Methodologies for measuring outcome delivery incentive risk

A report for Ofwat prepared by PwC

July 2022

Table of Contents

Executive Summary

PwC was commissioned by Ofwat to provide recommendations on methodologies to measure Outcome Delivery Incentive (ODI) risks for its next price review (PR24). This report presents an assessment of possible methodologies, including an assessment of their advantages and drawbacks.

In their consultation responses to Ofwat PR19 final determinations and PR24 consultations, some companies said they considered that in PR19 there was a lack of consistency in the approach to measuring ODI risk across companies, with companies making their own assumptions around downside scenarios (provided by P10 outcomes),[1] and upside scenarios (provided by P90 outcomes). Therefore, in PR24, Ofwat is seeking to develop a more consistent approach to measuring ODI risk.

Our approach

Our approach consisted of identifying three methodologies, aimed at addressing the issues raised by water companies. The proposed methodologies were then illustrated using a limited number of performance commitments (PCs) and a subset of companies, using historical performance data. The purpose of the exercise was to observe how the three methodologies would have performed had they been used in PR19 (we also used some 2020/21 data). As a final step, we completed an overall assessment using five assessment criteria.

The proposed methodologies

Methodology 1 is the “do minimum” option, and it is essentially the same methodology used in PR19. Companies estimate their possible range of risk (P10 and P90 reference points) for each of their PCs separately, using companies’ own methodologies, based on a mix of historical evidence, the company’s forecasts and expert judgement. There is no particular data requirement on the units of measure companies should deploy or Ofwat’s guidance on how the estimation should be performed (e.g. role of historical data or how expert judgement is used). Companies account for interdependencies across PCs by applying scaling factors calculated by Ofwat to the simple additive of their P10 and P90 estimates.

Methodology 2 is similar to Methodology 1, with a few refinements, aimed at resolving concerns raised by companies in their consultation responses and representations around comparability of the risk measured across companies. For example, the estimation of risk sources need not be limited to P10 and P90 risk reference points, but can be extended to median, expected performance, P5 and P95. Another difference relates to the data used for the estimations in that Methodology 2 can incorporate the use of industry-wide historical data, collected using standardised units of measure and submitted to Ofwat yearly to ensure consistency across companies. If Ofwat considers that this approach does not provide sufficient incentives for companies to improve their performance or does not target company specific needs sufficiently, it can intervene with adjustments implemented at the company level. In applying Methodology 2, Ofwat would need to provide clear guidance on risk aggregation approaches, such as explicitly stating how interdependencies between PCs should be incorporated. A final difference relates to the way in which companies account for interdependencies between PCs: Methodology 2 has three different variants for how Ofwat and water companies can calculate scaling factors based on sector-level information.

Finally, Methodology 3 differs from the other two methodologies in that it does not focus only on specific risk reference points (e.g. P10), but it is based on assumptions of the whole distribution of a company’s expected performance against its PCs. Output distributions are created using the Monte Carlo simulation method. Using this methodology, water companies have to provide justified assumptions for the performance distribution of each PC, specifications of any risk parameters involved and any correlation assumptions: the quality, precision and consistency of the inputs used is a necessary condition to obtain reliable estimates using this more complex methodology. The Monte Carlo analysis, performed with appropriate software (e.g.@risk), produces the full distribution of aggregated RoRE impacts: this means the output is a single distribution of impacts for all PCs. This distribution accounts for the interdependencies across PCs by using the matrix of correlations between PCs as inputs to the analysis.

[1] Ofwat defines P10 as the PC performance threshold at which there is only a 10% chance of outturn performance being worse.

Each method requires calculating notional performance under a range of performance uncertainties. This can be difficult, as this requires assessing an efficient level of performance, possibly removing any identified inefficiency. The benefit of the notional approach is that it should create the right balance between providing sufficient incentives to improve efficiency and excessive underperformance payment. Potential methods for assessing the notional performance level include taking the sector average or the upper quartile performance of a company or the sector over time, or using historical data, and excluding significant outlier events. This is a policy choice for Ofwat. As this work was based on PR19, we use a 60% gearing assumption for our calculations of RoRE.

Assessment criteria

In order to provide a qualitative assessment of the three methodologies, we identified 5 main criteria, against which we scored our proposed methodologies.

1. **Ease of implementation**: the statistical/analytical capabilities companies require in order to implement each methodology, and also how the methodology integrates with the wider PR24 risk analysis.
2. **Evidence/data requirements**: how a methodology strikes a balance between ensuring the robustness and reliability of the data required and the challenges for companies to provide these data and for Ofwat to assess.
3. **Consistency**: how a methodology enables consistent approaches by companies, easing the comparability of results.
4. **Rerunning/updating analysis**: how much effort is required (by companies or Ofwat) to conduct updates and reruns during the price review process as inputs (e.g. target performance levels) are revised.
5. **Verification/assurance**: how a methodology produces results that Ofwat can understand, use and interpret and / or be externally assured.

Illustrative examples

In order to illustrate how proposed methodologies work, we have run them on a subset of PCs and companies. PCs were selected in order to cover both water-only and wastewater companies' commitments, and to ensure the minimal impact of the COVID pandemic. Companies, instead, were selected to ensure sufficient variability in the observed levels of performance.

The results from our illustrative examples suggest that, in general, Methodologies 2 and 3 tend to provide more accurate ranges, providing a more appropriate view of the risk faced by companies. However, there are cases in which Methodology 2 provides estimates that are unnecessarily wide or that fail to capture the actual RoRE impact. This might be explained by two factors: first the fact that historical data might have high variability (given that it includes outlier events); second are totex interactions, in that a higher expenditure to target improvements in a certain outcome area would be expected to reduce the risk associated with it. Both issues are likely to be solved by applying Methodology 2 (in all its variants) at the notional level - as described in section 2.

Overall considerations and suggestions for next steps

Methodology 1 has a number of limitations which have been raised by companies in the consultation responses and representations to the CMA appeals. Due to inconsistencies in some of the data used by certain companies and the differences in the approaches used for the estimation of the reference points, it is difficult for Ofwat to provide guidance and implement a standardised approach. However, the primary advantage of Methodology 1 is that it is simple for companies to implement and easily allows them to account for company specific risks.

Methodology 2 provides a good refinement on Methodology 1, with a balance between setting a reasonable data requirement of water companies and more standardised, consistent estimates of risk, but does not require the complexity of Monte Carlo analysis. This means companies can perform the analysis with existing capabilities. In this report we set out how Ofwat can achieve greater standardised, consistent estimates of ODI risk.

Methodology 3 requires that all companies comply with the data requirements imposed by Methodology 2 to ensure analysis is run on a set of data that is standardised and consistent across companies. As this methodology uses Monte Carlo simulations, it also requires all companies to specify the assumptions used in their analysis (including distributions and correlations across all the PCs) and submit them to Ofwat. Monte Carlo analysis is complex (and at risk of being untransparent) and requires the right software to perform the simulations. There is therefore a risk that either this methodology becomes resource intensive, or the quality of statistical analysis could be compromised. A further drawback of Monte Carlo analysis is that, given it is based on a set number of random draws (iterations), it cannot be easily replicated or updated easily to account for adjusted parameters. However, if implemented correctly, this methodology is best placed to consider company specific aspects and interdependencies across PCs, delivering the most precise estimates.

Our suggestion to Ofwat is to adopt Methodology 2, which would:

- Address most of the concerns raised by companies in their consultation responses and representations around transparency, consistency and equality of treatment. This would, however, lead to higher data requirements;
- Require Ofwat to continue to harmonise common ODI definitions;
- Require Ofwat to develop a comprehensive record of historical performance variation which will grow over time.
- Require Ofwat to provide guidance on how to blend company historical performance variation with industry level performance variation, i.e. by setting a definition of notional performance. Our illustrative examples show that using company historical variation can result in widely different risk ranges. While there may be strong grounds for using an overall industry approach, there may also be a requirement to incorporate company specific factors (e.g. by adding the financial contribution of bespoke ODIs to the risk ranges calculated through the ODI risk measurement methodology);
- Require that Ofwat update the scaling factors, using either it’s own Monte Carlo approach or a correlation matrix across PCs).

1. Introduction

Purpose of this report

PwC was commissioned by Ofwat to provide recommendations to help inform its policy decision around the methodology to measure Outcome Delivery Incentive (ODI)[2] risks for its next price review (PR24). This has involved assessing possible methodologies for PR24, including an assessment of their advantages and drawbacks, to inform Ofwat's policy decisions around how it assesses risks in the PR24 ODI regime.

Context to this report

The purpose of ODIs within the water sector

Ofwat sets performance commitments (PCs) for water companies, along with target levels of performance (performance commitment levels, or PCLs). ODIs operate to adjust price controls, allowing companies to receive additional payments from customers when they exceed PCLs and requiring them to return money to customers when they fall short of PCLs. ODI rates were set at PR19 using the incremental benefits and incremental costs associated with changes in performance, making sure that customers are fairly compensated for any underperformance, and companies receive additional payments for outperformance where this is valued by customers.

The importance of accurately measuring ODI risk

ODIs are a key component of Ofwat's price control framework in that they align the interest of stakeholders, challenging companies to improve their performance level through outperformance and underperformance payments. However, excessive out/underperformance payments can introduce significant risk into the regime. Therefore, a robust understanding of the risk resulting from the overall ODI package is crucial to the price control process. It also ensures that the overall balance of risk is proportionate and appropriate.

In their responses to Ofwat PR19 final determinations and PR24 consultations, some companies said they considered that in PR19 there was a lack of consistency in the approach to measuring ODI risk across companies, with companies making their own assumptions around downside scenarios (given by P10 outcomes),[3] and upside scenarios (given by P90 outcomes).[4] In particular:

- As a response to Ofwat final determinations[5], 8 companies[6] flagged that they disagreed with Ofwat adjustments to companies' P10 and P90 estimates because of a lack of guidance regarding how the estimates should have been performed, i.e. it was not clear that P10 and P90 performance levels should be set on the basis of an efficient company;
- in response to the Risk and Return consultation issued in December 2021[7], 5 companies[8] stated that Ofwat had not been transparent in their adjustments to companies' P10 and P90 levels. However, some of them were not able to replicate their own calculations;
- In response to the ODI discussion paper[9], Affinity Water highlighted how companies would benefit from more comparability across companies on common parameters.

In PR24, Ofwat is seeking a more consistent approach to measuring ODI risk (as well as seeking to use more common PCs- all with financial ODIs).[10]

With a PR24 final methodology to be published in December 2022, there is a need for a practical and implementable approach to measuring ODI risk that can deliver consistent outputs across companies.

[2] Ofwat defines ODIs (Outcome Delivery Incentives) as the financial incentives for companies to outperform and avoid underperformance against each of their performance commitments.
[3] Ofwat defines P10 as the PC performance threshold at which there is only a 10% chance of outturn performance being worse
[4] Ofwat defines P90 as the PC performance threshold at which there is only a 10% chance of outturn performance being better
[5] Ofwat (2019), PR19 Final Determinations - Delivering outcomes for customers policy appendix
[6] Anglian Water, Wessex Water, South East Water, Yorkshire Water, Thames Water, South Staffs Water and Bristol Water
[7] Ofwat (2021), 'PR24 and beyond: Discussion paper on risk and return'
[8] Bristol Water, Wessex Water, Yorkshire Water, Southern Water and South East Water
[9] Ofwat (2022), 'PR24 and beyond: a discussion paper on outcome delivery incentives', responses to the consultation have been informally shared by Ofwat.
[10] Ofwat (2021), 'PR24 and beyond: Performance commitments for future price reviews'

Approach

Based on: (i) the concerns raised by the companies in their representations and consultation responses regarding the PR19 methodology; (ii) Ofwat's suggestion to consider a Monte Carlo simulation approach to measure ODI risks; and (iii) review of options used in other situations, we identified three possible methodologies for PR24

We then identified a set of criteria to assess the three methodologies against and used a subset of companies and PCs to provide illustrations of how the methodologies would work in practice.

We selected a subset of PCs to make sure both water and wastewater activities were represented: we included two water PCs (supply interruptions and water quality contacts) and two wastewater ones (internal sewer flooding and pollution incidents). We selected a subset of companies with varied historic performance levels. We also conducted regression analysis to analyse the interdependencies (correlations) between performance on the metrics over time.

To create the illustrative outputs for each methodology, we take the targets set at PR19 and then use historic data to create a performance distribution. We then apply the ODI rates for the relevant PCs to the underperformance and outperformance ranges to create the illustrative RoRE range. To assess the relative effectiveness of the methodologies, we looked at how well they aligned to companies' actual RoRE performance so far in the current price control period (2020-25, AMP7); better alignment was treated as an indicator that the methodology was a better predictor of actual performance[11].

Lastly, we completed our overall assessment across the criteria, which informed the conclusions of this report.

Structure of this report

The rest of this report is structured as follows:

- **Chapter 2** provides a description of three possible methodologies.
- **Chapter 3** describes our assessment criteria for assessing the relative strengths and drawbacks of each methodology.
- **Chapter 4** presents a detailed description of how we selected a subset of PC and companies to provide illustrations of the different methodologies in our assessment.
- **Chapter 5** describes the analytical steps we have applied to generate the illustrative RoRE ranges using the three methodologies.
- **Chapter 6** presents the findings of our assessment against our assessment criteria.
- **Chapter 7** summarises our conclusions and recommendations for Ofwat's next steps.

[11] It is worth noting that our analysis, aimed at providing an illustrative example, is only based on one year of data.

2. Methodologies for measuring outcomes risk

In this chapter, we provide a high-level description of three methodologies we test in this report. This includes a specification of the information requirements for the companies, both on an annual basis and at the point of business plan submission, and the guidance Ofwat would need to provide using each methodology.

One possible option, which should always be included in evaluation of policy decisions, is to maintain the existing approach; therefore, our proposed Methodology 1 is continuation of the existing methodology used to measure outcomes risk at PR19. A second possible methodology is a refinement of the existing methodology, aimed at addressing most of the concerns raised by the companies in consultation responses. We also build on Ofwat's suggestion to consider Monte Carlo simulations for ODI risk measurement.[12] So the third methodology we test is a full Monte Carlo analysis that companies would run on their, and industry, data, and Ofwat could use to calibrate ODI risk estimates.

Across all methodologies, the calibration of ODI rates is performed through customer willingness to pay and other evidence provided by each water company. Ofwat can decide to apply caps, collars and deadbands at an individual ODI level (and aggregate level) to limit the maximum upside and downside risk exposure.

In addition, it is worth noting that the proposed methodologies do not include bespoke ODIs as those can be highly company-specific. Therefore, each company can add the financial contribution of bespoke ODIs to the risk ranges estimated through the three methodologies: a simple addition may be appropriate, where it is assumed there is little or no correlation between bespoke ODIs and common PCs. Companies should provide a valid explanation for how they calculated the financial contribution of bespoke ODIs in their business plan submissions.

Finally, companies should apply methodologies to an efficient company with a notional financial structure: this assumes a 60% level of gearing, as in PR19, and an efficient level of performance. Calculating efficient performance can be difficult, as this requires removing any identified inefficiencies. The benefit of this approach is that it should provide a more comprehensive view of risks, and create the right balance between providing sufficient incentives to improve cost efficiency and excessive underperformance payments. Potential methods for assessing the notional performance level include taking the sector average or the upper quartile performance of a company or the sector over time, or using a wider set of historical data and excluding significant outlier events.

Overview of the three methodologies tested

Our approach is to describe and compare the proposed methodologies using the following:

- **Risk sources**. These are the estimations of the reference points (e.g. P10, P90, median, expected performance) used by the companies to measure performance against each PC. Depending on the methodology, risk sources can also refer to the specifications of the risk parameters included in the analysis and their correlations. Once calculated, risk sources are multiplied by a PC-specific ODI rate to obtain the payments companies expect in their estimated downside/upside performance scenarios.
- **Method of aggregation**. This refers to how companies should aggregate the payments calculated for each PC across the portfolio of PCs (and ultimately other cost and finance risks). This relates to how companies account for the interdependencies between commitments: for example, if two PCs are perfectly correlated, then payments should be simply summed. Conversely, if two PCs are not correlated, then the payments should be corrected for the fact that the performance in a certain parameter does not influence at all the performance in another one and therefore the combination of risks won't be as much as the simple sum of risk ranges.
- **Use of outputs.** Once total payments are calculated, then there is a choice on how these outputs should be interpreted and used. For example, total ODI payments can be expressed as a percentage of a company's regulated equity, and this shows how the whole ODI package impacts a company's returns. This could be

[12] Ofwat, "PR24 and beyond: Creating tomorrow, together", p.116, available at: https://www.ofwat.gov.uk/wp-content/uploads/2021/05/PR24-and-Beyond-Creating-tomorrow-together.pdf

expressed as a range from the worst to the best scenario, or as a full distribution of return on regulated equity percentages, based on the expected distribution of a company's performance. A fuller range of outputs can inform calibration across a number of priorities (e.g. skew, incentivisation, and resilience).

In the following section we provide an overview of the three methodologies.

Methodology 1

Methodology 1 is the "no change" option, and is essentially the same methodology used in PR19.

Companies estimate their possible range of risk (P10 and P90 reference points) for each of their PCs separately, including bespoke ODIs. These estimates are done using companies' own methodologies, based on a mix of historical evidence, the company's forecasts and expert judgement. There is no particular data requirement on the units of measure companies should deploy or Ofwat's guidance on how the estimation should be performed (e.g. role of historical data or how expert judgement is used).

The method of aggregation is a simple additive of the expected payments for each PC, to which companies can apply industry-wide scaling factors to account for ODI interdependencies. More specifically, once reference points are calculated, companies multiply them by ODI rates to obtain expected payments for each PC. Expected payments are then summed across PCs to obtain total payments from the whole ODI package. A benefit of this method is that it allows identification of the contribution of each PC to the overall result.

Applying a scaling factor is a way to account for the interdependencies between PCs: for example, the fact that a company has a bad performance in internal sewer flooding may make it more likely that it has poor performance in terms of pollution incidents. To account for these correlations, at PR19 Ofwat imposed industry-wide scaling factors, which were calculated based on a mix of historical evidence and expert judgement: they amount to 70% for the downside and 90% for the upside.[13] It is worth noting PR19 scaling factors were based on companies' submissions and their analysis of interdependencies between PCs in the PR19 regulatory period: therefore the values are likely to change for PR24.

Outputs of the analysis are primarily tables of potential risk impact (in £m) and Return on Regulatory Equity (RoRE) ranges, which are the sum of expected payments in the low (P10) and high (P90) cases divided by the amount of a company's regulated equity. RoRE ranges are a proxy for how the ODI package may impact a company's future financial performance and therefore returns. Outputs are generally expressed as stacked histograms, showing the contribution of each PC to the total RoRE range. They can be presented both in gross or net terms, i.e. with or without accounting for the scaling factors.

The information requirements for Methodology 1 are summarised in the following table:

Table 2.1: Information requirements for Methodology 1

Requirement type	Requirement
Annual information requirement for companies	Historical performance for each PC
Information requirement for companies at business plan submission	• Historical performance; • Historical out- and underperformance payments; • P10 and P90 estimates for each PC (over the upcoming period, i.e. 2025-30); • Explanation of how the estimates have been calculated (methodology used); • Proposed ODI rates, if required;
Guidance provided by Ofwat	• How companies should estimate their P10 and P90 figures for each PC: a mix of historical evidence, expert judgement and forecasts; • How companies should calculate the financial contribution of each ODI/PC as a proportion of their

[13] In 'PR19 Final Determinations - Delivering outcomes for customers policy appendix', Ofwat explained that "*For P10 estimates, we use 70% which is broadly in line with the average ratios provided by companies. For outperformance, we consider that this should be greater in order to correct the likely pessimism in company estimates. We use 90%.*" This was largely based on the evidence submitted by companies and Ofwat's expert judgement.

	RoRE; ● How companies should account for interdependencies across PCs to calculate the aggregate impact of the whole ODI package: simple additive of the financial contributions of each ODI to which companies apply industry-wide scaling factors (calculated by Ofwat based on historical information and applying expert judgement. ● Publishing industry-wide scaling factors during the price control process, including an explanation of how they have been calculated.

Methodology 2

Methodology 2 is similar to Methodology 1, with a number of refinements, aimed at resolving concerns raised by companies in consultation responses and representations around comparability of the risk measured across companies and consistency of the data used.

For example, the estimation of risk sources need not be limited to P10 and P90 risk reference points, but can be extended to median, expected performance, P5 and P95 (which could be used to provide an idea of the tail risk associated with each PC). Methodology 2 is only applied to common PCs, while the impact of bespoke ODIs needs to be added to the estimated risk ranges. For example, a company might propose a bespoke PC which contributes 0.5% downside to the aggregated RORE and Ofwat might decide to account for this by reducing its scaling factor by one percentage point on the downside.

Another difference relates to the data used for the estimations: Using Methodology 2, Ofwat can mandate the use of industry-wide historical data on ODI performance, collected using standardised units of measure and submitted to Ofwat yearly to ensure consistency across companies. Now that there is a growing historical record on company outcomes performance, it is possible to provide a more centralised, and consistent basis for measuring outcomes risk. If Ofwat considers that this approach does not provide sufficient incentives for companies to improve their performance or does not target company specific needs sufficiently, it can intervene with adjustments implemented at a sector, or company level. Companies could also request company-specific adjustments to the companies' risk ranges, similar to cost-adjustment claims.

In applying Methodology 2, Ofwat would need to provide clear guidance on aggregation approaches, such as explicitly stating interdependencies between PCs, companies and/or setting methodological and evidential requirements for correlations across PCs and time periods, as described in Table 2.2 below.

The method of aggregation is a simple additive of P10 and P90 payments, for each PC across the price control period. Adjustments, calculated as scaling factors applied to the P10 and P90 simple additives, are applied to account for correlations across PCs.

The method for incorporating scaling factors defines the three possible variants of Methodology 2. These are:

- **Variant 1 (Expert judgement)**: Ofwat calculates scaling factors based on expert judgement as at PR19, but ensuring sufficient transparency on the calculations and assumptions made;
- **Variant 2 (MC scaling factors)**: Ofwat calculates scaling factors using Monte Carlo analysis. This means that Ofwat collects all historical data from companies and runs an industry-wide Monte Carlo analysis to estimate the impact of interdependencies between PCs on the aggregated RoRE ranges. This draws from a part of the technique used in Methodology 3, but the application of Methodology 2 does not require any Monte Carlo analysis to be performed by individual companies.
- **Variant 3 (Correlation Matrix)**: Ofwat provides companies with the inputs to calculate their own scaling factors. After publishing the PR24 methodology, Ofwat calculates the sector-level aggregated P10-P90 level from historical data with notional adjustments. Then, companies calculate their scaling factors by taking the ratio between the sector-level P10 (or P90) and the simple additive of the P10 (or P90) estimated for each PC. The calculated scaling factors are used in business plans submissions The more detailed analytical steps are described as follows:

1. After the methodology for PR24 has been published, companies submit their estimates of P10 and P90 performance and historical RoRE ranges, based on historical data;

2. Ofwat assumes normal distribution[14] of the historical RoRE ranges across all companies and calculates the mean and a 20% confidence interval around the mean. The upper and lower bounds of the confidence interval would provide P10 and P90 values that account for interdependencies between PCs.

   The confidence interval is calculated using this formula:

$$[Mean(Y) \pm 1.28 * StDev(Y) = \sqrt{\Sigma_i^n w_i^2 Var(X_i) + \Sigma_{i \neq j} w_i w_j Cov(X_i, X_j)}]$$

   Where:

$$StDev(Y) = \sqrt{\Sigma_i^n w_i^2 Var(X_i) + \Sigma_{i \neq j} w_i w_j Cov(X_i, X_j)}]$$

   and the P10 and P90 figures are obtained using the following formulae:

$$\text{P90} = Mean(Y) + 1.28 * StDev(Y) = \sqrt{\Sigma_i^n w_i^2 Var(X_i) + \Sigma_{i \neq j} w_i w_j Cov(X_i, X_j)}$$

$$\text{P10} = Mean(Y) - 1.28 * StDev(Y) = \sqrt{\Sigma_i^n w_i^2 Var(X_i) + \Sigma_{i \neq j} w_i w_j Cov(X_i, X_j)}$$

   Where

   - Y is the RoRE range of the whole ODI package, excluding bespoke ODIs
   - X_i is the financial contribution of the single PC_i
   - W is the weight attributed to each PC, for simplicity these could be assumed as equal across PCs
   - Var (X) is the variance associated to each PC
   - Cov (X_i, X_j) is the correlation coefficient between PC_i and PC_j

3. The ratio between the simple additive of the P10 and P90 calculated at step 1 and the lower and upper bound of the confidence interval calculated in Step 3 gives the scaling factors for individual companies.

Companies apply the scaling factors calculated following Ofwat guidance explained in the 4 steps above and use them as inputs for their business plan submissions.

The use of outputs is the same as Methodology 1. The information requirements for Methodology 2 are summarised in the following table:

Table 2.2: Information requirements for Methodology 2

Requirement type	Requirement
Annual information requirement for companies	• Historical performance for each PC; • Historical out- and underperformance payments (in absolute terms and expressed as RoRE impacts).
Information requirement for companies at business plan submission	• P10-P90, P5-P95, P50, expected performance for each PC: these need to be submitted both in gross and net terms, i.e. before and after applying the scaling factors;

[14] Any more complex distribution than normal would require the use of Variant 2 (MC scaling factors) - Monte Carlo to calculate scaling factors.

	• Explanation of how the estimates have been calculated (including the calculation of the scaling factor - using Variant 3 (Correlation matrix). For Variant 1 (Expert judgement) Ofwat will provide the scaling factor using expert judgement. For Variant 2 (MC scaling factors), Ofwat will use historical information and Monte Carlo techniques to provide the scaling factor); • Inputs used for the estimates, i.e. historical data consistently stored; • financial contribution of bespoke ODIs for P10-P90, P5-P95, P50, if any; • explanation of how companies calculated the notional performance level; • ODI rates; • Graphical representation of the estimated risk ranges.
Guidance provided by Ofwat	• Guidance to ensure data consistency across companies and over time: standardised units of measure for each PC and a template of the data files; • How companies should estimate their P5, P10, P50, P90, P95 figures: main reliance on historical data, collated according to the guidance provided by Ofwat; • How companies should calculate the notional performance level and how efficiency is defined; • How companies should calculate the financial impact of the whole ODI package on their RoRE; • Which graphical representation of the results companies should adopt (e.g. stacked histograms) and how they are to be interpreted; For **Variant 1 (Expert judgement)**: • How companies should account for interdependencies across PCs to calculate the aggregate impact of the whole ODI package: simple additive of the financial contributions of each ODI to which companies apply scaling factors; • The actual level of the scaling factors for PR24 is calculated by Ofwat using expert judgement and shared with the companies during the price control process. For **Variant 2 (MC scaling factors)**: • Transparency around the assumptions used and the analytical steps taken to run the Monte Carlo analysis; • Using historical performance data submitted by companies, Ofwat will run its own Monte Carlo to obtain adjustment factors. These will be shared with companies during the price control process. For **Variant 3 (Correlation Matrix)**: • Ofwat will share an industry-wide ODI correlation matrix across during the price control process; • Transparency around how Ofwat calculates the inputs for the matrix of correlation coefficients between common PCs at the sector-level; • Transparency around how the sector-level P10 and P90 have been calculated (publication of the assumptions and analytical steps undertaken as part of the statistical method used); • After calculating the sector-level P10 and P90, Ofwat provides guidance to the companies around how to calculate the scaling factors, i.e. by calculating the ratio between the sector-level P10 (or P90) and the simple additive of the P10 (or P90) estimated by the companies.

Methodology 3

Methodology 3 differs from the other two methodologies in that it does not focus only on specific risk reference points, but it is based on assumptions of the whole distribution of a company's expected performance against its PCs. Distributions are simulated using the Monte Carlo simulation method.

Using this methodology, water companies would have to provide justified assumptions for the performance distribution of each PC, specifications of any risk parameters involved and any correlation assumptions, as described in Table 2.3 below.

Links between models will need to be far more specified, robust and automated for a complete Monte Carlo based approach. The robustness of the outputs is dependent upon complete and accurate assessment of input risks and their inter-relationships.

To run the Monte Carlo analysis, companies set up a model to calculate the RoRE impact of each single draw from the Monte Carlo analysis: this consists in multiplying each draw (simulated observation) by the relevant ODI rate and dividing it by a company's regulated equity.

The Monte Carlo analysis, performed with appropriate software (e.g. @risk), produces the full distribution of aggregated RoRE impacts: this means the output is a single distribution of impacts for all PCs. It is worth noting that Monte Carlo simulation will have to be informed by the matrix of correlations between the common PCs and to account for the interdependencies between common PCs and therefore there is no need to apply scaling factors.

There are two kinds of outputs which can be produced using Methodology 3: the first is a similar histogram produced using Methodology 1 and 2 showing P10 and P90 ranges, while the other one is a graphical representation of the full distribution of the possible RoRE impacts. The latter, otherwise known as the 'flag chart', was suggested by CEPA as a tool for providing clarity of different risk outcome points and calibration of asymmetric risk profiles, likely return at all points of the distribution, and overall RoRE range.[15] The ability to get a fuller picture of all forecast outcomes allows for refinement in analysis and greater detail of output, subject to the reliability of the inputs used.

The information requirements for Methodology 3 are summarised in the following table:

Table 2.3: Information requirements for Methodology 3

Requirement type	Requirement
Annual information requirement for companies	• Historical performance for each PC; • Historical out- and underperformance payments (in absolute terms and expressed as RoRE impacts).
Information requirement for companies at business plan submission	• Assumption on the distribution of the performance against all common PCs, including a justification for the selected assumption; • Assumptions used for any other risk parameter involved in the analysis, including a justification for each assumption; • Correlation matrices for all common PCs; • Software selected for the Monte Carlo analysis; • Clear description of the analytical steps of the analysis; • Graphical representation and rationale behind the interpretation of the results. • Full distribution of the resulting RoRE ranges, specifying the P10-P90 range.
Guidance provided by Ofwat	• Guidance to ensure data consistency across companies and over time: standardised units of measure for each PC and a template of the data files; • How companies should calculate the notional performance level and how efficiency is defined; • How companies should calculate the financial impact of the whole ODI package on their RoRE: this is the model that would feed into the Monte Carlo analysis; • How companies should account for interdependencies across PCs, i.e. how the matrix of correlation coefficients of the common PCs should be used as input for the Monte Carlo analysis; • The analytical steps companies should take to complete

[15] CEPA (2021), "Allocation of risk: Prepared for Ofwat" 18 June 2021

	the Monte Carlo Analysis; ● Which graphical representation of the results companies should adopt (e.g. flag charts) and how they are to be interpreted;

3. Assessment criteria

We use a framework to assess the three methodologies. This consists of a set of criteria to understand the range of policy considerations for comparing the suitability of the different options.

Based on the framework, we then provide qualitative scores for each of the methodologies against each criteria, in the form of a red, amber, green (RAG) rating, to analyse the strengths and drawbacks of each methodology.

Assessment

We use 5 main criteria to evaluate the three methodologies:

- **Ease of implementation**: this criterion primarily relates to the statistical/analytical capabilities companies need to have to implement each methodology, and also how the methodology integrates with the wider PR24 risk analysis.
- **Evidence/data requirements**: Robust additional data and evidence can improve the quality of the analysis and increase confidence in the results. However, methodologies that require substantially more or different data to be gathered compared to previous price controls (e.g. historical performance levels in comparison to PCs and larger probability ranges such as P1 through to P99 performance levels) may present challenges for companies to provide and Ofwat to assess.
- **Consistency**: a methodology should enable consistent approaches by companies, giving meaningful and comparable results that Ofwat can use to inform its PR24 regulatory risk assessment. In addition, a methodology should be able to account for bespoke ODIs, including interactions between these and common ODIs.
- **Rerunning/updating analysis**: it is important that a methodology takes into account the need to conduct updates and reruns during the price review process as inputs (e.g. target performance levels) are revised. A methodology requiring disproportionate efforts to update will challenge the deliverability of the PR24 price review.
- **Verification/assurance**: a methodology should provide results that Ofwat can understand, use and interpret and / or be externally assured.

In Chapter 6 we apply our criteria to the three methodologies and score them using RAG ratings. The RAG assessments are then aggregated into an overall judgement for each methodology.

4. Building the data for our illustrative examples

In order to assess the three methodologies it is helpful to show how they can work in practice, as applied to real PCs and actual companies. This chapter presents a detailed description of how we selected our subsets of PCs and companies, as well as our analysis of the correlations between PCs that underpin the illustrative methodology applications in our assessment.

PC selection

Our illustrative analysis focused on the PCs that were broadly comparable across companies and less likely to have been impacted by the COVID-19 pandemic, to reduce the number of outliers. The four PCs we have selected are: Supply interruptions, water quality contacts, internal sewer flooding and pollution incidents. The following figures 4.1 to 4.16 provide a graphical representation of:

- The distribution of companies' performance in actual terms;
- The distribution of performance deviation from the targets; and
- Companies' performance against their PCLs over time.

It is worth noting that not all companies had financial incentives or targets over the full 2012-20 period, which explains why in the next subsections some of the charts showing deviations from the targets present fewer observations.

Supply interruptions

Definition: Average supply interruption greater than three hours (minutes per property).[16]

Distribution: As shown in Figure 4.1, performances on supply interruptions are clustered around the 15 minute per property mark. This is to be expected, as the most common PCL between 2017 - 2020 was 12 minutes per property. This indicates that most companies performed slightly worse than the PCL. Figure 4.2 shows that the distribution of deviations from target is skewed to the downside. There are some noticeable outliers in the data, where companies significantly underperformed against the PCL targets.

[16] Ofwat PR19 definitions. Available at: https://www.ofwat.gov.uk/outcomes-definitions-pr19/

Figure 4.1: Distribution of Supply Interruptions (2012 - 2021)

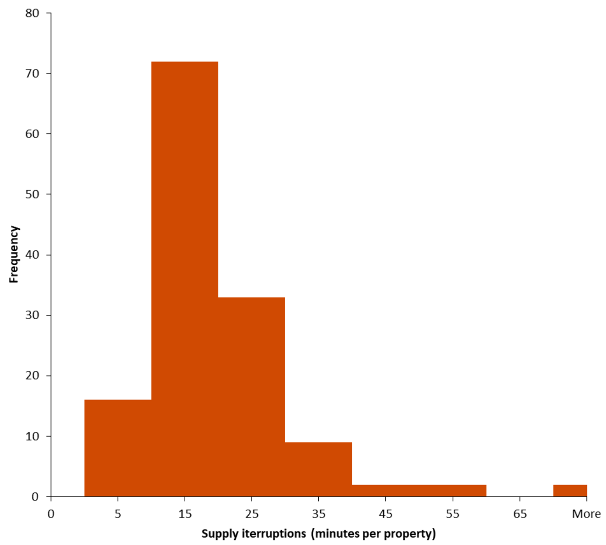
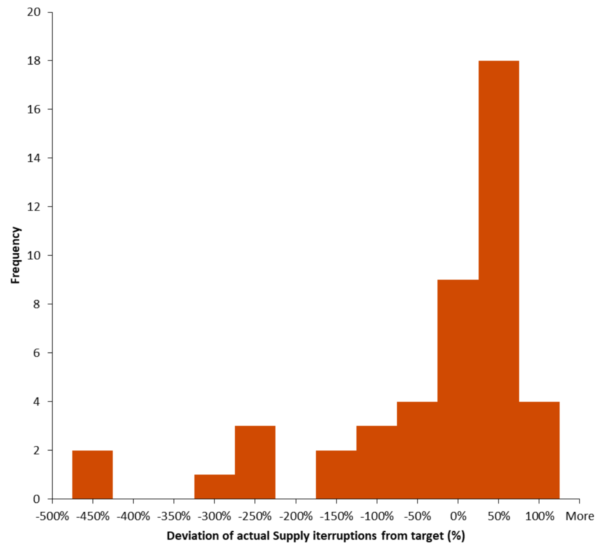


Figure 4.2: Deviation of supply interruptions from target (%) (2017 - 2021)[17]



Performance against PC over time: Figure 4.3 shows that, with the exception of a few outliers, the number of supply interruptions in minutes has been declining (i.e. the performance of companies has been improving) since 2017-18. This is further shown in Figure 4.4, as deviations from target are largest in the year 2017-18, before centering around the target in 2020-21.

Figure 4.3: Performance of Supply Interruptions between 2012 and 2021

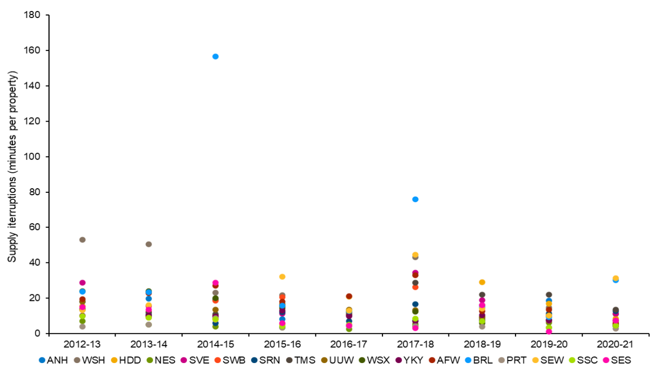
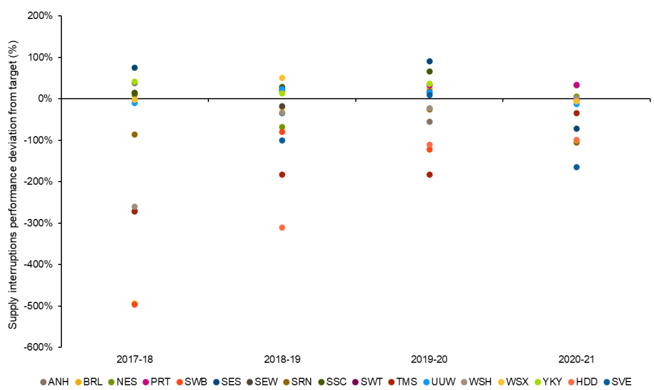


Figure 4.4: Deviation of Supply interruptions from PCL target between 2017 and 2021



Water quality contacts

Definition: The number of customer contacts about water quality (appearance; taste and odour).[18]

Distribution: Figure 4.5 shows the distribution of Water Quality Contacts. The distribution is centered around 2 contacts per 1,000 population. There are some noticeable outliers in the data, where companies recorded 5 or more contacts per 1,000 population about water quality. This pattern is also seen in the deviation from PCL targets (Figure 4.6), with the distribution centered around 20% below the target and some outliers as far as 200% from the target. The data suggests that the selected companies underperformed against the PCL targets between 2017 and 2021.

[17] Deviation is calculated by subtracting actual performance from the target. As a higher value in the PC suggests a worse performance, a <0% deviation indicates that the company has underperformed compared with the target.
[18] Ofwat PR19 definitions. Available at: https://www.ofwat.gov.uk/outcomes-definitions-pr19/

Figure 4.5: Distribution of Water Quality Contacts (rate per 1,000 population) (2012 - 2021)

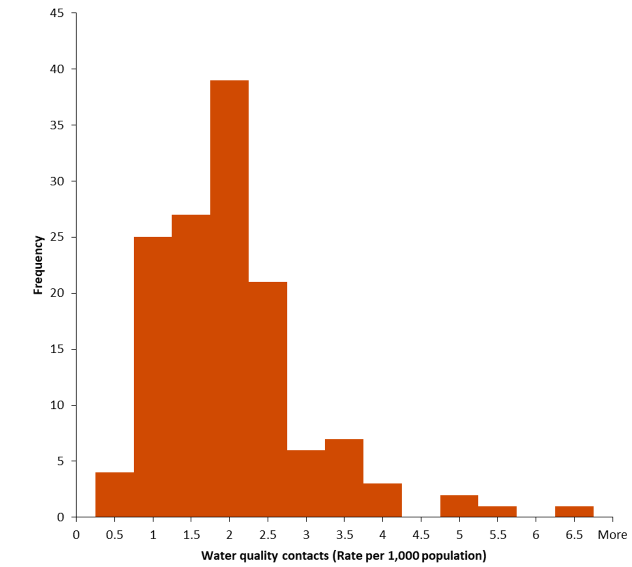
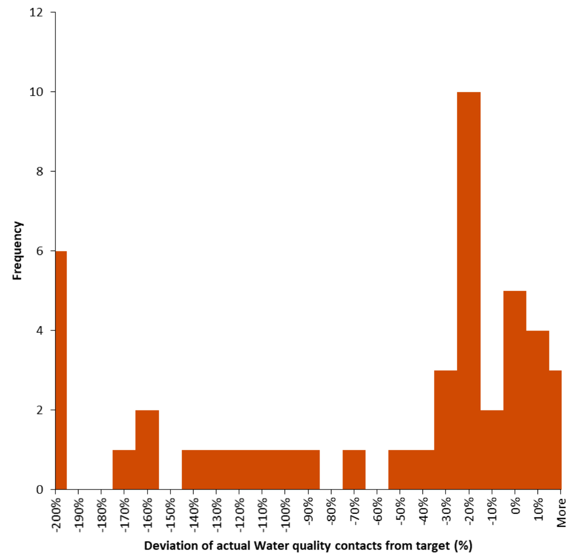


Figure 4.6: Deviation of Water Quality Contacts from target (%) (2017 - 2021)



Performance against PC over time: Figure 4.7 displays a generally declining trend in the number of contacts about water quality, with a few outliers, between 2012 and 2021. This suggests that companies are performing better on water quality over time. This is further shown in Figure 4.8, where the deviation of water quality contacts from the target declines between 2017 and 2021.

Figure 4.7: Performance of Water Quality Contacts between 2012 and 2021

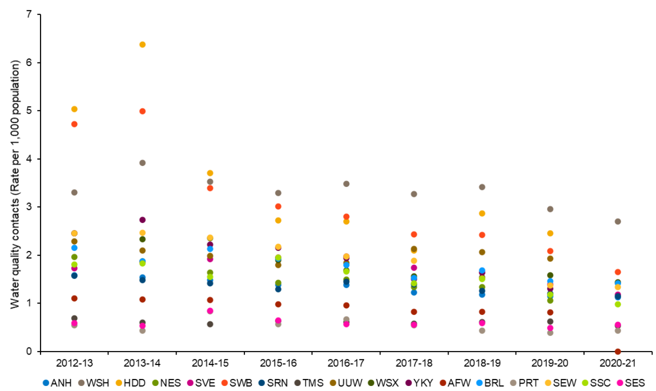
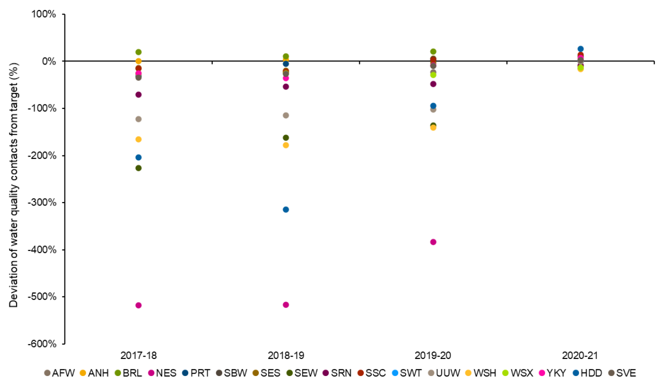


Figure 4.8: Deviation of Water Quality Contacts from PCL target between 2017 and 2021



Internal sewer flooding

Definition: The number of internal flooding incidents per year (sewerage companies only).[19] i.e. the number of properties flooded internally per year.

Distribution: Internal sewer flooding incidents from 2014 - 2021 follow an irregular and skewed distribution. As shown in Figure 4.9, most companies report 100 - 500 incidents of internal sewer flooding incidents each year. However, a considerable number of companies reported between 600 - 2,000 incidents of sewer flooding. This is also displayed in Figure 4.10, where most companies outperform their PCL by roughly 20%, but some underperformed by over 40% to 100% of their PCL. Observations have been normalised to show companies' performance over time .

[19] *ibid.*

Figure 4.9: Distribution of normalised Internal Sewer Flooding Incidents (2014 - 2021)

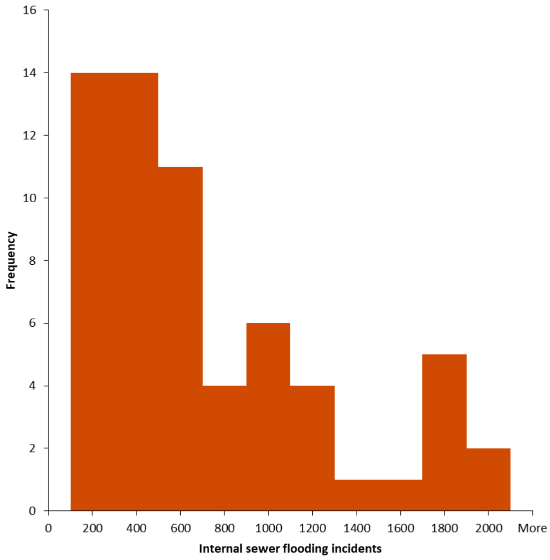
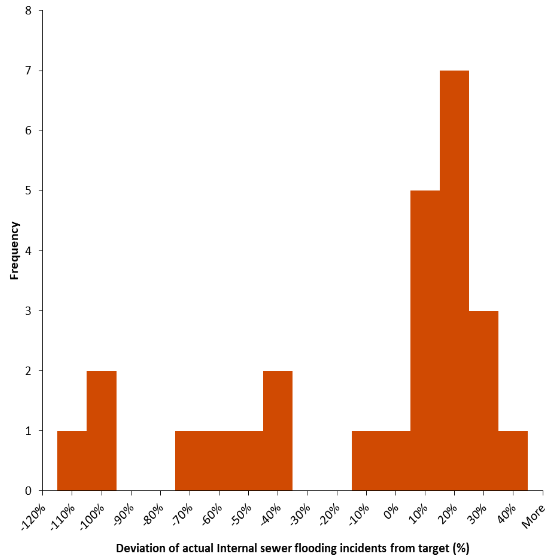


Figure 4.10: Deviation of Internal Sewer Flooding Incidents from target (%) (2017 - 2021)



Performance against PC over time: In Figure 4.11 and Figure 4.12, the historical data has been normalised to the number of sewer flooding incidents per 10,000 properties, to match the format of the 2020-21 data. Figure 4.11 displays a slowly declining trend in the number of internal sewer flooding incidents between 2014 and 2020, then a sharp decline in 2020-21, indicating improved performance on sewer flooding over time. Despite this improvement, Figure 4.12 shows larger negative deviations from the target in 2020 -21, indicating some companies are underperforming compared to their PCLs in previous years. This indicates that the move from AMP6 (PR14) to AMP7 (PR19) may have introduced more stretching PCLs and therefore a step-change in expected performance from AMP6 which some companies have yet to achieve.

Figure 4.11: Performance of Internal Sewer Flooding incidents between 2014 and 2021 - *normalised figures, floodings per 10,000 properties*

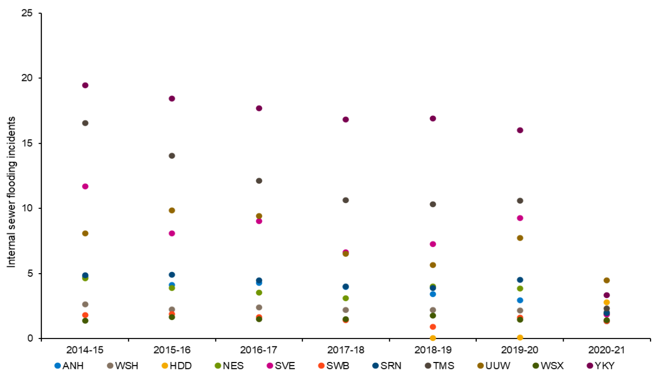
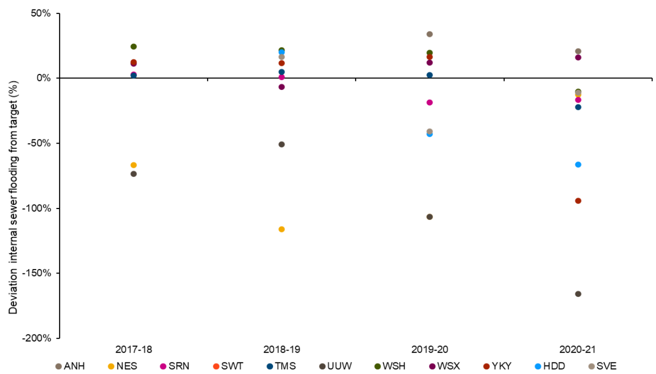


Figure 4.12: Deviation of Internal Sewer Flooding Incidents from PCL target between 2017 and 2021 - normalised figures, floodings per 10,000 properties



Pollution incidents

Definition: Category 1-3 pollution incidents, as reported to the Environment Agency and Natural Resources Wales.[20] For this PC, the historical data (2012 - 2021) had to be normalised to the number of pollution incidents per 10,000km of sewerage network to align with the 2020-21 PCL targets.

Distribution: Figure 4.13 shows the distribution of Pollution Incidents between 2012 and 2021, which is centered around 40 incidents per 10,000km of sewer. As with the other PCs, the distribution is skewed to the downside, with the worst performing companies recording over 100 incidents per 10,000km of sewer.

[20] *ibid.*

Figure 4.13: Distribution of Pollution Incidents per 10,000km of sewer (2012 - 2021)

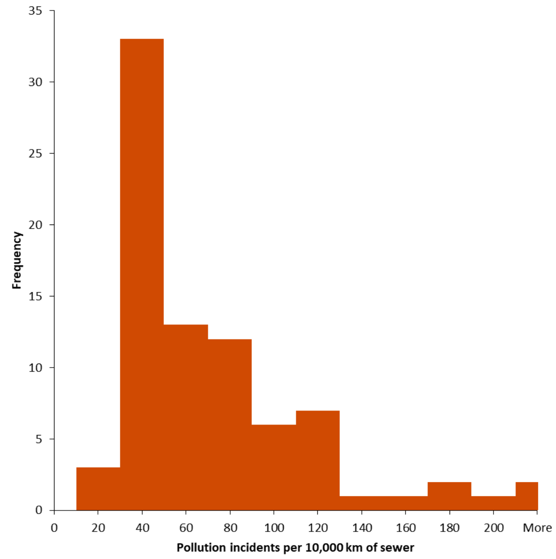
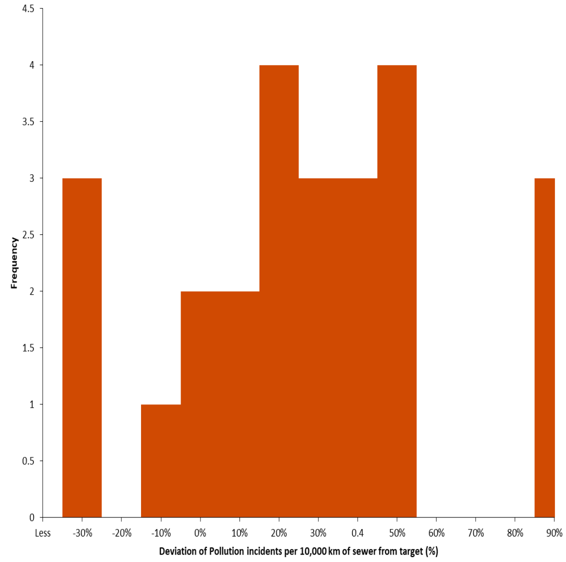


Figure 4.14: Deviation of Pollution Incidents from target (%) (2017 - 2020)



Performance against PC over time: In the historical performance of companies' pollution incidents, there is a general downward trend between 2012 and 2021, indicating that companies in general performed better on this PC over time (Figure 4.15). However, there are some outliers to this trend and the number of pollution incidents increased for some companies in 2019-21. Figure 4.16 shows that deviations from target stay largely consistent between 2017 and 2020, with a few extreme outliers in 2018-19 from companies which vastly underperformed against the PCL.

Figure 4.15: Performance of Pollution incidents between 2012 and 2021

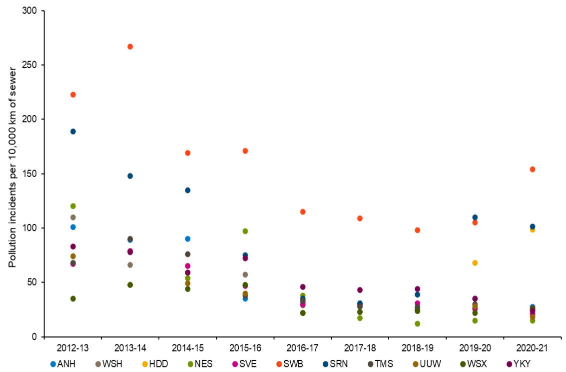
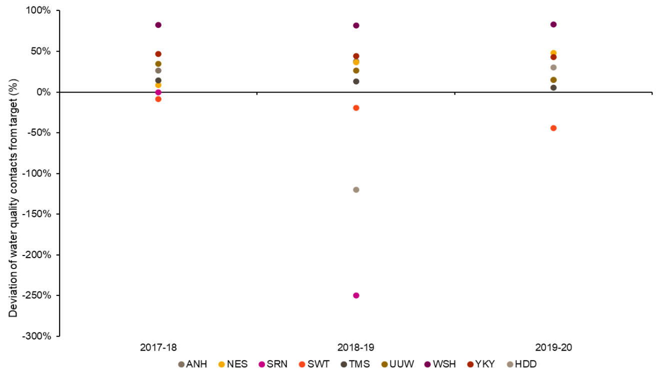


Figure 4.16: Deviation of Pollution Incidents from PCL target between 2017 and 2020[21]



Selection of companies

We present results for four companies, with the aim to show examples of good and bad performance against the selected PCs.

We selected Wessex Water, Anglian Water, Severn Trent and Southern Water as examples of companies that provide a variation in performance over the period 2012/13 to 2020/21.

[21] The 2020-21 deviation data is excluded here due to the inconsistency between the PCL and performance data.

It is important to note that performance on our selected PCs is not necessarily representative of the performance of a company across all PCs and ODIs. For example, in 2018-19 Severn Trent performed well against a number of PCs, such as external sewer flooding incidents, but underperformed against other selected PCs.[22]

Interdependencies across PCs

To aggregate risk sources, it is important to consider the relationships between the different PCs. In other words, how much the performance of the company against a certain performance commitment (e.g. supply interruptions) is likely to influence the performance against another one (e.g. water quality contacts).

A first step to understand the interdependencies between commitments is to check their correlations, across companies in a given year. Illustrative examples for 2017, 2018 and 2019 are shown in the following Tables 4.1 - 4.4.

To estimate correlation coefficients, we checked for consistency of targets and units of measures across companies and for the presence of structural breaks in the series. Then, we calculated the correlation coefficients by running an OLS regression of one commitment against the other, for all the combinations of PCs and including observations (normalised when appropriate) from all companies. We calculated correlation coefficients using information from all companies for each year, to account for the impact of time trends, and then take averages over time.

Finally, we consider that the analysis of interdependencies between PCs should be complemented by engineering judgement, to explain the rationale behind the estimated correlation coefficients.

[22] Severn Trent Annual Performance Report (2019), page 33, Available at: https://www.stwater.co.uk/content/dam/stw/regulatory-library/annual_performance_report_2018_2019.pdf

Table 4.1: 2017-18 Correlation matrix

	Supply interruptions	Water Quality contacts	Internal sewer flooding	Pollution incidents
Supply interruptions	1.00			
Water Quality contacts	**0.12**	1.00		
Internal sewer flooding	**-0.13**	**0.31**	1.00	
Pollution incidents	**0.04**	**0.72**	**0.07**	1.00

Table 4.2: 2018-19 Correlation matrix

	Supply interruptions	Water Quality contacts	Internal sewer flooding	Pollution incidents
Supply interruptions	1.00			
Water Quality contacts	**0.57**	1.00		
Internal sewer flooding	**-0.14**	**0.23**	1.00	
Pollution incidents	**0.14**	**0.76**	**0.35**	1.00

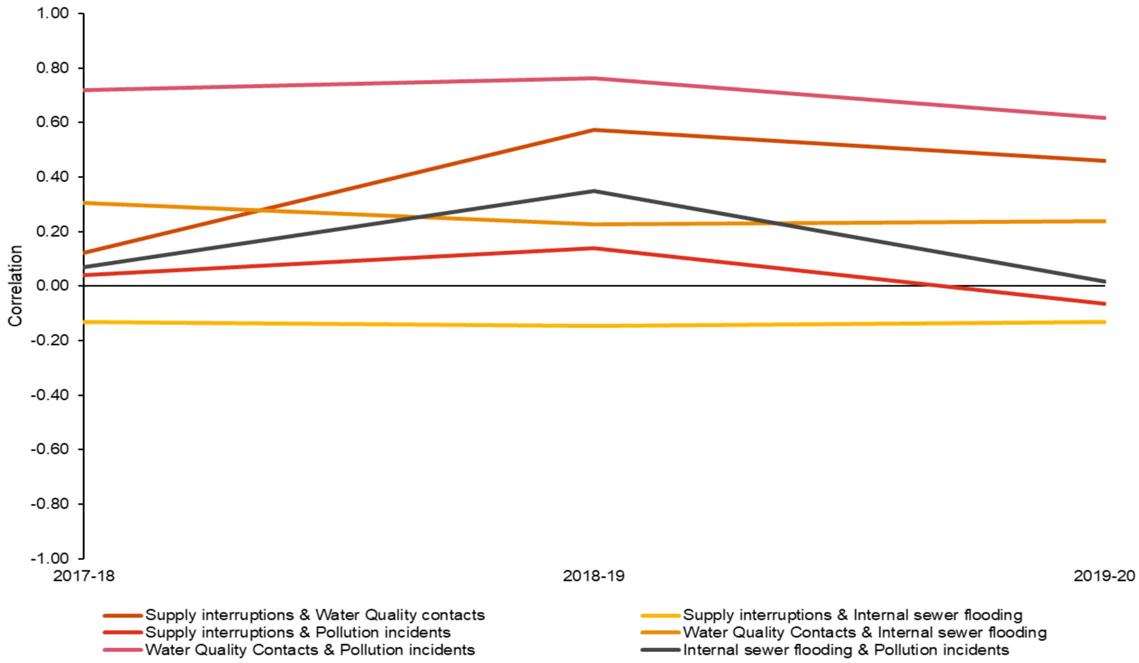
Table 4.3: 2019-20 Correlation matrix

	Supply interruptions	Water Quality contacts	Internal sewer flooding	Pollution incidents
Supply interruptions	1.00			
Water Quality contacts	**0.46**	1.00		
Internal sewer flooding	**-0.13**	**0.24**	1.00	
Pollution incidents	**-0.07**	**0.62**	**0.02**	1.00

While some correlations appear stable over time, such as supply interruptions and internal sewer flooding[23], others show considerable variability, such as pollution incidents with internal sewer flooding. We therefore plot coefficients to understand their dynamics over time, as shown in Figure 4.17, and then consider their average, as reported in Table 4.4.

[23]The correlation between supply interruptions and internal sewer flooding can be explained by how companies manage their network assets. Bad network maintenance is likely to result in longer and more frequent supply interruptions and a higher chance of internal sewer flooding.

Figure 4.17: Evolution of correlations over time



Averaged coefficients report generally low values within the range [-0.5, 0.5], with only one exception: 0.7 correlation between pollution incidents and water quality contacts (the correlation between PCs is likely to be driven by the quality of water treatment: bad water treatment is more likely to induce customer complaints, but also more likely to cause pollution incidents) and looks relatively stable over time. Therefore, when estimating their expected payments, companies should take into account the fact that a good (or bad) performance in one of the two PCs is likely to result in a good (or bad) performance in the other.

Table 4.4: Average of Correlations from 2017 - 2020

	Supply interruptions	Water Quality contacts	Internal sewer flooding	Pollution incidents
Supply interruptions	1.00			
Water Quality contacts	**0.38**	1.00		
Internal sewer flooding	**-0.14**	**0.26**	1.00	
Pollution incidents	**0.04**	**0.70**	**0.14**	1.00

It is worth noting that, as the number of PCs included in the analysis increases, the size of the matrix increases and it becomes more complex to account for all the relevant correlations. It is therefore important that a methodology only accounts for the meaningful interdependencies by, for example, simplifying the analysis by setting all small or insignificant correlations to zero.

5. Illustrative examples to support assessment of proposed methodologies

This chapter describes the analytical steps we have applied to generate illustrative RoRE ranges using the three methodologies, as well as the RoRE ranges themselves. Calculations are performed on a notionally geared basis, with a 60% gearing assumption, as at PR19. Data sources are the SDR (Service Delivery Reports) published on Ofwat's website companies' business plans supplied by Ofwat.

It is worth noting that our illustrative example is run on companies' historical performance and not on notional performance, as this depends on Ofwat's decision on the definition of notional performance: this means that for all companies we consider all historical observations, without excluding outlier events or taking a time average or the upper quartile performance. This might result in wider (or less accurate) risk ranges, as companies' actual historical performance may be outside of the notional range - in particular for methodology 2, that relies fully on historical data.

Methodology 1: PR19

To generate the illustrative RoRE range outputs under this methodology we take the following steps:

- **Step 1:** We take company estimates for P10 and P90 levels, based on companies' own methodologies. These were submitted to Ofwat, and subsequently refined as part of PR19.
- **Step 2:** Based on the P10 and P90 estimates, we calculate expected payment ranges. This is done by multiplying the relevant out/ underperformance rate with companies' P10 and P90 performance levels. We note that depending on the PC and ODI, the ultimate impact also accounts for deadbands[24], caps and collars[25], but for simplicity we do not include those in our illustrative example.
- **Step 3:** We sum the relevant out/ underperformance payments for each specific PC and then adjust them for the interdependencies across PCs. Adjustments are based on scaling factors, which were set by Ofwat to 70% for P10 and 90% for P90.
- **Step 4:** We transform payments into RoRE ranges: this is calculated dividing the payment by the forecasted regulated equity amounts for that specific year.

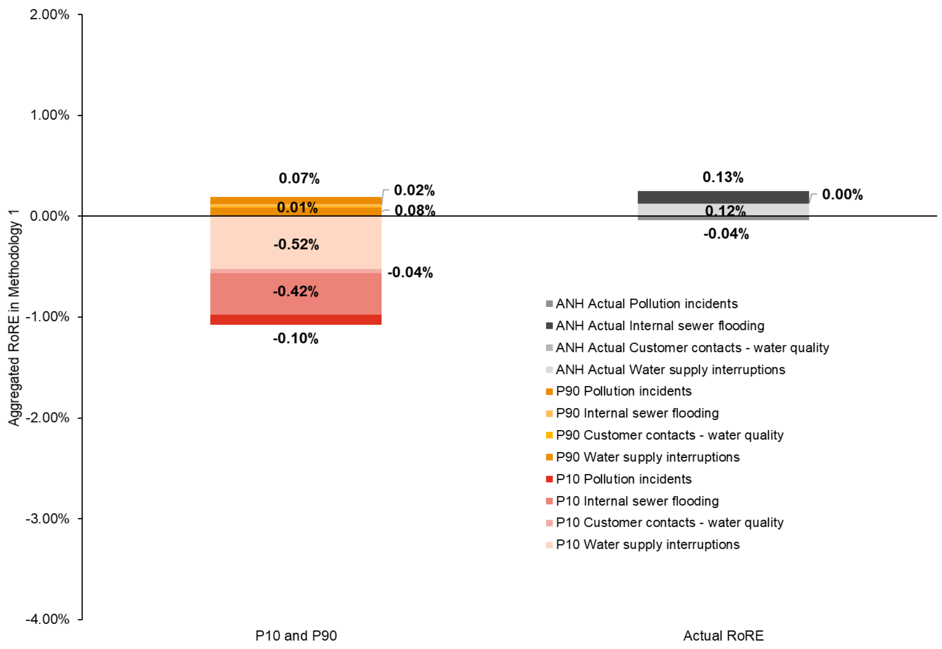
Anglian Water (ANH)

Table 5.1: Estimated RoRE range and Actual RoRE impact in Methodology 1 for Anglian Water

Estimated RoRE range	-1.08% to 0.19%
Actual RoRE impact[26]	0.21%[27]

[24] Ofwat define deadbands as a specified range of performance levels where the ODI underperformance or outperformance payment is zero
[25] Ofwat defines caps and collars as the limits on outperformance and underperformance payments for an ODI, respectively.
[26] Actual RoRE impact refers to 2020-21 actual performance, net of scaling factors.
[27] The actual 2020-21 performance of ANH's Water Supply Interruptions (5.03 minutes) and Internal Sewer Flooding Incidents per 10,000 properties (1.33), were below the projected P90 rates: 5.61 minutes and 1.62 respectively. Therefore, actual RoRE impact (0.21%) exceeds the P90 range (0.19%).

Figure 5.1: RoRE range for Anglian Water using Methodology 1



For Anglian Water, it is worth noting that the actual RoRE impact lies outside the range estimated through Methodology 1 by 0.02 percentage points, and the methodology estimates a downward skew, which did not materialise in practice, due to factors covered above.

Wessex Water (WSX)

Table 5.2: Estimated RoRE range and Actual RoRE impact in Methodology 1 for WSX

Estimated RoRE range	-0.53% to 0.58%
Actual RoRE impact	0.21%

Figure 5.2: RoRE range for Wessex Water using Methodology 1

For Wessex Water, we observe that the actual RoRE impact falls within the reasonably narrow estimated RoRE range.

Severn Trent (SVE)

Table 5.3: Estimated RoRE range and Actual RoRE impact in Methodology 1 for Severn Trent

Estimated RoRE range	-0.68% to 0.60%
Actual RoRE impact	-0.25%

Figure 5.3: RoRE range for Severn Trent using Methodology 1

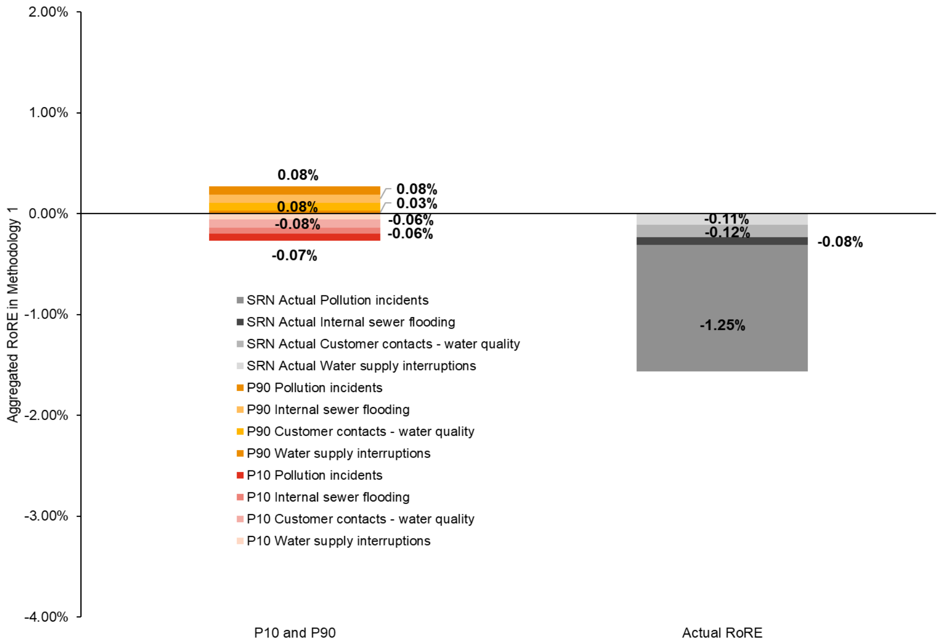
For Severn Trent, we observe that the actual RoRE impact falls within the reasonably narrow estimated RoRE range.

Southern Water (SRN)

Table 5.4: Estimated RoRE range and Actual RoRE impact in Methodology 1 for Southern Water

Estimated RoRE range	-0.27% to 0.27%
Actual RoRE impact	-1.56%[28]

Figure 5.4: RoRE range for Southern Water using Methodology 1



For Southern Water, we observe that the actual RoRE impact falls below the estimated RoRE range by almost 1 percentage point on the downside. This was caused by an especially poor performance on pollution incidents and illustrates that outturn performance can fall outside P10 and P90 reference points. In addition, it is worth noting that a collar is likely to have been triggered to limit the financial consequences of the unexpected poor performance.

Considerations

It is important to note that one of the limitations of Methodology 1 is a lack of consistency across companies' data which makes industry-wide comparison complex and require considerable effort to standardise units of measure.

Methodology 1 also does not specify companies' expected performance: whether it is at target, at P50, above or below one of the two.

[28] In 2020-2021, the SRN actual performance for Pollution incidents (101.52) was considerably weaker than the 2020-21 target (24.51). This is causing a large downside RoRE impact (-1.25%) which means overall RoRE impact (-1.56%) is lower than the P10 range (-0.27%).

Methodology 2: PR19 Refined

To generate the illustrative RoRE range outputs using this methodology we apply the following steps:

- **Step 1:** After standardising units of measure, we calculate P10 and P90 based on historical data (2012-20 period, subject to data availability) at the company level. Our illustrative example only estimates P10 and P90 figures, however other P numbers such as P5, P95 and P50 could be used to inform this analysis.
- **Step 2:** Based on the P10 and P90 estimates, we calculate expected payment ranges. This is done by multiplying the relevant out/ underperformance rate with the P10 and P90 performance levels calculated in step 1. Depending on the ODI, the calculation should account for deadbands, caps and collars - but we do not include those in our illustrative example.
- **Step 3:** We sum payments for each specific PC and then adjust the total to account for the interdependencies across PCs. For the purpose of our illustrative example we maintain the scaling factors at PR19 levels, as in Methodology 1 (70% for P10 and 90% for P90).
- **Step 4:** ODI payments are transformed into RoRE ranges, obtained by dividing the payments by the forecasted regulated equity amounts for the corresponding year.

Anglian Water (ANH)

Table 5.5: Estimated RoRE range and Actual RoRE impact in Methodology 2 for Anglian Water

Estimated RoRE range	-0.76%[29] to 0.85%[30]
Actual RoRE impact[31]	0.21%

Figure 5.5: RoRE range for Anglian Water using Methodology 2

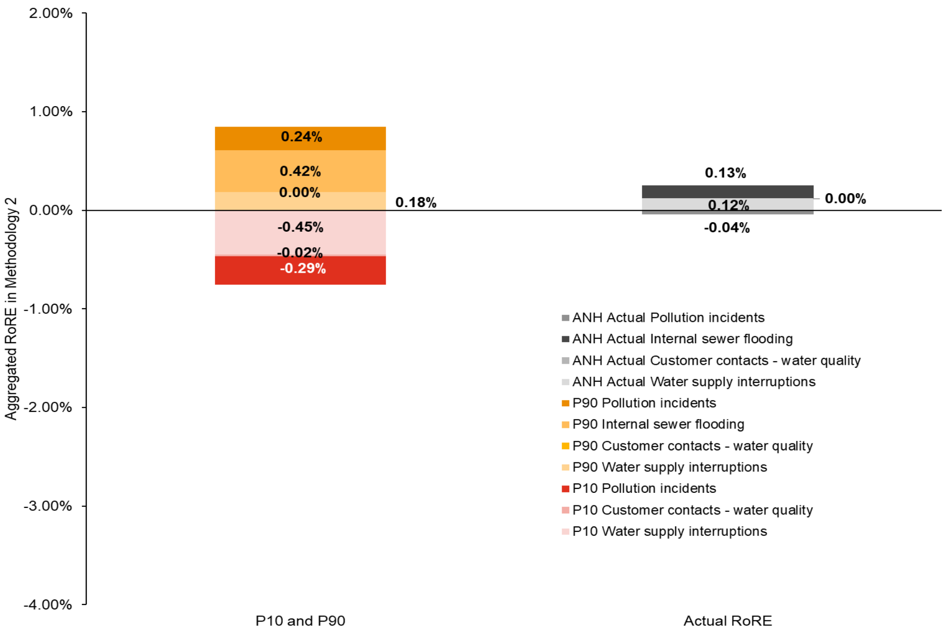


Figure 5.5 shows that the estimated range includes the actual impact, while being relatively narrow ([-0.76%;0.85%] range).

[29] Pollution Incidents per 10,000km P10 value is a positive number, so we have excluded it here. The positive P10 value is driven by ANH's historical P10 (247) being lower than the 2019-20 Target (298). This anomaly is driven by data limitations, as we only have access to data going back to 2017 for Pollution Incidents per 10,000km.
[30] Due to the data limitations outlined in the Footnote above, we consider the historical P50 (219) as the target for Pollution Incidents P90.
[31] Actual RoRE impact refers to 2020-21 net of scaling factors.

Wessex Water (WSX)[32]

Table 5.6: Estimated RoRE range and Actual RoRE impact in Methodology 2 for Wessex Water

Estimated RoRE range	-0.10% to 0.16%
Actual RoRE impact	0.21%

Figure 5.6: RoRE range for Wessex Water using Methodology 2

Figure 5.6 shows that Methodology 2 estimates a relatively narrow range. Furthermore, the actual RoRE impact lies outside the range estimated in Methodology 2 by a 0.05 percentage point difference.

Severn Trent (SVE)

Table 5.7: Estimated RoRE range and Actual RoRE impact in Methodology 2 for Severn Trent

Estimated RoRE range	-0.92%[33] to 0.69%
Actual RoRE impact	-0.23%

[32] Wessex Water had no PR14 PCL for Pollution Incidents. The PCL was to achieve EPA 4* star rating. Therefore, we use the historical P50 as a proxy for the Pollution Incidents PCL target.
[33] We have excluded P10 Internal Sewer Flooding Incidents here, due to the large deviation between the historical P10 (10.47) from the 2019-20 target (6.57).

Figure 5.7: RoRE range for Severn Trent using Methodology 2

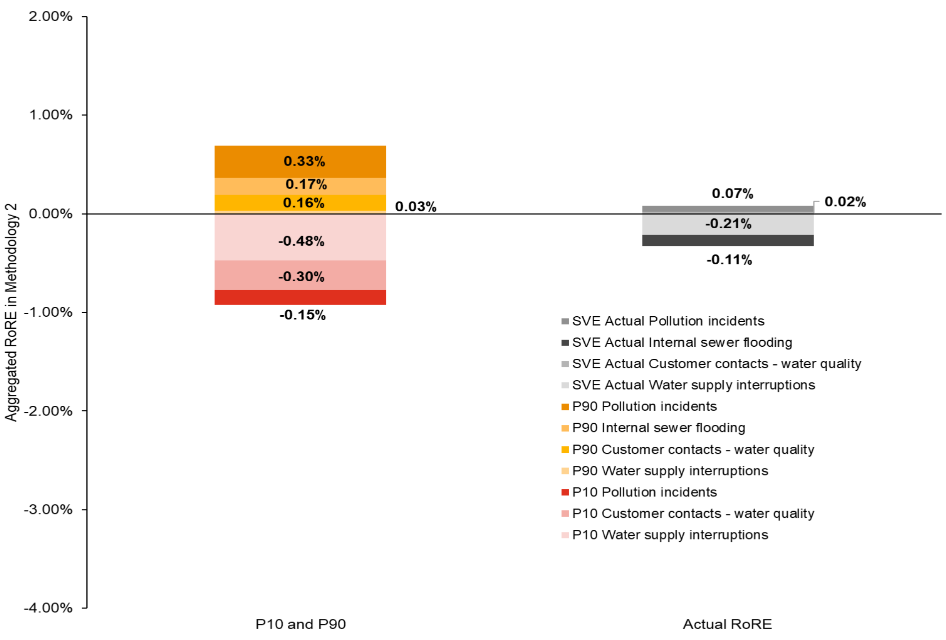


Figure 5.7 shows that the actual RoRE range lies within the estimated range in Methodology 2.

Southern Water (SRN)

Table 5.8: Estimated RoRE range and Actual RoRE impact in Methodology 2 for Southern Water

Estimated RoRE range	-3.47%[34] to 0.73%
Actual RoRE impact	-1.56%

[34] The large negative P10 value for Pollution Incidents is caused by a large deviation between the historical P10 (337.4) from the 2019-20 target (158.0). In 2019-20, SRN recorded 427 pollution incidents per 10,000km and as we only have access to data as far back as 2017, this has a large impact on the P10.

Figure 5.8: Stacked bar chart for results from Methodology 2 and Actual RoRE for Southern Water

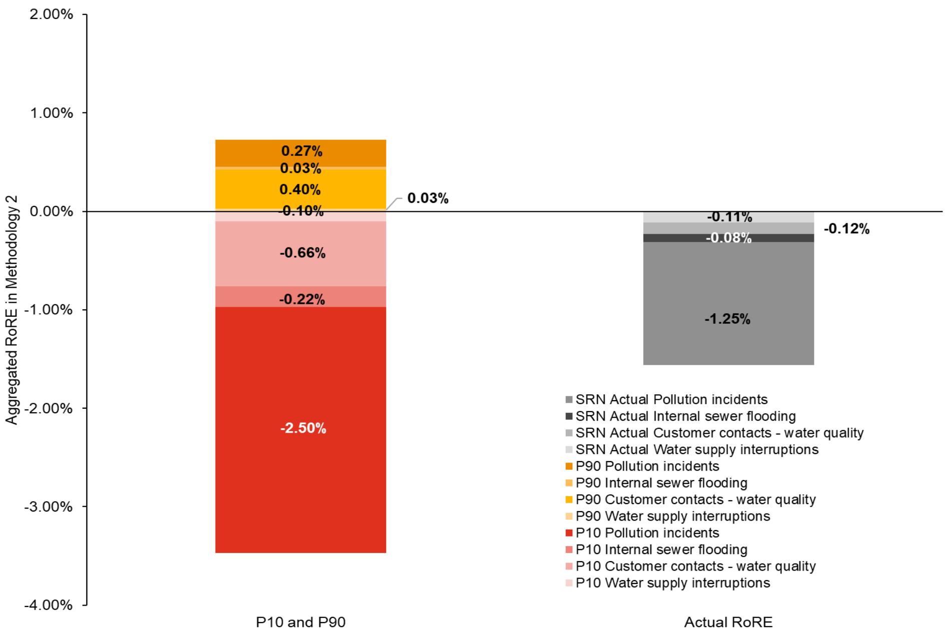


Figure 5.8 shows that Methodology 2 estimates a sufficiently wide range to capture the negative RoRE impact.

Considerations

Methodology 2 is a detailed, standardised and coordinated framework for assessing ODI risk and therefore could resolve compatibility issues. This methodology is able to solve the two main issues related to the other methodologies: it solves the inconsistency and lack of clarity problems of Methodology 1, while retaining some of the simplicity of the approach.

It is worth noting that also Methodology 2 still involves challenges: for example, in the case of Wessex Water, where the actual RoRE impact lies outside of the estimated range (but just for a small 0.05% difference[35]). Southern Water also shows how basing P10 and P90 on company historical performance can result in wide differences in risk ranges. As mentioned in the beginning of this Section, applying this methodology to the notional level of performance is likely to resolve these challenges.

Methodology 3: Monte Carlo

To generate the illustrative RoRE range outputs at the company level we apply the following steps:

- **Step 1:** We collect and sense-check available companies' historical data with standardised units of measure, to ensure consistency across companies.
- **Step 2:** Based on all the historical data collected, we make assumptions on performance distributions as inputs to run Monte Carlo simulations.
- **Step 3:** We set up a model that calculates payments from companies' performances and transforms them into aggregated RoRE ranges and then run the simulations using @risk software.
- **Step 4:** Monte Carlo outputs show the distribution of aggregated RoRE ranges. The P10-P90 range from that distribution is the result of the analysis using this methodology.

---

[35] This difference can be explained by multiple factors including totex sharing risk or the fact that the estimates refer to events with 80% probability of occurrence.

Anglian Water (ANH)

Table 5.9: Estimated RoRE range and Actual RoRE impact in Methodology 3 for Anglian Water

Estimated RoRE range	-0.07% to 0.44%
Actual RoRE impact[36]	0.21%

Figure 5.9: Monte Carlo Simulation for Anglian Water

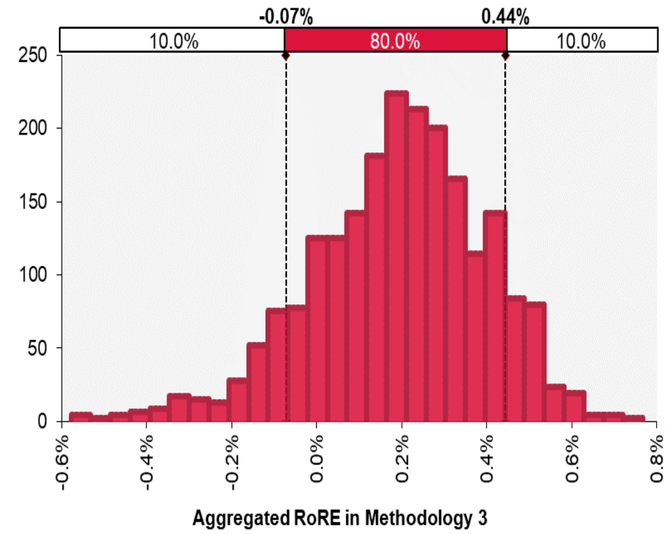
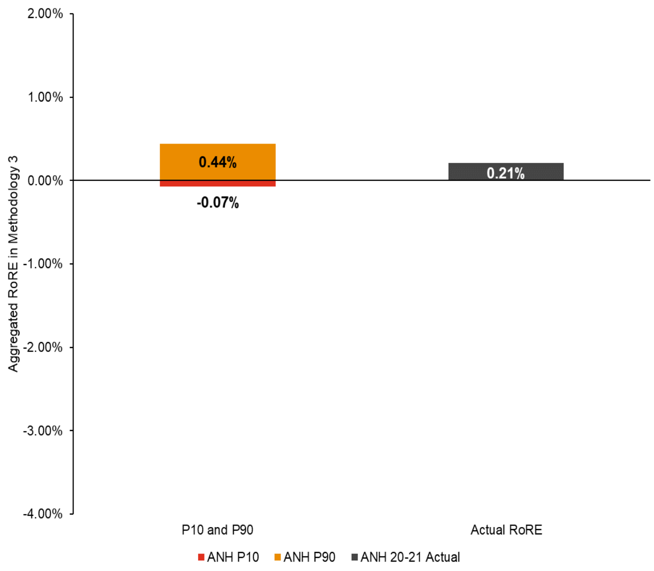


Figure 5.10: Anglian Water P10 - P90 and Actual RoRE



Wessex Water (WSX)

Table 5.10: Estimated RoRE range and Actual RoRE impact in Methodology 3 for Wessex Water

Estimated RoRE range	-0.03% to 0.41%
Actual RoRE impact	0.21%

[36] Actual RoRE impact refers to 2020-21 net of scaling factors.

Figure 5.11: Monte Carlo Simulation for Wessex Water

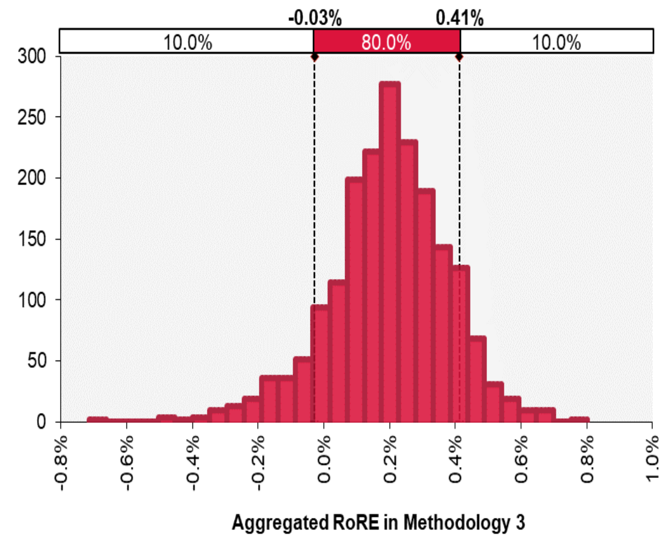
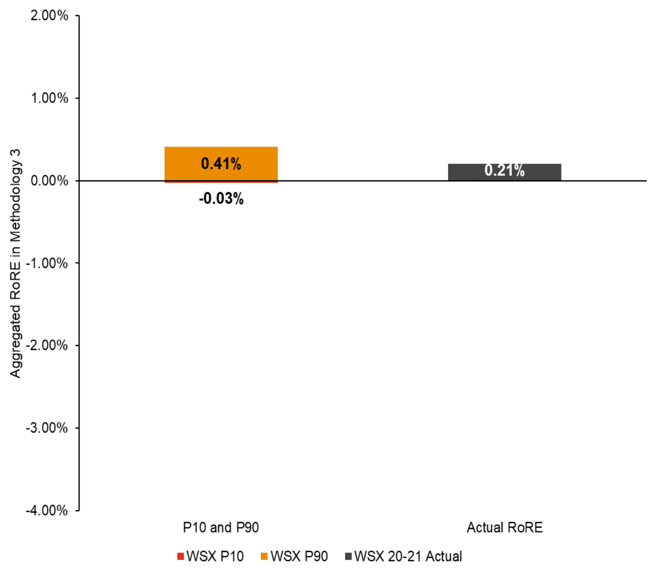


Figure 5.12: Wessex Water P10 - P90 and Actual RoRE



Severn Trent (SVE)

Table 5.11: Estimated RoRE range and Actual RoRE impact in Methodology 3 for Severn Trent

Estimated RoRE range	-0.58% to 0.07%
Actual RoRE impact	-0.25%

Figure 5.13: Results of Monte Carlo Simulation for Severn Trent

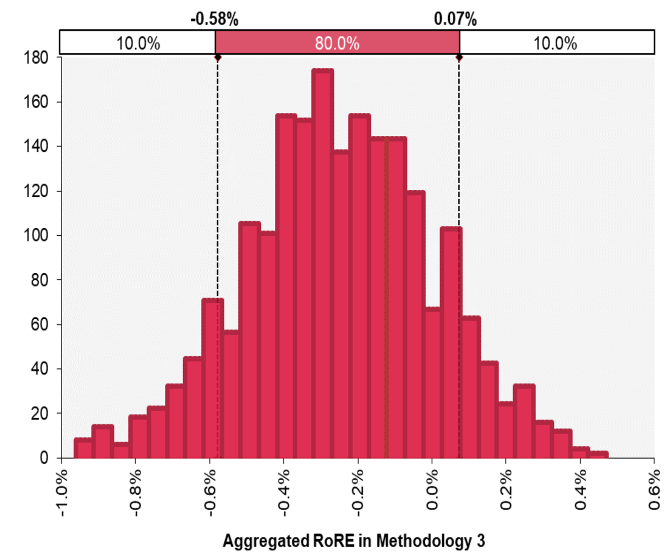
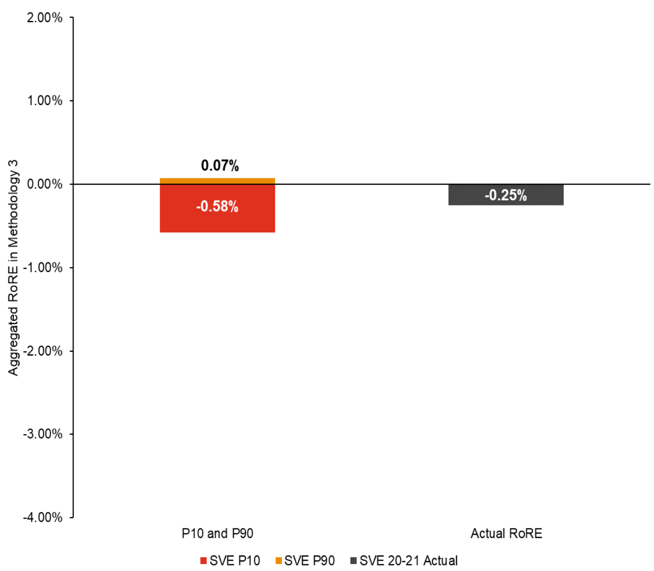


Figure 5.14: Severn Trent P10 - P90 and Actual RoRE



Southern Water (SRN)

Table 5.12: Estimated RoRE range and Actual RoRE impact in Methodology 3 for Southern Water

Estimated RoRE range	-3.22% to 0.09%
Actual RoRE impact	-1.56%

Figure 5.15: Monte Carlo Simulation for Southern Water

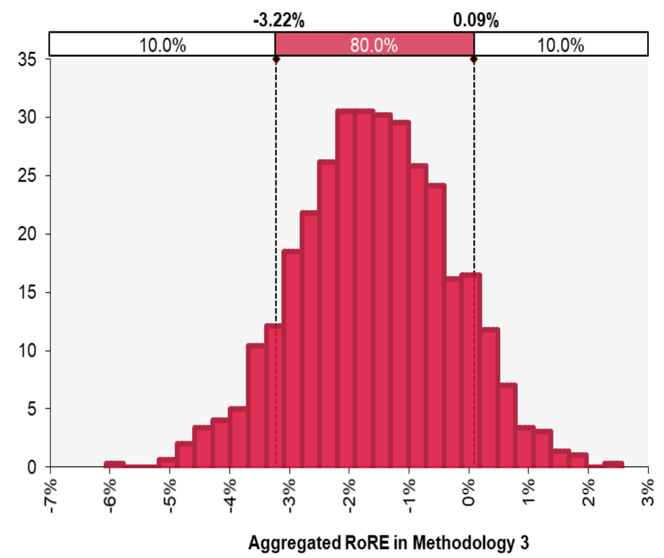
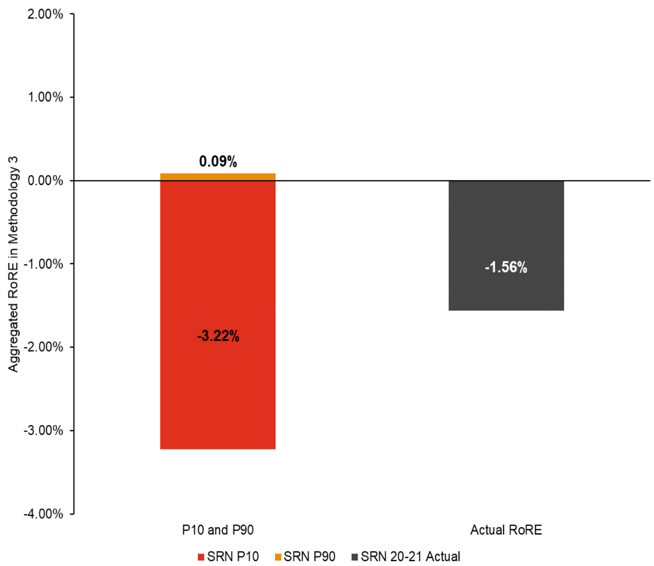


Figure 5.16: Southern Water P10 - P90 and Actual RoRE



Overall comparison of methodologies

In this section we compare the different methodologies for each company.

Anglian Water (ANH)

Figure 5.17: Methodologies 1 - 3 and Actual RoRE for Anglian Water.

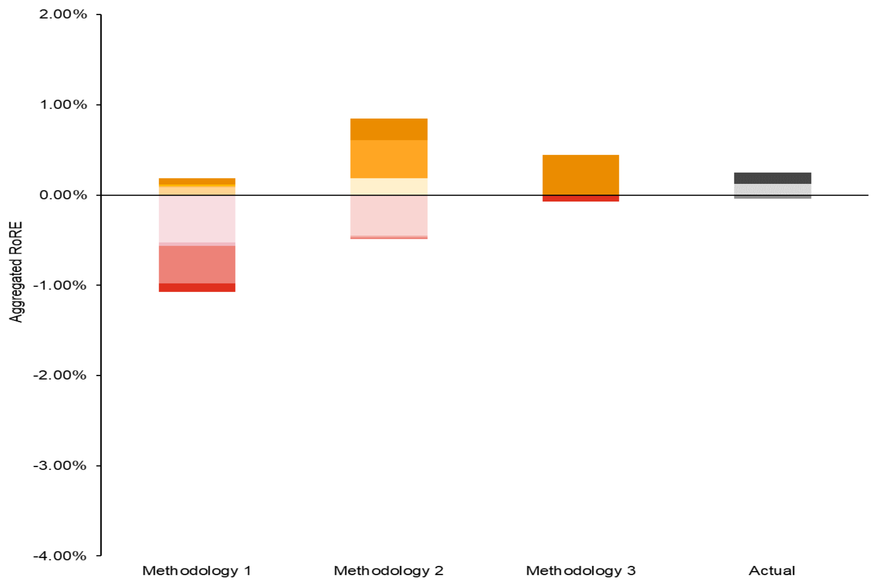


Figure 5.17 shows how RoRE range estimations compare across methodologies against the actual RoRE impact for Anglian Water. In Methodology 1, the actual RoRE impact lies outside the range estimated through Methodology 1 by 0.02%, and the methodology estimates a downward skew, whereas the actual RoRE range is skewed to the positive. On the other hand, the actual RoRE impact falls within the estimated ranges of both Methodology 2 and Methodology 3, with Methodology 3 producing a more precise range.

Wessex Water (WSX)

Figure 5.18: Methodologies 1 - 3 and Actual RoRE for Wessex Water

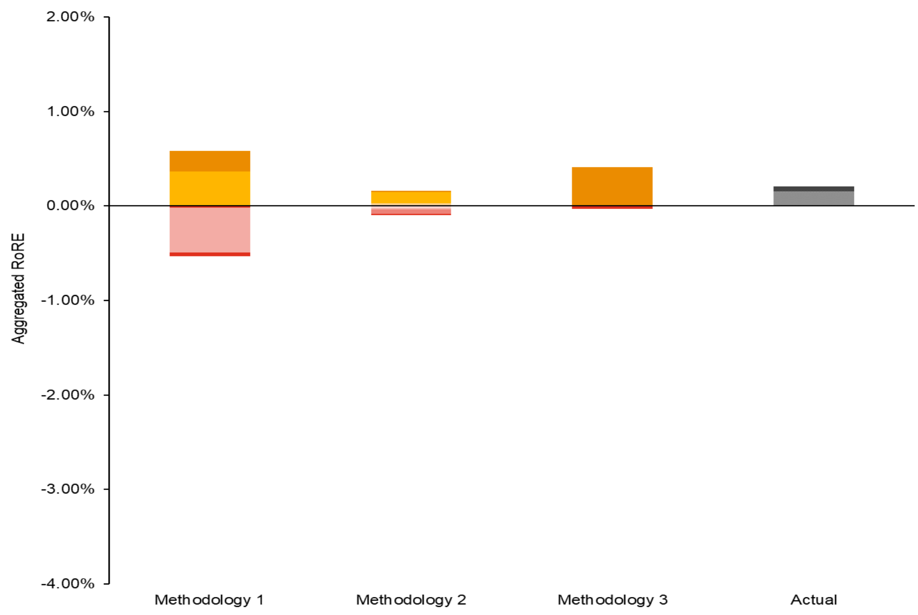


Figure 5.18 shows how RoRE range estimations compare across methodologies against the actual RoRE impact for Wessex Water. In Methodology 1 and 3, the actual RoRE impact lies within the reasonably narrow estimated RoRE ranges. However, in Methodology 2, the actual RoRE range lies outside the range estimated in Methodology 2 by a 0.05% difference.

This is an example that relying purely on historical data might sometimes lead to a failure to estimate an adequate range. However, by imposing data requirements that ensure consistency across companies and over time, the quality of Methodology 2 estimates will improve as companies develop their datasets following Ofwat’s guidance.

Severn Trent (SVE)

Figure 5.19: Methodologies 1 - 3 and Actual RoRE for Severn Trent.

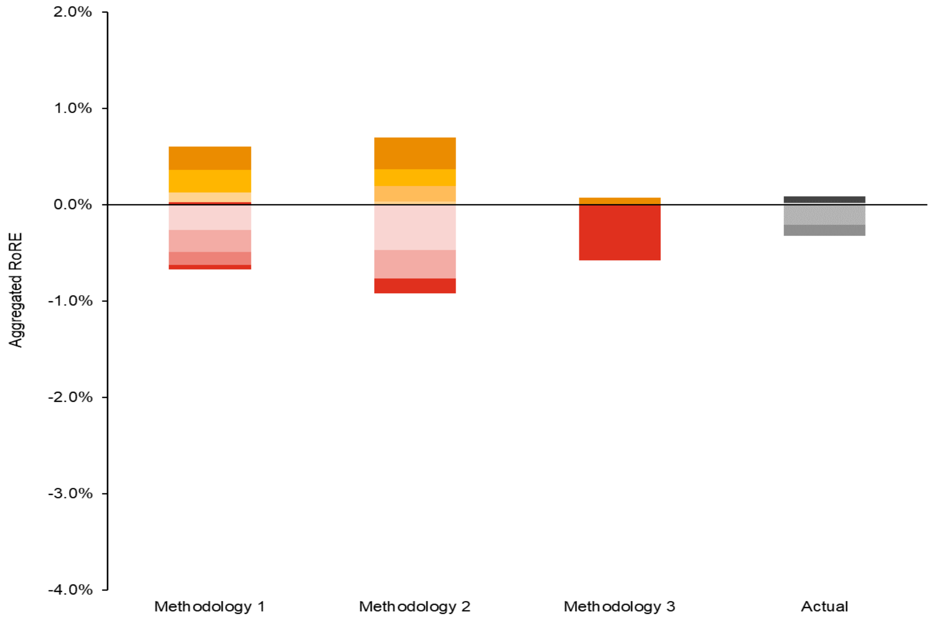


Figure 5.19 shows how RoRE range estimations compare across methodologies against the actual RoRE impact for Severn Trent. In all three methodologies, the actual RoRE impact lies within the estimated RoRE ranges.

In this case, it is worth noting that the range produced by Methodology 2 is wider (less accurate) than the one produced by Methodology 1 and this is likely to be driven by the quality of the data and the number of observations available for the estimate. Therefore, also in this case, by imposing data requirements that ensure consistency across companies and over time, the quality of Methodology 2 estimates will improve as companies develop their datasets following Ofwat's guidance.

Southern Water (SRN)

Figure 5.20: Methodologies 1 - 3 and Actual RoRE for Southern Water

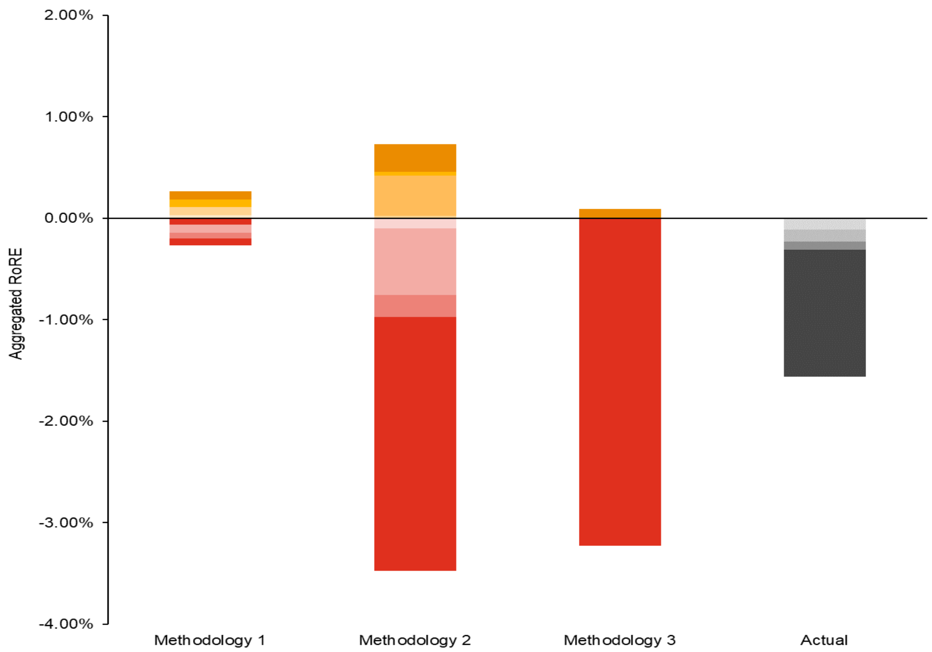


Figure 5.20 shows how RoRE range estimations compare across methodologies against the actual RoRE impact for Southern Water. In Methodology 1, the actual RoRE impact lies considerably outside the estimated range, representing a considerable amount of risk for the company. Conversely, Methodology 2 and 3 both estimate a sufficiently wide range to capture the negative RoRE impact.

This suggests that Methodology 1 might fail to show the appropriate level of risk faced by a company based on its best and worst-case scenario, as happened to Southern Water in 2020-21.

Considerations

The ranges estimated by Methodology 3 are narrower and more accurate than those estimated through Methodology 1 and 2. In addition, results appear to be more informative as they show the full distribution of RoRE impacts, allowing to target multiple points of the distribution such as P5-P95, P25, or P50.

However, there are drawbacks to consider for this methodology:

- It is impossible to identify the contribution of each PC to the overall RoRE range;
- The 'black box' nature of Monte Carlo analysis, which produces a set number of random draws, makes updates and re-runs difficult to undertake;
- It requires considerable statistical capability to correctly specify the assumption on the distributions of the different PCs as well as any other risk parameter included in the analysis and more time needs to be spent on specifying probability distributions and statistical linkages.

6. Summary Assessment

Drawing upon our illustrative examples, this chapter presents the findings of our assessment against our 5 key criteria:

- Ease of implementation
- Evidence / data requirements
- Consistency
- Re-running / updating analysis
- Verification / assurance

Details on our findings are reported in Table 6.1 below.

Table 6.1: Summary findings on our Methodologies

Criteria	Methodology 1 PR19	Methodology 2 PR19 refined	Methodology 3 Monte Carlo
Ease of implementation	Easy - This would involve rolling out the same methodology as used for PR19.	Moderate - This would involve making careful refinements to the PR19 methodology.	Hard - Implementing a full Monte Carlo approach would require substantially more guidance for companies, capability within companies to carry out Monte Carlo modelling. Ofwat would need to be able to incorporate a more complex approach into its PR24 methodology.
Evidence / data requirements	Low - Companies are given wide discretion to decide their own risk scenarios, own performance ranges, and make any adjustments. In our illustrative analysis, we found that the consistency of data between companies was poor (e.g. in different formats) and considerable effort was required to standardise them.	Medium - Data requirements are higher than PR19 approach (particularly for Ofwat), with emphasis on historical performance variation and standardisation across the industry. This may ultimately reduce data requirements for companies	High - Monte Carlo requires specific input on risk parameters, risk distributions, correlations and Monte Carlo running approach (e.g. fixed random numbers, number of simulations etc.). In our analysis, we show the impact of assumptions on performance distribution on the overall results. It is also apparent that risk distributions do not follow normal distributions, so simplifications to a Monte Carlo approach could reduce robustness in the results. It will therefore be challenging for companies and Ofwat to provide robust assumptions.

Criteria	Methodology 1 PR19	Methodology 2 PR19 refined	Methodology 3 Monte Carlo
Consistency	Poor - Without a clear methodology for companies to follow, there is a high risk of inconsistency. This is evidenced in the estimated RoRE ranges for Methodology 1: Anglian Water's and Southern Water's estimated ranges do not contain the actual RoRE impact.	Good - With a focus on core common ODIs, and risk ranges drawn from historical performance ranges, this methodology should deliver far more consistent outcomes compared to PR19. For example, in Figures 5.17 - 5.20 above, the estimated Methodology 2 RoRE ranges are generally more consistent with the actual RoRE impact than in Methodology 1.	Good - Because Monte Carlo requires a high degree of specification, the outputs should be consistent across companies. For example, this can be seen in Figures 5.17 - 5.20 below, where the estimated RoRE ranges are largely consistent with the actual RoRE impact.
Re-running / updating analysis	Moderate **-** Lack of clarity on how companies have run their risk analysis means it may not be easy to re-run or adjust analysis. This may therefore require multiple iterations of analysis. In our analysis we found that the P10 and P90 risk sources could not be tweaked given the inconsistent methodologies across companies.	Easy - Standardised requirements mean that it should be much easier to re-run analysis (e.g. changing a parameter consistently across the whole industry). Also, under one of the variants it requires less involvement / modelling from companies.	Hard - Monte Carlo runs can be different if using a different sequence of random simulations. Monte Carlo results can't easily be replicated without access to the model used to create them. In addition, our analysis shows that the results from the Monte Carlo simulation do not allow easy observation of the effects of each individual ODIs.
Verification / assurance	Moderate **-** Simple analysis should enable companies and Ofwat to verify.	Easy - Consistent guidance on data sources and inputs should mean that results can be easily replicated, verified and externally assured.	Hard - Monte Carlo results can't be easily verified (as the P10 (& P50,P90) points are simulation outputs that cannot be calculated independently.

Our assessment suggests that **Methodology 1 would be the easiest to implement**, and with the lowest data requirement for companies. However, **it is the hardest to update and verify**, due to its lack of consistency in terms of methodology used for estimations, but also data consistency (e.g. units of measure used for performance data).

Methodology 3, instead, ensures the highest consistency, both in terms of data required and method used to estimate the reference points, **but it has the drawback that it is hard to implement and to verify** - in that it suffers the risk of being a more of a 'black box', where it is difficult to verify the contribution of each individual PC to overall results.

Finally, **Methodology 2 strikes a good balance across the assessment criteria** in that it imposes reasonable requirements on companies in terms of data and methodology applied, while ensuring high consistency, replicability of the results and ease of external verification/assurance.

7. Conclusions and recommendations

In this chapter, we present our conclusions with respect to the three methodologies proposed and provide our suggestions and recommendations to Ofwat for its next steps.

Key Considerations

The application of the three methodologies to historical data suggests that, in general, Methodologies 2 and 3 tend to provide more precise ranges, showing a more appropriate picture of the risk faced by companies.

However, there are cases in which Methodology 2 provides estimates that are unnecessarily wide or that fail to capture the actual impact. This might be explained by two factors: first the fact that historical data might have high variability (given that it includes outlier events); second are totex interactions, in that a higher expenditure in a certain area may reduce the risk associated with it. Both issues are likely to be solved by applying Methodology 2 (in all its variants) at the level of an efficient company with a notional structure - as described in section 2.

Methodology 1: PR19 Methodology

Methodology 1 has a number of limitations which have been raised by companies on several occasions.[37] Due to inconsistencies in some of the data used by certain companies and the differences in the approaches used for the estimation of the reference points, it is difficult for Ofwat to provide guidance and implement a standardised approach.

As shown in Table 2.1, the primary advantage of Methodology 1 is that it is simple for companies to implement and easily allows them to account for company specific risks.

Methodology 2: Refined PR19 Methodology

Methodology 2 imposes higher data requirements such as standardised units of measure, and guidance on how to provide risk ranges and adjustment factors. This means Ofwat should provide companies with guidance regarding the format, units of measure and consistency of data. Due to the strong reliance on historical data in Methodology 2, companies must ensure that their historical data is accurate and available, for consistency over time. Furthermore, calculations are meant to be fully replicable, which implies Ofwat should require companies to submit their data to check and supervise data consistency before risk outputs are calculated.

Overall, Methodology 2 provides a good balance as it requires a reasonable burden on companies in relation to data requirements and better standardisation/consistency to estimate risk, but it does not require the complexity of a Monte Carlo analysis, as confirmed by Table 2.2. This means companies should be able to perform the analysis internally with existing capabilities. The next subsection compares the different variants of this methodology.

Methodology 3: Monte Carlo Analysis

Methodology 3 requires that all companies comply with the data requirements imposed by Methodology 2 to ensure analysis is run on a set of data that is standardised and consistent across companies. As this methodology uses Monte Carlo simulations, it also requires all companies to specify the assumptions used in their analysis (including distributions and correlations across all the PCs) and submit them to Ofwat.

Monte Carlo analysis is complex, lacks transparency and requires appropriate software to perform the simulations. There is therefore a risk that either this methodology becomes resource intensive, or the quality of statistical analysis could be compromised.

A further drawback of Monte Carlo analysis is that, given it is based on a set number of random draws (iterations), it cannot be easily replicated or updated easily to account for adjusted parameters.

However, if implemented correctly, this Methodology is best placed to consider company specific aspects and interdependencies across PCs, delivering the most precise estimates.

[37] See Section 1 for a more detailed description of the points raised by companies.

Overall considerations

Our suggestion to Ofwat is to adopt Methodology 2, which would:

- Address most of the concerns raised by companies in their consultation responses and representations around transparency, consistency and equality of treatment. This would, however, lead to higher data requirements;
- Presume Ofwat would continue to harmonise common PC definitions;
- Presume Ofwat would develop a comprehensive record of historical performance variation which will grow over time.
- Require guidance on how to blend company historical performance variation with industry level performance variation, i.e. by setting a definition of notional and efficient performance. Our illustrative examples show that using company-specific historical variation can result in widely different risk ranges. While there may be strong grounds for using an overall industry approach, there may also be a requirement to incorporate company specific factors (e.g. by adding the financial contribution of bespoke ODIs to the risk ranges calculated through the ODI risk measurement methodology);
- Require that Ofwat update the scaling factors, using Variant 2 (MC scaling factors) or Variant 3 (Correlation matrix).[38]
    - Using Variant 2 (MC scaling factors), Ofwat calculates the scaling factors by implementing a Monte Carlo analysis using historical performance information from all companies, as described in the example box below.

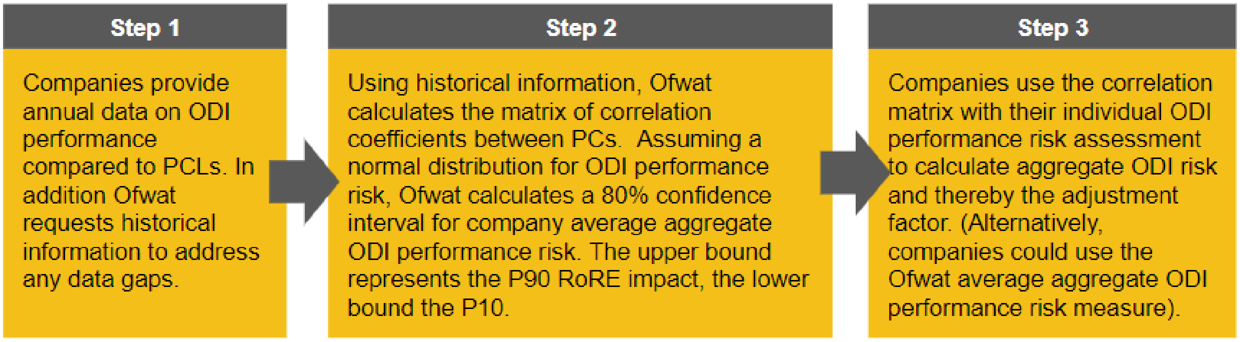
Figure 7.1: Example box for calculating scaling factors through Methodology 2 (Variant 2 - MC scaling factors)

Step 1	Step 2	Step 3
Companies provide annual data on ODI performance compared to PCLs. In addition Ofwat requests historical information to address any data gaps.	Ofwat compiles a consistent dataset of historical performance information. This includes company-level and industry average ODI performance ranges which are also expressed as RoRE ranges for individual ODIs. Ofwat calculates the P10 and P90 performance range, based upon historical information.	Ofwat uses Monte Carlo analysis, relying on historical performance distributions and correlations, to assess company average aggregate ODI risk. Ofwat may adjust distributions or correlations for outliers. The industry scaling factor is calculated by comparing Monte Carlo output to the addition individual ODI risk, calibrated at the P10-P90 distribution points.

- Using Variant 3 (Correlation matrix), Ofwat provides companies with the inputs required to calculate the scaling factors, based on the matrix of correlation coefficients between PCs of all companies and an assumption of normality for the distribution of the RoRE ranges. These are obtained using the formula specified in Section 2 and following the steps described in the example box below.

[38] It is not possible to present results for Variant 2 (MC scaling factors) and Variant 3 (Correlation matrix) of Methodology 2 from our illustrative example, as only considering our subset of PCs is likely to generate scaling factors that are materially different from the ones we would obtain by including all PCs.

Figure 7.2: Example box for calculating scaling factors through Methodology 2 (Variant 3 - Correlation Matrix)



Both exercises deliver results that would represent refinement from the PR19 approach (which had limited historical data to draw from). However, it should be noted that Variant 3 relies on the assumption of normality of distribution of the RoRE impacts. If, reviewing the historical data, an assumption of normal distribution may not be appropriate in all cases, we then suggest Ofwat uses Variant 2 (MC scaling factors) as Monte Carlo analysis can easily cater for different distribution assumptions.

In addition, we consider that only the common PCs should be included in the calculations, to ensure the scaling factors obtained can be applied consistently to all companies. Following this, companies could add the financial contribution of any bespoke ODIs to the estimated ranges, as those are highly company-specific and unlikely to be correlated to other common PCs. Any other company specific adjustment could be raised by companies, including a clear justification for it.

This document has been prepared only for Ofwat and solely for the purpose and on the terms agreed with Ofwat. We accept no liability (including for negligence) to anyone else in connection with this document, and it may not be provided to anyone else. If you receive a request under freedom of information legislation to disclose any information we provided to you, you will consult with us promptly before any disclosure.

This is a draft prepared for discussion purposes only and should not be relied upon; the contents are subject to amendment or withdrawal and our final conclusions and findings will be set out in our final deliverable.

© 2022 PricewaterhouseCoopers LLP. All rights reserved. 'PwC' refers to the UK member firm, and may sometimes refer to the PwC network. Each member firm is a separate legal entity. Please see www.pwc.com/structure for further details.