copperleaf inputs.

Table 5.6: Values (per unit) selected in response to Delphi Rankings

COMBINED SSC	Selected Value	Source (see Table 5.2)
Water not safe to drink (per property affected)	£5,983	NO NERA – HIGHEST Central value
Flooding from a burst pipe (per property affected)	£2,090	NO NERA – HIGHEST Central value
Unexpected temporary loss of water supply (per property affected)	£3,369	ALL – HIGHEST Central value
Water hardness (per property affected)	£378	ALL – HIGHEST Central value
Taste and smell of water (per property affected)	£2,116	ALL – HIGHEST Central value
Low water pressure (per property affected)	£437	LOW NERA AND ODI - ALL OTHERS HIGHEST Central value
Lead pipes (per property affected)	£42	PRE PR24 Central value
Water metering (per customer)	£8	ALL – HIGHEST Central value
Customer service (per customer)	£0	ALL – HIGHEST Central value
Temporary use ban (1% change in risk)	£685,465	ALL – HIGHEST Central value
Leakage (1 Mega Litre per day)	£170,328	ALL – HIGHEST Central value
Protecting wildlife habitats (per hectare)	£19,109	ALL – HIGHEST Central value

6. APPENDICES

Peer Review: Methodology and Application

For this phase of the research, Professor Iain Fraser of Kent University was requested to consider the following:

1. Review the calculations to ensure that the PJM approach from 2018 has been correctly implemented.
2. Challenge the RAG ratings that we have assigned to all the latest data sources and whether the sources themselves are credible.
3. Review the feedback from the Peer Review that was done on PJM's report in 2018 and then comment on this and on how we are approaching this at PR24.
4. Suggest any data source that might have been missed from outside the sector or within water that could be added to the triangulation approach.

In addition to these specific questions, Professor Fraser provided some wider thoughts on some of the issues affecting WTP research and how these should be considered in relation to triangulation. These are presented in the final section. In each section, further commentary from ourselves is also included.

Request 1   Review the calculations to ensure that the PJM approach from 2018 has been correctly implemented

Impact has adapted the approach used by Accent in PR19 to weight the various available WTP values and produce the central and high/low values against which the new results from the PR24 studies can be compared. Having reviewed the relevant workbooks, the implementation of the calculations appears to be in line with the PR19 exercise. For example, you can easily verify how the high and low values reported have been derived.

> Impact: We appreciate the time and attention that has been given to checking our work and the assurance that the approach has been implemented correctly

Request 2   Challenge the RAG ratings that we have assigned to all the latest data sources and whether the sources themselves are credible.

As part of undertaking this project for SSC, Impact has employed the Red/Amber/Green (RAG) ratings approach to subjectively assess the basis of theoretical/statistical/depth validity as part of the triangulation of evidence. The RAG is by design a subjective tool that can be used to undertake sensitivity analysis of evidence. As such it is essential that the RAG approach is adequately tested so as to determine the robustness of results being reported and to reveal any step changes that occur if evidence is weighted differently. Within the Excel files provided there is evidence of sensitivity analysis of the RAG ratings used. This is appropriate given the highly subjective nature of the process of weighting existing studies. As indicate by Impact, consideration of the techniques used by each study is implicit in the RAG ratings (eg the Satisfaction / Contacts etc are particularly marked down for this). The sensitivity analysis conducted by Impact is reported in Table 2.9 and Figure 3.1c of this report.

As part of this review, additional sensitivity analysis has been conducted. However, in this case the analysis does not change the weighting attached to the five RAG rating categories. Instead, the actual RAG rating has been changed. An example of this has been conducted for the household (HH) data to examine the extent to which the WTP values change. The base line RAG ratings and the revised ratings are shown in Table 6.1.

We then compare how the change in RAG rating, that is assuming all existing data sources and information are of the highest quality, impacts the combined unit WTP values derived. This is an extreme test of the robustness of the analysis and if the impact of this reweighting of existing evidence has an impact, we would expect it to be revealed as a large difference in the resulting combined unit WTP value derived.

The percentage difference between the two (Base and revised) is shown in Table 6.2 in the column headed % Diff WTP. The final column also reports if the change in the combined unit WTP value falls outside of the reported high and low values for all of the attributes shown.

Table 6.1: Sensitivity Analysis of RAG Ratings by Study

Sources / RAG ratings	Base		Revised	
WTP core_DCE	Green	100%	Green	100%
WTP core_Maxdiff	Amber	25%	Green	100%
WTPCore_DCE2	Green	100%	Green	100%
WTPCore_DCE2a	Red	0%	Green	100%
WTPCore_DCE2_LowBill	Red	0%	Green	100%
WTP core_DCE_Private	Red	0%	Green	100%
Priorities	Amber	25%	Green	100%
Contacts	Amber	25%	Green	100%
Satisfaction	Amber	25%	Green	100%
WTPPR14	Red	0%	Green	100%
WRMP online	Amber	25%	Green	100%
WRMP workshops	Amber	25%	Green	100%
ExternalWTP14	Red	0%	Green	100%
ExternalWTP19	Amber	25%	Green	100%
PC Slider	Amber	25%	Green	100%
WRMP MCDA	Amber	25%	Green	100%
SRO	Green / Amber	50%	Green	100%

Table 6.2: Sensitivity Analysis of Combined HH and NHH WTP from Change in RAG Ratings

Attributes	Unit	% Diff WTP		Within High/Low Range	
		HH	NHH	HH	NHH
Water not safe to drink	Property affected	-24.8%	-15.5%	Yes	Yes
Discoloured water	Property affected	2.0%	75.3%	Yes	Yes
Taste and smell of water	Property affected	13.2%	98.7%	Yes	Yes
Lead pipes	Property affected	-15.8%	36.7%	Yes	Yes
Water hardness	Property affected	-37.3%	105.2%	Yes	Yes
Unexpected temporary loss of water supply	Property affected	17.1%	27.8%	Yes	Yes
Low water pressure	Property affected	-0.5%	28.6%	Yes	Yes
Flooding from a burst pipe	Property affected	-19.6%	-48.1%	No	Yes
Temporary use ban	1% change in risk	-23.4%	-13.6%	Yes	Yes
Drought restrictions	1% change in risk	41.2%	19.4%	Yes	Yes
Leakage	ML/D	-9.8%	38.1%	Yes	Yes
Water metering	Household	-53.6%		Yes	
Protecting wildlife habitats	Household	-1.1%	38.5%	Yes	Yes
Managing impacts on rivers & streams	Hectare	5.2%	15.9%	Yes	Yes
Traffic disruption	Hectare	-12.1%	8.2%	Yes	Yes

The results in Table 6.2 reveal that the biggest percentage changes occur for the attributes that are the most sensitive to how the estimates have been derived. Importantly, even though some of the percentage differences are significant, the new estimates do not fall outside the range of the estimates presented by Impact for the base level analysis. The only exception is 'flooding from a burst pipe' for HH customers, and even in this extreme case the value is only just outside the range.

The task also raises the issue about credibility of sources. The most obvious response to this question would seem to be that even if certain sources are subject to a significant degree of uncertainty, the overall impact on the estimates appears to be minimal.

> Impact: the ability of the approach to reflect the comparative uncertainty around different sources while delivering results that are within the range of the base analysis is reassuring. The final ranges of triangulated values would therefore appear not to be unduly influenced by the subjective RAG ratings, while still reflecting the variability in the reliability of the sources.

Request 3   Review the feedback from the Peer Review that was done on PJM's report in 2018 and then comment on this and on how we are approaching this at PR24

The Peer Review of the Accent/PJM 2018 report was undertaken by Professor Giles Atkinson. The review is generally of a high quality. The review correctly identifies an important issue with the work: the *"modification of WTP values"*. Following the HM Treasury (2018) Green Book in terms of method is appropriate, but I think that there are issues with how the methods have frequently been employed and not the methods themselves. It is the application of the methods that has given rise to the proposed approach to adjusting estimates once the various methods have been employed. This raises two options:

i)   Adjust the estimates ex post, which is the method employed and reviewed by Atkinson; or
ii)   Reframe how the methods are being used in an effort to overcome the identified limitations and as such avoid the need for ex post adjustments.

The adjustment method used has now been published in a refereed article (Chalak and Metcalfe, 2022). The methodology is a re-interpretation of how a Value of a Statistical Life (VSL) is calculated (see Hammitt (2000) for details). The methodology implies a two-part valuation exercise. Both parts are credible methodologically. The first step is to undertake a Max Diff exercise, a type of Best-Worst Scaling (BWS). There is clear merit in using this approach when one is interested in the relative rank of attributes.

The use of the Max Diff exercise does in fact suggest a way to align the Delphi exercise and the triangulation exercise. If the same set of attributes are used in this as in the Delphi exercise, it is possible to see if the preferences of the sample of respondents matches those of the "experts" in the Delphi panel. This could be a useful addition to how the research activity is linked, as a means to examine if responses are "internally consistent."

Returning to the methodology employed, there is then a second stated preference survey undertaken to examine a limited number of service package options. This is akin to a standard CV or DCE (it is implemented a status quo versus a single package). Again, this task makes sense in and of itself and the analysis and results appear meaningful. The real issue with the method, is how the two sets of information are brought together. There are clearly theoretical assumptions being made that are not explicitly stated. It remains to be seen if these assumptions are important.

> Impact: It is reassuring that the approach adopted in PR19 and now extended to PR24 has gained wider recognition. Professor Fraser's comments relating to Max Diff and Delphi have been incorporated in our analysis of the Delphi responses, where we compare 'expert' rankings with the WTP results. The final comment on methodology picks up the general issue raised in the first paragraph about the application of the methods. That is, the theoretical foundations for combining the results of the max diff / DCE are not that clear. This is a general criticism that can be applied both to the current ODI approach (Accent/PJM 2022) and the NERA approach, both of which combine two different methods.

Request 4   Suggest any data sources, we might have missed from outside the sector or within water that could be added to the triangulation approach.

Currently I have no obvious suggestions other than reconsidering the Environment Agency work on rivers and water bodies that was published in 2007. The data are based on the National Water Environment Benefit Survey (NWEBS) (NERA Accent, 2007). I am aware that the Environment Agency are in the process of moving towards undertaking research to update the values and estimate.

Request 5   Observations on WTP for Price Reviews and the Triangulation of Values

The use of SP methods to measure WTP

The use of stated preference methods such as contingent valuation (CV) and discrete choice experiments (DCEs) is now widely accepted (i.e., HM Treasury (2022) Green Book) in the Price Review process, although there are clearly concerns being raised about the results being generated. These concerns stem in part from the good/service (an odd mix of private and public issues) being examined simultaneously and how methods are being implemented in practice.

For example, it would be useful to reframe a DCE as an insurance purchase: 'What would you be prepared to pay to insure against specific issues that directly relate to the household'. A task such as this would focus on only issues that directly relate to the individual/household. The task would much more clearly align with the standard application of stated preference methods, especially how they are employed in product or service choice contexts (e.g., buying a new food product). In addition, the public good issues being examined need to be clearly framed within the regulatory requirements that water companies are obliged to satisfy. Asking for customers to pay more to ensure regulatory compliance needs to considered against a company's existing environmental performance in the proceeding years. Offering public good choices without placing the improvements against actual performance should possibly be avoided. In regard to water quality, the Environment Agency (NERA and Accent, 2007) used both an increase in water bills and price of others goods in the research it conducted in 2007 on water quality. However, it is questionable if such a payment vehicle is appropriate in 2023 given the poor reputation of water companies when it comes to a range of environmental issues.

Triangulation

Much is being made about how this type of research can be considered as meaningful as a result of triangulation. There needs to be clear and unambiguous statement about the meaning of triangulation as it is being applied here. For example, Noble and Heale (2019) provide the following definition:

> *"Triangulation is a method used to increase the credibility and validity of research findings. Credibility refers to trustworthiness and how believable a study is; validity is concerned with the extent to which a study accurately reflects or evaluates the concept or ideas being investigated."*

In practice, Noble and Heale (2019) identify four types of triangulation:
1) data triangulation
2) investigator triangulation
3) theory triangulation
4) methodological triangulation

There are also known limitations with triangulation and these need to be recognised whenever a study alludes to the implicit validity conferred on research as a result of employing this type of method.

A potential way to understand how well results derived using different methodologies correlate is as follows. The MaxDiff (Best-Worst Scaling (BWS)) approach reveals how respondents rank a set of attributes. As shown in Chalak and Metcalfe (2022) the set of attributes given to household survey participants was composed of 19 attributes. It would be interesting to know if the Delphi study yields a similar set of ranking scores to those produced using this methodology.[12]

---

[12]   See 'Delphi rankings v measured priorities' in section 4 of this report, where some rankings were broadly in line, but others differed markedly

Future applications of SP methods

In terms of DCE implementation, there are several methods that already exist that can deal with a large number of attributes. The partial design DCE is now reasonably well established and would seem well suited to the type of problem being examined here. An example of this approach is provided by Kessels et al. (2011). Another new form of DCE introduced by Caputo and Lusk (2022) also warrants consideration.

As noted by Ukpong et al. (2019) including attributes that have private costs and benefits as well as clear social implications may involve a degree of altruism which has been the subject of some concern in the stated preference literature (Zhang et al. 2013). This issue has led some to argue that altruism has no place in valuation research (Jacobsson et al. 2007). Of course, if we exclude such attributes from a DCE, it implies that either respondents do not care about them, or they attributes are assumed to be constant. These assumptions may be unrealistic when faced with actual choices that are required in reality. A potential way to address this issue would be to have two separate DCE for water services and waste water services. For the water services the choice could be framed in terms of buying an insurance policy to avoid the set of potential impacts. Trying to reduce the set of attributes and to make the choice task easier and more meaningful is clearly what is driving the variation in methods being presented. And although this is to be expected, it is probably best to be avoided when the research is meant to be underpinning a regulatory exercise and as such really requires consistency of method across all studies.

There appears to be confusion in some stated preference applications about whether or not WTP estimates need to be greater than zero (i.e., non-negative). There is no reason to assume that specific attribute levels in DCE will be positive. Thus, expressing a negative WTP is akin to requiring a price reduction. It is not, however, a measure of Willingness to Accept (WTA). A measure of WTA requires a specific framing of the choice task and implicit change in how property rights are being assigned. For example, if offered a piece of chicken that has been produce using a chlorine wash, then we might expect many consumers to require a price reduction (i.e., a negative value placed on this attribute of the good) to consider buying the good. In contrast, if I propose an action, such as building a road that impacts the view from your house and increases noise and pollution levels, this implicitly impacts on an implied property right that one had in regard to view from your house, plus the change in noise and pollution levels. This type of change can be framed as a WTA task. Confusion regarding this issue seems to be occurring in some of the valuation results being reported.

> Impact: Regarding the work conducted for SSC by NERA for PR24, we note the comments regarding the potential problem of consistency of methods, where the approach used is notably different from other more conventional SP applications, albeit with the objective of making the exercise simpler for respondents. From the above discussion about separating water and wastewater, it is a positive that the NERA study only focuses on water services, whereas the Accent/PJM ODI study combines water and waste water services in one exercise. Finally, the examples regarding negative WTP are accepted, but we would argue that while some of the negative WTP values for some attributes in the NERA study correspond to these examples (e.g., smart metering is not unambiguously positive for all customers), others do not (e.g., water discolouration, customer service)

References

References for section 3

Chalak, A., & Metcalfe, P. (2022). Valuing water and wastewater service improvements via impact-weighted numbers of service failures. Journal of Environmental Economics and Policy, 11(1), 39-55.
Hammitt, J. K. (2000). "Valuing Mortality Risk: Theory and Practice." Environmental Science & Technology 34 (8):1396–1400.
Willis, K., & Sheldon, R. (2022). Research on customers' willingness-to-pay for service changes in UK water company price reviews 1994–2019. Journal of Environmental Economics and Policy, 11(1): 4-20.

References for Section 5

Caputo, V., & Lusk, J. L. (2022). The Basket-Based choice Experiment: A method for food demand policy analysis. Food Policy, 109, 102252.
Chalak, A., & Metcalfe, P. (2022). Valuing water and wastewater service improvements via impact-weighted numbers of service failures. Journal of Environmental Economics and Policy, 11(1), 39-55.
HM Treasury (2022). The Green Book. Central Government Guidance on Appraisal and Evaluation. www.gov.uk/official-documents

Kessels, R., Jones, B., & Goos, P. (2011). Bayesian optimal designs for discrete choice experiments with partial profiles. Journal of Choice Modelling, 4(3), 52-74.

NERA and Accent (2007) The Benefits of Water Framework Directive Programmes of Measures in England and Wales. Report to DEFRA.
Noble, H. and R. Heale, (2019). Triangulation in research, with examples. Evidence-Based Nursing 22:67-68.
Ukpong, I. G., Balcombe, K. G., Fraser, I. M., & Areal, F. J. (2019). Preferences for mitigation of the negative impacts of the oil and gas industry in the Niger Delta Region of Nigeria. Environmental and Resource Economics, 74(2), 811-843.
Zhang, J., Adamowicz, W., Dupont, D. P., & Krupnick, A. (2013). Assessing the extent of altruism in the valuation of community drinking water quality improvements. Water Resources Research, 49(10), 6286-6297.

Peer Review: Final Observations

Aggregating/Disaggregating Values

A major challenge faced in the triangulation exercise in Chapter 3 is that aggregation or unit for which the WTP (or WTA) is derived is not necessarily consistent across studies. So, a major issue that has been confronted in the report is how best to normalise values to allow comparison at a common unit level. In practice this means that heterogeneity (ie, variation) not only occurs in terms of methodological implementation but also what is measured and how it is measured. It should therefore come as no great surprise that there is a significant degree of variation reported in the values used in Chapter 3. This illustrates the point that all studies can be consistent with the HMT Green Book but how they are implemented can and does vary significantly.

The triangulation exercise also requires an assumption about aggregation/disaggregation of values for differences in scale. Essentially the assumption is that estimates can be increased/decreased given the change in scale. However, when employing contingent valuation methods, it has long been known that as the scope of a problem increases and/or decreases, the associated WTP estimates do not necessarily reflect the change in scope. In general, we would hope that as the scope of say an environmental issue increases, that the WTP estimates will also increase (although this need not be linear). So, the normalisation of estimates to enable a comparison between studies for identical/similar attributes is assuming that scope holds.

A related issue in regard to derivation of WTP estimates is in terms of public versus private attributes. The 'nature' attribute may be small at the individual household level, but you then aggregate by the population covered by the water company. This will always make the total WTP look very large. This will always be the case for an attribute which is a public good. For the private attributes estimating a total WTP estimate is obviously more complicated, especially as some are stated as percentage changes or probabilities.

As an aside the HH 'nature' WTP values in Table 3.5 are significantly larger (£,1800-£10,000 per ha/yr) than the payments that farmers obtain for current agri-environmental policy, such as the Countryside Stewardship Scheme (CSS). For example, if a farmer adopts an option such AB8 (Creation of wildflower meadow) then they are paid £673 per ha/yr for at least 5 years.

In summary, my comments so far are more about the complexity and inherent difficulty in undertaking the triangulation exercise, rather than specific observations on the approach (see previous review of Methodology and Application).

WTP versus WTA

The issue around how to frame the DCE all comes down to how property rights are being framed. So, the WTP question generally implies that you are paying for something. This also means that a negative WTP *is not equivalent to a WTA estimate*; it is actually the price reduction that is required as opposed to the price increase.

If a WTA question is employed, then the implication is that something is typically being taken away, which implies prior ownership. In many ways, attributes that are being framed as negative impacts on consumers can and probably should be framed as WTA type questions. The rationale for not using WTA reflects an historical legacy that it is not really appropriate. Indeed, as noted, there are now calls for employing WTA when it is appropriate - see the reference: Ando, A.W. Equity and Cost-Effectiveness in Valuation and Action Planning to Preserve Biodiversity. *Environ Resource Econ* **83**, 999–1015 (2022).

Compensating Behaviours

Looking forward to future Price Reviews, one obvious question to ask respondents would be: How would you respond to an increase in the per unit cost of water services? Unless the own price elasticity of demand is very (very) inelastic, we might expect to see some sort of reduction in demand. In future work, we could ask HH and NHH how they might change behaviour and water use.

Delphi Validation

In general, I like Chapter 4 and the comparison of the attributes by rank order. I think that the Delphi reveals some interesting issues around the importance of attributes as understood by experts and how some attributes have been framed and valued by the WTP/WTA studies. the most obvious being "hard water".

Tables 4.5 and 4.6 make for interesting reading. I like the direct comparison of attribute ranks. My only comment is then, given the difference in ranks between the methods can these differences be used to make any changes to the WTP/WTA estimates or at least inform a revised set of RAG weightings?

The Delphi respondents' comments about the NERA approach are interesting. This suggests that *how the tasks are presented* and not what they have revealed is considered more important. Again, these answers seem to indicate that any triangulation exercise is going to be complicated and subject to lots of practical challenges. I think the Delphi responses to the WTA probably reflect a lack of understanding for why a WTA is appropriate. This may be something that needs to be revisited in future Delphi like exercises.

I think this Delphi study reveals that the way problems are framed in the survey instrument has a bearing on how experts judge the survey. But the framing and the resulting implementation are not the same thing. If the water companies want consistent (with HMT Green Book) WTP/WTA estimates, then they need to take care with how these are being derived. I would strongly advocate that employing a methodology that has not been previously peer reviewed is a risky strategy to employ.

Impact response to Professor Frasers final observations:

The challenges of triangulating widely differing values derived from very different methods are indeed considerable. The approach we have applied on behalf of SCC is only one method, but we argue that it is supported by building on the established approach develop for PR19, expanding the scope with the Depth component of the RAG ratings and the addition of the Delphi panel and finally, transparency of process as outlined in this report and subjected to independent Peer Review.

We have attempted to address the question as to how Delphi feedback can be used to guide the selection of values that will go on to be used for investment appraisal. To this purpose we have constructed an additional table in Chapter 5: 'Total SSC Scenarios', where selected values are presented in response to the comparisons of Delphi v NERA/ODI made in Tables 4.5 and 4.6.

Finally, the diversity of methods used to derive customer values, and specifically the fundamental differences between WTP and WTA as ways to measure those customer values, needs to be more firmly addressed in future Price Reviews. The reviewer's support for WTA, as used in the ODI ratings research, is noted. Also noted is the implicit criticism of the approach used by NERA, with respect to it being a largely untested method prior to this study. In defence of this, we would point to the extensive peer review that was conducted during the development of their approach and the modifications that were made in response to criticism during the design and piloting phases.

However, the issue remains that because the NERA approach is new and innovative, there is limited evidence from previous study to provide support. When compared to other sources (both WTA and WTP values), the general output from the NERA appears to be greater extremes in values, either very large (as for water not safe to drink, leakage and flooding) when other studies show much smaller values, or nil (as for temporary use bans, unplanned interruptions and water discoloration) when other studies indicate at least some value.

Values Derived for PR24

Chapter 3 summarised the weighted WTP values derived from all sources, pre-PR24 and PR24. This section examines in more detail the two PR24 studies (NERA and ODI)

PR24 ODI Ratings

Tables 6.3a/b and 6.4 show the findings from the latest ODI Rating work[13]. Those attributes relevant to the SSC service areas are highlighted in bold. The values for SSC, both HH and NHH, are all higher than for the total UK sample.

Table 6.3a: HH WTA from Accent/PJM research, 2022, ODI ratings

Attribute	Unit	All SSC Households	All UK Households
Emergency drought restrictions (2 months)	*Per property affected*	*£292*	*£236*
Unexpected water supply interruption (24h)	Per property affected	**£307**	**£204**
Do not drink notice (48h)	Per property affected	**£211**	**£184**
Boil water notice (48h)	*Per property affected*	*£201*	*£148*
Unexpected water supply interruption (6h)	Per property affected	**£172**	**£121**
Water taste and smell (24h)	Per property affected	**£118**	**£81**
Discoloured water (24h)	Per property affected	**£134**	**£78**
Discoloured water (6h)	Per property affected	**£99**	**£71**
Water taste and smell (6h)	Per property affected	**£108**	**£71**
Planned water supply interruption (6h)	*Per property affected*	*£76*	*£60*
Unexpected low water pressure (6h)	Per property affected	**£70**	**£62**
Low flows in rivers nearby (2 months)	*Per property affected*	*£58*	*£54*
Low flows in rivers elsewhere (2 months)	*Per property affected*	*£67*	*£43*
Hosepipe ban (5 months)	Per property affected	**£48**	**£40**

---

[13] Accent/PJM Research, 2022, ODI Ratings, Ofwat and CCW

Table 6.3b: Vulnerable Customers[14] WTA from Accent/PJM research, 2022, ODI ratings

Attribute	Unit	Medical	Communi-cation	Life stage	Financial
Emergency drought restrictions (2 months)	*Per property affected*	*£410*	*£493*	*£380*	*£69*
Unexpected water supply interruption (24h)	Per property affected	**£436**	**£490**	**£371**	**£74**
Do not drink notice (48h)	Per property affected	**£314**	**£329**	**£273**	**£46**
Boil water notice (48h)	*Per property affected*	*£273*	*£316*	*£262*	*£45*
Unexpected water supply interruption (6h)	Per property affected	**£238**	**£287**	**£225**	**£40**
Water taste and smell (24h)	Per property affected	**£165**	**£194**	**£157**	**£29**
Discoloured water (24h)	Per property affected	**£181**	**£219**	**£174**	**£34**
Discoloured water (6h)	Per property affected	**£143**	**£154**	**£129**	**£25**
Water taste and smell (6h)	Per property affected	**£149**	**£181**	**£139**	**£27**
Planned water supply interruption (6h)	*Per property affected*	*£108*	*£122*	*£100*	*£18*
Unexpected low water pressure (6h)	Per property affected	**£95**	**£112**	**£88**	**£18**
Low flows in rivers nearby (2 months)	*Per property affected*	*£79*	*£96*	*£78*	*£15*
Low flows in rivers elsewhere (2 months)	*Per property affected*	*£92*	*£109*	*£84*	*£16*
Hosepipe ban (5 months)	Per property affected	**£69**	**£82**	**£65**	**£11**

---

[14] Financial vulnerability was defined as 'I usually / always struggle to pay my bills', which accounts for about 5% of the total UK sample. For the other types of vulnerability, the responses were collapsed into three variables (Medical, Communication, Life stage)'; the codes for this question were: Medical: Disabled or suffers from a debilitating illness, Has a learning difficulty, Relies on water for medical reasons, Visually impaired (i.e. struggles to read even with glasses), Deaf or hard of hearing; Communication: Speaks English as a second language; Life stage: Over the age of 75 years old,] or a new parent.

Table 6.4: NHH WTA values from Accent/PJM research, 2022, ODI ratings

Attribute	Unit	All SSC NHH	All UK NHH
Emergency drought restrictions (2 months)	*Per property affected*	*£22,071*	*£20,254*
Unexpected water supply interruption (24h)	Per property affected	**£29,140**	**£22,972**
Do not drink notice (48h)	Per property affected	**£14,669**	**£12,295**
Boil water notice (48h)	*Per property affected*	*£9,093*	*£9,926*
Unexpected water supply interruption (6h)	Per property affected	**£10,709**	**£16,217**
Water taste and smell (24h)	Per property affected	**£7,756**	**£4,813**
Discoloured water (24h)	Per property affected	**£5,540**	**£4,857**
Discoloured water (6h)	Per property affected	**£5,222**	**£5,379**
Water taste and smell (6h)	Per property affected	**£5,915**	**£4,393**
Planned water supply interruption (6h)	*Per property affected*	*£9,219*	*£8,342*
Unexpected low water pressure (6h)	Per property affected	**£4,238**	**£4,612**
Low flows in rivers nearby (2 months)	*Per property affected*	**£1,667**	**£1,885**
Low flows in rivers elsewhere (2 months)	*Per property affected*	*£1,731*	*£1,682*
Hosepipe ban (5 months)	Per property affected	**£1,353**	**£1,341**

NERA SSC PR24

Tables 6.5a/b and 6.6 show the findings from the latest PR24 SSC work[15]. Where negative WTP values were reported, these are represented by '-' and have been treated as zero on recommendation from NERA.

Table 6.5a: HH WTP values from Nera,2022, WTP PR24

Attribute		Unit	HH WTP (£ per unit per household)		
			SST	CAM	Total[16]
A	Customer service	reduction in the percentage of costumers that wait more than 10 minutes	-	£0.00	£0
B	Risk of temporary "do not drink" notice	reduction in number of properties that received "do not drink" notice	£0.74	£0.97	£0.79
C	Installing "smart" water meters	increase in the percentage of properties having an operational "smart" meter by 2030	-	-	£0
D	Hard water supply	increase in the number of properties that benefit from investment (thousands)	£0.00	£0.03	£0.01
E	Lead pipes	reduction in the percentage of properties that have a lead supply pipe by 2030	-	-	£0
F	Water lost to leakage from pipes	reduction in the percentage of water that is lost to leakage	£0.61	£1.40	£0.77
G	Issues with tap water colour, taste, or smell	reduction in the percentage of properties experiencing issues with tap water per year (tenth of a percentage)	-	£0.11	£0.08
H	Chance of property flooding from a burst pipe	reduction in the flooding incidents per year	£0.16	£1.03	£0.34
I	Low water pressure	reduction in the percentage of properties experiencing low pressure per year (tenth of a percentage)	-	-	£0
J	Supporting nature and wildlife	increase in the number of acres protected and enhanced (tens)	£0.03	£0.28	£0.08
K	Unplanned short interruptions to water supply	reduction in the percentage of properties experiencing a short interruption per year (hundredth of a percentage)	-	-	£0
L	Risk of temporary use ban, including hosepipes	reduction in the percentage chance of temporary use ban in a given year	-	-	£0

---

[15] NERA, 2022, Willingness to pay for water services at PR24, SSC, Table 2: Main Model WTP per Unit Change from SQ
[16] For consistency with pre-PR24 outcomes, we have taken the value from each region and calculated an average total value weighted by number of customers

Table 6.5b: Vulnerable HH WTP values from Nera,2022, WTP PR24

Attribute			HH WTP (£ per unit per household)		
			Financial Vulnerability	Social Vulnerability	Vulnerable (combined)
A	Customer service	reduction in the percentage of costumers that wait more than 10 minutes	-	-	-
B	Risk of temporary "do not drink" notice	reduction in number of properties that received "do not drink" notice	£0.80	£0.58	£0.59
C	Installing "smart" water meters	increase in the percentage of properties having an operational "smart" meter by 2030	-	-	-
D	Hard water supply	increase in the number of properties that benefit from investment (thousands)	£0.00	£0.00	£0.00
E	Lead pipes	reduction in the percentage of properties that have a lead supply pipe by 2030	-	-	-
F	Water lost to leakage from pipes	reduction in the percentage of water that is lost to leakage	£0.82	£0.15	£0.51
G	Issues with tap water colour, taste, or smell	reduction in the percentage of properties experiencing issues with tap water per year (tenth of a percentage)	-	-	-
H	Chance of property flooding from a burst pipe	reduction in the flooding incidents per year	£0.18	£0.00	£0.31
I	Low water pressure	reduction in the percentage of properties experiencing low pressure per year (tenth of a percentage)	-	-	-
J	Supporting nature and wildlife	increase in the number of acres protected and enhanced (tens)	£0.03	£0.06	£0.07
K	Unplanned short interruptions to water supply	reduction in the percentage of properties experiencing a short interruption per year (hundredth of a percentage)	-	-	-
L	Risk of temporary use ban, including hosepipes	reduction in the percentage chance of temporary use ban in a given year	-	-	-

Table 6.6: NHH WTP values from Nera,2022, WTP PR24

Attribute		Unit	NHH WTP				
			(% of bill per unit per property)		(£ per unit per property)		
			SST	CAM	SST	CAM	Total[17]
A	Customer service	reduction in the percentage of costumers that wait more than 10 minutes	-	-	-	-	£0
B	Risk of temporary "do not drink" notice	reduction in number of properties that received "do not drink" notice	0.013	0.015	£57.51	£76.80	£61.39
C	Installing "smart" water meters	increase in the percentage of properties having an operational "smart" meter by 2030	-	-	-	-	£0
D	Hard water supply	increase in the number of properties that benefit from investment (thousands)	-	-	-	-	£0
E	Lead pipes	reduction in the percentage of properties that have a lead supply pipe by 2030	-	-	-	-	£0
F	Water lost to leakage from pipes	reduction in the percentage of water that is lost to leakage	0.001	0.010	£4.42	£51.20	£13.84
G	Issues with tap water colour, taste, or smell	reduction in the percentage of properties experiencing issues with tap water per year (tenth of a percentage)	-	-	-	-	£0
H	Chance of property flooding from a burst pipe	reduction in the flooding incidents per year	0.002	0.005	£8.85	£25.60	£12.22
I	Low water pressure	reduction in the percentage of properties experiencing low pressure per year (tenth of a percentage)	-	-	-	-	£0
J	Supporting nature and wildlife	increase in the number of acres protected and enhanced (tens)	0.0002	0.003	£0.88	£15.36	£3.80
K	Unplanned short interruptions to water supply	reduction in the percentage of properties experiencing a short interruption per year (hundredth of a percentage)	-	-	-	-	£0
L	Risk of temporary use ban, including hosepipes	reduction in the percentage chance of temporary use ban in a given year	-	-	-	-	£0

Delphi Panel: Detailed Feedback on Service Attributes

Delphi: Household Customers

Attribute	Phase 1 HH insights		Phase 2 HH insights
Risk of temporary 'do not drink' notice	First thoughts: -Perceived impact of a do not drink notice is significant, more so for HH than NHH -'Essential service for HH and NHH respondents' -Importance of this might increase in the future if further evidence of water chemicals damages to humans and nature emerges 'This service deserves attention in water service planning' - water quality perceptions have been impacted by COVID (hypothesis) - want to see more information – can you still wash/ cook with the water? How many are actually impacted a year, 2 out of how many? - What about information on percentage of customers who have already installed water filter systems – these people might be less concerned about this disservice		-panelist 2 insight from overall HH customer WTP values from NERA: 'Relatively high WTP value on risk of temporary 'do not drink' notice. I would have thought customers would expect clean safe water as core service and therefore less WTP for it'.
Water quality: 'Issues with tap water colour, taste or smell'	-put high as customers expect clean, safe water from their water company. -'not a significant issue for customers but can exasperated when source changes' -need to be clear about facts about water quality, since there can be a lot of information in the media that might be untrue -good to know that it is still safe to drink. -low number of contacts for this and is still decreasing. But 'a very sensitive topic and fear and anxiety over water quality can be quickly triggered so information and transparency is a key factor to prioritize for this service.' - current habits on water consumption and info on installed water filter systems could help understand the necessity to monitor/improve this service		-panelist 3, after seeing industry comparison data, thought this could move further down the priority list (from a 4 to an 8), as it appears SSC does well compared to others -panelist 3 reaction to HH ODI values: 'Water taste and smell being higher than discoloured water is surprising'
Water quality: 'Hard water supply'	-not a major issue. Interesting that there is more demand for water softening in SSW than CAM, even though water is harder in CAM -picture is more nuanced than hard water just affecting appliances (health impacts of soft/hard water) -localised feature. 'the health affects of hard water are rarely considered and the main side effects are reported for appliances'		-panelist 2 elaboration on HH priorities ranking: 'Hard water - put not sure, as unclear as to the prevalence in the region and the demographic make-up. Many people care about the impact on appliances and skin and hair but taste wise it's a personal preference.'
Water quality: 'Lead pipes'	-'real concerns about lead pipes' -need to be clear about where responsibility sits for lead pipe replacement and who pays – i.e. householder or spread across customer base? -unclear how much of a health risk this really is. Who is most at risk? Symptoms? 'The benefit is expressed in pipes removed per customer not in the improvement in health which is the likely benefit to the customer -'this question is really about public health but has been calibrated in pounds and lead pipes'		-panelist 2 elaboration on HH priorities ranking: 'Lead pipes I put unsure, as I don't know how much of an issue that is in the company's area, health risks or levels of awareness.'
Unplanned interruptions to water supply	-customers need to be aware of the reality and likelihood (or not) of it happening. Explaining how interruptions can happen is also key, as not everything can be avoided. -how many properties would receive a short supply interruption multiple times.	-panelist 2 thought about putting supply interruptions lower as the company's performance is relatively good, but there has been a lot of national media coverage about interruptions, so decided on balance to keep it high - panelist 2 on reactions to how ODI questions were asked: '. Not defined vulnerable people - some may assume (as they do now as they are on the PSR) that they would get bottled water delivered	

---

[17] For consistency with pre-PR24 outcomes, we have taken the value from each region and calculated an average total value weighted by number of customers

Attribute	Phase 1 HH insights	Phase 2 HH insights
	-'water reliability is considered a key determinant of overall customer satisfaction and the expectation is that water must be reliable'. -what's the support package? When does it kick in, how is it accessed, how are vulnerable households targeted, etc -who will get bottled water? What about non vulnerable customers.	when they won't in practice. - None of these scenarios have considered multiple events. e.g. energy and water supply interruptions. Some people's adaption strategies may rely on using mobile phones or other activities that rely on energy.' - panelist 2 elaboration on HH priorities ranking: 'Unplanned supply interruptions I put high as I could 4th, though would have put it higher. It causes severe inconvenience to some people, safety issues for others, increasing coverage in the media. Growing body of evidence from wide variety of places that customers want more investment in long-term resilience.'
Chance of property flooding from a burst pipe	-need to be clear about the reality of a burst pipe and the support that would be provided -need to be clear about tech solutions available to support this area -no sense here about the often profoundly negative impact of a burst pipe on customers' lives or businesses and their employees when a property is flooded. -what about cost of drying the property, or living elsewhere? -where is the impact of bursts on transport picked up? E.g. burst pipes in streets results in roadworks and traffic problems – a high customer priority -not sure how easy it is to measure the importance of this service independently from the others – seems correlated with other services like water reliability and smart meters, so improving this infrastructure will be enough to minimize multiple disservices -framing of question is important – 51 properties seems like a high number but in context of 600,000 it seems small. If expressed as 0.0085% it seems much smaller and would illicit a lower WTP – could be considered a skewed question. -'would be interested to understand the impact of the 2020 Tipton incident and how that has changed perceptions'	-panelist 1, when talking about HH NERA values: for financially vulnerable groups, they were 'surprised that there was no prioritization of addressing flood risk since the impact of this is huge. Maybe the perception is that other authorities or bodies will fund any mitigation and redress should this happen?' - panelist 2 talking about reaction to ODI questions: 'says it takes 1 month to get back to normal. Experience of flooding suggests more like 3 months and some properties never fully recover'
Low water pressure	-cause and impacts need to be clearly detailed -can HHs take an action? -relatively low importance service since expectation is that water is or must be reliable -customers seem prepared to occasionally experience low pressure as long as notice is clear and in advance. Communication is a key element to manage this service -does it impact certain types of properties? How long would it take to fill a kettle, will the toilet flush, etc? -would schools have to close? Would be a consideration for parents.	-panelist 2 elaboration on HH priorities ranking: 'not sure as assuming it's a low number of people currently impacted and not sure what projections are for the future if investment isn't made'
Water lost to leakage from pipes	-need to detail support available to reduce customer-side leakage – identification, advice, repairs, funding, etc -SSC side or customer side leaks? -consumers would want to know how you can help them with their own leakage. -one of the most important factors with v high WTP in PR19. 'once informed [customers] are vocal in raising the importance to tackle this problem' -urgency of this service is probably also triggered by the need to save/preserve water -olympic swimming pool reference is useful -high media coverage	-panelist 1 when talking about HH customer WTP values from NERA: 'I was surprised about the lower water lost to leakage figure for SSW compared to CAM, given the Tipton incidents that are within recent memory. I imagine the higher CAM figure reflects the concerns about the aquifer in CAM and the environmental benefits that would stem from better leakage control….' -panelist 2 on elaboration of HH priority rankings: 'Leakage put highest, as is prominent in the media, people hate the waste, it has benefit for a number of the other areas e.g. environment, supply interruptions, water pressure.'
Supporting nature and wildlife	-'water companies are tainted by the issues arising from the current sewage discharge issue' -SSC say they have a legal duty to protect and enhance nature, so why are they asking this question? If	-Panelist 2 disagrees with a point raised by another expert in phase 1- that COL is driving environment down the priority list. Disagree that it is affordability vs environment. When people cant afford to go out, green/blue spaces become more important. Pollution disproportionately hits those on low incomes. Informed consumers know that if we protect biodiversity, it will keep costs down in the long run -When looking at NERA method: Participant 2 mentions the way that this attribute is presented, without explaining real life consequences e.g. ;what would this mean

Attribute	Phase 1 HH insights	Phase 2 HH insights
	it is a legal duty consumers might ask why they have to pay -no mention of lived experience of nature and wildlife, e.g. dog walks or water sports, which are important for customers to understand the impact. -less than half of customers think that water companies are doing a good job for the environment. -COL is shifting this down the priority list -no explanation of what protection and enhancement actually means	in terms of smell, look and feel of areas, resilience, pleasant places to walk dog/ with children, etc. Also the value of a hectare of biodiversity gain is not the same regardless, depends on where it is.' -panelist 2 when talking about robustness of NERA: 'There's potential for some double counting in terms of outcomes trying to achieve- query how this has been taken into account given how you've asked the question customers are likely to have undervalued environmental improvements.' More generally, higher values for CAM might be due to relative affluence. -panelist 1 when talking about HH WTP values from NERA: 'The supporting wildlife and nature was not surprising in CAM given the local focus, but I would have imagined a slightly higher focus in the SSW area'. 'Imagined a greater focus on supporting nature and wildlife' in CAM region. 'My conclusion is that consensus and support is for the economic use of water abstracted from the aquifer (ie through leakage reduction) rather than focusing on schemes that provide some mitigation. Solve the problem at the source rather than apply a sticking plaster later, etc.' - panelist 1 about NERA WTP values (about future customers): 'In previous insight from future customers (in the electricity distribution price review) there was a considerably higher figure even for environmental sustainability initiatives, and so somewhat surprised that there is only a 2p uplift for future consumers compared to the HH average (ie to 10p from 8p). Might be worth drilling into this further to understand where environmental support and sustainability is being picked upon elsewhere (such as reducing water out though leakage).' -panelist 2 elaboration on HH priorities ranking: 'while I'd love it to be 1, in practice there are significant number of people for whom this is a relative low priority. It is important as supports resilience, safety, health and has come higher up the agenda during covid where people are using blue green spaces/the outdoors more. When you have informed discussions and people understand the value of environment becomes a higher priority. '
Risk of temporary usage bans, including hosepipes	-when are these invoked, and what are the criteria to invoke them? -immediate vs longer term impact of TUBs? -perception vs reality of when TUBs have been implemented in the SSW and CAM areas. -just one part of their drought response plan. -TUBs terminology is not common outside the industry. -can be hard for consumers to engage with a risk of 1 in 40 type probabilities -clear info is needed to accept bans -customers seem prepared to expect more bans in the future -what if customers prefer more frequent, shorter restrictions. -Population growth -what mitigation is needed to avoid TUBs being seen to be negatively impactful? What support can SSC provide to home owners, e.g. encouraging use of brown water, using soak aways not drains, providing temporary water harvesting -one put low as numbers impacted are pretty low	-panelist 1 on ODI HH values: 'I had imagined that there would be a expectation for greater levels of compensation for hosepipe bans ... it would be interesting to see the split between SSW and CAM.' -panelist 2 elaboration on HH priorities ranking: 'Put temporary usage ban as lower given willingness to accept value and as had TUBs recently and for many not as bad as thought. Though there may be evidence suggesting this caused real problems for some.'
Installing smart water meters	-how does move to metering impact consumption and water bills? Support available for people starting the transition? -consumers may have a preconception of what a smart meter is, especially if they have an energy smart meter. What are the benefits to customer, SSC, and wider society? -consumers will be less willing to pay for new meters that don't give the granularity of data they expect -mixed feelings as customers would like to see this change in the next 10-15 years. Not as urgent as reducing leakage, disruption, bans… -can be seen as instructive and potentially unfair to vulnerable HHs	-panelist 2 elaboration on HH priorities ranking: 'Considered putting 3rd but in end put 9th - lower than PR19 as half of energy smart meters don't work, confidence in smart meters has declined. Not seen as value for money. Though people do want the data and control and properly smart meters could help with

Attribute	Phase 1 HH insights	Phase 2 HH insights
	-education campaigns are needed to increase acceptability for this innovation -how and by how much, if at all, would this investment reduce customer bills? -not all properties have a water meter, and some can be old technology. Read manually by someone visiting their road, which is expensive and time consuming. Could they be read remotely? More frequent usage info to help people manage their water use and budget more, and identify leaks?	managing water demand and resilience.'
Customer service	-what does customer service mean? -key principles – self serve, first call resolution, notice of interruptions, multi-channel contact, supply issue reporting and update, etc, 24/7 contact etc, and how SSC perform and costs/impacts of variation across the service. -question is about call waiting times, not customer service more generally. Call waiting times are important, but not a proxy for wider service or consumer satisfaction. -in CAM, customers are more demanding than SSW -10 mins seems a long waiting time compared to other service providers, e.g. DNOs 98% of calls answered within 60 seconds.	-panelist 2 disagrees with a point made by a panelist in phase 1 that phone is the preferred channel - its not the preferred channel for everyone, but it's an important channel that needs to be safeguarded so those on low incomes/no credit on phone can still use this channel - after seeing industry comparisons, panelist 2 considered putting customer service as a lower priority, especially for SSW, as its performance is relatively good. But kept it at 4, as most people won't have contacted SSC, but would want to know its good service when they need it -panelist 3 thought this can move further down priority list as SSC are perming better than other companies - panelist 2 elaboration on HH priorities ranking: 'Put customer service - 6th. People expect excellent customer service. Expectations are rising. They know it's possible with possibly relatively low investment and well designed services.'

Delphi: Non-Household Customers

Attribute	Phase 1 NHH insights	Phase 2 NHH insights
Risk of temporary 'do not drink' notice	-'Essential service for HH and NHH respondents'	-
Water quality: 'Issues with tap water colour, taste or smell'	-	-
Water quality: 'Hard water supply'	-	-
Water quality: 'Lead pipes'	-	-
Unplanned interruptions to water supply	-	-panelist 1, general reactions to NERA NHH values: 'No surprise really that continued, uninterrupted supply was key. This is an enabler for most businesses, or a simple hygiene factor.' -panelist 1, reactions to ODI NHH values: 'I had expected that an interruption to supply would over index given the criticality water supply is for most NHHs - either being used for the business' core business, or as an enabler for staff or customers. It is interesting to see how the compensation for a day long interruption is much greater than a part day interruption, although I would imagine that if the part day was during the day to the night then a higher level of compensation would be expected.'
Chance of property flooding from a burst pipe	-does SSC cover the loss of income and the income of employees if a business can't operate?	-
Low water pressure	-	-
Water lost to leakage from pipes	-	-
Supporting nature and wildlife	-	-
Risk of temporary usage bans, including hosepipes	-	-
Installing smart water meters	-	-panelist 2 would put this higher on the list, as they agree with a comment from Phase 1 that said businesses likely to want data on usage to help with managing costs but also to monitor water impacts for those that have to meet sustainability targets/ demonstrate how meeting ESG targets in the case of larger organisations.
Customer service	-	-panelist 2 would put customer service higher up for water dependent businesses in particular, as when things go wrong they would want quick resolution of problems given impact on daily operations.