BR-MeX Pilot Report

Ofwat / MOSL

10th April 2024

Contents

IFF Research

1 Key findings and recommendations

Introduction

The Water Services Regulation Authority (Ofwat)'s monitoring of the business retail market and its review of incumbent support for effective markets found that wholesalers need to improve their support for the non-domestic market. In particular, it was found wholesalers need a stronger focus on understanding the needs and requirements of their business customers and to tailor their wholesale service offerings accordingly. This led to calls for there to be a Business Customer Measure of Experience (B-MeX), to measure the quality of services delivered by wholesalers to business customers in the business retail market. The Retailer Measure of Experience (R-MeX) survey has been run by the market operator (MOSL) since 2018, and provides a qualitative and quantitative measurement of how water retailers perceive the service of the wholesalers that they work with. Ofwat and MOSL intend to use the findings of both surveys to better incentivise wholesalers to provide a good level of service to their business customers and water retailers.

The aim of this study was to design the B-MeX survey, and test and pilot all aspects of the B-MeX and R-MeX surveys to ensure the survey methodologies are robust for a shadow run in 2024, and full implementation from April 2025. Below we summarise the key findings and recommendations from the study.

B-MeX

Sample quality

- **A substantial proportion of sample records provided were of poor quality**: these records did not contain sufficient information to allow interviewers to reach the correct contact, and to help customers recognise the specific contact in question. Furthermore, a number of records were found to be ineligible once interviewers spoke to the respondent (e.g. passers-by, non-business contacts).
- **Discrepancies in the coding of sample data (particularly 'reason for contact') limited the options for setting quota targets and weighting the final survey data**; this could lead to the survey data not being representative of all customers who had contact with their wholesaler in the given period, which could skew the final scores.
- **Sample quality issues and discrepancies also required substantial cleaning and processing prior to fieldwork**; this leads to longer gaps between contacts taking place and customers being contacted for their feedback, which could affect their levels of recall; it would also reduce the efficiency of conducting fieldwork over a longer period.

Recommendations

- Wholesalers **must ensure full details are recorded for every contact**, and supplied to the survey agency (e.g. information such as customer name, business name and reason for contact cannot be left blank)
- Wholesalers must **exclude ineligible records** from their sample file, to avoid interviewers wasting time calling and screening contacts that are out of scope for the survey.
- Wholesalers must **use a consistent set of categories** to record 'reason for contact.'

Sample delivery frequency

- Sample for the pilot was based on a months' worth of contacts from each wholesaler; during pilot fieldwork, **respondents were generally able to recall the contact** they had had (either spontaneously or after prompting by the interviewer), **suggesting that a gap of between 1 and 2 months between contact and surveying is appropriate**.
- More frequent sample deliveries (e.g. weekly or fortnightly) would be more burdensome for wholesalers and the survey agency in terms of collating, delivering and preparing the sample; whereas a lower frequency would risk the gap between the contact and the interview being too long for customers to give meaningful feedback.

Recommendations

- **Sample should be drawn and delivered on a monthly basis by wholesalers**, meaning customers will be surveyed between one and two months after their contact.

Fieldwork methodology

- Some customers were initially unclear, or not confident, in differentiating between their wholesaler and their retailer; **telephone interviewers were able to read out definitions to explain the role of their wholesaler vs. retailer**, which helped ensure respondents were giving relevant feedback.
- Customers could also have difficulty recalling the exact contact they had with their wholesaler, and i**nterviewers played a valuable role prompting respondents with details** from the sample file (e.g. dates, reason for contact) to help them remember.
- **Telephone surveys are able to achieve higher response rates than online surveys**, and this will be essential to ensure that sufficient base sizes for robust analysis are achieved across all wholesalers, including those which much smaller contact populations to draw from.

Recommendations

- **The survey should be conducted by telephone** – telephone interviewers are able to ensure respondents understand the difference between their retailer and wholesaler, and prompt them to remember the specific contact we want them to give feedback on; an online survey would lack this live quality assurance and there would be a high risk of respondents skipping longer blocks of text, and therefore misunderstanding the survey / giving feedback relating to their retailer or other contacts/incidents. **A telephone methodology will also maximise response rates**, which will be essential to ensure a robust base size for all wholesalers.

Quota targets and weighting

- As the wholesalers vary substantially in size, **if the sample is allowed to fall out naturally, only a very small number of interviews would be completed with customers of the smaller wholesalers**, meaning their final scores would be less robust due to being based on only a small number of customers. To avoid this, quotas were set to cap the number of interviews with very large wholesalers and ensure that sufficient completes are achieved with customers of small wholesalers to allow robust analysis.

- In order to ensure final scores are representative of all contacts undertaken by each wholesaler within each month, **quotas within wholesaler can be set during fieldwork to ensure that the number of completed interviews are in line with the population**; for example, if 20% of a wholesalers contacts were indirect, we would set a quota to aim for 20% of interviews for that wholesaler to be with indirect sample. This approach was tested during the pilot and its effectiveness assessed during the weighting process.
- **Weighting is a technique used to adjust the results of a survey to make it more representative of the larger population** (in this case, the population is all contacts undertaken by each wholesaler). As it is unlikely that all quotas set during fieldwork will be met exactly, weighting can be used to adjust the survey profile to match the population. **A potential weighting strategy explored using the pilot data found that the quota management was effective**, and the achieved interviews closely matched the target population; therefore, the **weighting efficiencies for each wholesaler were high**, meaning only minor adjustments were needed.

Recommendations

- **To ensure sufficient interviews per wholesaler for robust analysis,** a target of 200 interviews per quarter / 800 interviews per year should be set for each wholesaler. Where wholesalers do not have sufficient sample to achieve this, a census approach should be taken (i.e. all sample will be loaded for fieldwork, in order to achieve as many interviews as possible).
- **Quota targets should be set during fieldwork** to ensure that the completed interviews, and therefore scores, are representative of the full population of business customer contacts for each wholesaler.
- **Final data should be weighted to population targets within each wholesaler.**

Questionnaire

- Respondents found the questionnaire clear and easy to understand, therefore we do not recommend any changes to the main body of the questionnaire.
- Some respondents, particularly those from the indirect sample, were not confident in differentiating their wholesaler and retailer, and in speaking about a specific contact.

Recommendations

- **Include additional information and prompts at the start of the survey** to help respondents differentiate between their wholesaler and retailer and identify the specific contact in question – particularly for indirect contacts.
- **Review answer codes for question A3 on how contact with the wholesaler came about** to tailor the options of direct vs. indirect sample.

R-MeX

Survey methodology

- In the cognitive and follow-up qualitative interviews, retailers expressed satisfaction with the methodology of the current R-MeX survey: most felt the Excel format and the current frequency of surveys works well.
- Responses were mixed in terms of possible options for weighting the data, and many were unsure; concerns expressed included that wholesalers could be incentivised to focus their efforts on areas that receive a higher weighting, or on relationships with retailers with a larger weighting.

Recommendations

- **Maintain the current survey methodology and frequency.**
- **Do not apply any weighting to the data**, as this could have the effect of artificially inflating certain scores and could encourage wholesalers to focus their efforts disproportionately.

Questionnaire

- In the cognitive stage retailers were positive about the questionnaire and therefore only minor changes were suggested for the pilot:
  - Some retailers were concerned that the satisfaction scale was not interpreted consistently across retailers, therefore labels were added to the top, bottom and middle points on the scale to provide additional clarification.
  - Retailers found the Excel spreadsheet with macros cumbersome to use, particularly when feeding back against multiple wholesalers. For the pilot, a new version of the spreadsheet was created without macros.
- Feedback from retailers following the pilot suggested the changes had worked well. They felt the scoring guidance gave them more confidence in scores being consistent across retailers, and the new spreadsheet was found to be easier and quicker to use.

Recommendations

- **Maintain changes from the pilot,** keeping the scoring guidance and revised spreadsheet without macros.

Analysis of scoring

- Average scores from the pilot broadly aligned with trends seen across previous waves, suggesting the questionnaire changes did not cause a marked change in the way retailers scored wholesalers. Rating scales can, if not carefully moderated, elicit difference response styles from different individuals. Over large samples, these differences, would tend to 'even out' naturally; however, as R-MeX only has a very small sample size, such differences, if present, could have a bigger effect on wholesalers' final scores.
- Statistical analysis was carried out on the pilot data to check the robustness of the final scores. This indicated that at least some variance in the scores given by different retailers

could be driven by such difference or variations in how individual retailers interpreted the scale (rather than simply being due to differences in wholesaler performance).

- Two different options for alternative scoring systems were considered, adopting a less granular scale, or ‘centering’ the existing raw overall satisfaction scores to neutralise the impact of any inconsistencies in retailer response style. The latter would allow the existing 0-10 scale to be retained, while facilitating a fair comparison of the ‘relative’ ratings of wholesalers.

Recommendations

- **Use the ‘centering’ scoring adjustment,** to adjust overall satisfaction scores for each wholesaler before feeding into the BR-MeX combined score.
- **Continue to feed back raw overall scores to wholesalers for reference**, in addition to the raw scores given for individual service elements.

2 Introduction

Background

The Water Services Regulation Authority (Ofwat) oversees the regulation of water and wastewater in England and Wales. Their remit includes ensuring water companies provide the best service to customers and communities, improve the environment, and make sure our water supplies are secure for future generations.

In April 2017, the water services market was opened to enable choice for non-household i.e. business customers, through the creation of the business retail market. Business customers in this market are able to choose their retailer for the provision of water and wastewater services.

It remains the case in this market that business customers' water supply is provided to them by a wholesaler. That is, each wholesaler remains responsible for maintaining water and wastewater pipe networks across its region, managing the supply of water to business customers and maintaining water meters, and collecting, treating and returning wastewater to the environment. Wholesalers also provide a number of other services to retailers and by extension to business customers.

Retailers are responsible for customer-facing activities, such as managing bills and payments, reading water meters, and customer service and water saving advice.

The market operator - Market Operator Services Limited (MOSL) - was established in 2017 and manages the business retail market in England. MOSL run a number of features of the market, including administration of settlement between wholesalers and retailers, and are guardians of the industry Codes that govern the market.

Background to B-MeX

Wholesalers play a key role in facilitating the delivery of a good business customer experience. However, Ofwat's monitoring of the business retail market[1] and its review of incumbent support for effective markets[2] found that wholesalers need to improve their support for the non-domestic market. In particular, Ofwat's findings included that wholesalers need a stronger focus on understanding the needs and requirements of their business customers and to tailor their wholesale service offerings accordingly.

Following this, there have been calls from the water industry for there to be a Business Customer Measure of Experience (B-MeX), to measure the quality of services delivered by wholesalers to business customers in the business retail market, which would be used to incentivise wholesalers to provide a good level of service to their business customers. Work has already been conducted into the concept and feasibility of a B-MeX incentive[3], and a working group was set up to develop a model for a B-MeX survey[4]. Ofwat commissioned IFF Research to further develop and conduct a pilot of the B-MeX survey, for full implementation from Spring 2025.

---

[1] State of the market report 2021-2022
[2] Project RISE
[3] The introduction of B-MeX into the Business Retail Market
[4] B-MeX Working Group

Background to R-MeX

R-MeX is a survey that provides qualitative and quantitative measurement of how water retailers perceive the service of the wholesalers that they work with and has run twice a year since August 2018. Again, Ofwat commissioned IFF Research to further develop the R-MeX survey questions and conduct a pilot, for full implementation from Spring 2025.

Aims and Objectives

The ultimate objective of this research was for IFF Research to design, test and pilot all aspects of the B-MeX and R-MeX surveys to ensure that the survey methodology is robust for a shadow run in 2024, and full implementation from April 2025. Within this remit, IFF Research were required to:

- Design a pilot methodology for B-MeX and R-MeX surveys.
- Conduct fieldwork to pilot and test both surveys and reflect on findings.
- Analyse findings and produce recommendations and guidance for a shadow year of B-MeX and R-MeX fieldwork in 2024.

These activities will provide a base to inform Ofwat's decisions for the BR-MeX final incentive design.

With regards to analysing findings and producing recommendations for B-MeX, IFF Research were asked to consider the following:

- The most appropriate way to source customer contact information based on direct customer contacts with wholesalers and indirect customer contacts made via transactions with the bilateral transactions hub[5].
- The most appropriate contact channel(s) for the survey considering the accuracy of customer responses, response rates and value for money.
- Sample design, including sample sizes, quotas, survey frequencies and sampling approach, so that the B-MeX survey is as statistically robust as possible.
- Questionnaire design, ensuring that business customers are as clear as possible on what feedback they should provide.

With regards to analysing findings and producing recommendations for R-MeX, IFF Research were asked to consider the following:

- The most appropriate channel(s) for the Retailers surveys which would obtain the most accurate and timely response.
- Sample design, survey frequencies and sampling approach so that the R-MeX survey is as statistically robust as possible.

[5] The bi-laterals transactions hub is a centralised mechanism enabling retailers and wholesalers to initiate and manage retailer requests for services from wholesalers, including on behalf of end customers, such as fixing a water meter.

- Recommendations on methods to ensure that retailers provide the most objective feedback.
- Questionnaire design, including amends to existing questions and length of survey if needed.

Report structure

The report structure is as follows:

- Chapters 3-5 focus on work around the B-MeX survey:
  - Chapter 3 outlines the methodology for work around the B-MeX survey, including insights on how the scoping stage, cognitive testing, the B-MeX pilot, and follow-up qualitative interviews were conducted;
  - Chapter 4 focuses on findings from the B-MeX pilot. As well as outlining survey results, it also provides insight on sample quality, a breakdown of who completed the survey, and an outline of survey outcomes (i.e. completes, refusals, and ineligible responses). It also provides insights from qualitative follow-up interviews;
  - Chapter 5 focuses on considerations for future iterations of the B-MeX survey, including thoughts on methodology, sample design, amends to the questionnaire, and where and how the survey aligns and differs from the already existing C-MeX survey[6] for domestic customers.
- Chapters 6-8 focus on work around the R-MeX survey
  - Chapter 6 outlines the methodology for work around the R-MeX survey, including insights on how the scoping stage, cognitive testing, the B-MeX pilot, and follow-up qualitative interviews were conducted;
  - Chapter 7 focuses on findings from the R-MeX pilot. As well as outlining survey results, it provides insight on whether any changes made to the R-MeX survey as part of the pilot may have impacted results, as well as insights from qualitative follow-up interviews;
  - Chapter 8 focuses on considerations for future iterations of the R-MeX survey, including thoughts on methodology, sample design, any amends needed to the questionnaire, and where and how the survey aligns with C-MeX;
- Chapter 9 is a conclusion, where findings will be summarised and recommendations for future iterations of B-MeX and R-MeX highlighted.

[6] Customer and developer services experience - Ofwat

3 B-MeX: Methodology

This chapter outlines the methodology for work around the B-MeX survey, including insights on how the scoping stage, cognitive testing, the B-MeX pilot, and follow-up qualitative interviews were conducted.

Scoping stage

As a starting point, a scoping phase was held for B-MeX, to understand more about who should be eligible to take part in the survey, what data was available to work with and what timescales would be suitable for the pilot.

The scoping phase started with an inception meeting with Ofwat, where some initial queries were raised by IFF, such as who would be eligible to take part in the B-MeX survey, and what types of contact, if any, would need to be excluded. It was discussed that contacts, rather than businesses, would likely be the main unit of analysis for B-MeX.

Other queries required the input of water wholesalers. Table 3.1 outlines the questions that IFF Research raised to wholesalers as part of the scoping process, and the reasons for why these questions needed to be answered.

Table 3.1 Research questions.

Key question	Why was this important?
What was the size of the eligible population by wholesaler?	IFF needed to know the eligible population to be able to: • Determine how many interviews could be achieved as part of the pilot by each wholesaler, and; • Make recommendations for the size of sample required for the mainstage for it to be considered statistically robust. For the pilot itself, IFF needed to know the eligible population by individual wholesaler to determine if and how to set quotas to ensure that the survey is robust but also feasible and fair.
What type of contact details did each sample source contain and what was the quality of this?	IFF Research recommended conducting the B-MeX pilot using CATI and to send a batch of emails to test if it would be feasible to have an online element too. The ability to conduct a CATI survey depended on the volume and quality of telephone numbers available for the business customers of each wholesaler. The number of online invitations sent to test response rates also depended on the sample availability.

What other information could be provided as part of the sample?	This information would have potentially informed the design of the survey itself (e.g., we could potentially have been able to reduce the burden on participants by not needing to ask them when they made contact and what that contact was about if information could be provided as part of the sample).
How frequently were sample sources updated and how quickly could they be transferred to IFF Research?	This was important in terms of informing our recommendations for the frequency of the main survey.

In order to get answers to these questions, IFF sent out a document to wholesalers with a list of sample fields that wholesalers would ideally provide for B-MeX fieldwork, with reasons for why each field was needed for fieldwork. Wholesalers were asked whether they could provide all the fields requested, and reasons for why any fields could not be provided.

Wholesalers were also asked to provide answers on questions around:

- Sample format (e.g. what format could they provide data in, what quality of information is held, whether they could provide any other information on each contact that would be useful)
- Types of contact (e.g. how different contacts are categorised, whether any particular contacts should be excluded)
- Sample size (e.g. how many contacts are received in an average month, whether there is any variation in this through the year)
- Sample delivery and timings (e.g. how long would it take to deliver sample after a request is made, what would burden of preparing and transferring sample be)

The full list of sample fields and questions sent to wholesalers can be found in Annex A.

Some key findings from these scoping questions are as follows:

- Overall, it was found that wholesalers could provide most sample fields requested. Most said that they would not be able to provide any firmographic information for contacts, and so it was decided that this information would not be requested going forwards.
- All wholesalers were able to provide sample in Excel or CSV format.
- Several wholesalers flagged that the quality of their sample was likely to be mixed, with inconsistent formatting due to free text entries.
- Several wholesalers flagged that they would not be able to provide contact name/job title, phone number *and* an email address for all contacts.

- The number of contacts received in a month varied greatly between wholesalers, ranging between 30 and 3029. Two wholesalers were unable to provide this information.

- Most wholesalers said that they would be able to transfer sample for the B-MeX survey within a few working days, as is the expectation for C-MeX data. However, some wholesalers were unable to give a timescale. One wholesaler stated that they would be able to provide sample within one calendar month.

Cognitive testing

At the same time as the scoping phase, IFF Research worked on developing the B-MeX survey for fieldwork. The format largely followed survey questions previously developed by the B-MeX working group, with some wording changes suggested to drive ease of understanding, and screener questions added to ensure, as much as possible, that the B-MeX survey would be completed by the correct people (i.e. contacts at a business that had been in touch with their wholesaler within a particular timeframe).

IFF Research undertook cognitive testing of the B-MeX survey to ensure that the survey questions were being understood in the way that they were intended to be, and to gather participants' thoughts on the survey overall, including survey length, question relevance, and whether anything could be added or developed to make the survey as easy to complete as possible. A total of 50 cognitive interviews were conducted between 12th September and 12th October 2023.

Pilot

To test the proposed B-MeX questionnaire and methodology, we conducted a full-scale pilot. This involved administering the survey as it would be for mainstage fieldwork. Direct sample was requested from wholesalers, and indirect sample was requested from MOSL, covering a period of one month (i.e. all contacts between 18 September and 18 October).

The aim of the pilot survey was to test the:

- **response rates to both surveys** – both at an overall and subgroup level, to determine the extent to which the survey has captured data from the diverse business population and to identify any key subgroup biases (e.g., by wholesaler / quality of sample / type of sample).

- **reliability of survey findings –** e.g., exploring item non-response, the proportion of don't know responses, any evidence of "positive acquiescence" vs. varied response patterns in business responses, and the role of the survey design on these.

- **experience of completing the survey –** by analysing response patterns, completion times, the type of feedback given in open-ended questions; and through re-contacting respondents for follow-up qualitative interviews to explore their experience of the survey in more detail.

We conducted pilot fieldwork by telephone (CATI) between 6th November and 8th December 2023, with businesses that had contacted their water wholesaler between 18th September and 18th October 2023 (1,188 of which were direct contacts and 220 indirect contacts sourced via the bilateral hub). A total of 1,408 interviews were completed, with all wholesalers represented.

Qualitative Research

Between the 15th December 2023 and 19th January 2024, we conducted follow-up qualitative depth interviews with 20 businesses who had completed the pilot survey. This allowed us to conduct a final test of the pilot survey questions to ensure changes made as a result of the cognitive testing had been successful and explore any additional issues that might have arisen during the pilot.

Qualitative interviews were purposively sampled to ensure specific survey responses and potential issues identified during the pilot were explored in further detail. A range of different types of businesses participated in this phase of the research.

Interviews were conducted virtually via Microsoft Teams. Each lasted approximately 30 minutes.

4 B-MeX: Pilot Analysis

Sample

Sample collection process

We received a total of 18,564 records from water wholesalers and MOSL to be used in the B-MeX pilot. The sample from water wholesalers is known as 'direct sample' (i.e. where business customers had contacted their water wholesaler directly), and sample from MOSL is known as 'indirect sample' (i.e. where business customers had made contact with their water wholesaler via their retailer). The indirect sample, sourced from MOSL, relates to business customer requests that retailers actioned – ie. requested a wholesaler service on behalf of the customer – via the "Bilateral hub". The records related to direct and indirect customer contacts that occurred over the period 18th September 2023 to 18th October 2023.

Table 4.1 provides a breakdown of the number of records received for each wholesaler, as well as how many of these came from direct vs. indirect sample.

Table 4.1 Number of records received by IFF, relating to wholesaler

	Direct Sample	Indirect sample	Total sample	% of overall sample
Affinity Water	258	58	316	2%
Anglian Water	1,633	296	1,929	10%
Northumbrian Water	794	123	917	5%
Portsmouth Water	83	9	92	1%
SES Water	90	21	111	1%
Severn Trent Water	6,252	788	7,040	38%
South East Water	548	129	677	4%
South Staffordshire Water	144	83	227	1%
South West Water[7]	924	245	1,169	6%
Southern Water	826	96	922	5%
Thames Water	905	422	1,327	7%
United Utilities	958	325	1,283	7%
Wessex Water	898	30	928	5%
Yorkshire Water	1,512	113	1,625	9%
Total	**15,825**	**2,738**	**18,563**	100%

[7] South West Water and Bristol Water are reported together for the purposes of the sample and outcomes sections of this report.

Sample cleaning

After sample was received, IFF undertook a manual process in which all contact data was collated and cleaned to remove unusable sample. This included:

- Records with no telephone number (4,692)
- Records for wholesalers that were not in scope of the research (1)[8]
- Records where the date was not in scope of pilot fieldwork (1,540)[9]
- Duplicate records (i.e. where a business had contacted their wholesaler more than once within the eligible fieldwork period)[10] (2,488)
- Records to be excluded under Section 1.3 of Ofwat's Request for Information for Pilot Data (216)[11]

0 provides a breakdown of exclusions by wholesaler, and the proportion of the wholesaler's total sample that each exclusion removed. Excluding those that were duplicates, collectively these records are referred to as ineligible throughout the rest of this report.

Table 4.2 Number and proportion of exclusions by wholesaler

	No telephone number	Date not in scope	Duplicate records	Exclusions under RFI	% of total wholesaler sample excluded
Affinity Water	10%	8%	13%	0%	**31%**
Anglian Water	58%	9%	9%	-	**76%**
Northumbrian Water	33%	6%	2%	11%	**51%**
Portsmouth Water	11%	3%	9%	38%	**61%**
SES Water	41%	4%	4%	9%	**57%**
Severn Trent Water	28%	8%	15%	0%	**51%**
South East Water	1%	11%	4%	-	**16%**
South Staffordshire Water	63%	18%	1%	-	**81%**
South West Water	17%	10%	20%	-	**46%**
Southern Water	8%	5%	9%	-	**21%**
Thames Water	23%	21%	11%	-	**56%**

[8] Dŵr Cymru Welsh Water.
[9] These were all within indirect sample records.
[10] Records were de-duplicated by telephone number. Where duplicates occurred, the record with the most recent date of contact was retained.
[11] This included: Any contacts related to household queries; Any contacts relating to non-wholesaler activities; Contacts dealt with or with regard to developer services; Wrong numbers, including calls where a customer has contacted the wrong company; Contacts where the customer has called about a non-appointed activity and the call has no connection with the appointed business; Contacts about recreational and amenity activities; Enquiries from CCWater on behalf of a customer; Any outbound contacts; Any contacts related to bilateral transactions.

	No telephone number	Date not in scope	Duplicate records	Exclusions under RFI	% of total wholesaler sample excluded
United Utilities	10%	7%	16%	-	**33%**
Wessex Water	25%	2%	9%	7%	**42%**
Yorkshire Water	8%	3%	27%	-	**38%**
Total	**25%**	**8%**	**13%**	**1%**	**48%**

Base: Affinity Water (n=316), Anglian Water (n=1,929), Northumbrian Water (n=917), Portsmouth Water (n=92), SES Water (n=111), Severn Trent Water (n=7,040), South East Water (n=677), South Staffordshire Water (n=227), South West Water (n=1,157), Southern Water (n=922), Thames Water (n=1,327), United Utilities (n=1,283), Wessex Water (n=928) and Yorkshire Water (n=1,635)

This resulted in a final pilot sample of 9,627 records. Table 4.3 provides a breakdown of the number of usable records for each wholesaler.

Table 4.3 Number of usable records by wholesaler

	Direct Sample	Indirect sample	Total sample	% of overall sample
Affinity Water	187	30	217	2%
Anglian Water	407	64	471	5%
Northumbrian Water	381	66	447	5%
Portsmouth Water	30	6	36	0%
SES Water	31	17	48	1%
Severn Trent Water	3,266	196	3,462	36%
South East Water	528	39	567	6%
South Staffordshire Water	2	41	43	0%
South West Water	510	118	628	7%
Southern Water	675	50	725	8%
Thames Water	460	130	590	6%
United Utilities	652	203	855	9%
Wessex Water	535	3	538	6%
Yorkshire Water	945	55	1,000	10%
Total	**8,609**	**1,018**	**9,627**	100%

Sample quality

During the sampling process, records that were usable (i.e. in scope and with a telephone number) but could be considered as 'poor quality' were flagged. These were classed as records that both had an unclear reason for contact, and either no named contact or company name on the sample. These records were not removed for the purposes of the pilot to help understand what, if any, effect these records would have on response rates. Due to the lack of information on these pieces of sample, IFF hypothesised that these may prove more difficult to engage in pilot fieldwork.

Table 4.4 shows a breakdown of 'poor' sample for each wholesaler. All records marked 'poor' originated exclusively from direct records received from wholesales, no indirect records received from MOSL were marked 'poor.'

Table 4.4 Number of poor records by wholesaler

	Total usable sample	Poor records	% of wholesaler sample
Affinity Water	217	98	45%
Anglian Water	471	-	-
Northumbrian Water	447	43	10%
Portsmouth Water	36	2	6%
SES Water	48	-	-
Severn Trent Water	3,462	2,731	79%
South East Water	567	257	45%
South Staffordshire Water	43	-	-
South West Water	628	176	28%
Southern Water	725	7	1%
Thames Water	590	328	56%
United Utilities	855	-	-
Wessex Water	538	3	1%
Yorkshire Water	1,000	-	-
Total	**9,627**	**3,645**	**37.86%**

Overall, 37.86% of usable sample records were flagged as being 'poor', although it should be noted that the majority (75%) of these records were from Severn Trent. If Severn Trent's sample were excluded entirely, 'poor' records would constitute 14.83% of the usable sample.[12]

Outcome data

All usable records were dialled at least once during fieldwork, with the exception of 128 records that were removed as they had been contacted for the cognitive fieldwork and thus were excluded to avoid overburdening respondents. Therefore, a total of 9,499 records were dialled at least once during fieldwork. In this section we explore the outcomes of all dialled records at the end of fieldwork; in addition to completed interviews, records were allocated to the following outcome categories:

- **Ineligible:** respondents that screened out due to being ineligible for the survey (e.g. not a business contact).
- **Refused:** where a respondent explicitly refused to take part in the survey.
- **Partial interviews:** any instance where an interview has started but not finished or ended with another definite outcome (e.g. a refusal).
- **Unobtainable:** where it was not possible to make any contact using the details provided, e.g. wrong numbers, dead lines etc.

[12] Severn Trent water was unable to distinguish between there inbound and outbound contacts and therefore sent both, hence their disproportionate sample size.

- **No definite outcome:** this includes all records that were technically still 'live' at the close of fieldwork, e.g. records where no contact had been made but the maximum number of call attempts had not been reached; it is likely many of these would have moved into the 'unobtainable' category had fieldwork continued for longer.

Table 4.6 shows the overall outcomes of all dialled sample. Interviews were completed with 15% of all records dialled. Just over half (51%) of records had no definite outcome at the end of fieldwork, and 7% were unobtainable. In the remainder of cases, at least some contact was made: 16% were direct refusals, 9% were found to be ineligible for the survey, and a small proportion (2%) started the survey but dropped out before completing.

Table 4.5 Overall sample outcomes (all dialled sample)

	Count	%
Complete	**1,408**	**15%**
Unobtainable	681	7%
Ineligible	809	9%
Refusal	1,540	16%
Partial interview	189	2%
No definite outcome	4,872	51%
Total	**9,499**	**100%**

Differences in sample outcomes by wholesaler are shown in Table 4.6 below.

Table 4.6 Outcomes by wholesaler (all dialled sample)

	Base	Complete	Partial[13]	Refused	Ineligible [14]	Unobtain -able	No definite outcome
Affinity Water	215	23%	1%	29%	7%	4%	37%
Anglian Water	455	23%	1%	22%	4%	9%	41%
Northumbria n Water	436	17%	2%	23%	6%	6%	47%
Portsmouth Water	36	31%	-	14%	3%	6%	47%
SES Water	47	38%	-	13%	9%	-	40%
Severn Trent Water	3,445	7%	3%	11%	13%	4%	62%
South East Water	564	7%	2%	21%	4%	27%	39%

[13] Currently, partial interviews include any instance where an interview has started but not finished or ended with a definite outcome (e.g. a refusal).
[14] Ineligible contacts include: S1 – Not a business / domestic organisation; Residential number. S2 – Not a business / non-domestic organisation. S7 – Matter not in regard to being supplied as business customer.

	Base	Complete	Partial[13]	Refused	Ineligible[14]	Unobtain-able	No definite outcome
South Staffordshire Water	43	23%	2%	14%	2%	2%	56%
South West Water	625	11%	2%	17%	3%	18%	48%
Southern Water	710	21%	2%	21%	8%	6%	43%
Thames Water	586	20%	2%	20%	7%	7%	45%
United Utilities	837	21%	1%	16%	5%	5%	51%
Wessex Water	531	23%	2%	19%	7%	5%	44%
Yorkshire Water	969	24%	2%	16%	6%	6%	46%
Total	**9,499**	**15%**	**2%**	**16%**	**9%**	**7%**	**51%**

Base: All dialled sample

Table 4.7 highlights that, as noted in the previous section, sample records classified as 'poor' were much less likely to complete the survey (7% compared with 19% of non-poor records). 'Poor' records were also more likely to be classified as ineligible compared with non-poor records (12% vs 6%) and were more likely to have no definite outcome at the close of fieldwork (58% poor records vs 47% 'non-poor' records). However, poor records were *less* likely to refuse to take part than non-poor records (13% vs 18%), indicating that the lower completion rate was driven by difficulties reaching a contact, or the contact not being suitable, rather than a lack of willingness to take part among these records.

Table 4.7 Outcomes by quality of record

	'Poor' records	% of 'poor' records dialled	'Better' records	% of 'Better' records dialled
Completes	**272**	**7%**	**1,136**	**19%**
Unobtainable	231	6%	450	8%
Ineligible	447	12%	362	6%
Refused	485	13%	1,055	18%
Partial interviews	88	2%	101	2%
No definite outcome	2,110	58%	2,762	47%
Total	**3,633**	**100%**	**5,866**	**100%**

Table 4.8 shows a breakdown of the reasons for refusal given by the 16% of records that explicitly refused to take part. As the table highlights, a high proportion of refusals (30%) were unable to recall any contact with their wholesaler. Nine per cent said the contact was initiated by a contractor on their behalf and 4% said they were themselves a contractor who contacted the wholesaler on behalf of the business.

Relatively few respondents refused to take part on the basis of not being able to speak in enough detail about the contact (8%), or not being clear on the distinction between wholesaler and retailer (1%). Just over a fifth of refusals (22%) did not give a reason for refusing.

Table 4.8 Reasons for refusal

	Number of refusals	% of refusals
No recollection of contact (S1/S2)	460	30%
Third party reported the issue on their behalf (S1/S2)	145	9%
Can't speak in enough detail about contact (S1/S2)	130	8%
Does not want to take part in any survey (S1/S2)	90	6%
They are a contractor who reported the issue (S1/S2)	66	4%
No longer works at the business (S1/S2)	66	4%
Company policy (S1/S2)	63	4%
Not happy to proceed with survey (A1)	46	3%
Not happy to proceed based on having to talk about interactions with wholesaler (S5)	25	2%
Not happy to carry on after definitions of wholesaler and retailer (S4)	12	1%
Taken part in recent survey (S1/S2)	10	1%
Another reason for refusal (S1/S2)	82	5%
No reason given (S1/S2)	345	22%
Total	**1,540**	**100%**

Table 4.9 below provides a breakdown of reasons records screened out as ineligible; the majority were due to the contact not being a business (57%) or the matter not being in regard to being supplied as a business customer (33%).

Table 4.9 Reasons for ineligibility

	Number of ineligible responses	% of ineligible responses
Not a business / domestic organisation (S1/S2)	459	57%
Matter not in regard to being supplied as business customer (S7)	271	33%
Does not think business contacted wholesaler (S7)	71	9%
Residential Number (S1)	8	1%
Total	**809**	

Survey completion

The following tables summarise completes by wholesaler, direct and indirect contacts, and record quality.

Whilst ideal targets were calculated per wholesaler based on the sample provided, no hard quotas were set to allow completes to naturally fall out and test the performance of the sample provided. Completes were monitored by sample type (direct vs. indirect), outcome (whether the issue was resolved or not at the point the contact took place), and service type (whether the contact related to water or waste water) to check that achieved interviews were falling out broadly in line with the population (i.e. the proportions of these factors in the full dataset of contacts).

Some wholesalers provided large volumes of sample, primarily due to their size, so completes by wholesaler were monitored to prevent under representation occurring amongst smaller wholesalers with fewer sample. Table 4.10 provides a breakdown of completes by wholesaler.

Table 4.10 Number of completes by wholesaler

	Ideal target	No. completes			% of overall completes	% of business customers in wholesaler region[15]
		Direct	Indirect	Total		
Affinity Water	34	44	5	49	3%	5.2%
Anglian Water	218	92	12	104	7%	9.4%
Northumbrian Water	102	58	14	72	5%	7.8%
Portsmouth Water	30	10	1	11	1%	1.1%
SES Water	30	11	7	18	1%	1.0%
Severn Trent Water	145	205	36	241	17%	14.5%
South East Water	79	34	7	41	3%	3.7%
South Staffordshire Water	25	-	10	10	1%	3.1%
South West Water	111	46	24	70	5%	5.8%
Southern Water	109	133	13	146	10%	4.0%
Thames Water	122	84	32	116	8%	15.0%
United Utilities	135	135	43	178	13%	13.4%
Wessex Water	105	123	1	124	9%	3.4%
Yorkshire Water	155	213	15	228	16%	10.1%
Total	**1,400**	**1,188**	**220**	**1,408**	**100%**	**100.0%**

Direct completes accounted for the majority of completes broadly in line with the proportion of direct vs indirect sample provided, as shown in Table 4.11. Indirect contacts were slightly over-represented in the final survey completes due to the better response rate achieved among indirect sample, and deliberate targeting of indirect records to ensure a sufficient base size for sub-group analysis in the pilot.

Table 4.11 Number of completes by direct/indirect contacts

	No. of dialled sample records	% of total dialled sample	Completes	% of overall completes
Direct	8,504	90%	1,188	84%
Indirect	995	10%	220	16%
Total	**9,499**	**100%**	**1,408**	**100%**

[15] As proxied by number of water SPIDs in each wholesaler area (figures provided by Ofwat)

The overall response rate for the B-MeX pilot was 15%.[16] Response rates for business surveys vary considerably due to a range of factors. We believe that this is a good response rate given the complexity of the wholesaler retailer relationship and the difficulties encountered with the sample.

Table 4.12 provides a breakdown of response rates by wholesaler. As shown here, response rates varied considerably. We believe the lower response rates could be improved considerably in the future if the quality of the sample collected and provided is improved.

Table 4.12 Response rate by wholesaler

	Total sample dialled	Completes	Response rate
Affinity Water	215	49	23%
Anglian Water	455	104	23%
Northumbrian Water	436	72	17%
Portsmouth Water	36	11	31%
SES Water	47	18	38%
Severn Trent Water	3,445	241	7%
South East Water	564	41	7%
South Staffordshire Water	43	10	23%
South West Water	625	70	11%
Southern Water	710	146	21%
Thames Water	586	116	20%
United Utilities	837	178	21%
Wessex Water	531	124	23%
Yorkshire Water	969	228	24%
Total	**9,499**	**1,408**	**15%**

A higher response rate was achieved with ‘better’ quality records (19%) compared with ‘poor’ quality sample records (8%), suggesting that sample quality is critical for achieving good response rates.

Table 4.13 Response rate by ‘poor’ vs. ‘better’ quality contacts

	Total sample dialled	Completes	Response rate
‘Poor’	3,633	272	7%
‘Better’	5,866	1,136	19%
Total	**9,**499	**1,408**	**15%**

This pattern holds across most wholesalers, as illustrated in Table 4.14.

[16] Response rate is calculated as completes as a proportion of sample that has been dialled at least once

Table 4.14 Response rate by wholesaler for 'poor'/non 'poor' contacts[17]

	'Poor' records response rate	'Better' quality records response rate
Affinity Water	24%	22%
Northumbrian Water	7%	18%
Severn Trent Water	6%	11%
South East Water	4%	10%
South West Water	8%	12%
Thames Water	18%	22%
Total	**8%**	**19%**

Indirect records, provided by MOSL, achieved a higher response rate (22%) than direct sample (14%) as illustrated in Table 4.15. This is likely due to the indirect sample being higher quality with all data fields fully populated, and due to a deliberate focus on calling this sample due to the lower volume available to ensure we could achieve a high enough base size for analysis.

Table 4.15 Response rate by direct/indirect contacts

	Total sample dialled	Completes	Response rate
Direct	8,504	1,188	14%
Indirect	995	220	22%
Total	**9,499**	**1,408**	**15%**

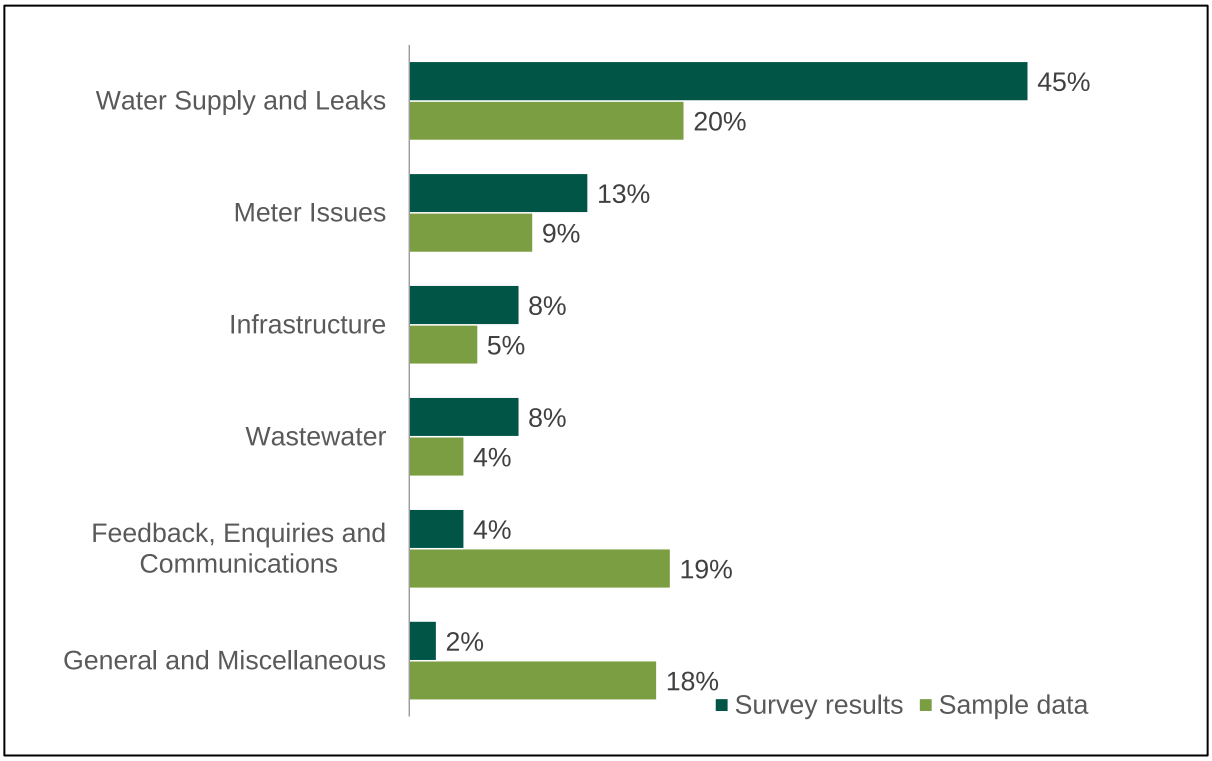
[17] Note: Some wholesalers are not included due to low base size for comparison or for having no 'poor' records; total figures are based on only the wholesalers shown.

Survey results & reflections from qualitative follow-up interviews

Reason for contact

The most common reasons for contact given by direct contacts in the survey were to discuss water supply and leaks (45% of respondents), or meter issues (13% of respondents). As shown in Figure 1, the reasons for contact given by direct contacts in the survey often differed from the reasons for contact listed on the sample file. The contrast in detail between the sample file, where the reason for contact from the different wholesalers (often vague or using internal acronyms) had to be manually aggregated into broad groups, and the individually coded response to the survey is likely to be a cause of this discrepancy. For example, 18% of the direct contacts were listed as “General and Miscellaneous” by wholesalers, where a more detailed coding process of survey results left only 2% in this category. Similarly, only 4% of direct contacts in the survey were coded into the more generic “Feedback, Enquiries and Communications” code, compared to 19% in the sample data.

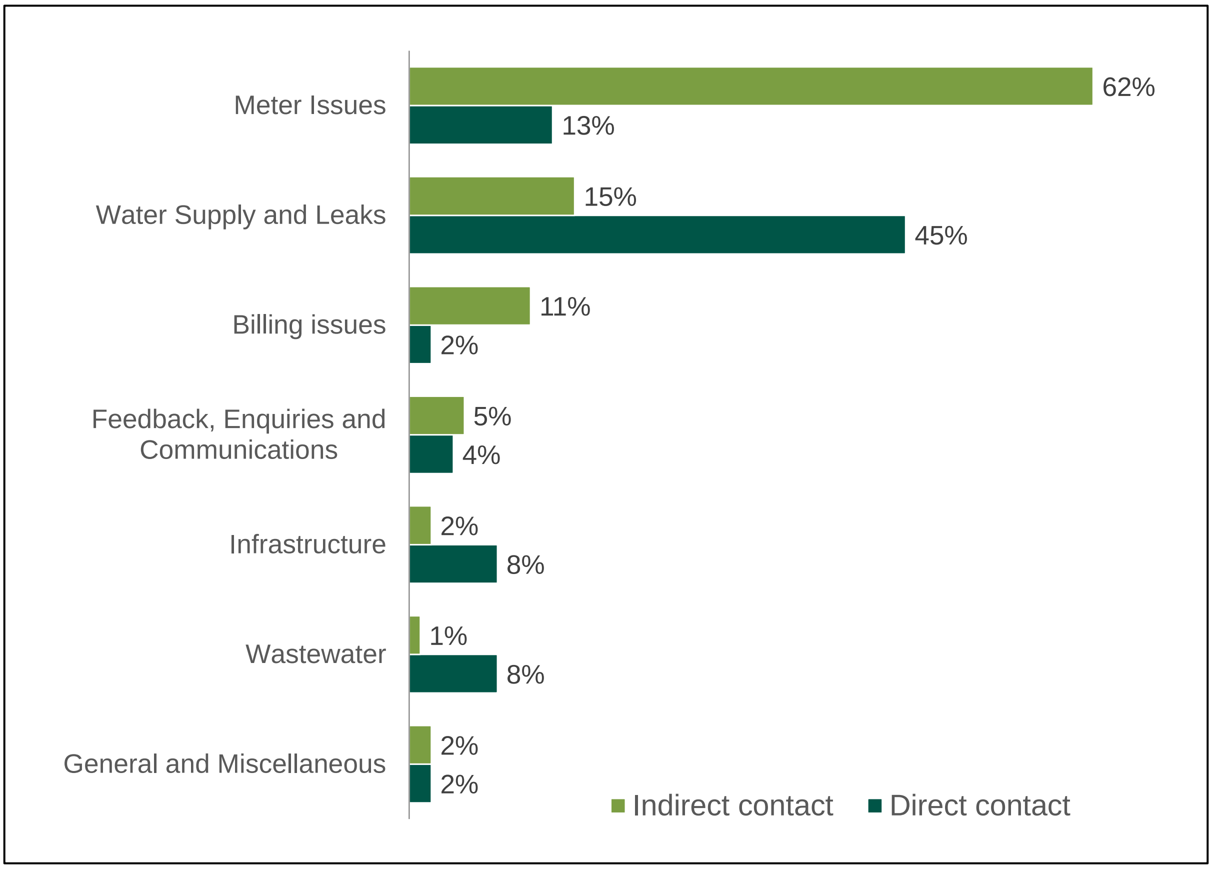
Figure 1 Reason for direct contact



S8. What was the main reason for making contact with your water wholesaler on that occasion?
Base: Direct contacts (1,188)

Indirect contacts came from a separate sample source and did not have a reason for contact listed on their record, so a similar comparison to the sample data is impossible. However, as shown in Figure 2, the reason for contact given in the survey by direct and indirect contacts varied considerably, as would be expected given indirect contacts are largely requests related to metering. Almost two thirds (62%) of indirect contacts were about meter issues, compared to only 13% of direct contacts. Indirect contacts were also more likely to be about billing issues than direct contacts (11% compared to 2%). Water supply and leaks was cited as the reason for contact by only 15% of indirect contacts compared with almost half of direct contacts (45%).

Figure 2 Reasons for contact (Direct vs Indirect)



S8. What was the main reason for making contact with your water wholesaler on that occasion?
Base: All (1,408)

There was also significant variation in the reasons for contact given by respondents in the survey amongst the different wholesalers. Almost seven in ten contacts from Affinity Water (69%) were about water supply and leaks, significantly higher than average amongst the other wholesalers (40%). The top reason for contact for amongst contacts from Anglian Water was to discuss meter issues (57% of contacts from Anglian Water) compared to an average of 20% across all wholesalers.

When asked specifically about whether their contact was about clean or wastewater, just over half of respondents (57%) said that their contact was about clean water, 22% said it was about wastewater, and 15% referenced both. Excluding those cases where service type was not included in the sample data (17% of sample), the service type referenced by respondents when prompted in the survey broadly matched information on the sample. Nine in ten respondents (89%) referenced either the service type that matched the information on the sample, or both clean and wastewater.

Just 2% of indirect contacts said that their contact was about wastewater only, compared to just over a quarter (26%) of direct contacts. Indirect contacts were more likely to say that their contact was about both clean and wastewater (30%) compared to direct contacts (13%).

When asked to consider how confident they felt during the survey answering if their contact was about clean or wastewater, all participants in the qualitative follow up interviews said they were very confident in the distinction:

"Clean comes in the property and waste is what's going out."

Direct contact, United Utilities

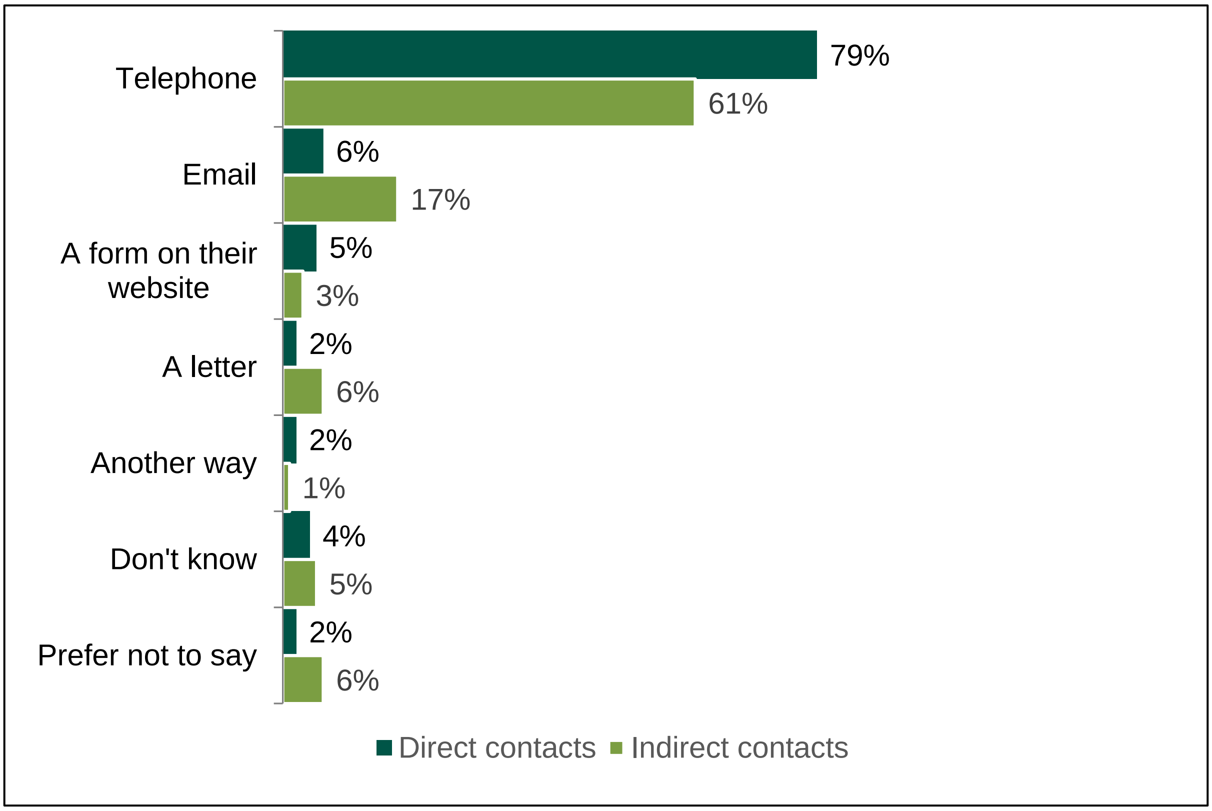
Method of contact

In the survey, three-quarters of respondents (75%) said they contacted their wholesaler directly. Smaller proportions contacted their retailer, who in turn asked their wholesaler to get in touch (13%), contacted by their wholesaler directly (6%), or contacted their retailer, who then passed on their wholesalers' details (2%).

Three-quarters (76%) of the contacts experienced by respondents in the survey were by telephone. The remainder of the contacts were either via email (7%), a form on the wholesaler's website (5%) or by letter (3%).

Nearly eight in ten (79%) direct contacts took place over the phone, significantly more than the 61% of indirect contacts through this method. Indirect contacts were more likely than direct contacts to have had their interaction over email (17% compared to 6%).

Figure 3 Contact channel



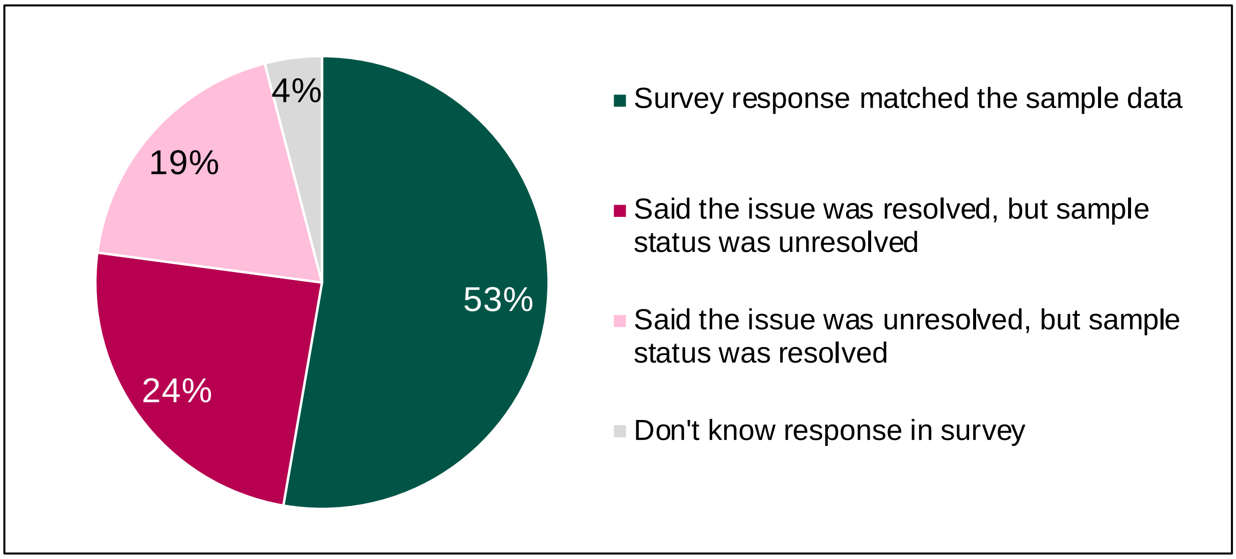
A4 & A5 Did you get in touch with your water wholesaler/your water wholesaler get in touch with you by…? Base: 1,356

Resolution of issue

Two-thirds of the respondents (66%) said in the survey that their issue had been resolved, whilst just under one third (31%) said that the issue remained unresolved. Two-thirds (68%) of direct contacts said that their matter had been resolved compared to only 50% of indirect contacts.

As shown in Figure 4, excluding all respondents whose issue status was unknown in the sample data (34%), only half (53%) of the responses given in the survey matched the status on sample. A quarter (24%) of respondents said in the survey that their issue had been resolved, even though the status of their issue in the sample was unresolved, suggesting that their case had reached a resolution in between the sample delivery date and the date of their interview. A further 19% said that their issue remained unresolved, even though their wholesaler had marked their case as resolved.

Figure 4 Survey status against sample status



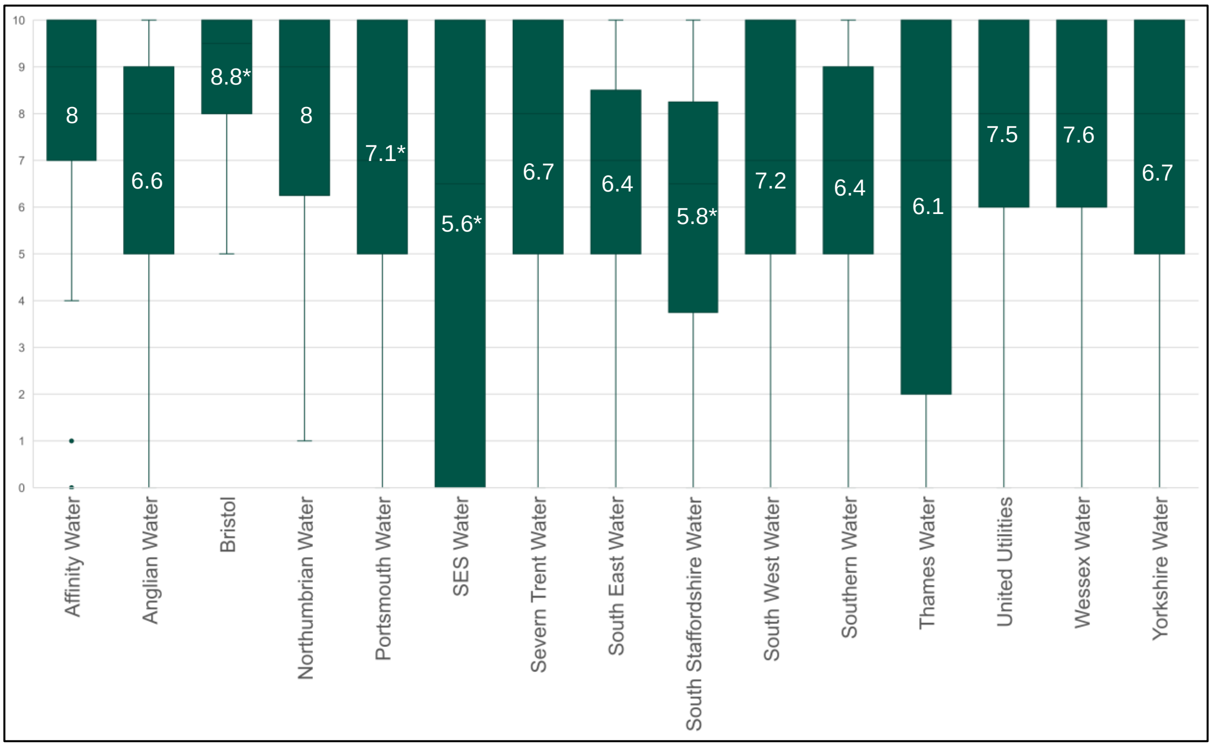
Base: All respondents with status data available on sample (935)

Satisfaction

In the survey, respondents were asked to use a scale of 0-10, where 0 represented extremely dissatisfied, 5 represented neither satisfied nor dissatisfied and 10 represented extremely satisfied, to rate their satisfaction of their experience of contact with their wholesaler. Respondents were reminded to think only about the specific contact that they had with their wholesaler on the date listed on the sample file.

Two-thirds of respondents (68%) were satisfied (a score of 6-10 out of 10) with the contact they had with their wholesaler, while a fifth were dissatisfied (19%) (a score of 0-4 out of 10). Of those who were dissatisfied, the most common score was 0 out of 10, with half of dissatisfied respondents giving this score (10% of all respondents). As shown in Figure 5, the mean satisfaction by wholesaler ranged from 5.6 to 8.8, however it should be noted that some wholesalers had very low base sizes in the pilot so these scores should be treated with caution. The mean score across all respondents was 6.9.

Figure 5 Satisfaction score by wholesaler



A7. Thinking just about your contact with your water wholesaler on [service date] how satisfied are you with your experience of your contact? Base: Affinity Water (49), Anglian Water (104), Bristol Water (17), Northumbrian Water (72), Portsmouth Water (11), SES Water (18), Severn Trent Water (241), South East Water (41), South Staffordshire Water (10), South West Water (70), Southern Water (146), Thames Water (116), United Utilities (178), Wessex Water (124), Yorkshire Water (228). indicates a low base size.*

In the qualitative follow-up interviews, most found the 0-10 scoring system easy to use and felt that it allowed for a fair and effective evaluation of the performance of the wholesaler. While many said they felt the description of the satisfaction scale made it clear how you should respond, there was confusion amongst some respondents as to what factors should be considered when giving a score. Some struggled to remember the specific interaction in enough detail, and others were influenced by other factors:

"At the time I thought they were just asking me about how the whole process... had I known it was just about my interactions with the wholesaler then I probably would have given 10 out of 10."

Indirect contact, Southern Water

While comparing the satisfaction of direct contacts against indirect contacts by each individual wholesaler would not be meaningful due to low base sizes, at the overall level, direct contacts gave higher satisfaction scores on average compared to indirect contacts (7.0 compared to 6.4).

It was clear that the status of the respondents' issue impacted the satisfaction score they gave in the quantitative survey. The mean score amongst those who said in the survey that their issue had been resolved was 8.1, compared to 4.3 amongst those whose issue remained unresolved. This was supported by feedback in the qualitative follow-up interviews, where respondents confirmed that unresolved issues had impacted their satisfaction and feedback. Where issues had been resolved, respondents seemed to be able to make a slightly more objective evaluation of their interaction, however it did still play a part for some.

“It's impacted it massively, it's hard to separate the two.”"

Direct contact, Yorkshire Water

"They identified the problem but didn't fix it. It is quite difficult to take that side out of the equation."

Direct contact, Affinity

“If they resolved the problem, it would affect the [satisfaction score] a lot, it would be a lot higher, you'd give them and 8 or 9 out of 10. But when they don't do anything, we can't give them a high score. If they don't resolve it, they don't deserve anything, they deserve a nil. “

Indirect contact, United Utilities

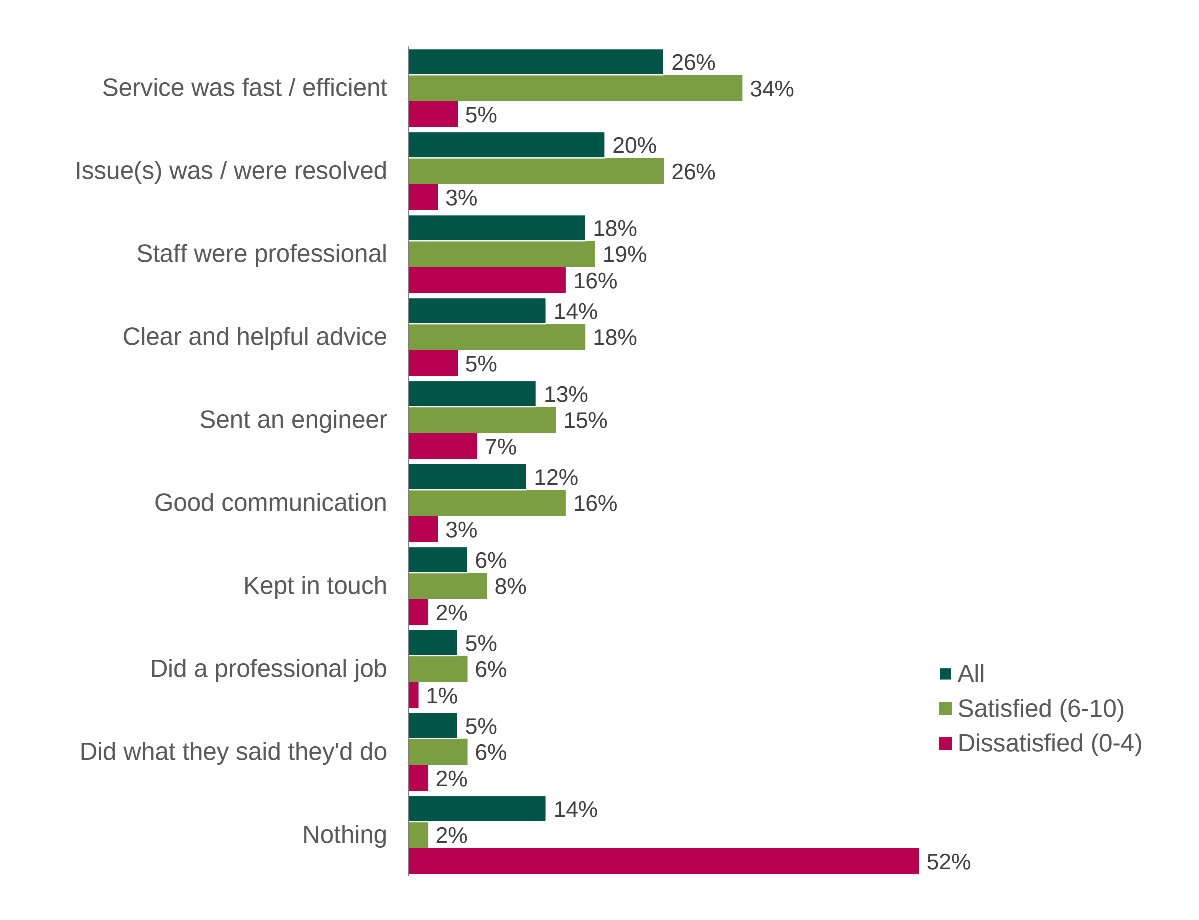
There were those who were able to separate the status of their wider issue with their wholesaler, focusing specifically on their interaction with the wholesaler on the date in question, but they tended to be in the minority:

“I was narked it hasn't been resolved but doesn't mean I couldn't have a decent conversation with the person on the 19th.”

Indirect contact, Thames Water

When asked to consider what their wholesaler did well, one quarter (26%) of all respondents said that the service was fast and efficient, As shown in Figure 6 What wholesalers did well, one in five (20%) referenced the fact that the issue was resolved,18% said that the staff were professional, and 14% felt that they provided clear and helpful information. Conversely, 14% of respondents said that there was nothing their wholesaler did well.

Figure 6 What wholesalers did well



A8. Again just thinking about your contact with your wholesaler on [service date] what would you say they did well? Base: All (1,408), Satisfied (964), Dissatisfied (263).

Direct contacts were significantly more likely to cite the fact that the service was fast and efficient than indirect contacts (28% compared to 14%). Other than this, direct and indirect contacts did not differ greatly in what they thought their wholesaler did well.

The table below shows the most common piece of positive feedback given by respondents from each of the 15 wholesalers:

Table 4.1 Top thing wholesaler did well by wholesaler

	Top thing wholesaler did well	Base size
Affinity Water	Issue was resolved	49
Anglian Water	Issue was resolved	104
Bristol Water	Service was fast/efficient	17*
Northumbrian Water	Service was fast/efficient	72
Portsmouth Water	Provided clear and helpful information	11*
SES Water	Nothing	18*
Severn Trent Water	Service was fast/efficient	241
South East Water	Service was fast/efficient	41*

	Top thing wholesaler did well	Base size
South Staffordshire Water	Good communication	10*
South West Water	Good communication	53
Southern Water	Service was fast/efficient	146
Thames Water	Service was fast/efficient	116
United Utilities	Service was fast/efficient	178
Wessex Water	Service was fast/efficient	124
Yorkshire Water	Service was fast/efficient	228
Total		**1,408**

** indicates a low base-size, findings should be treated with caution*

Clarity and good communication were the main themes that underpinned positive feedback given by respondents in the qualitative follow up interviews about the interaction they had with their wholesaler:

"He was professional and polite. He was understanding of the issues and looking to resolve the issues in the best way possible."

Indirect contact, Northumbria Water

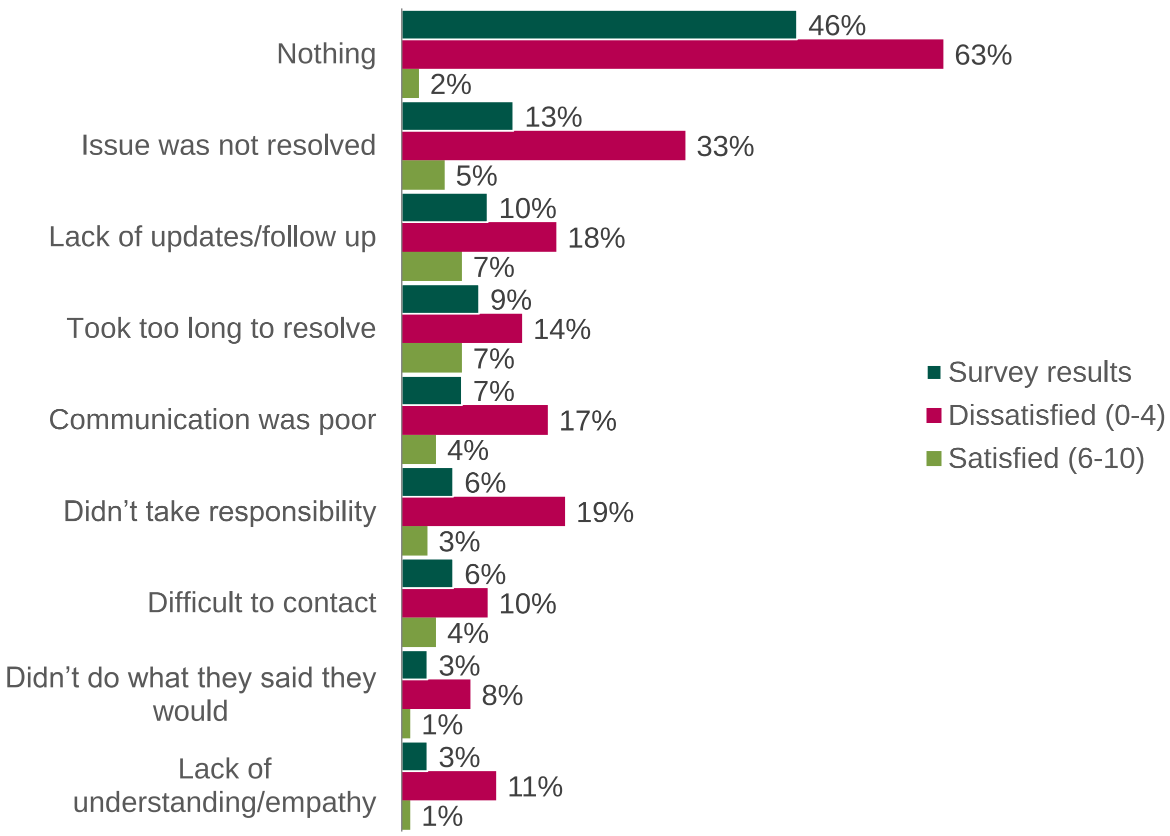
Where the wholesaler followed up with contact after the initial interaction, it was appreciated by respondents. Similarly, a lack of follow up communication was highlighted as something that wholesalers didn't do so well:

"They communicate fine, but it's what happens after. They'll say on the phone that they'll do something but then they don't. They're full of promises."

Indirect contact, United Utilities

In the quantitative survey, nearly half (46%) said that there was nothing their wholesaler did poorly. This was the most common response across all fifteen wholesalers. As shown in Figure 7, the top issue that respondents had with their contact with their wholesaler, referenced by just over one in ten respondents (13%), was that their problem had not been resolved. One in ten respondents felt there was a lack of updates or a follow up to their contact (10%), 9% felt that the issue took too long to resolve, and 7% felt that the communication was poor. There was very little significant variation between direct and indirect contact in terms of what their wholesalers did not do well.

Figure 7 What wholesalers didn't do well



A9. And again just thinking about your contact with your wholesaler on [service date] what would you say they did not do well? Base: All (1,408), Satisfied (964), Dissatisfied (263).

Of the participants who were dissatisfied (a score of 0-4) with their contact with their wholesaler (263 respondents), one third (33%) referenced the fact that their issue had not been resolved, and just under one in five (19%) felt their wholesaler didn't take responsibility for the problem.

When asked what wholesalers would need to do going forward to improve on the satisfaction score that was given, respondents most commonly mentioned resolving their issue (13%) or improving their communication (13%). Indirect contacts were significantly more likely than direct contacts to highlight the need for improved communication (20% vs 12%). Other suggestions included keeping in touch and providing updates (12%) and responding faster to issues (11%). Just under one in five (17%) had no improvements to suggest.

The follow up qualitative interviews revealed a similar theme, where recommendations on how to improve were often linked to the outcome of their interaction. Some respondents who had had indirect contact with their wholesaler did add that they thought a lack of continuity and communication between retailer and wholesalers negatively affected their interaction, and subsequently the score they gave.

Respondents recognised the importance of monitoring the service provided by wholesalers, and most would be happy to do the survey again, conscious that gathering feedback would help the wholesalers improve their service. One respondent felt it was important to talk to the same customers over several time periods to ensure continuity:

"When you have the same customer continuing to take surveys over a period of time, it really helps you get a better understanding because it's all coming from the same source."

Most participants in the follow-up qualitative interviews said that they were happy with the length of the B-Mex survey. Generally, the questions felt relevant to their interaction with their wholesaler, and respondents felt that the survey allowed them to give all the feedback they would like to give about their wholesaler:

"I was happy with everything included on the survey, I didn't think there was anything else I wanted to add."

Direct contact, United Utilities

Differentiating Wholesaler and Retailer

At the beginning of the telephone survey, the responsibilities of wholesalers and retailers were read to respondents to ensure the distinction in their roles was as clear as possible. In the qualitative follow-up interviews, participants were asked to reflect on how confident they felt in this distinction when they were answering the questions in the telephone survey.

Most respondents in the qualitative interviews said they felt confident in the distinction between a water wholesaler and a water retailer. There were those who didn't, for which the definitions read out by the interviewers proved useful:

"I just wanted to be clear on who was responsible for what... once it was explained it was perfectly understandable."

Direct contact, United Utilities

While most were comfortable with the definitions, there were certainly cases where interviewers spent considerable time before beginning the survey ensuring that the respondent was referring to contact with their wholesaler rather than their retailer.

There was for example confusion amongst respondents who had been in contact with Wave. In a couple of instances, respondents said that the contact in question (on the specified date) was in fact with Wave (a retailer), not their wholesaler. Wave is a retailer created in joint venture by Northumbrian Water and Anglian Water. It was unclear to interviewers whether Northumbrian Water and Anglian Water were announcing themselves as operating under the Wave name during contact with respondents, or if the respondent was thinking about the wrong interaction. With information on only the respondent's wholesaler, and not their retailer, it was difficult for interviewers to reassure respondents that they were talking about the correct interaction.

Looking forward to future iterations of the survey, making the name of the respondent's retailer and wholesaler available to interviewers is something to consider, to provide additional context on the interaction and to avoid confusion. Additional information around the actions the wholesaler took once the process had been initiated by the retailer would be particularly useful when speaking to indirect contact sample. Where information on the sample around the actions or contact made by the wholesaler is vague, for example if the sample indicates that the "retailer initiated operational process related to the Non Household Customer Enquiry," it was difficult for interviewers in the pilot survey to clarify the specific interaction for the respondent. More detail would help to avoid confusion, enable the interviewer to reassure the respondent, and ensure that the information gathered is accurate.

Ease of feeding back on a specific contact

Once the survey was underway, most participants in the follow-up interviews said that they found it relatively easy to talk about the contact that they had with their wholesaler. However, when prompted further, it became clear that some respondents who had been in touch with their wholesalers on multiple occasions struggled to pinpoint the specific contact in question. One participant initially said that the survey *"all seemed to be very straightforward",* but later in the conversation noted that they did in fact have some difficulty giving feedback about the specific interaction, as they had had *"a number of phone calls [with their wholesaler] over just a few days"*.

Again, the lack of information available to interviewers on each respondent's record made it difficult for them to clarify the situation. With only a service date and wholesaler name in some instances, interviewers could give little reassurance to respondents that they were thinking of the correct interaction, occasionally leading to some frustration:

"Unless I go back through my records, I can't determine what the contact was about. If people are going to do research, they need to ensure that the people they're asking the question of can reasonably have access to the information that would allow them to make a clear statement. I'm making an assumption about what that contact was about, but it could be about this other matter."

Indirect contact, South West Water

Similarly, when an interviewer's call would go through to a gatekeeper, without a named contact, it proved very difficult and time consuming for them to get through to the employee who had been in contact with their wholesaler, sometimes leaving the interview at an impasse.

Telephone interviewers in the quantitative survey reported a few incidents where respondents were only receiving wastewater services which led them to believe that they were not being supplied as a business customer. A similar issue occurred with the treasurer of a bowls club, who thought that they weren't eligible for the survey, not realising the club was a business. In both cases, the interviewer was able to reassure the respondent that they were a business customer, highlighting the benefit of a CATI setting over an online survey, to avoid unnecessary screen outs.

Weighting

The aim of the B-MeX survey is to derive a robust score for each wholesaler. As well as achieving a sufficient number of interviews with customers of each wholesaler, to ensure a good base for analysis, the robustness of the score will also depend on the extent to which completed interviews are representative of the population – i.e., the total contacts for each wholesaler within each month, and across the year. For example, as mentioned above, mean scores tended to differ for direct vs. indirect contacts, and depending upon whether the issue had been resolved; therefore, if a disproportionate number of interviews were undertaken with certain respondent types (due to natural variations in response rates and the fall out of the sample), this could impact a wholesaler's final score.

As discussed earlier in this report, during fieldwork interviews were monitored against ideal targets to ensure that achieved interviews broadly aligned with the population in terms of factors that could be identified across all wholesalers – the outcome of the contact, the service type the contact related to, and whether the contact was direct or indirect. However, in practice it is not usually possible to hit all quota targets exactly during fieldwork, as different customers may be more or less likely to take part; therefore, weighting may be needed in order to ensure the final results are representative.

Weighting is a technique used to adjust the results of a survey to make it more representative of the larger population. It involves assigning a 'weight' (i.e. a multiplier) to each response based on certain characteristics such as age, gender, ethnicity, or other relevant factors. The weights are used to balance out the results and provide a more accurate picture of the population's views. Weighting is necessary because not everyone in the population will respond to the survey, and the profile of those who do respond may not match the larger population.

The survey data above has been presented unweighted. However, we explored a potential weighting approach using the pilot data, to inform possible weighting strategies for the shadow year and beyond. Below we outline the approach taken.

Weighting method

Wholesalers did not have a uniform framework and protocol for recording contacts. This meant very few data points could be used for weighting which were consistently coded across wholesalers. We worked with two derived variables (coded from the sample information received):

Outcome: Whether an issue is:

1. 'Not resolved/open'
2. 'Resolved closed'
3. 'unknown'.

Servicetype: Whether enquiry was about:

1. 'Water'
2. 'Waste water'
3. 'Other/unknown'.

As the pilot survey only included a months' worth of contacts for each wholesaler, for some smaller wholesalers the total number of completes was too low for their data to be weighted; therefore, these wholesalers (Portsmouth Water, SES Water and South Staffordshire Water) were excluded from this weighting exercise. Bristol and South West Water were combined into one wholesaler to make up a big enough base.

Random Iterative Method (Rim) weighting is a form of weighting where targets are set for multiple variables separately. Statistical software is used to calculate final weights based on these targets, by applying adjustments to the data across multiple iterations which are repeated until the sample matches the population targets as closely as possible.

Two Rim weight targets were derived for each included wholesaler: one for Outcome and one for Servicetype. The entire available universe of contacts was used to derive targets, including eligible contacts missing telephone numbers, as shown in Table 4.2. Complete interviews for wholesalers were then individually rim weighted to these targets, producing weighted distributions for Outcome and Servicetype (% of enquiries in each category) that exactly matched the targets.

Table 4.2 Universe of contacts – totals by wholesaler

	Universe of contacts (pilot month)	Estimated annual universe of contacts	% of total universe of contacts
Affinity Water	250	3,000	1.7%
Anglian Water	1,590	19,080	11.1%
Northumbrian Water	748	8,976	5.2%
Portsmouth Water	46	552	0.3%
SES Water	93	1,116	0.6%
Severn Trent Water	5,443	65,316	38.0%
South East Water	574	6,888	4.0%
South Staffordshire Water	185	2,220	1.3%
South West Water	823	9,876	5.7%
Southern Water	795	9,540	5.6%
Thames Water	894	10,728	6.2%
United Utilities	982	11,784	6.9%
Wessex Water	766	9,192	5.3%
Yorkshire Water	1,130	13,560	7.9%
Total	**14,319**	**171,828**	**100%**

Weighting efficiency is a measure of how well the process of weighting survey data has worked to align the survey sample with the target characteristics of the overall population. A high weighting efficiency shows that the weighted profile of the completed interviews closely matches the desired targets.

Following the weighting approach described above, the weighting efficiencies for wholesalers were all high (the lowest was 79%, with most others in the 90% - 100% range). This suggests that the quota management undertaken was effective - the profiles of complete interviews by these variables were close to these targets, requiring only fairly light-touch weighting.

Impact of weighting on scores

Table 4.3 below shows the impact of the weighting on satisfaction scores for each wholesaler. Generally, the weighted scores remained very in line with the unweighted scores, with most seeing an increase or decrease of less than 0.1; the only exceptions being South West Water, whose score increased by 0.58 as a result of the weighting, and Thames Water, whose score increased by 0.11. The weighting does lead to some small shifts in the ranking of wholesalers, with South West Water moving up by 2 places, Northumbrian Water moving up 1 place, and Affinity Water, United Utilities and Wessex Water all moving down 1 place. Note we have not calculated confidence intervals for these mean scores due to the relatively low base sizes in the pilot, however in the following recommendations section we present some indicative confidence intervals for potential annual interview volumes. Due to the low base sizes for some wholesalers, and the fact they are based on a single month of contacts only, these scores should be treated as indicative only.

Table 4.3 Comparison of raw and weighted mean scores by wholesaler

	Base	Mean (indicative) score*(raw)	Mean (indicative) score* (weighted)	Difference	Rank (raw)	Rank (weighted)	Rank change
Affinity Water	47	8.04	7.97	-0.08	1	2	-1
Northumbrian Water	72	7.99	8.04	0.05	2	1	1
Wessex Water	124	7.61	7.65	0.04	3	4	-1
United Utilities	178	7.54	7.55	0	4	5	-1
South West Water	68	7.19	7.77	0.58	5	3	2
Yorkshire Water	226	6.72	6.75	0.03	6	6	=
Severn Trent Water	238	6.67	6.63	-0.04	7	7	=
Anglian Water	103	6.58	6.51	-0.07	8	8	=
South East Water	41	6.41	6.38	-0.04	9	9	=
Southern Water	145	6.4	6.35	-0.05	10	10	=
Thames Water	116	6.09	6.2	0.11	11	11	=

5 B-MeX: Recommendations for the Shadow Year

Methodology

We recommend that B-MeX is conducted as a **telephone survey**. There are two key reasons for this:

- A telephone methodology will provide **more accurate and relevant responses** – this will be critical for the mainstage of the survey when financial incentives and penalties are based on the scores.

- Telephone surveys also **achieve better response rates** than online surveys, which will ensure the results are more robust, including for smaller wholesalers who will have less sample available to survey.

Feedback from the pilot fieldwork and subsequent qualitative interviews shows that **some customers are not likely to confidently understand the difference between their retailer and wholesaler without the definition read out by the telephone interviewer**. Therefore, this element is critical to ensure the accuracy and relevancy of customer responses. Although this definition could be provided as text in an online survey, in our experience participants will often skip larger blocks of text. Telephone interviewers are also able to take a live 'quality control' role, by not proceeding with interviews if it is clear that the respondent does not understand the role of their wholesaler specifically.

The telephone **interviewers also played a valuable role helping to prompt respondents to remember the specific contact** for which we were seeking their feedback; again, there is a risk that with an online survey respondents might skip this information and erroneously give feedback about a different instance of contact or give more general feedback about their water supply.

The limitation of a telephone methodology is that customers must have a telephone number to be eligible for the survey, which could exclude some customers who made contact via online channels. We would recommend that wholesalers attempt to record a telephone number for all customers, regardless of contact channel, for the reasons outlined above; however, we also suggest trialling an online version of the survey among contacts for which only an email address is recorded during the shadow year, should sample volumes allow. (An online survey was not tested as part of the pilot as only a very small proportion of sample records had an email address but no telephone number).

Sample design

Sourcing sample

The provision of good quality sample on a regular basis **will be critical for the success of the B-MeX survey**. Analysis of the pilot shows that sample quality has a clear impact on response rates and the ability of respondents to recall and give feedback on specific contacts with their wholesaler. Additionally, poor quality sample has further implications on the costs and quality of the survey: additional time would be required by the Agent to clean and code the supplied sample, which creates a delay between the sample delivery and start of fieldwork, and means longer gaps between contacts taking place and customers being contacted for fieldwork; there will also be implications for quotas and weighting, as we cannot ensure we achieve interviews with a representative group of customers if the information in the sample used for targeting is incomplete or unreliable.

Our key recommendations are that:

- **Sample from all wholesalers should have ineligible contacts excluded** (e.g. passers-by, non-business contacts). The inclusion of ineligible contacts would lead to wasted interviewing time, thus damaging response rates, and making the survey more expensive to run. All wholesalers must implement methods to flag these types of ineligible contacts in their databases so they can be identified and excluded from the sample.
- **Sample files must contain all data fields deemed essential**, and essential fields **must be fully populated and in the correct format**. If it is not possible to provide information for any of these fields for an individual customer (for example, if the customer refuses to give their name) this should be clearly marked (e.g. entering "Refused" in the applicable field).
- As much as possible, non-essential fields should also be populated, and in the correct format – these will help to improve response rates and ensure the results are representative.
- Should any essential data fields be 'refused' (e.g. a customer refuses to give their name), as much data as possible should be collected in other data fields (e.g. job title, business name) to allow interviewers to reach the correct contact for the survey.

Essential data fields

We believe **the following fields should be regarded as essential:**

- Contact name (full name of the person the contact is with)
- Contact telephone number (field should contain an 11-digit number)
- Service type (Water / Wastewater)
- Date of contact (DD/MM/YYYY)
- Outcome of contact (Resolved / Not resolved / Unknown)
- Primary Reason for contact (coded categories)
- Secondary / supporting reason for contact – detailed (referred to in the pilot sample request as "lower level reason for contact") – these data will be used to jog a respondent's memory at the start of the survey, and to check that the reason for contact is correct; therefore the text / categories used in this field would need to make sense when read out to a customer over the phone (e.g. no numeric codes, industry/company jargon etc.).

The 'Primary Reason For Contact' field will need to use **pre-set codes or categories which are consistent across all wholesalers**. These data can then be used to feed in the sampling and weighting of the survey – we believe this should be **tested during the shadow year** (this was not possible during the pilot, as the data was not supplied in a consistent format). IFF devised the following categories for pilot fieldwork, based on a coding exercise of the 600+ categories/labels received across all the pilot data:

- Meter Issues

- Feedback, Enquiries and Communications
- Water Supply and Leaks
- Company Work and Maintenance
- Location and Boundaries
- Water Quality
- Infrastructure
- General and Miscellaneous
- Wastewater

However, we believe that as wholesalers will naturally have more expert understanding of their customer base and the types of contacts they receive, **a final list of categories (with clear definitions for each) should be devised and agreed on by wholesalers and Ofwat**. These could be further refined and tested during the shadow year. Note these categories do not need to replace any existing internal categories – wholesalers could undertake a re-coding exercise before delivering their sample to derive these categories from their own system codes.

We suggest the definitions for the ‘service type’ field should align with those used for the C-MeX survey (see Annex 2 of the C-MeX Guidance for 2020-25[18] – note the ‘billing’ category is not applicable for the B-MeX survey as this billing enquiries would be handled by the retailer).

Non-essential data fields

The following fields are not essential, as we believe the survey can be run successfully without this information if necessary; however, **including these additional fields would greatly enhance the success rate of the survey**:

- Business name (name of the business contact is with)
- Business address (split across standard multiple fields: address line 1, address line 2, town/city, postcode)
- Contact job title (job title of the person the contact is with)
- Contact email address (to be populated whenever possible)
- Contact channel (Email; telephone; web / social media; internal / operational; physical / paper)
- Retailer (if it is possible for wholesalers to provide this information, this will help the telephone interviewers to clarify with respondents who is their retailer vs. wholesaler, to ensure they are feeding back about the correct organisation).

MOSL should supply the following fields from the bilateral hub:

- Retailer
- Wholesaler
- Orid

[18] https://www.ofwat.gov.uk/wp-content/uploads/2020/03/C-MeX-guidance-for-2020-25.pdf

- Process
- DateRaised
- processname
- RequestStatus
- ActivityStatus
- SupplyPointId
- cusotmername
- cusotmerbannername
- ADDRESS_LINE_1
- ADDRESS_LINE_2
- ADDRESS_LINE_3
- ADDRESS_LINE_4
- ADDRESS_LINE_5
- POSTCODE
- CustomerContactRequired
- ConsentToContactCustomer
- CustomerContactName1
- CustomerContactNumber1
- CustomerExtension1
- CustomerContactName2
- CustomerContactNumber2
- CustomerExtension2
- CustomerContactEmail
- CustomerAwareOfServiceRequest
- CustomerPreferredMethodOfContact
- CustomerPreferredContactTime

Due to the greater potential for respondents from the indirect sample to have difficulty differentiating between their retailer and wholesaler when thinking about the contact they had, we suggest adding a field that gives more information on the role of the wholesaler in the contact/interaction; this can be inferred by some codes in the "processname" field, for example "Retailer initiated operational process related to the Meter Repair or Replacement work" could be used to prompt the customer that the wholesaler would have undertaken the repair or replacement of their meter. However, other codes are less useful, e.g. "Retailer initiated operational process related to the Non Household Customer Enquiry", which would not allow the telephone interviewer to help prompt the respondent on their contact with the wholesaler.

Eligibility and Exclusions

Wholesalers should provide data for all business customer contacts. The sample should exclude the following:

- Any contacts related to household queries
- Any contacts relating to non-wholesaler activities – e.g. calls regarding a customer's retailer
- Contacts dealt with or with regard to developer services
- Wrong numbers, including calls where a customer has contacted the wrong company
- Contacts where the customer has called about a non-appointed activity and the call has no connection with the appointed business – e.g. insurance services and plumbing, private septic tanks/cesspits, highway gullies, hot water issues where it is confirmed there is no issue on the cold water supply
- Contacts about recreational and amenity activities – e.g. water skiing or angling facilities at visitor sites
- Enquiries from CCWater on behalf of a customer
- Any outbound contacts
- Any contacts related to bilateral transactions

Duplicates should not be removed – these will be checked and removed by the Agent.

An accompanying table should be provided giving the total contacts contained in the sample file, and a table of exclusions, listing the total number of records that have been excluded within each of the above categories.

Sampling dates and frequency

We recommend that sample is **delivered on a monthly basis**, with all eligible contacts for the previous calendar month provided to the Agent within 5 working days of the end of that month. MOSL should provide sample from the bilateral hub on the same schedule and applying the same eligibility criteria.

We believe a monthly delivery provides a good balance between the need to minimise the gap between the initial contact and the survey interview, ensuring we can obtain enough interviews with smaller wholesalers to generate robust results at an annual level, while also ensuring we do not over-burden wholesalers. The pilot suggests that a monthly sample delivery will be appropriate in terms of customers being able to recall their contact and provide meaningful feedback.

Reducing the frequency would mean that for smaller wholesalers we would either a) need to interview customers more than a month after the contact took place, risking their ability to recall the interaction and provide meaningful feedback; or b) not be able to achieve a sufficient number of interviews across the year for robust results.

We recommend that once an individual customer has been interviewed within a month, they should then be excluded from the sample for the following two months, to ensure customers are not overburdened. However, based on the qualitative findings, customers indicated they were happy to be surveyed repeatedly, as they feel it is important to feed back on their wholesaler; therefore, we believe it will be acceptable to contact individuals once every three months if necessary (all individuals can of course opt out of the survey at any point).

Targets – by wholesaler

We recommend setting quotas to ensure that all wholesalers achieve a sufficient base for analysis over the course of the year, and to ensure that results are representative of the contacts received by each wholesaler within the contact period.

We recommend aligning targets per wholesaler with C-MeX, with a target of 200 interviews per quarter / 800 per annum per wholesaler, where contact volumes allow, as a minimum. Among wholesalers with insufficient sample volumes to achieve 800 interviews, we recommend taking a census approach, where all sample is loaded for fieldwork, to achieve as many interviews as possible.

However, based on projections using the pilot sample volumes and response rates, we believe for some wholesalers it would be possible to achieve more than 800 interviews per year; this would reduce the confidence intervals for the satisfaction scores for these wholesalers, giving us greater confidence in the results. For some of the smaller wholesalers, we have also suggested potential increased targets based on more optimistic projections (e.g. assuming those with response rates lower than the average in the pilot may be able to increase these as a result of improved sample quality in the shadow year / mainstage).

We recommend setting more ambitious targets if the budget allows; for the smaller wholesalers these should be set on a 'best effort' basis, as achieving these assumes an increase in response rates which will only be possible if the sample quality is improved.

We also suggest liaising with smaller wholesalers to sense check the likely sample volumes, based on their knowledge of the number of eligible contacts they are likely to receive each year; should the estimates below be underestimates, we recommend increasing targets accordingly (up to a cap of either 800 or 1000 per wholesaler).

The tables overleaf shows our suggested annual targets per wholesaler, both overall (Table 5.1) and split by direct and indirect sample (Table 5.2).

The annual sample volume estimates assume that the month of pilot data represents a "typical" month, and therefore these figures have been reached by multiplying the usable pilot sample by 12. However, the following caveats should be considered:

- As noted previously, Severn Trent were unable to separate inbound and outbound calls, therefore the volume of sample supplied for the pilot is an overestimation. Nevertheless, the volume of completed surveys achieved in the pilot suggests these suggested targets would be achievable.
- The majority of direct sample records provided by South Staffordshire Water had no telephone number and were therefore not usable; should it be possible to include telephone numbers for all contacts in future sample deliveries, a target of around 100 – 130 direct completes per year may then be achievable, based on the average direct response rate.

Table 5.1 Suggested B-MeX annual responses and targets

	Total pilot sample	Pilot completes	Basic pilot response rate (RR)	Estimated annual sample available	Estimated maximum annual responses (using pilot RR)	Estimated maximum annual responses (using average RR)	Suggested target	Suggested increased target
Affinity Water	217	49	23%	2,604	599	391	500	600
Anglian Water	471	104	23%	5,652	1,300	848	800	1,000
Bristol Water	234	17	7%	2,808	204	411	300	400
Northumbrian Water	447	72	17%	5,364	912	805	800	900
Portsmouth Water	36	11	31%	432	134	65	100	130
SES Water	48	18	38%	576	219	86	150	200
Severn Trent Water	3,462	241	7%	41,544	2,908	6,232	800	1,000
South East Water	567	41	7%	6,804	476	1,021	700	1,000
South Staffordshire Water	43	10	23%	516	119	77	100	120
South West Water	394	53	13%	4,728	636	691	375	450
Southern Water	725	146	21%	8,700	1,827	1,305	800	1,000
Thames Water	590	116	20%	7,080	1,416	1,062	800	1,000
United Utilities	855	178	21%	10,260	2,155	1,539	800	1,000
Wessex Water	538	124	23%	6,456	1,485	968	800	1,000
Yorkshire Water	1,000	228	24%	12,000	2,880	1,800	800	1,000
Total	**9,627**	**1,408**	**15%**	**115,524**	**17,329**	**17,329**	**8,750**	**10,950**

Table 5.2 B-MeX projected annual responses and targets – direct vs. indirect sample

	Usable starting sample (pilot month)		Completes		Response rate		Estimated annual sample		Estimated max. annual responses (pilot RR)		Estimated max. annual responses (average RR)		Suggested max. annual target (rounded)	
	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect	Direct	Indirect
Affinity Water	187	30	44	5	24%	17%	2,244	360	528	60	310	78	400	100
Anglian Water	407	64	92	12	23%	19%	4,884	768	1,104	144	674	166	640	160
Bristol Water	222	12	17	0	8%	0%	2,664	144	204	-	368	31	275	25
Northumbrian Water	381	66	58	14	15%	21%	4,572	792	696	168	631	171	630	170
Portsmouth Water	30	6	10	1	33%	17%	360	72	120	12	50	16	75	25
SES Water	31	17	11	7	35%	41%	372	204	132	84	51	44	80	70
Severn Trent Water	3,266	196	205	36	6%	18%	39,192	2,352	2,460	432	5,408	508	730	70
South East Water	528	39	34	7	6%	18%	6,336	468	408	84	874	101	630	70
South Staffordshire Water	2	41	-	10	-	24%	24	492	-	120	3	106	3	97
South West Water	288	106	29	24	10%	23%	3,456	1,272	348	288	477	275	240	135
Southern Water	675	50	133	13	20%	26%	8,100	600	1,596	156	1,118	130	715	85
Thames Water	460	130	84	32	18%	25%	5,520	1,560	1,008	384	762	337	555	245
United Utilities	652	203	135	43	21%	21%	7,824	2,436	1,620	516	1,080	526	540	260
Wessex Water	535	3	123	1	23%	33%	6,420	36	1,476	12	886	8	790	10
Yorkshire Water	945	55	213	15	23%	27%	11,340	660	2,556	180	1,565	143	730	70
Total	**8,609**	**1,018**	**1,188**	**220**	**14%**	**22%**	**103,308**	**12,216**	**14,256**	**2,640**	**14,256**	**2,640**	**7,033**	**1,592**

The table below shows the standard error and confidence intervals associated with different sample sizes; these have been calculated using the average satisfaction score (6.9) and standard deviation (3.2) from the pilot survey. We have also estimated the likely weighting design effects, assuming an effective sample size of 80% of the achieved sample size (assuming the weighting used for the shadow year and subsequent running of the survey may be more complex than in the pilot).

Table 5.3 Estimated standard error and confidence intervals for proposed sample sizes, using the average pilot score of 6.9

Target sample size	Likely effective sample size*	Standard error	Confidence interval
1000	800	0.11	6.68, 7.12
900	720	0.12	6.67, 7.13
800	640	0.13	6.65, 7.15
700	560	0.14	6.63, 7.17
600	480	0.15	6.61, 7.19
500	400	0.16	6.59, 7.21
200	160	0.25	6.40, 7.40
150	120	0.29	6.32, 7.48
130	104	0.31	6.28, 7.52
120	96	0.33	6.25, 7.55
100	80	0.36	6.19, 7.61

This shows that for the largest proposed sample size (within our 'increased targets' option) would result in a standard error of 0.11 for an average score of 6.9, meaning that we can be 95% confident the actual score is between 6.68 and 7.12. At the opposite end of the scale, for the smallest proposed sample size of 100, we could be 95% confident the same score is between 6.19 and 7.61.

Should the shadow year need to run for less than one year due to timing constraints, these targets would be reduced accordingly, although for the largest wholesalers we could increase the monthly targets to still achieve a more robust base size. However, for wholesalers with a smaller number of monthly contacts, we would be unlikely to generate robust scores.

Targets – within wholesaler

We recommend setting targets by key variables to ensure that the achieved interviews align as closely as possible to the population (i.e. all contacts for that wholesaler in each month). We believe that setting targets in line with the population, rather than using any pre-set values (e.g. 50% direct, 50% indirect), will lead to the most robust and defensible scores, as the final score can be taken as representative of all actual contacts undertaken, rather than particular customer or contact types being artificially over- or under-emphasised.

We recommend the survey agent undertaking the shadow year reviews the sample provided by wholesalers to determine the most appropriate variables to set quotas on, as this will depend on the quality and comprehensiveness of the data fields provided. Based on the pilot analysis, and recommendations for sample collection and delivery going forwards, we would anticipate these may include the following:

- Direct vs. indirect sample

- Service type (water vs. wastewater vs. both/other)
- Outcome (whether issue was resolved at point contact was recorded)
- Reason for contact (assuming all wholesalers are able to code reasons to agreed list of categories)

Drawing sample

The sample will be checked for duplicates and any invalid records (i.e. records with no telephone number or contact name) will be removed. Where duplicate individuals are found within the file, the most recent contact will be selected.

For wholesalers where a census approach is taken, all valid sample will be loaded for fieldwork. For the remainder, contacts will be drawn at random within selected strata (based on the pilot this would be service type and outcome, however we suggest this is reviewed during the shadow year as this approach was necessitated by the inconsistent nature of the pilot sample).

Fieldwork

Weighting

As noted in the section on targets above, we believe survey responses should be representative of the population, i.e. all contacts undertaken by a wholesaler each month. This will be partly achieved by setting quotas on key variables for each wholesaler, which should lead to achieved interviews closely aligning with population proportions. However, in practice it is unlikely that all quotas will be achieved exactly during fieldwork, and therefore we recommend weighting the data to bring the final results in line with the population.

For the weighting to be meaningful, it will be necessary for wholesalers to make the recommended improvements to the recording and coding of contacts, so that this is harmonized and made consistent across wholesalers. The variables we derived and worked with for the pilot weighting exercise (Outcome and ServiceType) were used mainly for convenience as there was very little consistent data to work with. It is likely that better (and more) variables than this will be available as targets to weight to at wholesaler level if harmonisation, coordination, and management of the contact recording process takes place.

We would advocate, overall, a similar approach to that taken in the pilot: Rim weighting to population targets by wholesaler. These targets are likely to be the same targets set as quotas during fieldwork. We would recommend setting them the targets at the same level of granularity as the sample source used for each wave: so, if the population of contacts is produced monthly, targets should be set on a monthly basis.

Quota management will ensure the achieved interviews are as close to the target profile as possible, meaning only light-touch weighting should be required .

As with setting quotas, we would suggest reviewing and revising the strategy for weighting during the shadow year, when hopefully wholesalers will have been able to harmonise and improve the sample data supplied.

Questionnaire

We would suggest making the name of the respondent's retailer available to interviewers along with the name of the wholesaler, to provide additional context on the interaction and to avoid confusion. For indirect sample we recommend also adding an additional prompt to let them know their initial contact was via their retailer, however we are interested in their interaction with the wholesaler only. If it is possible to add a field in the indirect sample giving more information on the interaction the customer would have had with their wholesaler specifically, we recommend this is also added in as a prompt for indirect sample, to help them differentiate between their retailer and wholesaler, and understand the specific contact we want them to provide feedback on (as opposed to the entirety of their contact regarding the issue, which may have involved actions and contact with the retailer as well).

We also recommend reviewing question A3 ("How did the contact with <Water_Wholesaler> on <Service_Date> come about?") so that different options are provided to indirect vs. direct contacts (as we know that indirect contacts made contact via their retailer); however, we suggest retaining this question to obtain more detail about how the contact took place, and to check respondent understanding.

Aside from this we believe that the questionnaire worked well in the pilot, and we do not recommend making any further changes.

6 R-MeX: Methodology

Objectives

The objectives of this strand of the research were to design, test and pilot all aspects of the Retailer Measure of Experience (R-MeX) survey to ensure the survey methodology is robust for a full-scale shadow run in spring 2024. To achieve this the specific objectives were:

1. Design a pilot methodology for R-MeX.
2. Conduct fieldwork to pilot and test the R-MeX survey and reflect on findings.
3. Analysis of findings and presentation of recommendations and guidance for the shadow year.

Methodology

To meet the objectives outlined above, a mixed methods approach was adopted consisting of:

- Scoping phase
- Cognitive testing
- Pilot exercise
- Follow up qualitative research.
- Analysis and reporting

Scoping phase

An initial inception meeting took place with MOSL and Ofwat to provide IFF with greater context and understanding of the R-MeX survey, how it had worked so far, whether the existing R-MeX questions gave MOSL and Ofwat the information needed, and initial ideas for improvements to the survey. Owing to the relatively small number of retailers who are sent the survey to complete, 20, it was agreed a census approach would be taken.

Cognitive testing

IFF conducted a total of 10 cognitive interviews between 17th August and 19th September 2023. Interviews were conducted virtually via Microsoft Teams. They lasted 55 minutes on average (ranging from 47 minutes to 78 minutes each).

During the qualitative interviews retailers were asked about their overall thoughts of the R-MeX survey, their prior experiences of completing it, their views on each of the specific questions included in the survey, and their thoughts on whether their engagement with the R-MeX survey might change if scores collected in future waves were to feed into the BR-MeX incentive.

Ultimately, the cognitive testing exercise was conducted to help ensure that the R-MeX survey questions are clear and unambiguous, and that the survey is conducive to capturing retailers' views and experiences in an accurate and consistent manner.

The retailers that participated in cognitive interviews were largely positive about the R-MeX survey. Retailers said that the survey is important to them because it collects feedback on areas that are key to their organisation that is not collected through other vehicles. They felt that the questions were mostly clear, easy to understand, and easy to answer. Retailers also felt that the survey covered all the key topics that they wanted to give feedback on. Retailers all took steps to ensure that their use of the 11-point satisfaction scale was consistent *internally*, where multiple individuals were involved in the scoring; however, they did express concerns that other retailers might be interpreting the score in different ways.

Therefore, our recommendations from the cognitive phase of this research were fairly limited in scale; these were:

1. To issue the survey with some 'scoring guidance'. Based on the feedback collected, this could help reduce any potential inconsistencies in scores given across retailers, which could, in turn, help reduce the burden associated with completing the survey by reducing the time that retailers need to spend moderating responses across their firm.

2. To remove the macros from the Excel spreadsheet. Several retailers reported that they found the macros in the current Excel spreadsheet to be too restrictive. They felt removing them could make the process of completing the survey easier and less time consuming. Although a few retailers suggested that MOSL should consider a return to an online format, we do not think that this is feasible given how MOSL plans to administer the R-MeX in the future.

We also recommended adding a catch-all optional free text response box to the pilot survey only to capture retailers' views on how they found completing the revised survey.

At this stage of the research, we did not propose making any major changes to the question wording. Few retailers were unsure about what to include or exclude when responding to those questions. The vast majority of retailer feedback on the survey questions was positive and retailers generally felt that the questions were clear, easy to understand, and easy to answer. Retailers also felt that the questions covered all of the areas that were important to provide feedback on.

Pilot

The pilot also formed the 6th wave of the R-MeX survey, hence it involved a full rollout of the pilot R-MeX survey to all eligible retailers in October 2023. The recommendations from the cognitive stage were implemented in the questionnaire, including the optional free text box inviting comments on the revised survey itself.

To mimic actual data collection and conditions, the survey was administrated by MOSL in line with previous waves. 17 out of 20 eligible retailers took part in the pilot with MOSL providing the full results to IFF for analysis.

Qualitative research

Following the completion of the survey, IFF conducted a total of 5 qualitative post-pilot follow-up interviews with retailers between the 19th of December 2023 and 12th of January 2024. Interviews were conducted virtually via Microsoft Teams and lasted 30 minutes on average with both retailers who had and had not taken part in the cognitive research.

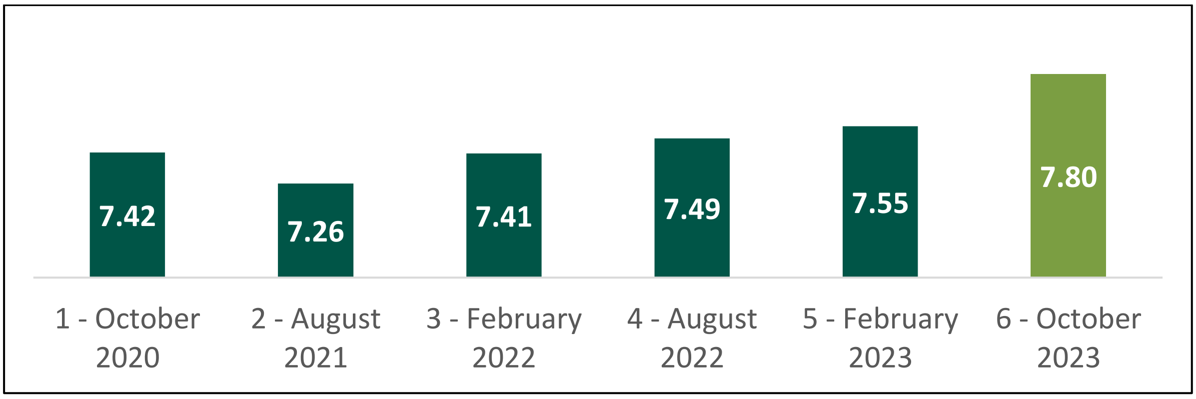
Following the qualitative interviews, analysis was conducted of the findings of these interviews and the R-MeX results, including the new optional feedback question at the end.

7 R-MeX: Pilot Analysis

Survey results

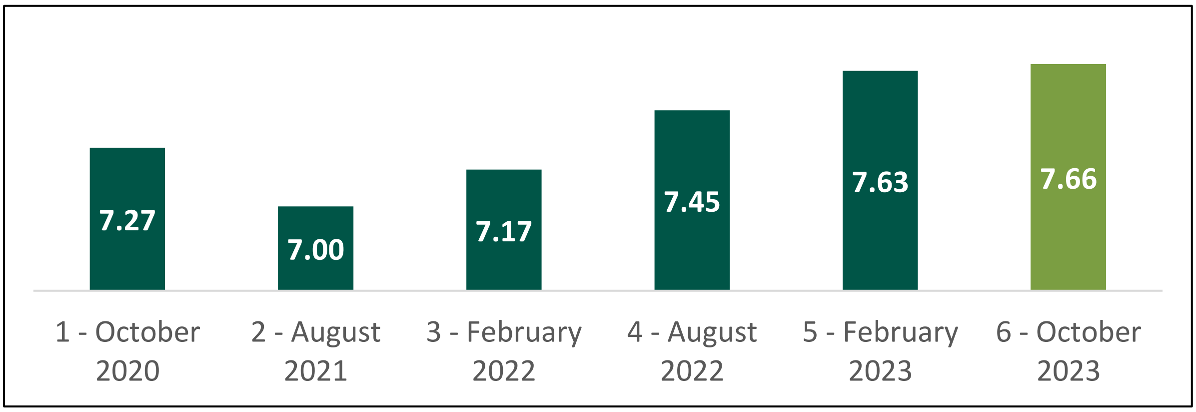
Average scores for overall service, and each of the 6 elements of service measured, have broadly remained stable with some upward trends since the launch of R-MeX.[1] The latest wave including the changes made for the pilot mostly aligns with this trend, as shown in 0 to Figure 7.6.This suggests the questionnaire changes did not cause a marked change in the way retailers scored wholesalers.

Figure 7.1 Average score for overall service.



Base: 1 – October 2020 (n=14), 2 – August 2021 (n=14), 3 – February 2022 (n=13), 4 – August 2022 (n=16), 5- February 2023 (n=15) and 6 – October 2023 (n=17)

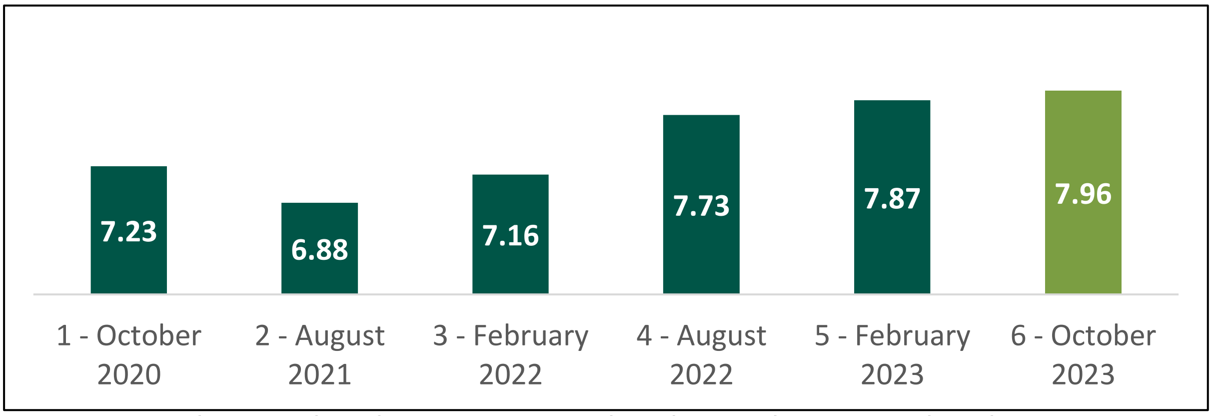
Figure 7.2 Average score for responsiveness.



Base: 1 – October 2020 (n=14), 2 – August 2021 (n=14), 3 – February 2022 (n=13), 4 – August 2022 (n=16), 5- February 2023 (n=15) and 6 – October 2023 (n=17)

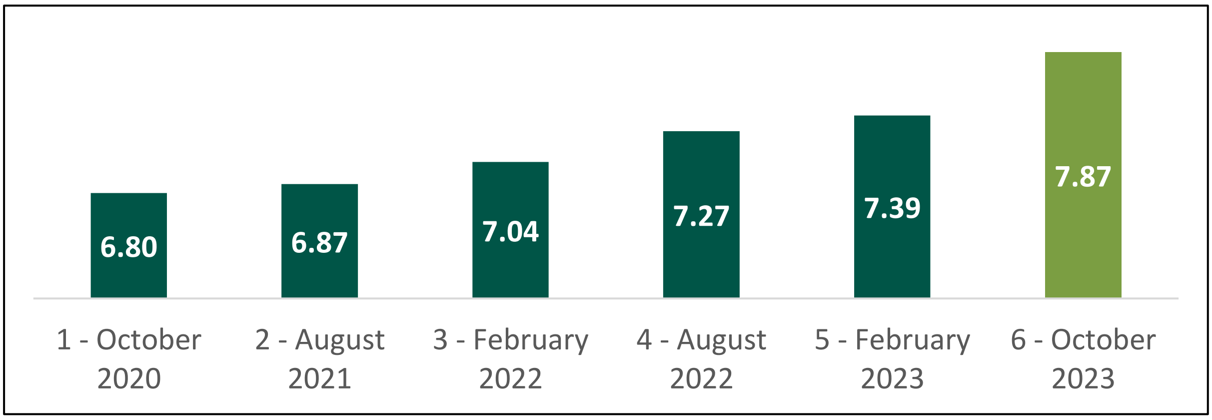
[1] MOSL have published R-MeX results here.

Figure 7.3 Average score for communication during incidents.



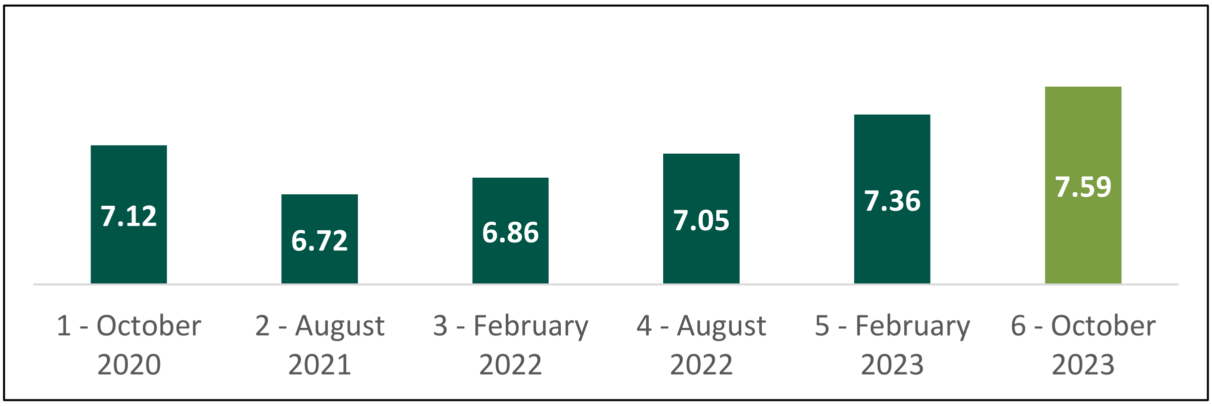
Base: 1 – October 2020 (n=14), 2 – August 2021 (n=14), 3 – February 2022 (n=13), 4 – August 2022 (n=16), 5- February 2023 (n=15) and 6 – October 2023 (n=17)

Figure 7.4 Average score for quality of data maintenance and improvement.



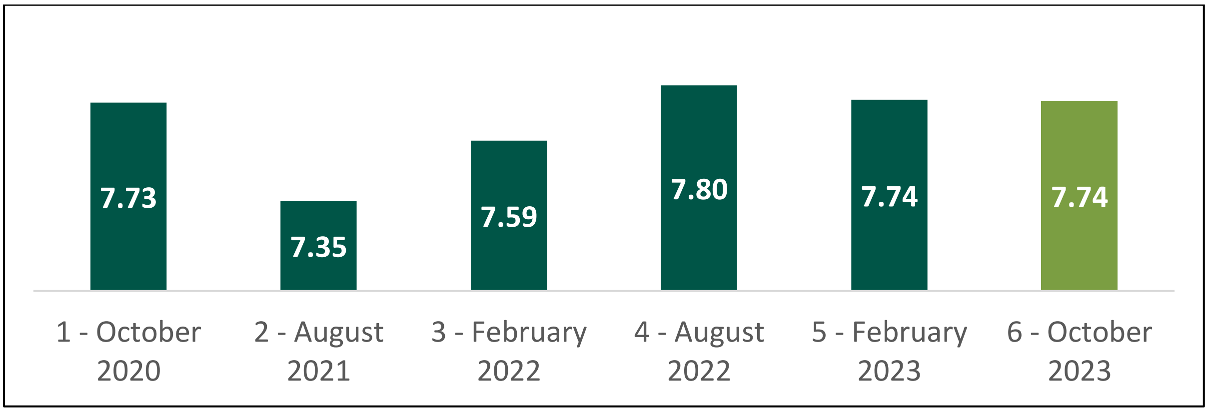
Base: 1 – October 2020 (n=14), 2 – August 2021 (n=14), 3 – February 2022 (n=13), 4 – August 2022 (n=16), 5- February 2023 (n=15) and 6 – October 2023 (n=17)

Figure 7.5 Average score for effectiveness of systems and notifications.



Base: 1 – October 2020 (n=14), 2 – August 2021 (n=14), 3 – February 2022 (n=13), 4 – August 2022 (n=16), 5- February 2023 (n=15) and 6 – October 2023 (n=17)

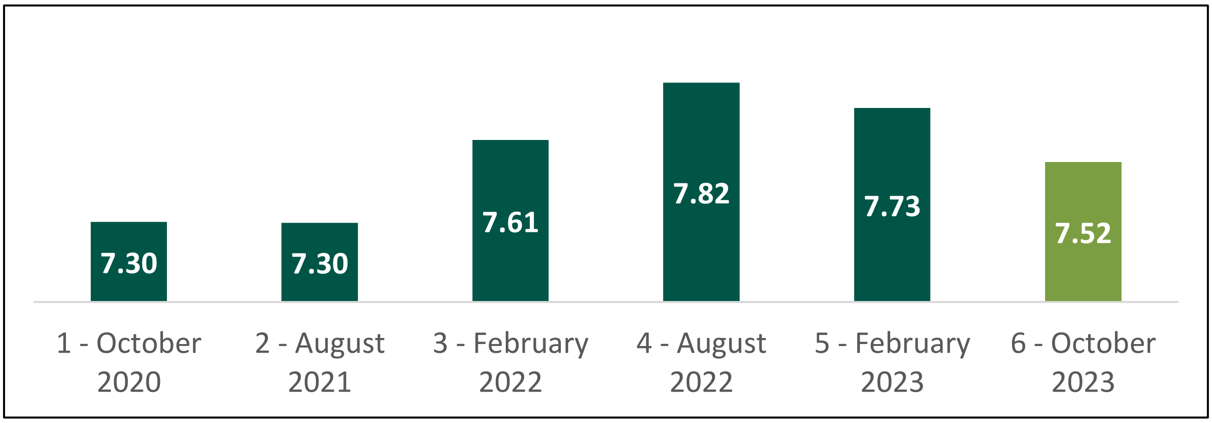
Figure 7.6 Average score for engagement and support.



Base: 1 – October 2020 (n=14), 2 – August 2021 (n=14), 3 – February 2022 (n=13), 4 – August 2022 (n=16), 5- February 2023 (n=15) and 6 – October 2023 (n=17)

In contrast to the other measures, the score for effectiveness of financial policies decreased compared to the previous waves, although this appears to be continuing a downward trend since the August 2022 survey, as shown in Figure 7.7. This is consistent with the view that the changes made to the questionnaire for the pilot R-MeX did not have the effect of comprehensively increasing or decreasing retailer scores across the board.

Figure 7.7 Average score for effectiveness of financial policies.

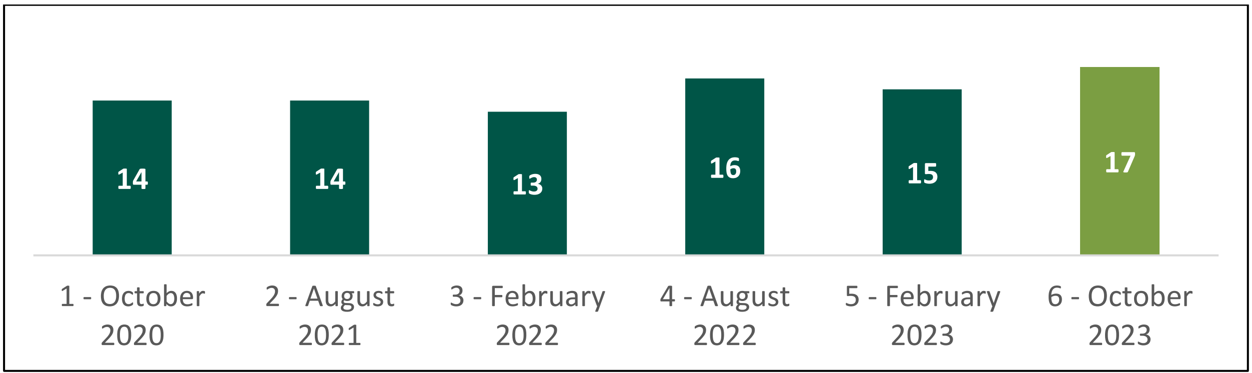


Base: 1 – October 2020 (n=14), 2 – August 2021 (n=14), 3 – February 2022 (n=13), 4 – August 2022 (n=16), 5- February 2023 (n=15) and 6 – October 2023 (n=17)

Participation in pilot

Participation in the pilot survey also remained high and consistent with previous surveys as shown in Figure 7.8 overleaf.

Figure 7.8 Number of retailers taking part in R-MeX.



Base: 1 – October 2020 (n=14), 2 – August 2021 (n=14), 3 – February 2022 (n=13), 4 – August 2022 (n=16), 5- February 2023 (n=15) and 6 – October 2023 (n=17)

There is a small upward trend in the number of retailers completing R-MeX surveys each wave overall with only 55% of potential responses completed in survey 2[2] compared with 71% in the latest and pilot wave of the survey.

Figure 7.9 Percentage of R-MeX survey responses achieved by retailers per wholesaler.

Survey responses received.	1 – October 2020	2 – August 2021	3 – February 2022	4 – August 2022	5 – February 2023	6 – October 2023
Affinity Water	47%	56%	44%	71%	71%	77%
Anglian Water	59%	67%	50%	77%	77%	82%
Bristol Water	47%	56%	44%	65%	59%	71%
Northumbrian Water	53%	50%	50%	71%	65%	77%
Portsmouth Water	31%	38%	41%	56%	63%	50%
SES Water	22%	44%	38%	56%	50%	44%
Severn Trent Water	50%	56%	63%	83%	72%	78%
South East Water	29%	53%	39%	53%	47%	65%
South Staffordshire Water	35%	50%	53%	71%	65%	71%
South West Water	60%	40%	53%	71%	65%	71%
Southern Water	25%	44%	50%	65%	71%	77%
Thames Water	53%	67%	61%	82%	71%	82%
United Utilities	53%	72%	61%	77%	77%	82%
Wessex Water	41%	44%	44%	59%	53%	59%
Yorkshire Water	53%	50%	50%	77%	71%	82%
Total	**45.3%**	**54.7%**	**50.0%**	**69.2%**	**65.4%**	**71.3%**

[2] Note that Survey 1 was sent to all retailers in the market, this was then revised for the second and all future surveys; therefore, this completion rate for Survey 1 is not comparable.

Reflections from qualitative follow-up

Awareness of changes

Among the 5 post-pilot qualitative interviews, most retailers reported they had noticed the changes introduced, although some had not. The majority agreed that both changes were beneficial, including those who hadn't noticed the changes spontaneously.

"[Removal of macros] made it so much easier. There were no formula issues. Historically we've had real problems."

Retailer

One comment left in the pilot survey at the additional optional survey feedback question also acknowledged one of the changes around the submission method and the removal of the macro.

"Thank you for changing the submission method!"

Retailer

Feedback on the scoring system

Feedback in the qualitative interviews suggested the scoring system works well and indicated that adding scoring labels has given retailers more confidence that their scores will be consistent with other retailers. One retailer mentioned they had already developed their own scoring labels prior to the pilot, and these matched the labels added into the pilot survey. This reinforced findings from the cognitive stage that retailers were mostly confident with the 10-point scale.

"We'd already come up with our own [scoring labels] and definition of the scores, which was in line with what came out."

Retailer

The minimal changes witnessed in average scoring in the pilot survey both overall and across individual service areas suggests most retailers did not change their own scoring in response to the new labels, despite perceptions in the cognitive interviews that retailers had not previously interpreted the scale in the same way.

Feedback on the Excel spreadsheet

Most retailers said they had noticed the changes introduced to the spreadsheet and the majority agreed that these were beneficial, including those who hadn't noticed the changes spontaneously.

"From an admin perspective, when you're pulling this together, you're doing it for [multiple] wholesalers. It was quite a cumbersome, time-consuming process to do that, being able to collate it all and in one Excel document with no fancy macros there was just so much easier… I'd say we are saving a couple of hours."

Retailer

Impact of changes

The overall impact of the changes was mostly perceived to be small, in line with the relatively minor nature of the changes made. Nonetheless retailers found both changes beneficial and therefore the overall impact was positive; there was low appetite for further changes.

"You already get a decent sort of view on a wholesaler performance, so I think it's adequate."

Retailer

Suggestions for further improvements

Only 5 of the 17 retailers who completed the R-MeX pilot survey left comments at the additional optional survey feedback question. Four of the comments included suggestions, these were: a request for a spell check feature, further clarity in scoring (unspecified), an alternative method to submit via an online form, and requesting an additional question on the importance of engagement and collaboration around water efficiency.

In the qualitative interviews, some retailers suggested changing the document further by enabling the spellcheck function in the spreadsheet, a feature pre-built into excel but currently unavailable in protected worksheets.

"It's a professional document, you've got pride in its presentation. I copy and paste it into word, spell check it and then paste it back into excel. That shouldn't be needed. It's protected in such a way that you can't do the spellcheck, I don't see why it needs to be."

Retailer

Whilst the scoring labels improved confidence, some retailers also noted that there is still some subjectivity around these and what each label could mean to each retailer. Therefore, whilst the retailers we spoke to reported understanding the labels and how to score they believed it could still leave some potential for inconsistency amongst others. To overcome this, some suggested further guidance could be created such as an example scoring guidance document.

"What does a model answer look like, what are the scores that MOSL are expecting wholesalers to achieve? I think if we had that it would give some sound guidance to retailers as to whether they're underscoring or overscoring and take away some of the ambiguity that exists around that."

Retailer

Statistical analysis of scoring and participation

Although retailers were happy with the 0-10 scale as a scoring mechanism, findings from the pilot were inconclusive as to whether the scale was being interpreted consistently across retailers. Findings from the cognitive stage suggested that different retailers may be taking different approaches to assigning scores on the scale. While feedback on the labels introduced in the pilot was positive, the fact that average scores remained in line with previous R-MeX scores and trends suggests the labels did not impact retailers' interpretation of the scale, and therefore differences in approach could still remain.

If retailers do not all use the scale in a consistent way, this could open potential risks in terms of the reliability of the final scores; for example, if some retailers tend to give higher or lower scores due to differences in their interpretation of the scale, this could skew the final scores for individual wholesalers, depending on which retailers they receive scores from.

Therefore, IFF undertook further statistical analysis of the pilot R-MeX data to review the robustness of the scores for use in the final BR-MeX incentive.

Rating scales can, if not carefully moderated, elicit difference response styles from different individuals:

- **The Extreme or 'E' effect**: Tendency to rate closer to the extremes of the scale (0 and 10) vs the middle (5). This effect results in some respondents having a higher standard deviation in their scores across objects (in this case, wholesalers) being rated.

- **The Level or 'L' effect:** Tendency to rate on average lower or higher on the scale (i.e. a tendency to use the top, upper middle, lower middle or bottom part of the scale). All other things equal, this could result in some respondents giving averagely higher or lower scores than others, regardless of the performance of the object being rated.

These response styles tend to be more noticeable in more granular scales, and when specific descriptions are not attached to each part of the scale. Over large samples, these differences would tend to 'even out' naturally. We note here that as R-MeX only has a small sample size, such differences, if present, could have a bigger effect on wholesalers' final scores.

For example, if Retailers A and B always tend to give higher scores, and Retailers C and D always tend to give lower scores, a wholesaler who is scored by only Retailers C and D would end up with a lower average score than a wholesaler scored by Retailers A and B (and a wholesaler scored by all four would be in the middle), even if their actual performance was the same.

Our analysis of the results dataset suggested that at least some variance in the scores given by retailers could be driven by some differences or variations in how individual retailers interpreted the scale (rather than simply being due to differences in wholesaler performance). Therefore, we would recommend using alternative options for moderating scores rather than using the raw R-MeX satisfaction scores as a primary input into the final BR-MeX incentive,. We explore some options below.

Alternative scoring methods

Alternative scoring systems were considered and tested on the R-MeX pilot data. Based on this, we would recommend one of two options:

- **Option 1: Less granular scale:** Adopt a less granular scale with detailed descriptions.

- **Option 2: 'Centering':** Adjust the raw overall satisfaction score to neutralise the impact of any inconsistencies in retailer response style.

For **Option 1**, we would suggest a four-point scale, and move away from comparing mean scores and instead focus on proportions in top box, or a combination of top and second top box. A detailed key is recommended to provide guidance to retailers on when to rate in each box, to ensure consistent use of the scale across retailers. A suggested four-point scale would be as follows:

Segment	Score
Outstanding	1
Good	2
Requires improvement	3
Unsatisfactory	4

The advantages of this approach are that forcing a limited number of options with more prescriptive instructions would facilitate consistent use of the measuring instrument and minimise any response style differences, and therefore provide a more stable, robust method of comparison. The disadvantages would be a lack of discrimination between the performance of wholesalers due to the less granular scale, the need to be more prescriptive in guiding retailers on how to respond, and moving away from the 0-10 score used across the other MeX measures.

For **Option 2,** we suggest adjusting the raw 'overall satisfaction' score by 'centering' the score for each retailer (of each wholesaler). This is done by calculating the mean average rating for each retailer across the scores for all the wholesalers it has rated, and then transforming the score for each individual wholesaler by subtracting the mean from their raw score, effectively converting it into a 'difference from mean' score. Wholesalers are then ultimately compared in terms of how far above or below average they rated for each retailer.

For example, Table 7.1 below shows an example of calculating adjusted scores for four wholesalers given by a single retailer. The mean average score given by this retailer is 7.25; therefore, the adjusted score for each wholesaler is created by subtracting that mean (7.25) from their raw score. For a wholesaler with a raw score of 9, this results in an adjusted score of 1.75; for a wholesaler with a raw score of 5, this results in an adjusted score of -2.25.

Table 7.1 Example score adjustments (for a hypothetical single retailer)

	Raw satisfaction score	Calculation *(mean score subtracted from raw score)*	Adjusted score
Wholesaler A	9	9 – 7.25	1.75
Wholesaler B	7	7 – 7.25	-0.25
Wholesaler C	5	5 – 7.25	-2.25
Wholesaler D	8	8 – 7.25	0.75
Mean average	**7.25**	-	-

The final adjusted score for each wholesaler would then be the mean average of their adjusted score from each retailer. For example, as shown in Table 7.2, Wholesaler A received scores from four retailers, ranging from 1.75 to 2.25, with the mean average of those scores being 1.56 (that is, Wholesaler A received scores that were on average 1.56 points higher than the average for each retailer who scored them). Wholesaler C received scores from three retailers, and their final score is the average of those three scores.

Table 7.2 Example calculation of final adjusted scores per wholesaler

	Adjusted scores				
	Retailer 1	Retailer 2	Retailer 3	Retailer 4	Mean average
Wholesaler A	1.75	1.25	1.0	2.25	**1.56**
Wholesaler B	-0.25	-0.5	-1.25	1.25	**-0.19**
Wholesaler C	-2.25	-1.75	N/A	-2.0	**-2.00**
Wholesaler D	0.75	1.0	0.75	1.25	**0.94**

These adjusted scores can then be indexed and weighted for inclusion in the overall BR-MeX score.

The advantages of this approach are that it allows the retention of the 0-10 rating scale and facilitates a fair comparison of the 'relative' ratings of wholesalers. The disadvantage is that the differences in 'absolute' scores for a specific wholesaler given across retailers are lost (although these differences may not be reliable in some cases if they result from different response styles).

Impact of adjusted scoring on pilot R-MeX scores

Table 7.3 below shows the raw scores and adjusted scores for each wholesaler,[3] alongside an indexed version of each score, based on the October 2023 R-MeX data. The table is ranked according to the raw overall mean scores, while the rank column under 'adjusted scores' gives the rank for each wholesaler based on the adjusted scores.

Table 7.3 Untransformed R-MeX scores and ranking

	Raw scores			Adjusted scores			
Wholesaler	Mean score	Indexed score	Rank	Mean score	Indexed score	Rank	Rank change
Affinity Water	8.77	91	1	0.93	89	1	=
United Utilities Water	8.36	50	2	0.56	53	2	=
Portsmouth Water	8.25	39	3	0.42	38	3	=
Yorkshire Water	8.14	29	4	0.35	32	4	=
Anglian Water	8.00	14	5	0.11	8	8	-3
Northumbrian Water	8.00	14	6	0.17	14	5	+1
South West Water	7.92	6	7	0.11	8	9	-2
Wessex Water	7.90	4	8	-0.03	-6	10	-2
Bristol Water	7.83	-2	9	0.12	9	7	+2
Southern Water	7.77	-9	10	-0.07	-11	11	-1
Sutton and East Surrey Water	7.71	-14	11	0.14	11	6	+5
South Staffordshire Water	7.58	-27	12	-0.21	-24	12	=
South East Water	7.18	-67	13	-0.56	-59	13	=
Severn Trent Water	6.86	-100	14	-0.84	-87	14	=
Thames Water	6.79	-107	15	-1.01	-104	15	=

The top scoring wholesaler and bottom four scoring wholesalers remain consistent using either scoring method. Overall, 10 of the 15 wholesalers either kept the same rank or moved up or down by only 1 place as a result of the score adjustments.

[3] Dwr Cymru Welsh Water and Hafren Dyfrdwy were included in the survey but are not included in this table as they did not receive enough responses.

Among the retailers dropping more than one place:

- Anglian Water rated well with ‘high score’ retailers and low among the more numerous ‘mid score’ retailers, therefore centering markedly reduces its relative score;
- South West Water and Wessex Water rated generally high with ‘high score’ retailers, therefore centering slightly reduces their relative score.

Two wholesalers were promoted by more than one rank after centering:

- Bristol Water was rated well with both high and low scoring retailers, therefore centering raises its relative score;
- Sutton and East Surrey Water rated well with the more numerous ‘mid scoring’ retailers and relatively poorly with the ‘high scoring’ retailers, therefore centering markedly raises its score.

8 R-MeX: Recommendations for Future Iterations

Methodology

Feedback collected from the retailers that participated showed they are largely content with the current R-MeX survey. They reported that is fit for purpose with the minor changes made for the pilot and that both changes were beneficial. Retailers reported that the R-MeX survey covers all the areas that it is important to give feedback on. Moreover, several felt the R-MeX survey was very important because it is the only vehicle through which a lot of this feedback is collected. Based on this, **we recommend that MOSL maintains the current methodology.**

Sample design

Survey frequencies

Most retailers were positive and in favour of the current frequency of the survey. Most expressed concerns about being asked to complete the survey more frequently because they felt this would be too burdensome. Some also felt that administering the survey more frequently would reduce the window in which wholesalers could make improvements, therefore undermining the ability of the R-MeX process to enable wholesalers to action feedback in a meaningful way. **We therefore recommend that the frequency of the R-MeX survey is not changed.**

Weighting

A range of views were given by retailers when they were asked whether they felt results from the R-MeX survey should be weighted. Most said they were unsure.

Those who were in favour of the results being weighted felt that questions about customer service and unplanned events should receive a higher weight. On the other hand, those against weighting said that this could result in wholesalers disproportionately focusing their efforts on the areas that receive a higher weighting going forward. Similarly, retailers of all sizes were also unsure if weighting should be applied at a retailer level for example by giving retailers who shared a larger customer base with a given wholesaler a greater weight; and cautioned that any weighting by retailer may cause wholesalers to focus attention on retailers with a larger weighting and neglect others.

We would recommend not weighting the R-MeX survey, as any of the proposed weighting scheme would be artificially inflating certain scores and could encourage wholesalers to focus their efforts disproportionately. We feel the scores will be more robust and less open to dispute if weighting is not applied.

Questionnaire

We recommend against making any further changes to the question wording. The vast majority of retailer feedback on the survey questions was positive with limited feedback for further changes. They generally felt that the questions were clear, easy to understand, and easy to answer. Retailers also felt that the questions covered all of the areas that they felt important to be able to provide feedback on.

Scoring adjustments

To the extent that concerns remain that not all retailers are interpreting the 0-10 scale in a consistent way and noting that any differences in approach to scoring are less subject to evening out across a large sample, we would recommend use of the ‘Option 2’ ('centering') scoring adjustment , whereby the overall satisfaction scores for each wholesaler are centered before being fed into the BR-MeX score. We highlight that even with this option, wholesalers should still be able to see their ‘raw’ overall score for reference, and the raw scores given for the individual service elements (which will not feed into the BR-MeX incentive), alongside all feedback comments given by retailers.

9 Annex A.

BR-MeX Pilot - Initial Sample Queries / Points for discussion

To conduct the B-MeX pilot, we believe we would like to receive sample with the following fields for **each direct contact** within a set timeframe:

Field	Reason for request	Notes
Any unique customer ID	For deduping	
Business name	For deduping; To assist reaching the relevant person for the survey	
Business address	For deduping; To assist reaching the relevant person for the survey	
Firmographic information	To assist setting survey quotas	Any information on business size, sector etc.
Contact name	To assist reaching the relevant person for the survey	
Contact job title	To assist reaching the relevant person for the survey	Not essential but helpful if available
Contact phone number	For deduping; To assist reaching the relevant person for the survey	
Contact email address	For deduping; To assist reaching the relevant person for the survey	
Mode of contact	To prompt respondent recall during the survey; potentially to determine initial survey mode of contact	e.g. telephone, online
Date of contact	To determine eligibility; to prompt respondent recall during the survey	Ideally in a standard date format
Type of contact	To determine eligibility	Categories TBD
Reason for contact	To determine eligibility; to prompt respondent recall during the survey; potentially to use in analysis (to give more context to responses / satisfaction)	Ideally this would be in a coded format (rather than free text)
Who the customer spoke to	To prompt respondent recall during the survey	
Outcome of contact / whether issue resolved	Potentially to use in analysis (to give more context to responses / satisfaction)	Not essential but helpful if available / if applicable

Questions we would like to ask:

Sample fields and format

- Could you provide all of the above fields? If not, which would / would not be feasible, and why?

  - Do you think other wholesalers could provide this information?
- What format could you provide this in? (e.g. Excel file or other format; coded data or free text).
- Where information is held, what is the quality of this? (e.g. held for 100% of records, only held in some cases, held in an inconsistent format, etc?
  - This is particularly critical for the contact fields (contact name or job title, phone number email address) – what proportion of records contain a) ALL this information, b) SOME of this information, c) NONE of this information?
- Do you have any other information stored for each contact that could be useful for us – what and why?

Types of contact

- What different types of contact do you hold? How are these categorised, if at all?
- What types of contact do you feel should be eligible for the survey, and why?

Sample size

- How many contacts do you receive in an average month?
  - How much does this vary across the year?
  - How much is this likely to vary between wholesalers, to your knowledge?

Sample delivery and timings

- How soon would you be able to deliver a sample file after we make a request?
- What would be your preferred method to securely transfer the sample to us?
  - Any GDPR concerns? Any preferences / requirements re: data sharing agreements?
- Would it be feasible to prepare and transfer sample twice per year?
  - How much burden would it place on the business to prepare and transfer this sample?

Any other comments / questions?

"

IFF Research illuminates the world for organisations businesses and individuals helping them to make better-informed decisions."

Our Values:

1. Being human first:

Whether employer or employee, client or collaborator, we are all humans first and foremost. Recognising this essential humanity is central to how we conduct our business, and how we lead our lives. We respect and accommodate each individual's way of thinking, working and communicating, mindful of the fact that each has their own story and means of telling it.

2. Impartiality and independence:

IFF is a research-led organisation which believes in letting the evidence do the talking. We don't undertake projects with a preconception of what "the answer" is, and we don't hide from the truths that research reveals. We are independent, in the research we conduct, of political flavour or dogma. We are open-minded, imaginative and intellectually rigorous.

3. Making a difference:

At IFF, we want to make a difference to the clients we work with, and we work with clients who share our ambition for positive change. We expect all IFF staff to take personal responsibility for everything they do at work, which should always be the best they can deliver.

IFF Research

5th Floor
St. Magnus House
3 Lower Thames Street
London
EC3R 6HD
Tel: +44(0)20 7250 3035
Website: iffresearch.com

Contact details: