FROM INSIGHT TO INFLUENCE

SSC LTDS Triangulation

Technical note

Prepared for SSC

Prepared by Impact Research

14 September 2023

Project No. 1345

South Staffs Water

Cambridge Water

Contents

1. Introduction

Objective

South Staffs Water (SSC) commissioned Impact Research to develop a **decision-making framework** for SSC to evidence that its long-term delivery strategy (LTDS) ambition and strategy reflects customers' priorities.

The key focus was on **phasing of investments to deliver on customer priorities** – *when* should investments be made across the 10 ambition areas tested in the LTDS research to unlock the benefits for customers and the environment?

Overall Approach

Initially, our intention was to standardise the results from all the different studies relevant to the LTDS from SSC's research programme and wider sources onto a common 0-100 scale, where the highest value (whether it is expressed as rating scores, allocations of points, estimates of utilities or willingness-to-pay values) is set to 100 and the lowest to zero. The contribution of each study would be calculated in relation to its 'Red-Amber-Green' (RAG) rating of the theoretical, statistical and depth validity.

When asked to assess the proposed approach, Professor Iain Fraser, also the peer reviewer for the PR24 WTP triangulation, gave the following feedback to the original proposal (see embedded report in ANNEX for full details):

- The RAG by data source was fine, as we can always examine how a change in RAG for any one data source impacts on the overall view of the data.
- The standardisation methodology was not at all clear. It attempted to combine closed scale data, real number line data (+/- infinity), percentages etc, all values that have very different meanings that required subjective assumptions about the properties of these scales.
- It was also unclear how we would calculate true confidence intervals for each data point.

After further discussion, it was agreed that an approach similar in principle to the RAG ratings used in the PR24 WTP triangulation should be used to also represent the priorities indicated by each source (see reports '**SSC08 PR24 Technical triangulation – Phase 1 Methodology**' and '**SSC09 PR24 Technical triangulation – Phase 2 Results'**). This would also further aid the consistency of approach across all SSC's triangulation. The priority weightings are therefore all based on expert user interpretations of the sources.

Caveats

The challenge of combining diverse/heterogeneous data types and that subjectivity inherent in the application of user-defined weightings mean that all results from this exercise need to be treated cautiously. The method has been developed to offer practical value as a means of drawing together diverse information in a common format, but its limitations need to be recognised and understood. The user should weigh the outputs against their wider understanding of SSC's ambitions. It is legitimate to change a rating or attribute definition to brings the results closer to expectations, but the user should be careful to show that this has been a conscious judgment on their part and not an independent validation. The value of the tool is therefore in assessing what changes in user ratings might be needed so as to match expectations, as well as observing instances where the results already fall in line with expectations.

An example we have observed relates to removal of lead pipes, where the lack of a strong direct measure requires the user to link lead pipes to water quality when using the customer priorities tracking priority index data. This inevitably leads to a strong value for this issue, despite other sources (LTDS research) showing that it is of relatively lesser importance. In discussions with SCC, the decision was made to 'downgrade' the value of this attribute to the lower end of its value range, as this was considered a fairer reflection of the actual value attached to lead pipes when customers are giving their preferences for investment.

While this use of ratings is clearly a subjective process, it has the advantage of being (a) transparent and (b) a practical way of combining very different data types into a common evaluation framework. Sensitivity testing can also be undertaken to assess the impact of changing the weighting of different data sources.

Future Improvement

Possible improvements that could considered in advance of PR29 are:

- SCC could consider incorporating measures of customer priorities to the 10 ambitions into its customer priorities tracker. This would need to be on a simplified basis compared to the LTDS research undertaken, but would allow for more explicit comparisons of the ambitions with the regular priority ratings within the tracker, as well as against LTDS.
- The use of subjective (but informed) weightings could be strengthened through the convening of panel of experts to come to a consensus view on what these weights should be, both in relation to the use of RAG ratings to verify the sources and the ratings used to represent the different source data on a common scale. This could take the form of a Delphi approach, similar to the method used in SSC's PR24 WTP Triangulation work.

Deliverables

The key output was an Excel-based reporting tool for SSC to consider customer preferences and support decision making in its LTDS, the key components of which are covered in this technical note. This was accompanied by a short internal PowerPoint presentation reporting the key outputs.

2. Approach in Detail

Sources

The following sources were used to provide the basis for the summary outputs:

Table 2.1: Sources

Insight data source	Specific insight data point	Detail of research study / insight	Reference document
SSC Customer priority tracker	Max-diff 0-100 priority rankings	20 priorities. Uses the priority scores from "Priorities household tracker" (derived using Max-diff technique). The Max Diff results are used as the basis for giving each of the 20 attributes a score from 1-10, where 10 is of maximum importance	3410PRE07_YEAR 3 QUANT INSIGHTS_V14.pptm
SSC LTDS research 1	0-100 points allocation exercise	Uses the results from "SSC PR24 LTDS research" which explored the prioritisation of 10 long-term "ambitions" using a points allocation method. The points allocation is used as the basis for giving each attribute a score from 1-10, where 10 is of maximum importance	SSC LTDS PR24 Presentation July 2023 v2.pptx
SSC LTDS research 2	% of customer selecting ambition targets to be met by dates – 2035, 2040, 2045, 2050	Uses the question relating to when respondents want the long-term ambitions to be achieved by – from study "SSC PR24 LTDS research"	
SSC LTDS research 3	Assessing agreement between points allocation and slide preference between investment and keeping bills low.	Uses the results from "SSC PR24 LTDS research". Explores degree of association between priority points allocated to each "ambition", and level of agreement between 2 opposing statements. Results are used to create "modifier values" that are used to focus the final priority ratings towards shorter or longer term goals	
SSC WTP 2022 and SSC valuations from ODI study (Ofwat)	Pull out Willingness to Pay (WTP) and (Willingness to Accept) WTA data points, where they can be mapped	The WTP values are used as the basis for giving each attribute a score from 1-10, where 10 is of maximum importance.	SSC09 PR24 Technical triangulation – Phase 2 Results.docx
SSC WRMP24 themes 1 and 3	Data points for leakage ambition support and achieving drought resilience and achieving environmental destination.	Uses findings from the WRAP's (Water Resources Advisory Panel) and the subsequent quantitative work (n=1,180) Respondents assigned a priority rating (1=High, 2=Med, 3=Low) to areas of potential action. These are then used as the basis for giving each attribute (where available) a score from 1-10, where 10 is of maximum importance.	SSC WRAP Theme 1 Research Findings 16.08.pdf Accent Quant themes 1 and 3 Study - Mar 2022.docx

The Calculations

1.RAG Ratings

The first step was to rate each insight source in terms of theoretical, statistical and depth validity, using the following scale:

Table 2.2: RAG rating scale

Rating	Value
Green	1.00
Green/Amber	0.50
Amber	0.25
Amber/Red	0.10
Red	0.00

The numeric weights are used to weight across the outputs from different sources. While the ratings can be varied by the user of the Excel tool, we used the following:

Table 2.3: Rating of Sources

Data source:	Theoretical	Statistical	Depth
Customer priority tracker	Green/Amber	Green	Green/Amber
LTDS research 1 (Priorities)	Green	Green	Green/Amber
LTDS research 2 (Quant)	Green	Green	Green/Amber
LTDS research 2 (Workshops)	Green	Amber/Red	Green
WTP 2022 and ODI	Amber	Green	Green/Amber
WRMP24 themes 1 and 3	Amber	Amber	Green/Amber

The justification for these weights is given in the annex. This gave a range of weights to use to allow robust sensitivity testing: the mean value across the three 'dimensions' of validity, the minimum and the maximum. For example, LTDS research 2 (Workshops) ranged from a minimum weight of 0.10 (low statistical validity), a mean weight of 0.7 (=[1.00+0.10+1.00]/3) and a maximum weight of 1.0 (high theoretical and depth validity).

2.User Scores

For each source, we took the results it provided for each service attribute / ambition and converted them to a score on a 1-10 scale, where 1 = low priority and 10 = highest priority (ties were allowed). Examples are given in Table 2.4 below.

Table 2.4: Examples of user priority scores

Source: SSC Customer Tracker	Result (Max Diff)	Priority Rating (1-10)
RELIABLE SUPPLY HIGH QUALITY WATER	16.02	10.0
BILL AFFORDABILITY	11.58	7.0
LEAKAGE Reduction	9.66	6.0
LONG-TERM PLANNING FOR WATER SUPPLY	7.26	5.0
PROTECTING WATER RESOURCES	5.34	3.0
FINANCIAL BILL SUPPORT	5.64	4.0
SENDING INCIDENT NOTIFICATIONS	5.58	3.0
PROVIDING ACCURATE AND INFORMATIVE BILLS	3.91	2.0
WATER HARDNESS	3.84	2.0
SERVICE SUPPORT – PSR	3.95	2.0
WATER PRESSURE	3.37	2.0
SUSTAINABILITY	3.22	2.0
QUICK RESOLUTION - EASY TO DEAL WITH	3.36	2.0
WATER EFFICIENCY - SUPPORT/INCENTIVES	3.31	2.0
SCHOOLS EDUCATION	2.04	1.0
SUPPORTING LOCAL COMMUNITY	1.44	1.0
IMPROVE LOCAL ENVIRONMENT	3.31	2.0
WATER RECYCLING / RE-USE	3.27	2.0
WIDE RANGE OF WAYS TO CONTACT	2.02	1.0
MORE REGULAR METER READINGS	1.88	1.0

Source: LTDS 1	Result (Max Diff)	Priority Rating (1-10)
Drought Resilience	5.90	6.0
WINEP	3.85	4.0
Achieving Net Zero Carbon	3.60	4.0
Leakage Reduction	9.18	9.0
Lead Pipe Removal	6.78	7.0
Reducing how much water we use at home and work	5.51	6.0
Reducing Supply Interruptions	4.93	5.0
Offering better and smarter customer service	3.66	4.0
Improving Water Quality	10.00	10.0
Tackling Water Poverty	6.93	7.0

NB: items in grey were not considered relevant to the LTDS ambitions

Table 2.4 also illustrates the issue of having different service attributes / ambitions measured by different sources. The 10 ambitions, tested directly in the LTDS study, have some representation in other sources (such as the Customer Priorities Tracker shown in the same table), but a way of mapping these was required. Table 2.5 shows how this this was done for the customer tracker. Each cell containing a value was used to weight the scores across the customer tracker attributes in each row to calculate an average value representing the corresponding ambition in each relevant column. A separate table was used for mapping the WTP/WTA measures onto the ambitions.

Table 2.5: Comparison table

		Source: LTDS 1									
		Drought Resilience	WINEP	Achieving Net Zero Carbon	Leakage Reduction	Lead Pipe Removal	Reducing how much water we use at home and at work	Reducing Supply Interrup-tions	Offering better and smarter customer service	Improving Water Quality	Tackling Water Poverty
Source: SSC Customer Tracker	RELIABLE SUPPLY HIGH QUALITY WATER					1		1		1	
	BILL AFFORDABILITY										1
	LEAKAGE Reduction	1			1						
	LONG-TERM PLANNING FOR WATER SUPPLY	1									
	PROTECTING WATER RESOURCES	1	1								
	FINANCIAL BILL SUPPORT								1		1
	PROVIDING ACCURATE AND INFORMATIVE BILLS								1		
	WATER HARDNESS									1	
	SERVICE SUPPORT – PSR								1		
	WATER PRESSURE							1			
	SUSTAINABILITY			1							
	QUICK RESOLUTION - EASY TO DEAL WITH								1		
	WATER EFFICIENCY - SUPPORT/INCENTIVES						1				1
	SCHOOLS EDUCATION								1		
	IMPROVE LOCAL ENVIRONMENT		1								
	WATER RECYCLING / RE-USE	1					1				
	MORE REGULAR METER READINGS						1		1		
	WIDE RANGE OF WAYS TO CONTACT								1		

3. Calculations

An illustrative example for a single ambition calculated for a single time (2035) is give in Figure 2.1 below. In step A, the RAG is defined for each source[1]; in step B the customer priority scores for the ambition are shown. These are then combined in step A x B to produce a final blended score in step C, together with its standard deviation.

Step D then introduces a time element, a result from the main SSC LTDS study that indicates the proportion of customers who would want the ambition delivered by 2035 (in this example, 50%). This takes the overall priority score and uses these % results to 'share' it over the different time horizons (2035, 2040, etc). Finally, in step E, a further result from the LTDS research is applied, representing the short v long term nature of the priority. This is based on questions in the LTDS study that ask whether customers would prioritise reductions in bills over investment in a given ambition.

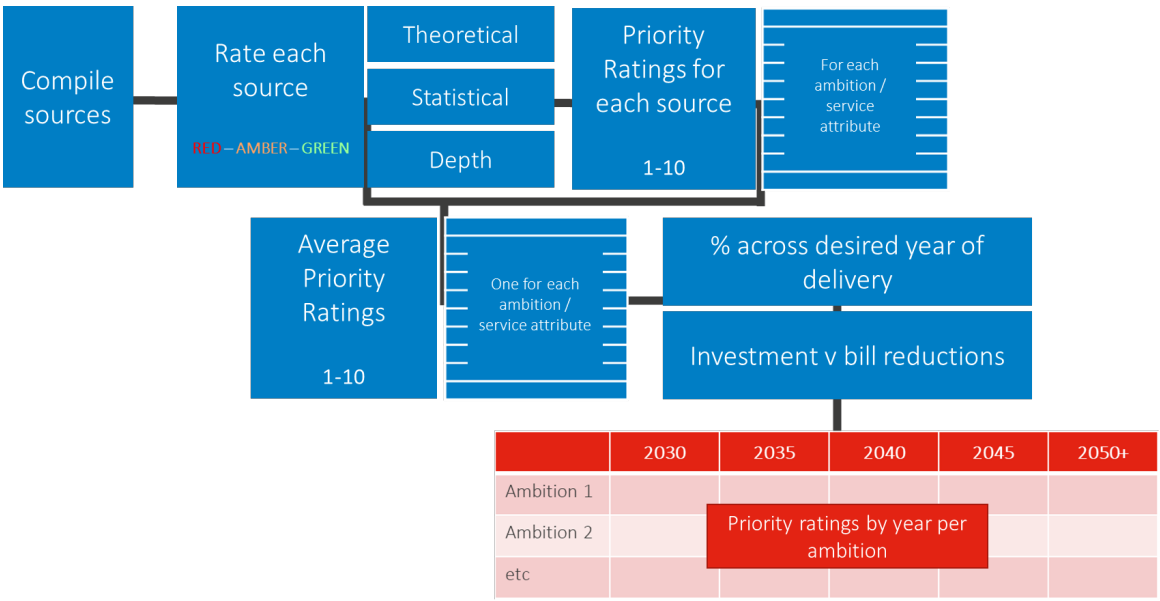
Figure 2.1: Illustrative example of calculations

		A		B	A x B		C
	Data source	RAG	Wgt	Priority Rating (user defined, 1-10)	Calculations		Score for 2035
	Customer priority tracker	Amber	0.5	7.0	0.5 x 7.0 = 3.5		
	LTDS research 1	Green	1.0	5.0	1 x 5.0 = 5.0	(3.5 + 5.0 +3.5 +0.6) / (0.5 + 1.0 +0.5 +0.1)	
	WTP 2022 and ODI	Amber	0.5	7.0	0.5 x 7.0 = 3.5	= 12.6 / 2.1	6.0 *(5.0 – 7.0)*
	WRMP24 themes 1 and 3 quant	Amber / Red	0.1	6.0	0.1 x 6.0 = 0.6	= 6.0 (range 5.0 – 7.0)	
D	LTDS research 2	% Ambition 2035		50%		X 0.5	3.0 *(2.5 – 3.5)*
E	LTDS research 3	Invest v bills		-10%		-10%	2.7 *(2.25 – 3.15)*

Final priority score for 2035

Schematic:

[1] In reality, this would be three sets of values for which an average weight is calculated, but for illustration we show a single set of ratings.

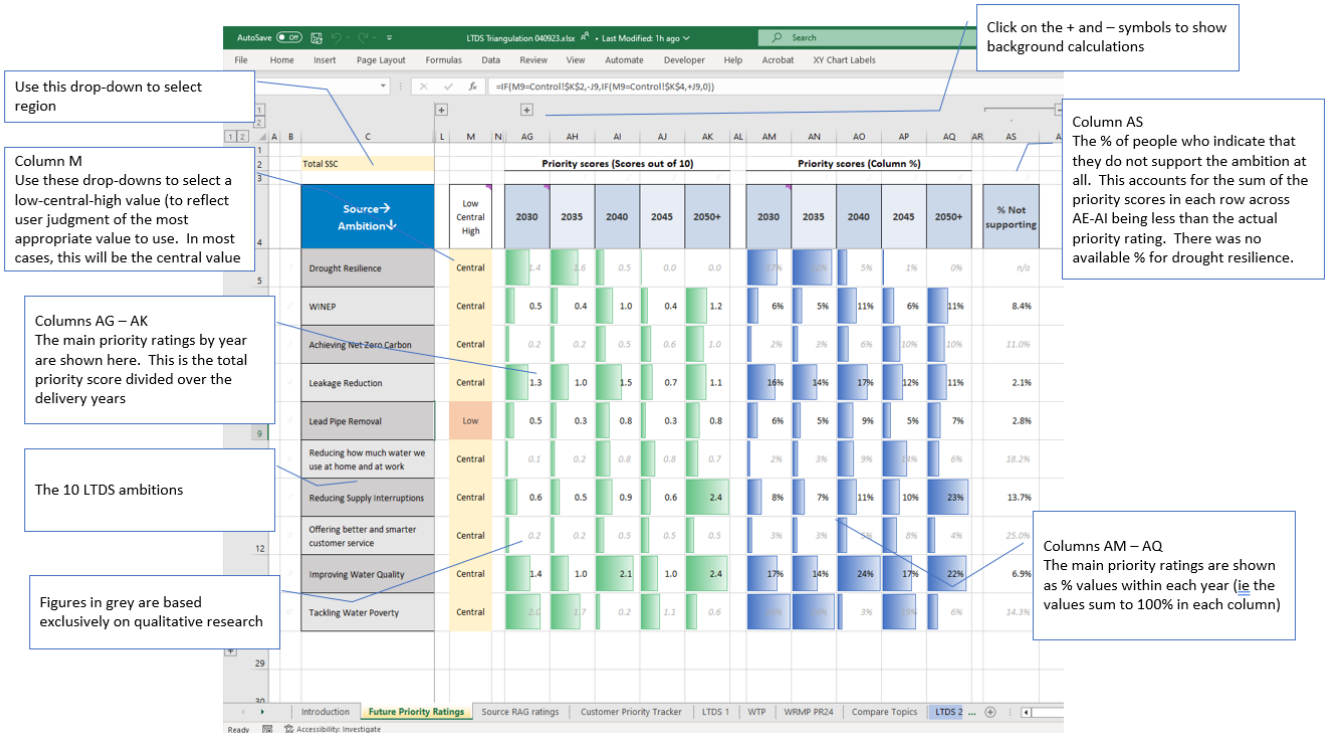

Compile sources
Rate each source
RED – AMBER – GREEN
Theoretical
Statistical
Depth
Priority Ratings for each source
1-10
For each ambition / service attribute
Average Priority Ratings
1-10
One for each ambition / service attribute
% across desired year of delivery
Investment v bill reductions
2030
2035
2040
2045
2050+
Ambition 1
Ambition 2
etc
Priority ratings by year per ambition

3. Outputs

Main Outputs

The main outputs from the tool are shown in Figure 3.1, where the priority ratings are shown across the delivery years 2030 – 2050+. This provides SSC with a decision making framework for assessing how customers would prefer investments phased to 2050 to deliver ambitions in its LTDS. All the information is based on HH customers, the only sub-division available is by region.

Figure 3.1: Main outputs ('Future Priority Ratings' sheet)



Sensitivity Tests

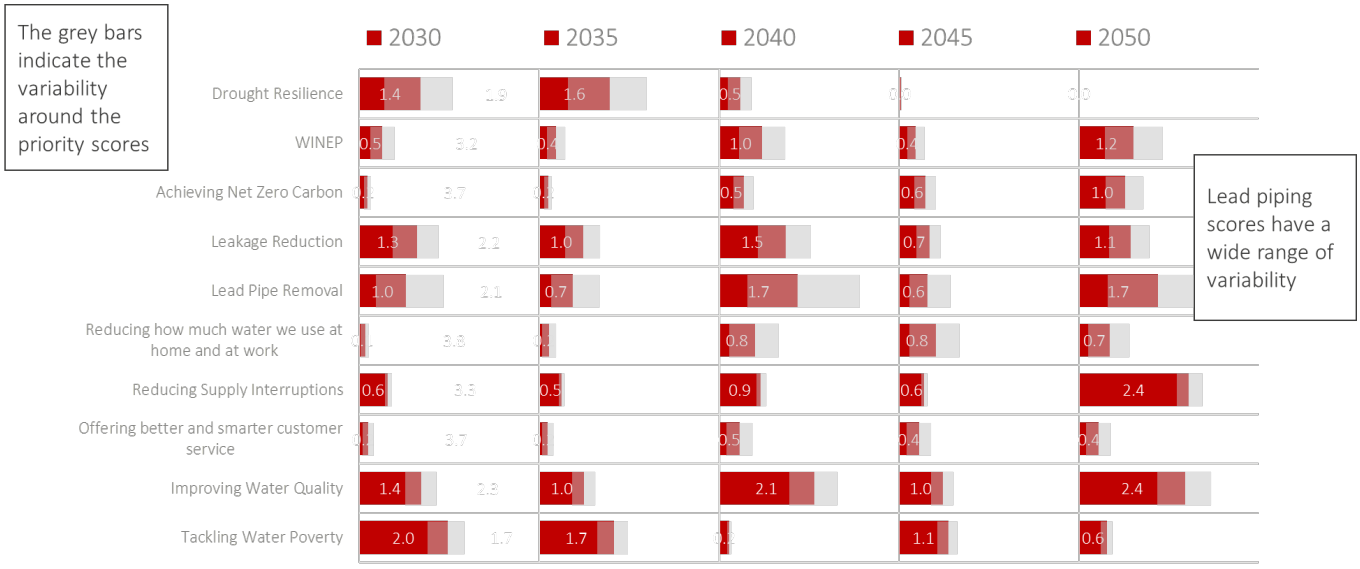
Table 3.1 shows the weighted standard deviation of the average scores from across the different source values available for each ambition.

Table 3.1: Weighted mean average scores and standard deviations of the estimates

Ambition	Weighted Mean	Standard Deviation
Drought Resilience	*4.0*	*2.0*
WINEP	*4.0*	*2.1*
Achieving Net Zero Carbon	*3.6*	*1.4*
Leakage Reduction	*6.2*	*2.3*
Lead Pipe Removal	*6.3*	*3.5*
Reducing how much water we use at home and at work	*3.3*	*2.1*
Reducing Supply Interruptions	*5.7*	*0.6*
Offering better and smarter customer service	*2.8*	*1.5*
Improving Water Quality	*8.0*	*1.9*
Tackling Water Poverty	*5.8*	*1.2*

Figure 3.2 indicates how these standard deviations can be used to indicate the range of variability around each reported value for each future year.

Figure 3.2: Error ranges around the average (central) values



Grey bars indicate the variation around the main results, driven by the standard deviations of the scores and alternative RAG ratings
Results for 2030 are extrapolated from 2035-2050 results, except for Drought resilience and Net Zero

NHH and Future Customers

The analysis is driven by the priorities of Household (HH) customers, because these comprise the bulk of the LTDS survey with sufficient sample sizes to support the priority (drivers) analysis. In Table 3.2 below, we draw out the qualitative differences for other important groups, Non-Household (NHH) customers and Future (FC) customers:

Table 3.2: Variations for NHH and FC customers (relative to HH)

Ambitions (in rank order for HH)	**NHH**	**FC**
Improving Water Quality		*Want delivery sooner (all by 2040)*
Lead Pipe Removal		A lower priority, *but those who do consider it want delivery sooner (all by 2040)*
Leakage Reduction		
Tackling Water Poverty		
Reducing Supply Interruptions		
Drought Resilience		
WINEP	*Slightly less urgent*	*Less urgent*
Achieving Net Zero Carbon	A higher priority (ranked third highest), but *slightly less urgent in terms of when delivered*	A higher priority, second only to improving water quality
Reducing how much water we use at home and work		
Offering better and smarter customer service		*Want delivery a little sooner*

4. Annex

Table 4.1: Summary of Peer Reviewer Correspondence on LTDS triangulation

Date	Peer Review Comment	Impact Response
31/07/23	I can see the reason for the research. My concern about the proposed method is that you have significantly differing data types and somehow you want to "aggregate" this information. • The RAG by data source is fine - and you can always examine how a change in RAG for anyone data sources impacts the overall view of the data. • However, the standardisation methodology doesn't make sense to me. You have closed scale data, real number line data (+/- infinity), percentages etc. You then take a subjective weight and multiply against values that have very different meanings - so when you add the resulting values, at least to me you are adding "apples and oranges" - you're adding a monetary amount to closed scale values...also the multiplicative part of the formula...why? • You also mention prior to the example slide that you will calculate confidence intervals - it isn't obvious to me how you can do this. I fully understand the end goal, but I really don't like what is currently being proposed!	Our initial idea was to have a way to represent each set of results onto a common scale (eg a 'share' out of 100 points) and then combine them with a suitable RAG weight to produce a cumulative or average score that combined priorities with respect to different time horizons. So perhaps it's a case of taking some steps back and asking first whether you think there could be a more credible way to combine such different types of information? It's an extension of the RAG approach used in the triangulation work, where we used a (subjective) series of weights to combine very different results from a variety of sources, where even the measures of monetary values were very different (WTP v WTA, budget allocation v Discrete choice) and non-monetary values were also adapted to fit into the same framework (eg the PJM method that applied their max diff results and the SSC tracker Shapley regression outputs to the WTP results).
03/08/23	*A revised proposal was developed in response to the feedback*	
03/08/23	Happy with slide 4 content. So, on slide the calculations are still somewhat ad hoc but they at least make clear how the information is being used and the final "score" is being arrived at. Slide 6 - somewhat odd that you are using the weights twice in the calculation - so akin to double counting - would seem to make more sense to simply divide the weighted score achieved by the maximum attainable which in your example would be 40. 12.4/40 = 0.31 (These values must lie between zero and one. The range I assume occurs as you change the weights and/or priority ratings? (Not clear) You might ask senior people to input their own weights/RAG and priority scores and let the variation generate the "range". This could become an interactive exercise - possibly.	Regarding the calculations on slide 6, what we were aiming for was a weighted average priority rating. The RAG weighting (how reliable/relevant the source is) and the priority rating (a user score given for each source, indicating how important that attribute is to consumers) are independent from one another. Your suggestion of basing the result on the total of 40 would imply that each source has an equal RAG weighting. The suggested approach in slide 6 means that, for example, the LTDS Research 1 source with its RAG weighting of 1.0 has twice as much influence on the average priority rating as, say, the Customer Priorities Tracker (RAG weighting 0.5) and 10 times as much influence as WRMP24 (RAG = 0.1). The final weighted rating value could be expressed as 0.6 on a scale of 0-1 (=6/10, where 10 is the highest possible average weighted rating, if all four sources had a rating of 10 each). Changing the RAG weights and priority scores would both impact on the average rating. Your suggestion of using this as a basis for drawing together different viewpoints is helpful – I'll suggest that in the proposal.
03/08/23	Ok, so the RAG weighting is a between measure of belief about plausibility of the data source/results - but you are also using it as a within measure as well - it does not really matter whether you report 12.6 or 6 - it just requires you to define the max/min of the scale - you can calculate the % contribution of each data source by dividing by 12.6 and multiplying by 100'. The formula you are using is - score = (sum (weight*priority)) / sum(weights) Basically, the calculations are ad hoc, it is just can they help reveal something useful for the user - the specific "number" range/scale may matter less.	Yes, the tool is purely a device for embodying the user's beliefs about the robustness of the sources and what they are saying. Transparency around the assumptions and sensitivity testing will therefore be important when we report the outcomes.
21/08/23	*The final internal report and Excel tool were sent to the Peer Reviewer*	
22/08/23	The approach as it is explained and the worked example all look ok. I have one observation - it is over the use of the word "defendable" - ppt slide three last bullet point - not sure I like it or what it implies. I would suggest describing the method as practical (given data limitations) but that the results generated need to be treated with a degree of caution.	

Table 4.1: Summary of Peer Reviewer Correspondence on LTDS triangulation

Date	Peer Review Comment	Impact Response
	To that effect, I would also include a caveat around the inherent difficulty of combining diverse/heterogeneous data types and that any results need to be treated cautiously. Basically, I'd keep on reminding the user that the method has practical value but that its limitations do need to be recognised and understood.	

Table 4.2: Rating of Sources

Data source:	Theoretical	Statistical	Depth
Customer priority tracker	Green/Amber	**Green**	Green/Amber
LTDS research 1 (Priorities)	**Green**	**Green**	Green/Amber
LTDS research 2 (Quant)	**Green**	**Green**	Green/Amber
LTDS research 2 (Workshops)	**Green**	Amber/Red	**Green**
WTP 2022 and ODI	**Amber**	**Green**	Green/Amber
WRMP24 themes 1 and 3	**Amber**	**Amber**	Green/Amber

Table 4.3a: Customer Priority Tracker

Description	Validity	Criteria	Comment	RAG Rating
Regular tracking research that quantifies customer priorities through a Max Diff approach	Theoretical	Are definitions of candidate and target measure the same?	The ambitions are not directly measured but they are largely represented by a broad range of measures	Green/Amber
		Are contextual conditions the same between candidate and target measures?	Customers are informed about each measure to a reasonable degree	
		If no to either of these, what issues do the differences give rise to?	-	
	Statistical	How large is the sample?	Statistically robust samples are covered for households in all waves of the research	Green
		How representative are the sample / timings?	Quotas and the subsequent weighting of data to Census profiles ensured a representative profile of customers.	
		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?	Max Diff is a well established method for measuring priorities	
	Depth	Extent of explorative and developmental work?	The survey design drew on extensive qual research and other sources	Green/Amber

Table 4.3b: LTDS Research - Priorities

Description	Validity	Criteria	Comment	RAG Rating
Regular tracking research that quantifies customer priorities towards LTDS ambitions	Theoretical	Are definitions of candidate and target measure the same?	The ambitions are directly measured	Green
		Are contextual conditions the same between candidate and target measures?	Customers are informed about each measure to a reasonable degree	
		If no to either of these, what issues do the differences give rise to?	-	
	Statistical	How large is the sample?	Statistically robust samples are covered for households	Green
		How representative are the sample / timings?	Quotas and the subsequent weighting of data to Census profiles ensured a representative profile of customers.	
		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?	Points allocation is a well-established method for expressing priorities, though Max Diff is a stronger approach	
	Depth	Extent of explorative and developmental work?	The survey design drew on extensive qual research and other sources	Green/Amber

Table 4.3c: LTDS Research - Quant

Description	Validity	Criteria	Comment	RAG Rating
Regular tracking research that quantifies customer priorities towards LTDS ambitions	Theoretical	Are definitions of candidate and target measure the same?	The ambitions are directly measured. Customers are requested to specify preferred year of delivering each ambition	Green
		Are contextual conditions the same between candidate and target measures?	Customers are informed about each measure to a reasonable degree	
		If no to either of these, what issues do the differences give rise to?	-	
	Statistical	How large is the sample?	Statistically robust samples are covered for households	Green
		How representative are the sample / timings?	Quotas and the subsequent weighting of data to Census profiles ensured a representative profile of customers.	
		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?	Points allocation is a well-established method for expressing priorities, though Max Diff is a stronger approach	
	Depth	Extent of explorative and developmental work?	The survey design drew on extensive qual research and other sources	Green/Amber

Table 4.3d: LTDS Research - Workshops

Description	Validity	Criteria	Comment	RAG Rating
Regular tracking research that quantifies customer priorities towards LTDS ambitions	Theoretical	Are definitions of candidate and target measure the same?	The ambitions are directly measured. Customers are requested to specify preferred year of delivering each ambition	Green
		Are contextual conditions the same between candidate and target measures?	Customers are informed about each measure to a reasonable degree	
		If no to either of these, what issues do the differences give rise to?	-	
	Statistical	How large is the sample?	Qualitative research – small groups of customers	Amber/Red
		How representative are the sample / timings?	Small numbers of groups representing regional and demographic variations	
		How wide are the confidence intervals?	-	
		Have the results been derived using best practice techniques?	Experienced moderators convened the groups	
	Depth	Extent of explorative and developmental work?	Strong insights gained from the qualitative format	Green

Table 4.3e: WTP 2022 and ODI

Description	Validity	Criteria	Comment	RAG Rating
Comprehensive triangulation of PR24 and earlier willingness-to-pay / accept research	Theoretical	Are definitions of candidate and target measure the same?	The ambitions are not directly measured but they are to various degrees represented by a broad range of measures	Amber
		Are contextual conditions the same between candidate and target measures?	Customers are informed about each measure to a reasonable degree	
		If no to either of these, what issues do the differences give rise to?	-	
	Statistical	How large is the sample?	Statistically robust samples are covered for households and non-households in all waves of the research	Green
		How representative are the sample / timings?	Quotas and the subsequent weighting of data to Census profiles ensured a representative profile of customers.	
		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?	Application of the RAG method adopted in PR19 and ratified by external assessors	
	Depth	Extent of explorative and developmental work?	The range of sources is broad and dependent on the decisions made when constructing the RAG approach	Green/Amber

Table 4.3f: WRMP24 themes 1 and 3

Description	Validity	Criteria	Comment	RAG Rating
Regular tracking research that quantifies customer priorities towards LTDS ambitions	Theoretical	Are definitions of candidate and target measure the same?	The ambitions are not directly measured.	Amber
		Are contextual conditions the same between candidate and target measures?	Customers are informed about each measure to a reasonable degree	
		If no to either of these, what issues do the differences give rise to?	-	
	Statistical	How large is the sample?	Statistically robust samples are covered for households and non-households in all waves of the research	Amber
		How representative are the sample / timings?	Quotas and the subsequent weighting of data to Census profiles ensured a representative profile of customers.	
		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?	Simple indicators of priority (rankings represented as average scores	
	Depth	Extent of explorative and developmental work?	Useful for covering items that are most weakly represented in the LTDS work – net zero and drought resilience	Green/Amber