

Accent

 PJM economics

Outcome Delivery Incentive Research: Testing and Development

Stage 2 report

Contact: Paul Metcalfe
E-mail: Paul.metcalfe@pjmeconomics.co.uk
Telephone: +44 (0) 7786 656834

File name: 3524_Stage2Report_v2.docx



Registered in London No. 2231083
Accent Marketing & Research Limited
Registered Address: 30 City Road, London,
EC1Y 2AB

Contents

Executive Summary	3
1 Introduction	5
1.1 Background	5
1.2 Objectives and scope	5
2 Overview of Survey Design	6
2.1 Core components	6
2.2 Impact exercise	6
2.3 Compensation exercise	8
2.4 Additional aspects tested	9
3 Overview of Testing and Development	10
3.1 Stages of testing and development	10
3.2 Cogs 1	11
3.3 Pilot 1	12
3.4 Cogs 2	16
3.5 Pilot 2	18
3.6 Cogs 3	20
4 Conclusions	22
APPENDIX A	Cognitive Testing (Wave 1) Report
APPENDIX B	Pilot 1 Report
APPENDIX C	Cognitive Testing (Wave 2) Report
APPENDIX D	Pilot 2 Report
APPENDIX E	Cognitive Testing (Wave 3) Report

Executive Summary

Introduction

Accent and PJM economics were commissioned jointly by Ofwat and CCW to develop and test a survey instrument and methodology for obtaining the customer evidence needed to support outcome delivery incentive (ODI) rate setting for common performance commitments (PC) at PR24.

This is the final report on Stage 2 of the study. Its focus is on documenting the testing and refinement undertaken prior to the implementation in the main survey. For each stage of testing, it details the key issues tested, the methodology used, the key findings, and the revisions made as a consequence. Full reports from each stage, plus the survey instrument versions used, are included as appendices to this report.

Overview of testing and development

Stage 2 of the study included the following rounds of testing with customers:

- **Cogs 1:** Cognitive depth interviews with 13 households and 5 non-household customers
- **Pilot 1:** Quantitative pilot with 1,058 household and 80 non-household customers
- **Cogs 2:** Cognitive depth interviews with 4 household and 2 non-household customers
- **Pilot 2:** Quantitative pilot with 402 household and 150 non-household customers
- **Cogs 3:** Cognitive depth interviews with 10 household and 8 non-household customers

A number of revisions were made at each stage, based on the findings from the research, and as a result of comments received from companies, the Environment Agency, Stantec (with regard to storm overflows and pollution incidents), and three academic peer reviewers/advisors: Profs. Ken Willis, Giles Atkinson and Stephane Hess. Their comments were invaluable in helping to refine the questionnaire and show materials.

Ofwat and CCW were highly engaged throughout the process via regular meetings to discuss comments received, findings from the customer testing, and other issues as needed.

Conclusions

The testing and development work undertaken throughout Stage 2, as described in this report, has been successful. Notwithstanding some difficulties uncovered during the testing process, it has resulted in a survey instrument that be considered to have performed sufficiently well to support its use within a main survey for obtaining the

evidence needed to support the setting of ODI rates for common PCs at PR24. Given the findings reported here, the survey can be expected to produce a strong and comparable evidence base on customer values for water and wastewater services for all companies.

1 Introduction

1.1 Background

Accent and PJM economics were commissioned jointly by Ofwat and CCW to develop and test a methodology for obtaining the customer evidence needed to support outcome delivery incentive (ODI) rate setting for common performance commitments (PC) at PR24. In commissioning a collaborative national study, Ofwat and CCW aimed to ensure comparability of results by applying a common methodology, and thereby identify genuine differences. Moreover, the study was believed to provide an opportunity to undertake an in-depth review of potential options and identify a best-practice approach that results in estimates that are high quality and fit for purpose.

The project has been structured as a two-stage study. The first stage was a review of methodology options to deliver research to inform ODIs, and the development of a preferred option/s for PR24 ODI research. The second stage was the development and testing of materials for use in this research. The delivery of the customer fieldwork constitutes a third stage, which did not form part of this commission.

1.2 Objectives and scope

The purpose of this report is to document the suite of testing work that has been undertaken for Stage 2 to develop and refine the survey instrument initially based on the recommendations from Stage 1.

- Section 2 presents an overview of the survey instrument, as recommended by the Stage 1 report.
- Section 3 summarises the testing and development work undertaken, including the issues tested, the methodologies used, and the findings and recommendations that resulted.
- Section 4 presents our conclusions from Stage 2 overall.

Additionally, a suite of appendices contain the full research reports, and survey instruments, from each wave of testing.

2 Overview of Survey Design

This section provides an overview of the survey design, including the various aspects of the survey instrument that were tested as part of Stage 2.

2.1 Core components

The Stage 1 report recommended that the survey questionnaire should be constructed to include two linked stated preference exercises:

- **Impact exercise**

A pairwise stated preference choice exercise, measuring the relative impacts of a wide range of service issue scenarios.

- **Compensation exercise**

A contingent valuation exercise asking participants to choose between experiencing a service issue and being compensated for it, or not experiencing the issue and not receiving any compensation.

These two exercises were intended to be linked together in the analysis to derive customer valuations for all of the service issues needed.

The remainder of this section describes the form of each exercise in turn, followed by an overview of the additional elements of the questionnaire that were tested as part of Stage 2. Note that certain features of the exercises, and service issue descriptions, were refined over the course of development. Indications are given where changes have been made.

2.2 Impact exercise

The first choice exercise focused on the impact of a range of different scenarios, including service issues at the customer's property, and environmental damages. The scenarios were to be combined in an experimental design to obtain the sequences of choices that participants would face in the survey. In each question, participants would be shown two scenarios, and asked to indicate which one would have the most impact on their household/organisation.

Participants would each see ten questions such as the one shown in Figure 1.

Figure 1: Impact exercise: example choice card

Which of these would have the most impact on your household?

Option A	Option B
<p>PLANNED water supply interruption (6 hours)</p> <ul style="list-style-type: none">▶ Your water company sends you a notice in the post that in 2 days' time your water supply will stop for 6 hours, affecting all taps, toilets, dishwasher, etc▶ This is due to planned maintenance in your local area▶ As planned, it then stops between 12:00 and 18:00 on a Wednesday afternoon   <p>Planned, 6 hours</p>	<p>Water taste and smell (6 hours)</p> <ul style="list-style-type: none">▶ Your tap water starts tasting or smelling different, without warning▶ This is due to chlorine, and the taste and smell is like a swimming pool▶ The water is safe to drink, and for use in the dishwasher or washing machine▶ This happens for 6 hours, between 12:00 and 18:00 on a Wednesday afternoon   <p>6 hours</p>

Source: Final (main stage) version of questionnaire

By varying combinations of service issues across the sequence of question, and sample, it is possible to measure an index of relative impact via econometric analysis.

To begin with, there were 25 service issues, all linked, either individually or in combination, to anticipated common PCs. This was subsequently revised to 26 service issues during the course of Stage 2 as a result of stakeholder testing, by dropping one of the service issues and adding two new ones.

The service issue descriptions also evolved as a consequence of customer and stakeholder testing. The full sets of showcards used at each stage of testing are included in the appendices to this report.

Key design issues for testing and refinement over the course of Stage 2 included:

- Wording of the introduction.
- Layout of the exercise and customers' ability to understand and choose.
- Descriptions of the service issues. (This included testing with companies and other stakeholders as well as with customers.)
- Performance of econometric models (quant pilots only).
- Consistency of results with expectations, i.e. in terms of the relative impacts of the various service issues (quant pilots only).

2.3 Compensation exercise

The purpose of the compensation exercise was to obtain an estimate of the amount of money required to compensate customers for service issues – a measure of the value of avoiding that service issue. A minimum of one service issue is needed for this exercise in order to obtain valuations for all attributes, with this service issue referred to as a ‘pivot’ because the valuations for all other service issues are obtained by pivoting off the values obtained from this service issue using the relative impacts derived from the impact exercise. However, the Stage 1 report recommended that two service issues should be included in the compensation exercise in order to provide two sets of estimates for sensitivity testing.

Figure 2 shows an example of a choice card from the survey. This figure, drawn from the final (main stage) survey instrument, shows a planned water supply interruption (6 hours) as the service issue being valued. Initially, the two service issues recommended for inclusion were: a short, unexpected supply interruption, and an external sewer flooding incident. These two were replaced during Stage 2 by a short planned water supply interruption and a boil water notice. This change was made as a consequence of the testing, described in the following section.

The figure also shows a compensation level of £25. The levels of compensation varied across customers in the testing of the survey instrument and were refined in the course of Stage 2 based on statistical criteria.

The exercise included four questions per participant: two for each of the two service issues. Amounts for an initial question in each case were varied across the sample, and this compensation amount was either halved or doubled in a follow-up question, depending on the response to the first question. This is the so-called ‘double-bounded contingent valuation’ method.

The compensation amounts were shown in pounds for households and as a percentage of the annual water and wastewater services bill for non-households.

Figure 2: Compensation exercise: example choice card

Which option would you prefer?

Option A	Option B
<p>PLANNED water supply interruption (6 hours)</p> <ul style="list-style-type: none">▶ Your water company sends you a notice in the post that in 2 days' time your water supply will stop for 6 hours, affecting all taps, toilets, dishwasher, etc▶ This is due to planned maintenance in your local area▶ As planned, it then stops between 12:00 and 18:00 on a Wednesday afternoon   <p>Compensation paid: £25*</p>	<p>No service issue</p> <ul style="list-style-type: none">▶ There would be no issue affecting the water service at your property

* Compensation would be paid automatically, and within 7 days, by crediting your bank account, if you have a direct debit set up, or by sending you a cheque otherwise

Source: Final (main stage) version of questionnaire

Key issues for testing and refinement over the course of Stage 2 included:

- Wording of the introduction.
- Layout of the exercise and customers' ability to understand and choose.
- Descriptions of the service issues, and their perceived realism. (This included testing with companies and other stakeholders as well as with customers.)
- Performance of econometric models (quant pilots only).
- Consistency of results with expectations, i.e. in terms of the relative value of the two service issues, and the consistence of this relative value with the relative impact derived from the impact exercise (quant pilots only).

2.4 Additional aspects tested

In addition to the two stated preference exercises described above, the survey instrument included the following elements that were also tested as part of Stage 2:

- Recruitment and screening
- Usage and attitudes
- Demographics

3 Overview of Testing and Development

This section provides an overview of the stages of testing and development undertaken throughout Stage 2, including the issues tested, the methodologies adopted, and the key findings and recommendations.

3.1 Stages of testing and development

Stage 2 of the study included qualitative testing via small samples of cognitive depth interviews, and quantitative testing via larger samples of pilot interviews. These provided complementary means to test the survey instrument.

- **Cognitive interviews** give an opportunity for the interviewer to probe participants on their thoughts and perceptions as they proceed through the interview, which can be a good way of identifying issues with particular forms of wording, questions, or show materials.
- **Quantitative pilot surveys** provide a means of testing the proportions of participants that feedback in positive or negative ways at key stages, whether the diagnostics and statistical properties of the econometric models are sufficiently good, whether the results conform with expectation or appear anomalous, and whether the design needs to be amended to improve its statistical properties.

In summary, the following rounds of testing with customers were undertaken:

- **Cogs 1:** Cognitive depth interviews with 13 households and 5 non-household customers
- **Pilot 1:** Quantitative pilot with 1,058 household and 80 non-household customers
- **Cogs 2:** Cognitive depth interviews with 4 household and 2 non-household customers
- **Pilot 2:** Quantitative pilot with 402 household and 150 non-household customers
- **Cogs 3:** Cognitive depth interviews with 10 household and 8 non-household customers
- **Welsh Cogs:** a Welsh translation of the survey was tested cognitively as a separate exercise by ORS research, focusing on the Welsh language itself, rather than the design.

A number of revisions were made at each stage, based on the findings from the research, and as a result of comments received from companies, the Environment Agency, Stantec

(with regard to storm overflows and pollution incidents), and three academic peer reviewers/advisors: Profs. Ken Willis, Giles Atkinson and Stephane Hess. Their comments were invaluable in helping to refine the questionnaire and show materials.

Ofwat and CCW were highly engaged throughout the process via regular meetings to discuss comments received, findings from the customer testing, and other issues as needed.

3.2 Cogs 1

Issues tested

The objectives of the exercise were to ensure that:

- The survey instrument was easily understood
- The language was straightforward and clear
- The survey instrument was supported by good visuals.

Methodology

Accent interviewed 13 household customers (HH) and 5 non-households (NHH). Interviews were conducted online, via MS Teams and Zoom video meeting platforms.

For HH, interviewers shared their screen, and control of their computer, and participants were asked to complete the survey questionnaire themselves, as though they were completing online, whilst simultaneously ‘thinking aloud’ to provide the interviewer with an insight into their thought and decision-making process as they progressed.

The 5 NHH participants were interviewed with the interviewer’s camera off, to replicate the telephone-based (CATI) interviews that were likely to occur in the main survey. Rather than completing the survey themselves, NHH participants were asked the questions in the style of a CATI interview. The images to support the stated preference exercise were shared with 2 of the NHH participants, however the remaining 3 NHH participants were read the attribute text information only and asked to respond based on this information. This allowed for an initial test of the importance of seeing the information versus having it read out.

At certain points, interviewers asked additional cognitive questions to assess how well the questionnaire and wording had been understood, and how easy or difficult it was to provide an answer. Care was taken to ensure that the cognitive questioning did not educate participants as this might lead to them approaching later questions differently than would have been the case had they not been informed or asked to think about certain elements of the study in more detail. Therefore, some of the HH cognitive interviews were carried out upon completion of the questionnaire, rather than simultaneously with self-completion.

Summary of findings and recommendations

Overall, participants perceived the survey as easy to complete and an acceptable length, with some commenting that they found the questions interesting and thought-provoking. The ‘think-aloud’ commentary helped to confirm that participants were understanding what was expected of them. Some slight hesitation was encountered at certain points during the cognitive interviews, which are detailed in Appendix A. This resulted in a set of minor recommendations for revision throughout the questionnaire. (See Appendix A for details.)

With regard to the impact exercise, participants tended to find it easy to select between scenarios, with stated reasons for choices all suggesting that the format worked as intended. Whilst the images and service issue descriptions generally worked very well, a number of comments were made about individual service issue descriptions and images, which resulted in a set of recommendations for revision.

The compensation exercise was also found to be easy to understand overall, with no changes recommended.

3.3 Pilot 1

Issues tested

The purpose of the first pilot was to test the stated preference designs and the fieldwork methodology. With respect to the stated preference designs, this included examination of:

- Feedback from participants
- Diagnostics on choice data
- Performance of econometric models
- Consistency of results with expectations
- Appropriateness of compensation levels used

Testing the fieldwork methodology focused on examination of:

- Response rates by customer group and characteristics, by mode, particularly focusing on the novel (to water sector) Postcode Address File (PAF) approach for households
- Feedback and variance of responses with respect to whether the non-household survey worked by telephone only, without the participant being able to see information on screen.

Methodology

The Pilot 1 sample contained 1,058 households, split 50:50 between Postcode Address File (PAF) and Panel methods, and 80 non-households (NHH).

For the HH PAF method, address sample was purchased for each of the 17 water companies according to a matrix of water and wastewater company combinations based on recommendations set out in the Stage 1 report. With an initial assumption of a 7.5% response rate 6,000 letters were sent out to achieve the target 450 interviews. Each letter, addressed to 'The occupier', explained the purpose of the survey and included an online link and QR code to access the survey. Incentives were offered to encourage participation, with half starting at £5, moving to £10 on reminder; and half pitched at £10 throughout. For those who were unable or didn't wish to respond online, Accent offered the option of a paper version of the questionnaire along with another pre-paid envelope for its return.

For the HH panel method, the Kantar online panel was used. Kantar was provided with target and maximum number of interviews for the required water / wastewater company combinations, and a list of the postcodes for each.

For the NHH pilot, Accent purchased telephone contact details from a sample company for 3,000 contacts, distributed according to the same water/wastewater matrix as used for the HH survey. Potential participants were phoned and told that they would be offered a series of choice pairs which were hosted online. They were offered a short link to these, or they were read out over the phone.

Summary of findings and recommendations

Survey methodology

The HH PAF response rates were 9.1% for the initial £5 incentive that changed to £10 for the reminders and 11.1% for the initial £10 incentive. The overall response, including the paper questionnaires, was 10.5% - significantly higher than the 7.5% expected.

Both PAF and Panel samples had a somewhat older age profile than the Census, and showed a similar level of ethnicity bias towards white participants. The PAF sample was somewhat closer to the Census on gender, but had a bias towards the highest social grade occupations. However, both samples had similar vulnerability and affordability status levels.

Since the findings did not conclusively point to an advantage of one HH survey method over the other, and given the pre-known pros and cons of both methods, the recommended approach taken forward was to split the main HH sample 50:50 between Panel and PAF methods rather than choose one over the other.

Stated preference design

The results of the Pilot 1 tests provided mixed evidence with respect to the stated preference design approach.

In support of the approach:

- Participant feedback was very good for both the impact and the compensation exercise.

- There were very few instances of non-trading behaviour in the impact exercise (where participants always chose the same alternative throughout the exercise).
- The econometric models were well estimated, especially considering the small sample size for non-households.
- The impact rankings derived from the econometric models were highly consistent with prior expectations.
- Also as expected, participants were more likely to take the compensation offered when it was high than when it was low.

However, set against these positive findings, two key issues were identified:

- At the highest compensation levels shown there were still substantial proportions of participants choosing not to take the compensation offered, indicating that higher compensation levels were needed, particularly in the case of external sewer flooding.
- Estimates of the relative value of the two 'pivot' scenarios included in the compensation exercise were significantly different when obtained from the compensation exercise than when obtained from the impact exercise. This indicated that participants were not choosing consistently across the two exercises.

Based on these findings, and those from Cogs 2 reported below, and following discussions with Ofwat, CCW and the three academic peer reviewers, it was agreed that a second pilot, and third set of cognitive interviews, should be undertaken containing two different versions of the stated preference approach for testing:

■ **Version 1**

Same as that used for Pilot 1 except for:

- Revised pair of pivot service issues within the compensation exercise, from Unexpected interruption (6h) and Sewer flooding outside your property (1 week), to Planned supply interruption (6h) and Boil water notice (48h)

This change was made for the following reasons:

- Sewer flooding is a high-impact incident, meaning that required compensation is higher, perhaps stretching into implausibly high numbers, and weakening precision.
- There is more ambiguity in the description over the severity of the impact of external sewer flooding in comparison to other service issues, particularly given that properties vary substantially across the population. This potentially links the perceived impact with the level of compensation shown, causing a bias.

The benefits of this new version were perceived as including the following:

- It would involve more questions involving customers trading off money, and thereby result in more statistically precise estimates of the marginal utility of money – a necessary component of valuation.
- It would prevent any discrepancy between relative impact and relative compensation required, and thereby avoid the problem identified at Pilot 1 of a difference in relative values between the impact exercise and the compensation exercise.
- It would prevent there being a lot of reliance on the required compensation levels for two specific service issues such that any error in measurement has an impact on the values estimated for all the other service issues.

Set against these, a potential disadvantage was identified, which was that combining the service issue and compensation into one question would impose an additional cognitive load on participants, which could affect the validity of the results.

3.4 Cogs 2

Issues tested

Whilst Cogs 1 and Pilot 1 were in progress, companies and other stakeholders were consulted on the survey instrument that was being used in Cogs 1. Responses were received from Thames Water, South Staffs Water, Welsh Water, Southern Water, United Utilities, Wessex Water and the Environment Agency. Additionally, helpful advice on storm overflows and pollution incidents was provided via an email exchange with Stantec, who led the recent national study on storm overflows.¹

Responses were discussed with Ofwat and CCW, alongside the results from Cogs 1 and Pilot 1, and a number of additional changes were made to the survey instrument as a result. These included changes to individual question wordings and changes to the descriptions of some of the service issues to ensure they better reflected average, or typical, incidents of each type. A core purpose of Cogs 2 was to test out the revised survey instrument cognitively once all these changes had been made.

Furthermore, based on Pilot 1 findings, compensation levels for the compensation exercise were deliberately set at very high levels in Cogs 2 in order to provide a strong test of whether such high levels could be used within the main design without weakening the credibility of the exercise, and how the wording of the exercise might need to change to accommodate such levels.

The performance of the two new service issues recommended for inclusion within the Version 1 compensation exercise (see above) were also tested in Cogs 2.

¹ Stantec (2021) Storm Overflows Evidence Project. Final report to DEFRA, Water UK, Environment Agency, Ofwat, Wildlife and Countryside Link and Blueprint for Water. November 2021.

Methodology

Accent interviewed 4 HH and 2 NHH customers for this wave. Interviews were conducted online, via the Zoom video meeting platform.

For HH interviews, interviewers shared their screen and control of their computer, and participants were asked to complete the survey questionnaire themselves, as though they were completing online. Participants were asked to work through the questionnaire autonomously for the most part but were asked to highlight areas of the questionnaire that were inaccessible, difficult to understand, or troublesome to navigate or complete. As extensive cognitive testing had been undertaken in the earlier phase, this approach was deemed appropriate to identify any areas for further change.

NHH participants were interviewed with the interviewer's camera off, to replicate as closely as possible the telephone-based (CATI) interview mode. Rather than completing the survey themselves, NHH participants were read the questions in the style of a CATI interview. The images to support the stated preference exercises were shared onscreen with NHH participants at the appropriate times, again replicating the anticipated approach for stimulus-assisted CATI interviews.

Summary of findings and recommendations

The Cogs 2 phase again showed high levels of accessibility and good comprehension. Questions and task instructions were clear and well understood, with participants able to respond as intended. Participants were encouraged to provide feedback at all points of the questionnaire, outside of specific cognitive probes, yet this resulted in very little need for improvement.

Stated preference task instructions were found to be clear and informative, successfully preparing participants for the choices they would be asked to make in both the 'impact of service issues' and 'compensation for service issues' tasks.

Some minor feedback and suggestions for improvement were captured on the service issues attributes that were presented at both stated preference tasks.

With regard to the compensation levels tested, some participants expressed concerns about their believability, particularly with regard to external sewer flooding. This finding was particularly true of the two NHH participants in the research.

Early feedback from a NHH participant suggested that the sums being offered did not match his definition of 'compensation' but instead constituted 'incentives'. The levels of compensation offered were considered too generous to make good the inconvenience that might be experienced because of a service issue.

Following this feedback, a probe was added to the topic guide to test the appropriateness of the word 'compensation', and the suitability of an alternative term 'one-off payment'. Most participants from this point onwards agreed that 'compensation' had been used

appropriately within this stated preference task and that its use met with their definition. Participants tended to believe that ‘compensation’ was a more appropriate label than ‘one off payment’, but none suggested that the label would impact their decision making in the task.

These findings suggested to us that the term ‘compensation’ was causing a link in participants’ minds between the severity of the issue and the amount of money being offered. For this reason, a switch to the term ‘one-off payment’ was recommended to be taken forward for further testing.

3.5 Pilot 2

Issues tested

The core purpose of Pilot 2 was to test the stated preference designs from the two versions recommended in response to the findings from Pilot 1 and Cogs 2. (See Section 3.3.) This again included examination of:

- Feedback from participants
- Diagnostics on choice data
- Performance of econometric models
- Consistency of results with expectations
- Appropriateness of compensation levels used

Methodology

The sample contained 402 households and 102 non-households, split between Versions 1 and 2.

For the HH survey:

- The Dynata online panel was used (for Pilot 1, Kantar panel was used)
- Dynata was provided with target numbers of interviews but with no quotas.
- 202 HH completed Version 1; 202 completed Version 2

For the NHH survey:

- 100 B2B panel interviews were conducted for Version 1
- 40 B2B panel interviews and 10 CATI interviews were conducted for Version 2.

Summary of findings and recommendations

The average completion time was 9 minutes, significantly shorter than the 11 minutes taken on average by the Pilot 1 panel sample. This may be partly explained by the fact that Pilot 2 samples were somewhat younger than in Pilot 1 and contained more males than females.

Version 1 findings

The results provided strong evidence in support of Version 1 overall, with the key issues identified at pilot stage entirely overcome.

■ Like in Pilot 1:

- Participant feedback remained very good for both the impact and the compensation exercise.
- There were very few instances of non-trading behaviour in the scenario impact exercise.
- The impact rankings derived from the econometric models were highly consistent with prior expectations.
- Also as expected, participants were more likely to take the compensation offered when it was high than when it was low.

■ Plus, new in Pilot 2:

- At the highest compensation levels shown the vast majority of participants now chose to take the compensation offered, meaning that mean and median valuation estimates could be well estimated.
- Estimates of the relative value of the two ‘pivot’ scenarios included in the compensation exercise were now much more closely aligned when obtained from the compensation exercise as when obtained from the impact exercise.

■ However:

- The use of the term ‘one-off payment’ caused some problems, with some misinterpreting the question as requiring a payment from the customer.
- The econometric models of impact choices were less well estimated in terms of goodness-of fit, and participants were quicker through the exercise than last time, leading to a much flatter, i.e. less diverse, set of impact scores.

Version 2 findings

The results suggested that Version 2 was more complex overall, and was unable to measure values.

■ Participant feedback was generally good, but not as strong as Version 1

■ There were more indications of non-trading behaviour, but still not seemingly a significant problem.

■ The impact rankings derived from the econometric models were again highly consistent with prior expectations.

■ However, the results suggested that participants were completely insensitive, on average, to the compensation amount shown. This implies that it was impossible to derive valuations from the estimated models.

We posited two possible explanations for this finding:

- First, some participants viewed the ‘one-off payment’ as a fee/charge to be paid instead of as compensation.
- Second, some participants may have focused on the service issues involved in each option ignoring compensation.

Although it should be possible to improve the survey materials to try and prevent these two issues, without a further pilot, there was no guarantee that the changes would have the desired effect. On this basis, and given the lack of time available in the programme to undertake a third pilot, we recommended abandoning this new version exercise in favour of the first version, which was found to work effectively.

Based on these findings, and those from Cogs 3 (reported below), the following recommendations were made:

- Version 1 should be adopted for the main stage, with linked Impact and Compensation choice exercises as in the original Stage 1 proposal.
- Planned supply interruption (6h) and Boil water notice (48h) should be retained as pivot service issues in the compensation exercise.
- The term ‘Compensation’ should be used rather than ‘One-off payment’.
- A question should be added after the first Impact choice question asking, ‘Why did you choose that option?’ (This was primarily recommended to slow down potential speeders, and encourage them to consider their responses more carefully. It would also serve the additional benefit of providing further information for analysis.)
- Analysis of the data should identify ‘speeders’ as well as other potentially invalid choice responses, and sensitivity test removing them from the sample.

Given these changes, and a few other minor edits, the survey was considered to be ready for main stage fieldwork.

3.6 Cogs 3

Issues tested

The purpose of the Cogs 3 wave, which was conducted at the same time as Pilot 2, was to test changes to the questionnaire since the 2nd round of depth interviews including, in particular, the new stated preference exercise adopted for Version 2.

Methodology

Cognitive interviews were conducted with 15 customers in total:

- 6 x Version 1 (HH and NHH)

■ 9 x Version 2 (HH and NHH)

Interviews were conducted on Zoom and participants were paid £30 (£45 for NHH). CATI (NHH) and CAWI (HH) modes were replicated during interview.

Summary of findings and recommendations

Version 1 findings

Refinements to the SP task instructions were successful and appeared to have added clarity where it was lacking previously. A few minor recommendations were made to further improve clarity.

Version 2 findings

The introduction to the SP exercise was considered very wordy by participants (but could not be reduced without omitting important info)

Compensation levels were generally considered credible, although the credibility of the highest payment of £30,000 for an internal sewer flooding incident was questioned in one case.

For both households and non-households, decision-making was primarily driven by a desire to avoid or minimise disruption and inconvenience. One-off payments played a relatively minor role.

One participant misinterpreted one-off payment as being to, rather than from, the water company.

The findings overall were consistent with those from Pilot 2.

4 Conclusions

The testing and development work undertaken throughout Stage 2, as described in this report, has been successful. It has resulted in a survey instrument that be considered to have performed sufficiently well to support its use within a main survey for obtaining the evidence needed to support the setting of ODI rates for common PCs at PR24. Given the findings reported here, the survey can be expected to produce a strong and comparable evidence base on customer values for water and wastewater services for all companies.