

PERFORMANCE COMMITMENTS AND OUTCOME DELIVERY INCENTIVES AT PR19

A report prepared for Ofwat

28 March 2017



Annabelle Ong

 020 7031 7000

 annabelle.ong@frontier-economics.com

Rob Francis

 020 7031 7000

 rob.francis@frontier-economics.com

Frontier Economics Ltd is a member of the Frontier Economics network, which consists of two separate companies based in Europe (Frontier Economics Ltd, with offices in Brussels, Cologne, Dublin, London & Madrid) and Australia (Frontier Economics Pty Ltd, with offices in Melbourne & Sydney). Both companies are independently owned, and legal commitments entered into by one company do not impose any obligations on the other company in the network. All views expressed in this document are the views of Frontier Economics Ltd.

CONTENTS

Executive Summary	5
Background and project objective	5
Key requirements for the PR19 framework	5
Recommendations: general improvements	6
Recommendations: options for key elements of PR19 framework	7
Overview of packages	8
Assessment of options	11
1 Introduction	15
1.1 Introduction	15
1.2 Project objective	16
1.3 Report outline	16
2 Key requirements for PR19 framework	17
2.1 Lessons learned from PR14	17
2.2 Outcomes consultation	24
2.3 Wider policy context	25
2.4 Summary	27
3 Recommendations: general improvements	28
3.1 Improvements to CBA method	28
3.2 Increasing the effectiveness of Ofwat's role	37
3.3 Increasing the effectiveness of the CCG challenge	39
3.4 Summary	40
4 Options for key elements of PR19 framework	42
4.1 High level instruments	42
4.2 Detailed methodology for the implementation of PCs	44
4.3 Detailed methodology for the implementation of ODIs	50
4.4 Implementing a measure-specific approach	58
4.5 Summary	58
5 Overview of packages	60
5.1 Package 1 – Focus on PR14 improvement	61
5.2 Package 2 – Focus on innovation	63
5.3 Package 3 – Focus on tailored rewards	64
5.4 Package 4 – Focus on integrated assessment	65
6 Assessment of options	68
6.1 Assessment against criteria	68
6.2 Ofwat's key choices for PR19	70
ANNEX A Data Envelopment Analysis	77
ANNEX B Implementing a measure-specific approach	80
ANNEX C Detailed assessment of packages	89

EXECUTIVE SUMMARY

Background and project objective

At PR14, Ofwat introduced an outcomes-based framework, which involved extensive customer engagement and companies setting their own outcomes, measures, targets and incentives alongside challenge from Customer Challenge Groups (CCGs). Overall, the PR14 framework has been viewed as a success as it is considered to be a truly innovative approach, and other regulators are considering how it could be incorporated into other sectors.

As the outcomes-based framework was a new approach at PR14, Ofwat's ambition is now to build on the lessons learned from PR14 to refine the outcomes framework for PR19. As part of the Water 2020 publications Ofwat has identified a number of ways in which the PR19 framework could be developed, and recently Ofwat has published a specific consultation on the PR19 outcomes framework, "the outcomes consultation".¹ While the outcomes consultation indicates Ofwat's potential approach to outcomes, measures and common targets, it does not provide detail on how PCs and ODIs may be set at PR19.

The objective of this project is to support Ofwat's development of the PR19 methodology, by providing options for the way that:

- PCs and commitment levels are set by companies (for bespoke PCs and common PCs where Ofwat does not set common commitment levels);
- ODIs are set for both common and bespoke PCs.

The way that Ofwat might set common commitment levels for some common PCs is outside the scope of this project.

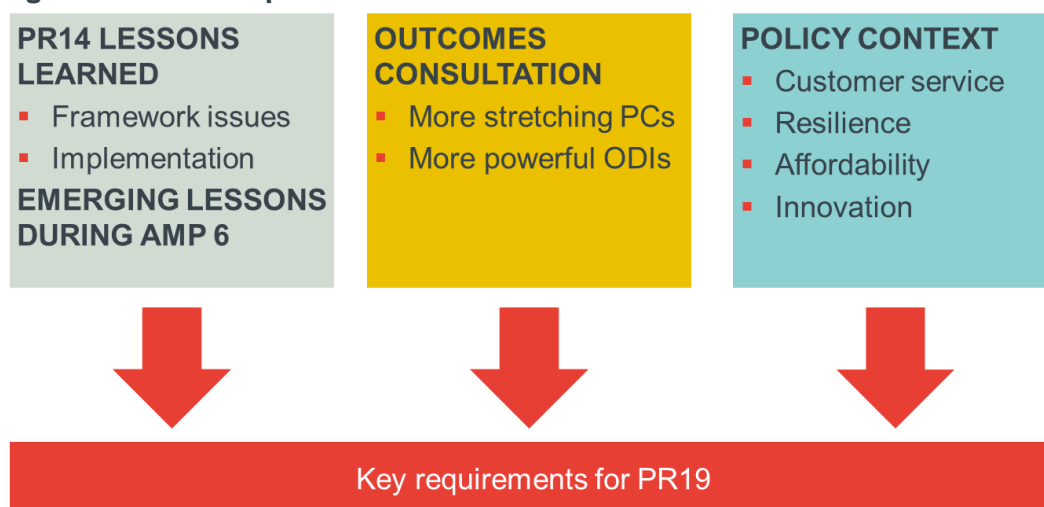
The aim of this project is to help Ofwat develop the PR19 methodology and the expectations that it will set for companies. The output from this project will enable Ofwat to make a confident choice regarding its methodology for PR19.

Key requirements for the PR19 framework

The PR19 outcomes framework needs to reflect the lessons learned from PR14, the objectives of the outcomes consultation, and the wider policy context. The diagram below illustrates how these points should be reflected in the development of the PR19 framework.

¹ Ofwat (2016), A consultation on the outcomes framework for PR19

Figure 1 Development of the PR19 framework



Source: Frontier Economics

Lessons learned from PR14 include specific issues with the detailed framework (e.g. reliance on single willingness-to-pay values), implementation issues (e.g. data requirements for filling in data tables) and emerging lessons during AMP6. In general the lessons indicate that the use of cost benefit analysis can be improved and companies, CCGs and Ofwat could make greater use of comparative information. The two key aspects of the outcomes consultation that are relevant for this project are Ofwat’s desire to introduce more stretching PCs and more powerful ODIs. The PR19 framework for PCs and ODIs also needs to recognise, where relevant, the four key themes of PR19 (customer service, resilience, affordability and innovation). This means that the PR19 approach needs to identify how these policy areas can be captured by the methodology.

Recommendations: general improvements

Based on the key requirements for PR19, we have identified a set of recommendations that focus on general improvements. These are clear improvements on the PR14 framework and so it is recommended that they should be made, regardless of what other choices are made about the PR19 framework. The box below summarises these recommendations for the PR19 framework.

RECOMMENDATIONS FOR PR19: GENERAL IMPROVEMENTS

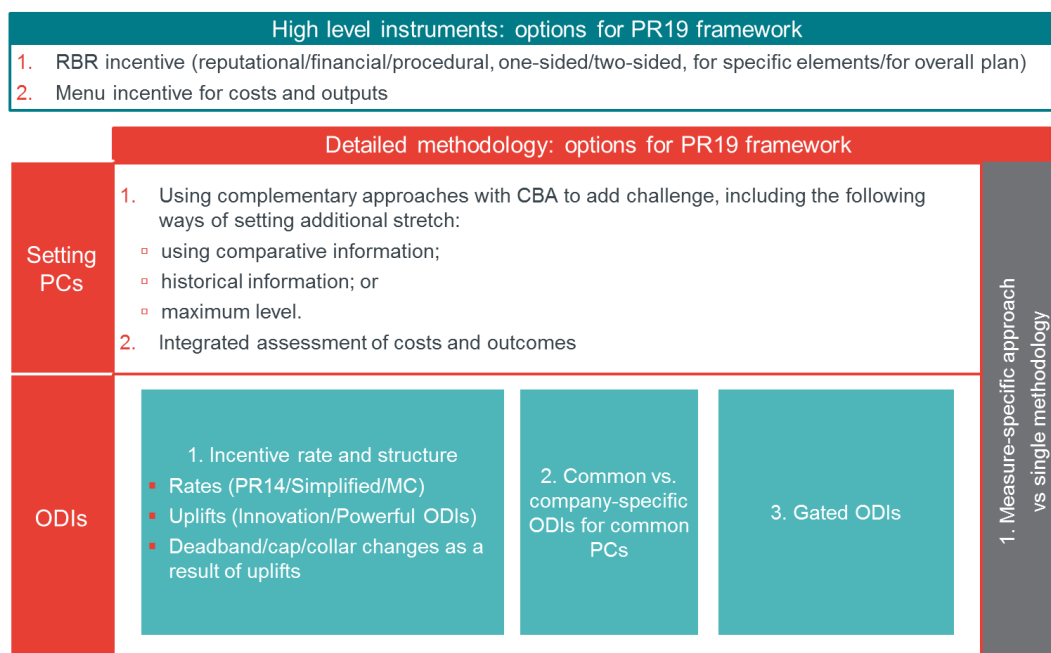
- Set higher expectations on the quality of the data and processes around Cost Benefit Analysis (CBA) to set PCs levels (i.e. both customer valuations and marginal cost data). This includes expectations for companies to carry out sensitivity tests to assess the impact of different levels of customer valuations and marginal costs. This should improve the robustness of the CBA, and result in more informed choices on where to set targets.
- Improve the effectiveness of Ofwat's role in assessing business plans by:
 - drawing on the lessons learned from PR14 to improve the effectiveness of process benchmarking of companies' plans;
 - committing to undertake deep-dives (i.e. more in-depth assessments) on a subset of PCs (that is selected after business plan submission); and
 - setting clearer expectations on how P10 and P90 performance levels should be calculated as part of the Return on Regulated Equity (RoRE) range, and be clearer on how interactions between PCs should be reflected in the range.
- Improve the tools CCGs have available to better enable them to play their role and challenge companies as effectively as possible:
 - workshops could be offered to new CCG members, and existing members that are interested, to explain the role the CCGs are expected to play and how they can best challenge / ask companies questions;
 - the availability of comparative information on companies' performance could be improved, including through the development of the Discover Water dashboard; and
 - best-practice from PR14, and updated during PR19, on how CCGs applied challenge, including how the questions asked could be shared between all CCGs, through the quarterly CCG chairs meetings and other means.

Recommendations: options for key elements of PR19 framework

In addition to the general improvements, we have identified a number of options that capture the key choices for how to set the incentives going forward. For each key element of the framework we have considered a number of different options available and the associated trade-offs. For example, some options trade off stronger incentives for innovation against a higher risk of unintended consequences, or a higher degree of complexity against improved robustness against data issues.

The figure below summarises the key elements of the methodology for PR19. It shows that Ofwat needs to decide on the high level instruments for PR19 as well as the detailed methodology for PCs and ODIs.

Figure 2 Summary of decisions for Ofwat on the detailed methodology



Source: Frontier Economics

Overview of packages

As there are many different choices to be made, we have developed four plausible packages (in addition to a base case, the PR14 framework). The rationale for the packages is to manage this complexity and facilitate the evaluation of Ofwat’s choices.

Each package represents a different combination of options, and is focused on achieving a specific policy objective. Within those policy objectives, we have selected the boldest set of options for each package. In the full report, we have identified changes that could be made to the packages that would reduce the risk of unintended consequences. The figure below summarises the four packages, as well as the base case package, and below the figure we provide a summary of each package.

Figure 3 Summary of packages

Base case

	High level instruments	Methodology
PCs	Rating of PCs and ODIs as part of overall RBR, high level review of all ODIs	PR14 (CBA for all bespoke, CCG provides challenge)
ODIs		PR14 (Company-specific ODIs on common PCs) PR14 (ODIs for all measures based on PR14 formula)

1. Focus on PR14 improvements

	High level instruments	Methodology
PCs	Same as base case (with general improvements)	CBA for all bespoke, more expectations set for CBA approach, rely on comparative info given to CCGs for extra challenge
ODIs		Common ODIs incl. uplifts for common PCs Based on PR14 formula for bespoke PCs with uplifts only on some measures

2. Focus on innovation

	High level instruments	Methodology
PCs	RBR incentive	Measure specific approach, proportionate application of CBA, incl. additional challenge
ODIs		Common ODIs with innovation uplift for common PCs Based on PR14 formula for bespoke PCs with innovation uplifts

3. Focus on tailored rewards

	High level instruments	Methodology
PCs	RBR incentive	Measure specific approach, proportionate application of CBA, incl. additional challenge
ODIs		Company specific ODIs for common PCs with uplifts Uplifts for ODIs on bespoke PCs where evidence is strong Gated ODIs

4. Focus on integrated assessment

	High level instruments	Methodology
PCs	Menu incentive for costs and outcomes	Assess costs and outcomes together combining results from multiple methodological approaches (e.g. SFA, DEA) Companies to collect data on customer valuations to apply to results from integrated benchmarking
ODIs		Rewards on each measure depend on performance on multiple measures, to reflect customer preferences across measures Innovation uplifts available for some measures

Source: Frontier Economics

- Base case.** We have designed the base case as the approach that was used at PR14. The outcomes framework would feature as part of the risk based review (RBR), but there would be no specific incentive for PCs and ODIs. CBA would be used to set PCs, with challenge from the CCGs, and the PR14 formula would be used to set ODIs.
- Package 1: Focus on PR14 improvements.** There would be no additional high level instruments in this package, relative to the base case. CBA would be used to set bespoke PCs, and CCGs would have access to comparative information to inform their challenge. Common ODIs would be used for common measures, and would include uplifts on rewards relative to the PR14 formula. ODIs on bespoke PCs would be based on the PR14 formula for ODIs, and would also include uplifts to rewards on some measures (where there was confidence that commitment levels are stretching). The rationale for these uplifts is that PR14 rewards were not always seen to be sufficient to encourage investment to improve performance above commitment levels.

- **Package 2: Focus on innovation.** This package would include a specific RBR incentive for PCs and ODIs. A measure specific approach would be used to set PCs, where CBA is largely used, including additional challenge such as a minimum improvement based on historical improvement. Common ODIs would be used for common PCs, and would include innovation uplifts. These innovation uplifts could be set in a number of ways. One option would be to set them in a way that reflects industry-wide customer valuation for innovative performance, and would be justified on the basis of positive externalities for innovation. This would be the maximum possible level for the uplifts. An alternative would be to set the uplifts based on expert judgement, which could be informed by an assessment of the level of benefit that would be brought about by improved performance, perhaps similar to the sort of analysis used in merger cases. Expert judgement could also be used to assign an uplift to the customer valuations collected through customer valuations, which could then result in company specific ODIs that reflect the size of the company as well as the benefits from innovation. As these uplifts reflect positive externalities of innovation, the rationale for including them is strongest on common PCs (as all companies measure those and have the same commitment levels). However, as we have defined this as the boldest package possible under the innovation objective, we have also suggested that innovation uplifts could apply to some bespoke PCs, but largely where at least some other companies measure the same PC and there is some chance that positive externalities would occur from innovation.
- **Package 3: Focus on tailored rewards.** This package would also include a specific RBR incentive on PCs and ODIs. It would be based on a measure-specific approach to setting PCs, using CBA and including additional challenge, as in package 2. ODIs would be set in a tailored and proportionate way. So company specific ODIs would be used on common PCs, and would include uplifts to provide additional incentives to improve performance above the commitment level, relative to the base case. Uplifts could also be applied to ODIs on bespoke PCs, but only in cases where there was sufficient confidence that the commitment level was stretching. Gates would be applied to some PCs, but they would only be applied to some measures and at levels just below the commitment level. This package would therefore go further than package 1, as it would include a specific RBR incentive, use a measure-specific approach, and include gates.
- **Package 4: Focus on integrated assessment.** A menu incentive would be set that reflects both costs and outcomes. Costs and outcomes would be assessed together using a range of methods, including for example data envelopment analysis. Ofwat would use the results from these assessments to define the efficient frontier, and companies would use insights from customer engagement to choose a place on that frontier. Rewards would depend on a company's performance across multiple measures, to reflect customers' preferences across a range of measures, and would include innovation uplifts on some measures.

As noted above, we have defined these packages so that each one consistently addresses a particular policy objective, and also reflects the boldest options within that policy context. That being said, we also recognise that there are

multiple options, and that changes could be made to the packages we have defined, while still combining options in a coherent way. For example, while we feel that gated ODIs fit well within package 3, it would also be possible to use the other elements of that package without any gates. Similarly, while we feel that it would be most justified, appropriate, and pragmatic to use common ODIs in package 2, it would theoretically be possible to use company specific ODIs with innovation uplifts.

Assessment of options

We assessed these packages against Ofwat's assessment framework. Our assessment is shown in the figure below.

Figure 4 Summary of the options assessment

	Package 1 – Focus on PR14 improvement	Package 2 – Focus on innovation	Package 3 – Focus on tailored rewards	Package 4 – Focus on integrated assessment
Protect customer interest				
Promote efficiency				
Maintain resilience	n/a	n/a	n/a	n/a
Protect the environment	n/a	n/a	n/a	n/a
Ensure affordability	n/a	n/a	n/a	n/a
Ensure financial viability	n/a	n/a	n/a	n/a
Pro-market approach	n/a	n/a	n/a	n/a
Better regulation principles:				
1) Proportionate and targeted				
2) Broader range of regulatory tools				
3) Flexibility and responsiveness				
4) Transparency and predictability				
Focus on customer impacts:				
1) Impact on customer bills				
2) Impact on service performance				
Focus on efficiency:				
1) Effective incentives				
2) Ownership and accountability				
3) Fostering innovation				
What resources are required to implement it:				
1) For Ofwat				
2) For companies				
How much will it cost	n/a	n/a	n/a	n/a
How quickly will we see change				

We concluded that package 1 would not sufficiently address the lessons learned from PR14, while package 4 would be difficult to implement at PR19. Our assessment therefore concluded that, out of the four packages we defined, package 2 or package 3 would be most suitable for PR19.

The choice between these two packages – one focused on innovation and the other focused on tailored rewards – depends on Ofwat’s policy direction for PR19. We identify the main considerations for Ofwat in choosing between these two packages in the figure below.

Figure 5 Key choice between packages 2 and 3

Package 2 (focus on innovation) is more appropriate when:	Package 3 (focus on tailored rewards) is more appropriate when:
<ul style="list-style-type: none"> ■ The objective is to shift the efficient frontier through innovation 	<ul style="list-style-type: none"> ■ The objective is to improve performance across the board
<ul style="list-style-type: none"> ■ There are concerns about the ability to test the level of stretch in commitment levels, but relative comfort in assessing the level at which companies would receive innovative rewards 	<ul style="list-style-type: none"> ■ There is comfort in being able to test the level of stretch in PC targets
<ul style="list-style-type: none"> ■ It is felt the level of PR14 rewards are appropriate in general, but would not incentivise genuine innovation / reward for the risks involved in innovation 	<ul style="list-style-type: none"> ■ It is felt that the level of rewards at PR14 are not sufficient to incentivise gradual improvement beyond PC levels
<ul style="list-style-type: none"> ■ Increasing the ODI proportion of the overall return is not a priority so there is no need for an approach that is consistent with that 	<ul style="list-style-type: none"> ■ Ofwat would like to increase the ODI proportion of the overall return
<ul style="list-style-type: none"> ■ Ofwat is comfortable with multiple companies delivering some improvement but only a few companies delivering innovative levels of performance 	<ul style="list-style-type: none"> ■ Ofwat is comfortable with multiple companies delivering gradual improvement beyond stretching PC levels, but not necessarily innovative levels of performance
<ul style="list-style-type: none"> ■ Ofwat is comfortable with some companies receiving material rewards for innovative performance 	<ul style="list-style-type: none"> ■ Ofwat is comfortable with multiple companies receiving rewards greater than at PR14 for gradual improvement beyond stretching PC levels
<ul style="list-style-type: none"> ■ Ofwat considers companies are likely to respond to innovation rewards 	<ul style="list-style-type: none"> ■ Ofwat considers firms likely to respond to more gradual rewards

Source: *Frontier Economics*

Finally, we note that Ofwat could choose to implement one of packages 2 or 3, with some modifications. This is because although we have defined coherent packages, there are still options within the packages and changes could be made to one or more of the detailed aspects of the packages, without losing the policy focus of the package. For example, Ofwat could choose to implement package 2 with company-specific ODIs, or it could choose to implement package 3 without

gated ODIs. In addition, the package design also depends on the way it is implemented so the detailed approach may require some modifications. For example, if package 3 was implemented with relatively small uplifts, this could reduce the need for gated ODIs.

1 INTRODUCTION

1.1 Introduction

At PR14, Ofwat moved to an outcomes-based framework for regulating water companies in England and Wales. Based on extensive engagement with customers and customer challenge groups (CCGs), water companies proposed their own outcomes², measures³, performance commitments (PCs)⁴ and outcome delivery incentives (ODIs⁵). While outcome-based regulation has been introduced in a number of sectors, Ofwat's approach took the concept a step further and PR14 is now viewed as a truly innovative approach that other regulators are trying to learn from.⁶

The overall approach is viewed as a success throughout the industry as companies were more customer-focused at PR14 compared to previous price controls and PR14 business plans reflected customers' views to a greater extent. Nevertheless, there is always room for improvement. For PR19, Ofwat is challenging itself to evolve the outcomes-based approach to address some of the issues that were raised at PR14. Ofwat has already undertaken significant work and has signalled in its Water 2020 publications that companies will be expected to use innovative ways of engaging with customers, and to make greater use of comparative information. In November 2016, Ofwat published its consultation on the PR19 outcomes framework, "the outcomes consultation", and Ofwat plans to publish its PR19 methodology in July 2017.⁷

In the outcomes consultation, Ofwat has indicated that it is considering the following:

- using ten common PCs (four from the five comparatively assessed PCs at PR14 plus six others);
- applying stretching common commitment levels to six of the common PCs⁸;
- developing guidance that should encourage companies to set stretching bespoke PCs;
- developing options for how ODIs can be set, including how to make use of a wider range of information on customers' preferences (including designing the detailed elements of ODIs, such that they become more powerful, including whether deadbands should be used and whether gated ODIs could be used);

² High level areas of service / impacts that customers value

³ Metrics that value the outcomes

⁴ To mirror other industry publications, we use the term PC broadly, to mean both the measure and also the commitment level.

⁵ The financial incentive rates (both rewards and penalties) that will be applied to the PC if the company performs above or below that target

⁶ For example, the CAA has reviewed the outcomes-based approach extensively and is considering how to implement a similar approach for Heathrow Airport. Similarly, the Essential Service Commission in Victoria, Australia has considered the approach.

⁷ Ofwat (2016), A consultation on the outcomes framework for PR19

⁸ In the remainder of this report, we use the word "stretching" in the same way as it is used in the outcomes consultation.

- whether industry standard ODIs could be applied to common PCs with common commitment levels, with more powerful rewards and penalties; and
- the implications for ODIs of a variable cost of equity approach, where the allowed financial return partly depends on a company's commitments.

For clarity, common PCs are those where Ofwat requires companies to monitor and report their performance against the metric.⁹ It is likely that for a subset of these common PCs (e.g. six), Ofwat will set common commitment levels for all companies. Companies will be able to set their own commitment levels for the other common PCs (e.g. four). In contrast, bespoke PCs are those where only a subset (or sometimes only one company) may record performance against the measure. If more than one company records performance on a given bespoke PC, there would be some comparative information available, but not a full data set.

1.2 Project objective

The objective of this project is to support Ofwat's development of the PR19 methodology, by providing options for the way that:

- PCs and commitment levels are set by companies (for bespoke PCs and common PCs where Ofwat does not set common commitment levels);
- ODIs are set for both common and bespoke PCs.

The way that Ofwat might set common commitment levels for some common PCs is outside the scope of this project.

The aim of this project is to help Ofwat develop the methodology that it should apply at PR19 and the expectations that it should set for companies. The output from this project will enable Ofwat to make a confident choice regarding its methodology for PR19.

1.3 Report outline

This report is structured as follows:

- Section 2 discusses the key requirements for PR19 including lessons learned from PR14, the outcomes consultation and the wider policy context;
- Section 3 identifies general improvements that should be made to the PR19 methodology regardless of the detailed methodology choices discussed in section 4;
- Section 4 provides the key options for the PR19 methodology;
- Section 5 provides an overview of packages; and
- Section 6 provides an assessment of those options.

Annex A discusses how data envelopment analysis can be applied. Annex B provides a detailed description of the measure-specific approach. Annex C provides the detailed assessment of the key options.

⁹ Ofwat (2016), A consultation on the outcomes framework for PR19, p. 14

2 KEY REQUIREMENTS FOR PR19 FRAMEWORK

Before developing options and recommendations for PR19, it is important to set out clearly the key requirements of the PR19 outcomes framework. In this section, we consider the:

- lessons learned from PR14;
- objectives from Ofwat's outcomes consultation; and
- wider policy context at PR19.

2.1 Lessons learned from PR14

At PR14 Ofwat introduced a new form of outcomes-based regulation, which involved companies setting their own outcomes, measures, PCs and ODIs. The introduction of this framework was seen to be innovative and the impact on the industry has generally been positive. At the same time, it is clear that there are important lessons that can be learned from PR14 both in terms of the implementation of the framework and the methodology. To inform our assessment of the lessons learned from PR14, we reviewed the following material:

- UKWIR project¹⁰ on PCs and ODIs and Ofwat's report¹¹ on the learning from PR14);
- the data tables and commentaries specific to PCs and ODIs that companies were required to submit as part of their PR14 business plans;¹²
- companies' actual performance on outcomes in 2015/16, for example the material on the Discover Water dashboard¹³ and also in Ofwat's PCs spreadsheet; and
- Ofwat's draft determinations on the in-period ODIs that some companies have in 2015-20.¹⁴

This section summarises the lessons learned in three categories:

- lessons learned with regard to the PR14 framework;
- lessons learned regarding the implementation of the PR14 framework; and
- emerging lessons learned during AMP6.

¹⁰ UKWIR (2016), Setting performance commitments and incentives to deliver best value for money, (UKWIR Report Ref No 16/RG/07/39)

¹¹ Ofwat (2015), Reflections on the price review – learning from PR14

¹² Ofwat shared this information with us on a confidential basis, and only in cases where companies had given approval to Ofwat to share this information with us.

¹³ <http://www.discoverwater.co.uk/>

¹⁴ Three companies (Anglian Water, Severn Trent Water, and South West Water) have in-period ODIs, meaning that they receive the annual financial impact of some ODIs each year, as opposed to receiving the full impact of ODIs at the end of the AMP. Ofwat published its draft determinations for these three companies' in-period ODIs in November 2016.

2.1.1 Lessons learned with regard to the PR14 framework

These lessons relate to the way the framework was set up and the impacts that it achieved. The PR14 framework was based on a bottom-up methodology. It required companies to collect information on marginal costs and on customer valuations for service improvements. Companies were then required to use this detailed data to apply cost-benefit analysis (CBA) to identify optimal targets for PCs and also to base ODIs on costs and customer valuations.

The table below summarises the advantages and disadvantages of the PR14 framework. Below the table we provide more detail on the key advantages and disadvantages.

Figure 6 Advantages and disadvantages of PR14 outcomes framework

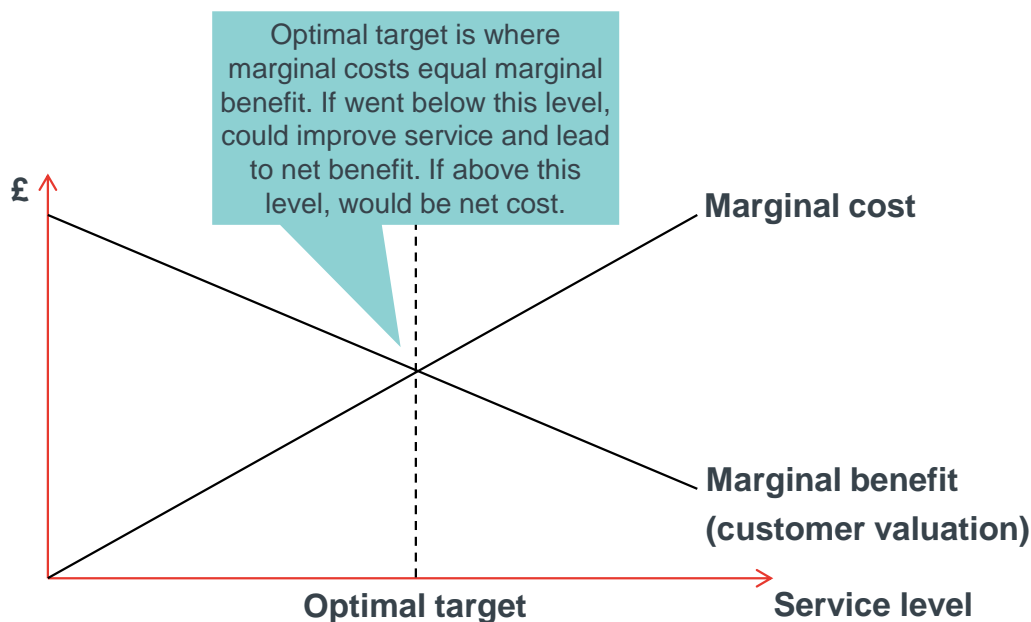
Advantages	Disadvantages
Approach was underpinned by economic theory	The detailed approach to setting PC levels might not have taken sufficient account of affordability concerns
Targets and incentive rates were based on data, rather than subjective judgements	Reliant on single values for costs and customer valuations
Stated preference WTP is a pragmatic approach to evaluating the marginal value of service performance	Potential inaccuracies in stated preference WTP values
Marginal costs for some areas, e.g. large capex projects, were clear cut	Allocating common costs was challenging in some areas
Encouraged companies to focus more on customers	Did not take sufficient account of uncertainty (in data, and what might happen in the future)
Companies owned plans, and they reflected customer preferences and specific circumstances of the company	Focused on individual PCs, and not on the outcomes package as a whole
Balance in favour of penalties, which was in-line with customer preferences reported at PR14	Incentives for companies to set stretching PC levels limited

Source: *Frontier Economics*

Note: *WTP=Willingness to pay*

Key advantages of the PR14 framework

One of the key advantages of the PR14 framework is that it was grounded in sound economic principles, as the bottom-up approach, in theory, should have resulted in efficient outcomes. The approach that companies were encouraged to follow is illustrated in the following diagram.

Figure 7 Principle of CBA

Source: *Frontier Economics*

As illustrated in the diagram above, companies were expected to set targets where the marginal cost of service improvement was equal to the marginal benefit (customer valuation). Setting targets in this way should have meant that companies prioritised their spending according to customer preferences, and therefore used resources in the most efficient way (i.e. achieving allocative efficiency).

In addition, as the framework was company-led, companies were in control of developing their own business plans, and PC levels. This meant the overall framework was less interventionist than other regulatory frameworks, where targets are often set based on regulatory judgement. The nature of the framework led to a culture shift in the way that companies developed their business plans, in particular in how they led customer engagement, and used the insights from that engagement to inform their business planning.

Key disadvantages of the PR14 framework

Although the PR14 framework did allow for customer, CCG and Ofwat challenge of companies' proposals and Ofwat could intervene where appropriate (and did so in several instances), there was no explicit comparison of marginal costs or any direct efficiency assessment of marginal costs. The scale of the costs may therefore not have been efficient (i.e. productively efficient). There was also no significant incentive for companies to reduce these marginal costs over time (i.e. the framework may not deliver dynamic efficiency).

Also, in practice the allocative efficiency of the PC levels heavily depended on the availability and quality of the underlying data. In practice, there are a number of challenges to overcome when developing inputs for the bottom-up approach, because the economic assumption of perfect information and simple cost structures does not hold in reality.

- There are a substantial number of **common costs**, as a result of service often being produced jointly, and it was not straightforward to allocate these to different PCs in an appropriate way. Often the allocation of common costs was implicit, as companies allocated costs to the most important cost driver. For example, attempts to reduce leakage may have increased pipe replacement rates, and all costs associated with these extra pipe replacements might have been allocated to leakage, even though these additional costs may also have impacted on the level of supply interruptions. The issue with these implicit allocations was that it was not clear that companies followed a consistent approach, and therefore whether their outcomes targets were comparable.
- Some companies found it **challenging to estimate the costs** they would save if their service performance deteriorated, as this is not commonly estimated. The best example for this is the cost reduction if companies lowered their mean zonal compliance performance. It is not clear that companies can estimate this cost in a meaningful way. In addition, there was also uncertainty around how much it would cost to deliver service improvements, particularly more challenging improvements and on some measures that had not been used for long. As a result, there was some uncertainty around the scale of the costs involved, as the companies did not have perfect information.
- As companies had to conduct customer research early in the PR14 process, they generally **relied on a single WTP survey**. This meant the results would be expected to be generally less reliable than were they to have used multiple research methods and tested the sensitivity of different results. Also, the outcomes and scale of improvement that were assessed in these early surveys were not always aligned with what emerged later as being the appropriate PCs. This meant that conversions, using inter- and extrapolations, were required to obtain customer valuations for the targets that were ultimately included in the business plan.
- In addition, as already noted by Ofwat, there are a number of **limitations with WTP surveys**. Ofwat has already encouraged companies to consider a wider range of evidence at PR19¹⁵. We do not cover these limitations with WTP surveys in depth here, as this topic has been discussed in many reports previously, and we assume that the content is familiar.

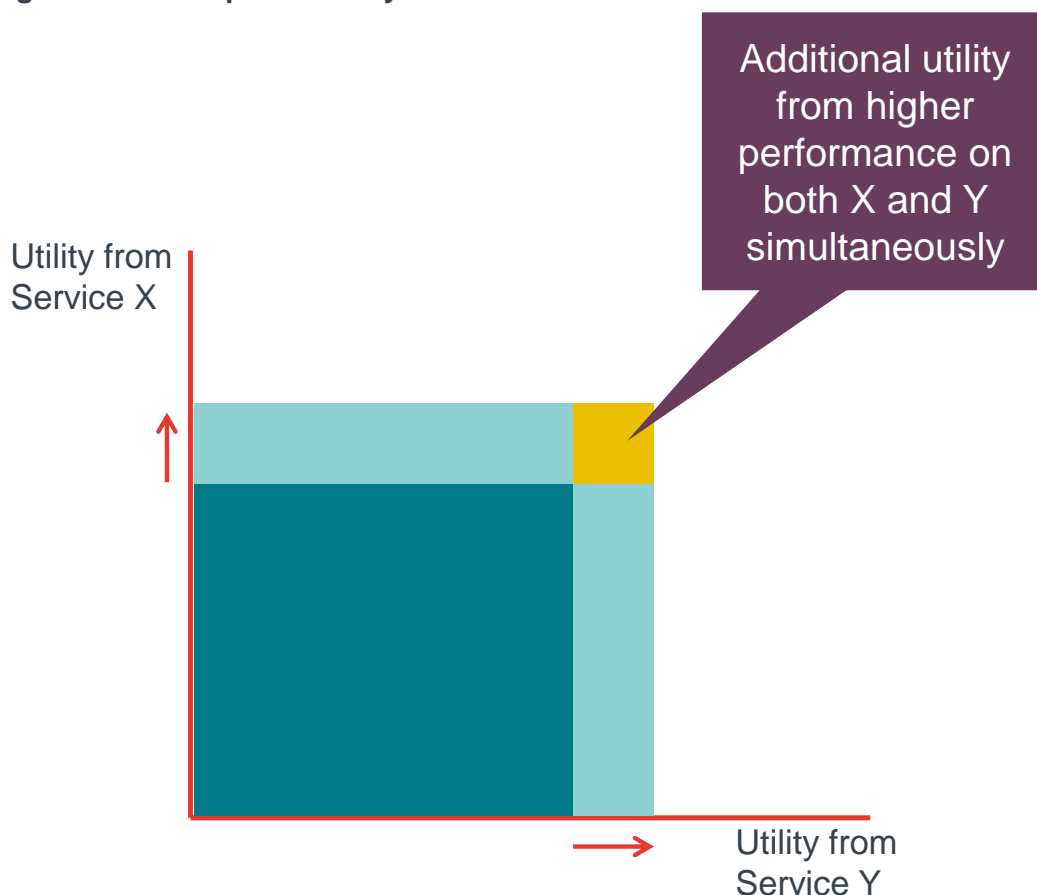
As a result, if one or both of cost data and customer valuations were poor quality, then the outcome of the CBA would not have been accurate, and therefore may not have resulted in the efficient targets that the framework was trying to achieve.

Due to these data limitations, in practice Ofwat and several other industry stakeholders recognised that the original PR14 framework might not have achieved the impact that it was intended to, which led to amendments such as the introduction of common PCs with common PC levels to ensure the PC levels were sufficiently stretching for the PCs customers thought were most important. The PR19 framework also needs to reflect these practical issues and provide a more flexible approach for companies and Ofwat from the start.

¹⁵ See page 14 of http://www.ofwat.gov.uk/wp-content/uploads/2015/12/pap_pos20160525w2020cust.pdf

In addition to data quality issues, the framework also did not reflect two important aspects of customers' preferences. First, elements of the service that water companies provide to customers may be complementary. This means that the value that customers receive from one service element positively depends on the water company's performance on another service element. This concept is illustrated in the diagram below.

Figure 8 Complementarity of service elements



Source: Frontier Economics

Second, customers' overall WTP for service improvements (or their overall acceptability of bill increases) was often lower than the sum of the individual WTP figures that were found in customer surveys. This may reflect the survey design as customers may express higher valuations when asked about specific measures rather than the full set of service elements. If customers are only considering a subset of measures, this could lead to an "overestimate" for the individual WTP, as they may be implicitly assuming they would not be paying for service elements that are not in that part of the survey.

When faced with this issue, companies considered whether it was appropriate to use the WTP values on individual service elements, as that would have over-estimated customers' overall WTP. Some companies chose to scale down the individual WTP figures to reflect the overall WTP for service improvements. However, as the PR14 framework was generally focussed on individual measures, there was only limited information on what companies should do in

this situation. Similarly, the framework did not explicitly require companies to reflect wider affordability concerns, which we consider below.

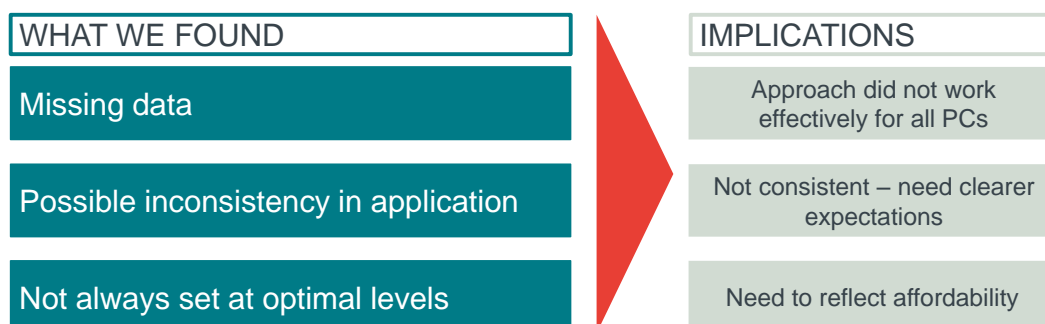
Ideally the PR19 framework would reflect these potential features of customer preferences more explicitly.

2.1.2 Lessons learned on implementing the framework

In addition to the lessons learned from PR14 on the framework, there are also important lessons regarding the implementation of the framework. These points relate to the requirements that were imposed on companies, and the practical difficulties that the companies faced in developing their business plans and setting PCs. It also reflects the regulatory burden that Ofwat faced when reviewing the material that was submitted by companies as part of the framework.

The diagram below summarises our findings from reviewing companies' data tables, and the implications of these issue on the effectiveness of the framework. We provide further detail on these points below the diagram.

Figure 9 Findings from reviewing companies' data tables



Source: Frontier Economics

We found the following issues. These are all manifestations of the lessons learned regarding the framework that were identified earlier.

- Missing data.** We found that in some cases companies were unable to provide any data for certain PCs, while for other PCs companies were only able to provide cost and customer valuation data for one level of service quality.

This implies that the effectiveness of the framework in identifying the efficient target for some PCs is likely to have been low. This may suggest that the framework did not work effectively for all PCs, and that in developing the PR19 framework, consideration should be given to what is the appropriate approach for different types of PCs.

- Inconsistency in application.** We observed that in some cases companies had recorded the marginal cost and customer valuation for the target as zero. This could be because companies had set the target at the current level, and had defined marginal costs and marginal customer valuations as being relative to the current service level, such that these values were zero at the target level of service. The reported zero values could also point to an inconsistency in the way that the underlying data was collected, or a

misunderstanding that these data points should be automatically set to zero at the target level.

Overall, we conclude that this implies the PR14 framework was not always consistently applied or easily understood by companies. The PR19 framework should look to address this, so it needs to be more accessible and easy to implement.

- **Targets not always set at optimal levels (as implied by data).** The data tables suggest that targets were not always set at optimal levels. We observed instances where higher levels of service quality (than the target) appeared to deliver more customer benefits, based on the information that companies provided. There were also examples where targets appeared to have been set at levels where marginal costs exceeded marginal benefits.

These cases, where the target was not set at the optimal level according to the underlying data, could be due to a number of reasons including affordability concerns or statutory obligations.

We conclude from this that the PR19 framework should more explicitly consider affordability constraints, and could set expectations on how companies should deal with those constraints. We also consider that the PR19 framework should explicitly recognise when targets are likely to be highly influenced by factors other than CBA, such as statutory obligations.

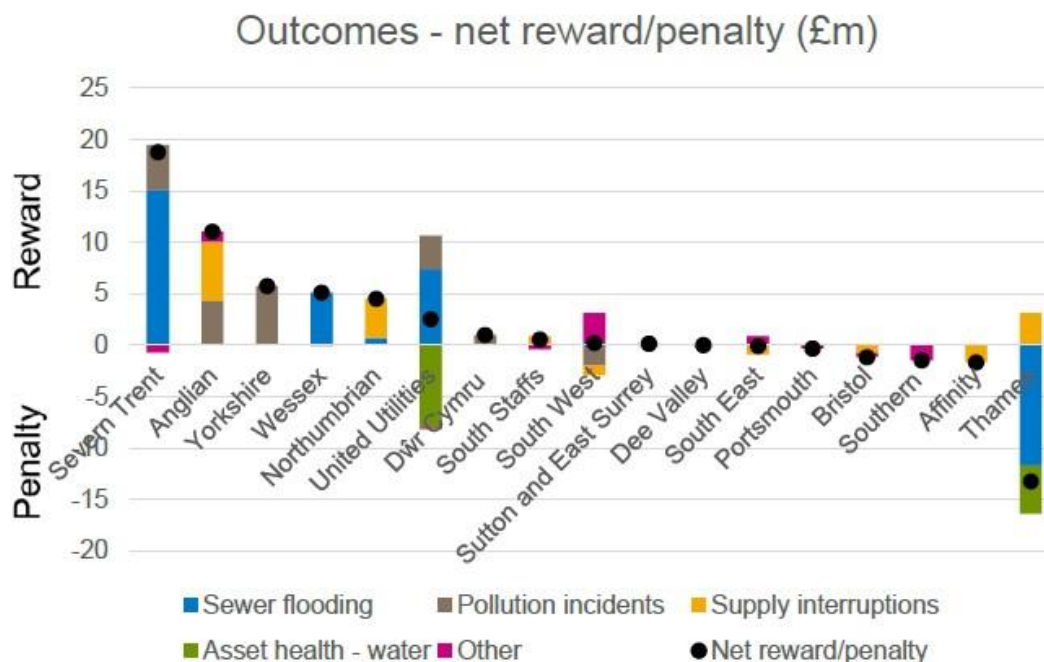
2.1.3 Emerging lessons learned during AMP6

As there is one year of available data on how companies have performed against their targets, we can see if there are any emerging lessons on how the PR14 framework appears to have worked in practice. Before doing this, we note that the return on regulatory equity (RoRE) ranges for companies at PR14 were asymmetric with allowance for larger penalties than rewards, even though Ofwat's January 2014 Risk and Reward guidance suggested a symmetric RoRE range. Despite this asymmetric RoRE range, conceptually we might not necessarily have expected larger penalties than rewards for the following two reasons.

- Companies have an incentive to respond to the ODIs, and to improve their outcomes performance.
- There is information asymmetry between Ofwat and companies. As the PR14 framework allowed companies to set their own PC levels on bespoke PCs, it is possible that companies set PC levels on bespoke PCs such that they were confident that they would achieve them (this was not the case for common PCs where the PC levels were set at the sector historical upper quartile level).
- Ofwat also introduced the totex (total expenditure) approach so companies have greater flexibility in trading-off capex and opex which may influence ODI performance.

Taking this conceptual view into account, we looked at the performance data from the first year of AMP6. The diagram below summarises companies' 2015/16 performance on ODIs.

Figure 10 Companies' 2015/16 performance on ODIs



Source: Ofwat

Overall, this figure shows a relatively mixed picture of performance, but on balance, there appear to be more rewards than penalties. As with any regulatory incentives, it is difficult to tell what is driving the rewards after they have been achieved, as it could be due to: companies improving their performance in response to ODIs; or companies setting commitment levels in a way that allows them to relatively easily achieve them. Overall, the ODI approach appears to be working as customers appear to be receiving higher levels of performance, and this evidence can be used to set more challenging targets next time around, which will bring further benefits to customers. It is also worth being aware that the PC levels for the common PCs become more stretching by 2017-18 when they reflect historical upper quartile performance.

We also note that companies' actual performance has generally been presented at an overall level as well as by PCs. This may be because there is a view that overall performance is important in and of itself, or it could be because it is simpler to summarise comparative overall performance. The PR14 framework did not explicitly take account of overall performance / performance across measures, and it may be beneficial if the PR19 framework more explicitly reflects the importance of overall performance.

2.2 Outcomes consultation

The options for the PR19 framework need to be consistent with the direction Ofwat has outlined in the outcomes consultation. The outcomes consultation had four key themes:

- making performance commitments more stretching;
- more powerful outcome delivery incentives;

- better reflecting resilience in outcomes; and
- making performance commitments more transparent to customers.

Ofwat is keen to have more stretching performance commitments at PR19, as that should deliver higher customer service to current and future customers, and better protect the environment. However, Ofwat recognised that it would be important to consider how this can best be achieved without simply resulting in higher bills. It suggested that a higher proportion of a company's overall return could be linked to ODIs, and also that powerful ODIs should only be applied where a company delivers outperformance against genuinely stretching PC levels. Ofwat also noted that the reputational impact of outcomes performance could be enhanced by making PCs more transparent to customers and other stakeholders.

2.3 Wider policy context

In addition to the lessons from PR14 and the objectives in the outcomes consultation, the wider policy context for PR19 needs to be reflected in the PC and ODI framework. Ofwat has four policy key themes for PR19 that should be recognised when the outcomes framework is developed.

- **Customer service** - the first of Ofwat's key themes for PR19 is customer service. As part of its work on retail market opening, Ofwat found that customer service in the water sector is generally considered to be behind that in other sectors. It is Ofwat's focus on customer service and on improving outcomes for customers that explains the direction taken in the outcomes consultation: a need for more stretching PC levels and more powerful ODIs to incentivise companies to deliver strong service performance.
- **Resilience** - the Water Act 2014 added a new duty to Ofwat's set of primary responsibilities¹⁶: to further the resilience objective in England and Wales. Ofwat's resilience objective is to:
 - "to secure the long-term resilience of water undertakers' supply systems and sewerage undertakers' sewerage systems as regards environmental pressures, population growth and changes in consumer behaviour, and
 - to secure that undertakers take steps for the purpose of enabling them to meet, in the long term, the need for the supply of water and the provision of sewerage services to consumers."¹⁷

As a result, the PR19 framework will need to incorporate resilience considerations more explicitly. The outcomes consultation suggested that one of the common PCs at PR19 may be a resilience measure. The wider PR19 framework will therefore need to factor in this new PC, and also other resilience concerns, such as the need for the definitions of the other common PCs to reflect resilience.

- **Affordability** - The affordability of water bills is a concern for Ofwat and the Welsh and UK Governments.¹⁸ Although there is no agreed definition of

¹⁶ Water Act 2014, Chapter 3, Section 22

¹⁷ Water Act 2014, Chapter 3, Section 22

affordability, it has often been measured in terms of the proportion of households that spend more than 3% of their income on water and sewerage bills. Ofwat found in 2014/15 that 11% of households in England, and 15% in Wales, were spending more than 5% of their income on water and sewerage bills, and 23% of households in England, and 32% in Wales, were spending more than 3% of their income on water and sewerage bills.¹⁹

The UK and Welsh Governments support schemes such as WaterSure / Welsh Water Direct and social tariffs. However, the Governments also recognise that water bills are dependent on the efficiency of the industry. It is therefore important that the outcomes framework reflects affordability concerns, in the way that targets are set but also in the scale of ODIs (particularly rewards).

Within affordability there is a specific question around how distributional concerns should be addressed. Companies' investment decisions are informed by the results of CBA, which in turn depend on customer valuations. Typically, the data used on customer valuations reflects the *average* customer valuation. By definition, some customers are willing to pay more for service improvements than this average, while others are only prepared to pay less than the average. As companies are limited in their ability to tailor service levels to specific households, and in any case charging is based on regional averaging, this means that some customers will receive improved service levels even though they would not be prepared to pay for the improvements, while others will receive improvements in service levels that they are more than willing to pay for. It is therefore important to understand the impacts of average bills in the context of service improvements.

Finally, we note that although the charging principles mean that customers pay average bills, there is at least one element of the regulatory framework that means customers are compensated for the actual performance they receive. The Guaranteed Standards of Service (GSS) mean that customers who suffer from particularly poor performance receive some financial compensation.

- **Innovation** - We understand that Ofwat is reflecting on how innovation could be incorporated into the outcomes framework. One of the lessons learned from the PR14 experience is that rewards may not be sufficient to deliver genuine innovation in service delivery. We understand that Ofwat is considering how it could amend the framework so that companies are incentivised to deliver a frontier shift in service performance and more innovation is delivered in future. Similarly, we understand that Ofwat is keen to provide incentives to develop new, innovative measures. These new measures could result in more stretching outcomes packages for some companies at PR19, and if they provide useful insight at PR19, they could become more central to all companies' outcomes packages in future, which would deliver service improvements for a wider customer base further in the future.

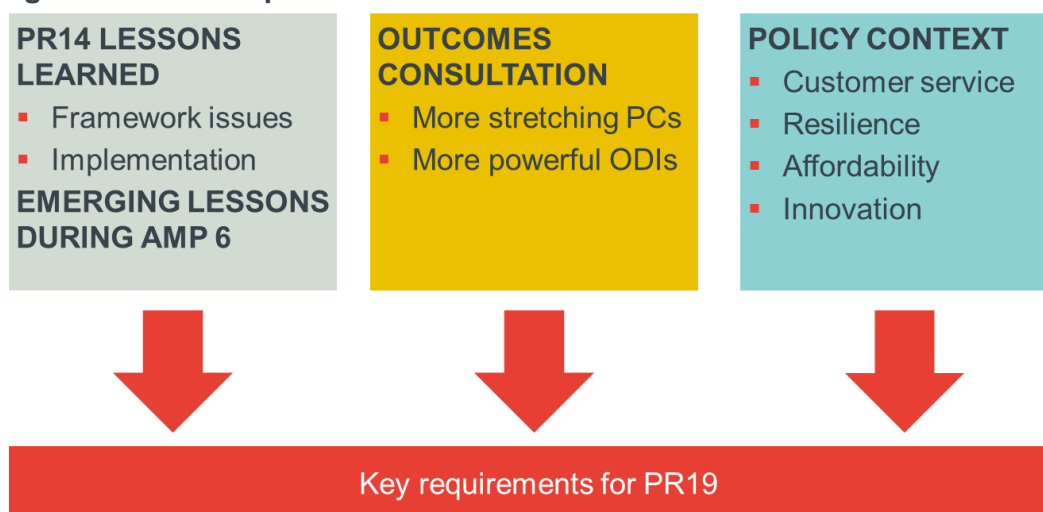
¹⁸ House of Commons briefing paper (8 August 2016), Water bills – affordability and support for household customers

¹⁹ Ofwat (2015), Affordability and debt 2014-15, p. 4

2.4 Summary

The PR19 outcomes framework should reflect the lessons learned from PR14, the objectives of the outcomes consultation, and the wider policy context. The diagram below summarises how these points should be reflected in the development of the PR19 framework.

Figure 11 Development of the PR19 framework



Source: *Frontier Economics*

In the remainder of this report we set out our recommendations for the PR19 outcomes framework, and the options that could be used to design the detailed methodology at PR19. Where relevant, we highlight how the recommendations and options should deliver improvements on issues considered in this section.

3 RECOMMENDATIONS: GENERAL IMPROVEMENTS

We have identified a number of changes that should be included in the PR19 outcomes framework, regardless of what other choices are made. These changes would deliver clear improvements to the way that the framework is implemented and to the customer impacts that should result from it. We also note that these recommendations are in line with Ofwat's customer engagement policy statement and expectations for PR19.

We have identified three areas for general improvements:

- improvements to CBA method;
- increasing the effectiveness of Ofwat's role; and
- increasing the effectiveness of CCG challenge.

In the remainder of this section we provide further detail on our recommendations. Within those recommendations, we make the distinction between guidance from Ofwat and expectations that Ofwat may set. This is an important distinction and so we provide a short explanation of the difference between the two below.

- Guidance issued by Ofwat would effectively be a requirement for companies to follow a specific approach. For example, Ofwat could set guidance on how it wants companies to calculate incentive rates. While guidance would make Ofwat's requirements clearer for companies to interpret, it would reduce the scope for companies to be innovative in the approaches that they use at PR19.
- Ofwat could alternatively set expectations for companies, which would mean that Ofwat would indicate what it considers to be of high quality, but Ofwat would be explicitly recognising that its expectation is not the only way of doing something. Setting expectations should provide companies with clear insights on what they are expected to do, but would also allow room for innovation in approaches.

3.1 Improvements to CBA method

The fundamental principle behind using CBA to set outcome targets is that companies' resources should be used in the most efficient way to deliver what customers want. This principle is a good one to maintain at PR19, however improvements should be made to the way that CBA is done, wherever it is relied on to set outcome targets.

In the remainder of this section, we provide detail on our recommendations to improve CBA at PR19 in the following areas:

- multiple data sources on customer valuation and triangulation approaches;
- improving the quality of cost data;
- affordability; and

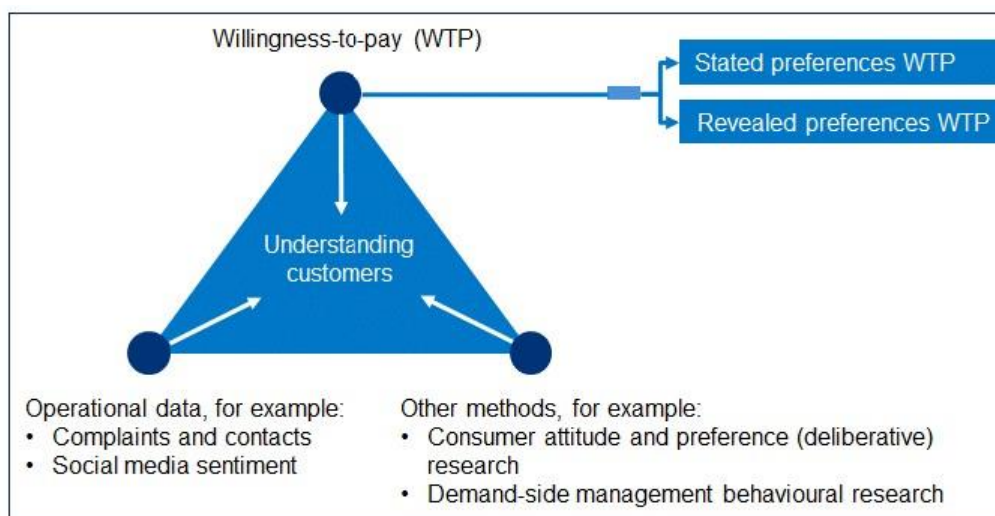
- distributional concerns.

3.1.1 Multiple data sources on customer valuation and triangulation approaches

As identified in Section 2, CBA relies heavily on the quality of the underlying data in order to be effective. Data on customer valuations is one of the key inputs to the CBA that is used to set outcome targets, and it is therefore important that the data is as robust as possible. One way to improve the quality of the customer valuation data is to use multiple data sources, rather than relying exclusively on a single measure, where possible and proportionate.

Ofwat has already stated that it is keen for companies to use multiple data sources for customer valuation at PR19. In its May 2016 customer engagement policy statement, Ofwat said “we encourage companies to explore alternative and complementary tools to validate and test results from stated preference WTP surveys”.²⁰ The diagram below illustrates the range of different sources that companies could use to collect insights on customer valuations, and how multiple methods could improve companies’ understanding of customers.

Figure 12 Use of multiple customer valuation methods



Source: Ofwat (2016), Customer engagement policy statement and expectations for PR19, p. 15

The diagram suggests that Ofwat is expecting companies to “triangulate” on a customer valuation figure, using a range of research methods. For example, companies could collect stated preference WTP data and combine this with operational data, and data from behavioural research. Companies could also use data from revealed preference studies and benefits transfer analysis.

Ofwat has stated that it expects companies to use a proportionate approach when determining which methods to use for which PCs, and how many methods to rely on when estimating customer valuations. This is because there is a trade-off to using more research methods to estimate customer valuations. While the use of more research methods may result in more robust information on

²⁰ Ofwat (2016), Customer engagement policy statement and expectations for PR19, p. 14

customer valuations, it is likely to lead to higher costs in carrying out customer engagement as part of the business planning process. For this reason, when companies are choosing which customer engagement methods to use to estimate customer valuations, they should identify:

- PCs that are associated with high cost levels; and
- PCs for which they are planning to deliver a significant improvement in service levels.

Companies should assign more time and resources to these PCs, as it is most important that the customer valuations are robust for these PCs. While the quality of data that is used to set targets on other PCs is still important, any uncertainty around the data in those cases is less likely to have a material impact on the business plan. If companies follow this approach, they should rely on multiple research methods to estimate customer valuations, but in a proportionate way.

In addition, it may be the case that certain research methods are best suited to particular PCs. For example, it may be more pragmatic to carry out behavioural trials for some PCs, and more difficult for others. Companies should also bear this in mind when developing their customer engagement strategies.

Once companies have developed their customer engagement strategies and have collected information on customer valuations, there is then a question as to how data on customer valuations are combined to use in CBA. Ofwat could set clear guidance on how companies should combine these values, or it could expect companies to do it, and allow them to find the best way of combining the values. The second approach should lead to more innovation in the way that companies combine customer valuations. We have identified three potential ways that companies could combine multiple data sources. If Ofwat chooses to set guidance on this, it could require companies to use one of these three approaches, or alternatively it could just be aware that these are some of the options available to companies and be aware of this when reviewing plans.

1. **A mechanistic rule:** this approach would consist of a rule for selecting between the different values or combining them into a final value. The rule would be defined before the data was collected and wouldn't change depending on the data obtained. There are a number of rules that could be used, for example:
 - a straight average of the values; or
 - a weighted average of all of the values (where the weights were defined in advance).

A straight average would be transparent and pragmatic to implement, however it may not make the best use of available information. For example, there may be reasons why one research method is expected to be more robust/reliable than another one, and in those cases, it may be more appropriate to apply a higher weighting to the method that is expected to be more robust. Setting a pre-defined rule however requires some form of ex- ante judgement on how the values should be combined, which may require further thought and justification from the companies. It

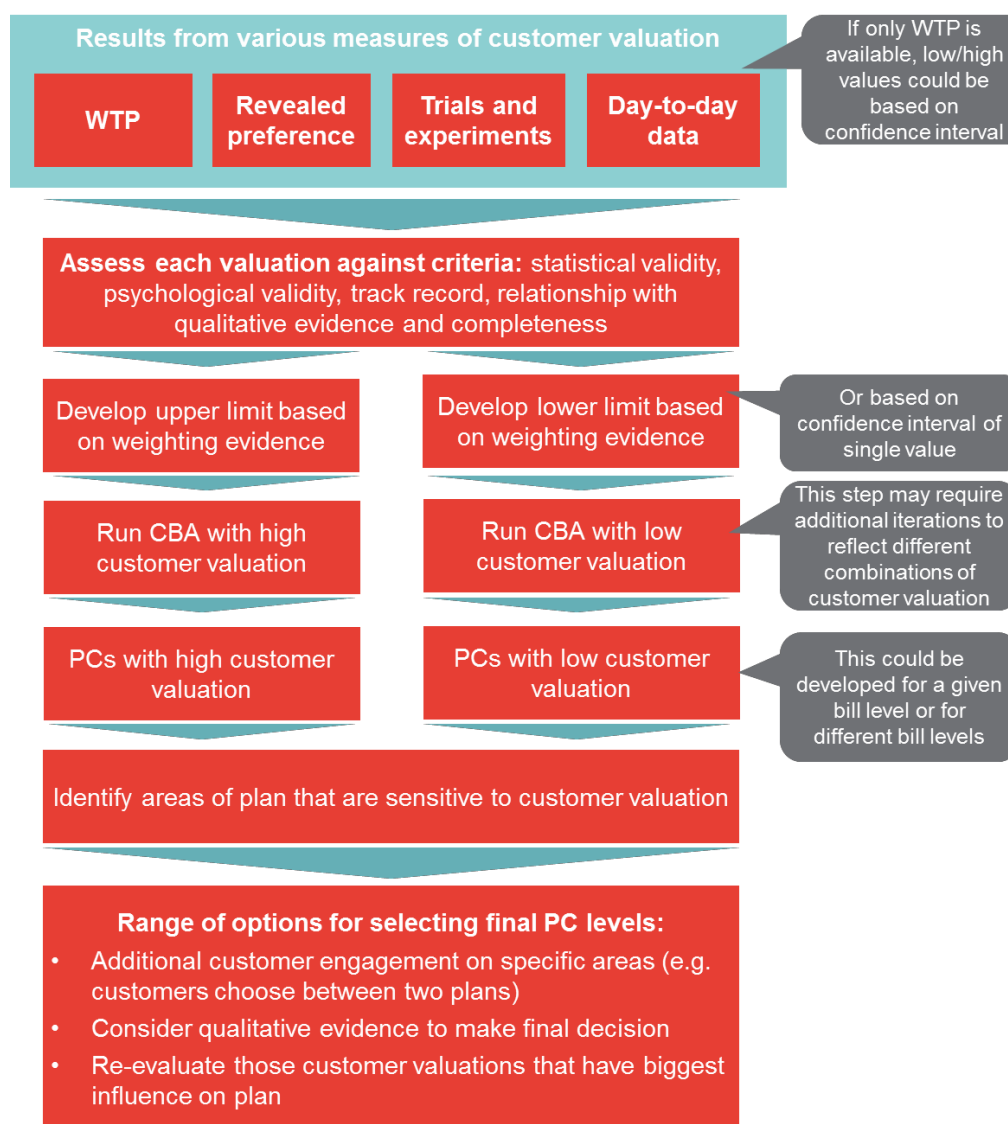
is also likely to require a more in-depth assessment from Ofwat to judge how companies have combined data.

2. **Systematic judgement:** this approach would be based on a reasoned judgement, informed by a system that is pre-defined, at least to some extent. For example, it may take account of the fact that revealed preference does not reflect the full value that a service improvement brings to customers as, it does not capture the “inconvenience” of an interruption. This may imply that revealed preferences should consistently be used as a lower bound. This reasoned judgement could be tested with and reviewed by the company’s CCG.
3. **Multi-input CBA:** this approach would test the sensitivity of the overall CBA to customer valuations. As this approach is more complex, and there are multiple options within it, we provide more detail on this below.

Multi-input CBA

The rationale behind multi-input CBA is that the question of which customer value should be used needs to be informed by how sensitive the CBA results are to different valuations. The figure below summarises how this could work in practice.

Figure 13 Multi-input CBA



Source: Frontier Economics

The process could include the following steps.

- In the first step, companies could use a range of valuation sources to develop a lower and upper bound estimate of customers' valuation. We expect that for some PCs stated preference WTP may be the only source available, but even if that is the case, all WTP estimates should have confidence intervals that can be used to define an upper and lower bound. For those PCs where a set of values are available from a range of data sources, we would expect companies to use a number of criteria to develop an upper and lower bound (e.g. companies could use the criteria set out in the UKWIR report²¹).
- This should result in a range of PC levels. Companies could test this for each PC separately, or simultaneously. The detailed choices could be left to companies, but the overall objective would be to identify those parts of the

²¹ UKWIR (2016), Setting performance commitments and incentives to deliver best value for money, (UKWIR Report Ref No 16/RG/07/39), Table 9

plan that are sensitive to customer valuation. Companies could tailor the detailed approach to align with their internal investment optimisation systems.

- Once those parts of the plan that are sensitive to customers' valuation have been identified, a range of options could be used:
 - companies could develop two or three versions of the plan and test these explicitly with customers or the CCG;
 - companies could review the most important customer valuations again and refine their views of the most reasonable values; or
 - companies could use qualitative information and prioritisation of service aspects to inform their final PCs.

During this step, companies should take account of potential differences between the approaches used to estimate customer valuations, and how those differences may affect the results. For example, revealed preference may not identify the whole cost/benefit of service, and as such it is likely to be an underestimate of customer valuation, if anything. A company should reflect this in how it sets its PCs, i.e. if it is choosing between two customer valuations, and the revealed preference one is the lower bound, it should probably go with the higher valuation.

Finally, a sophisticated modelling tool could also include a Monte Carlo simulation of different customer valuations and provide a probability distribution of the CBA outcomes. This could also incorporate uncertainty over financing assumptions, rainfall, or demand, and it could account for historic variations in performance.

While this approach may appear resource-intensive, the idea is that it would be based on companies' existing investment optimisation programmes to test how outcome targets would change in response to different customer valuations and provide the information as part of their evidence base. The methods suggested therefore provide a practical and proportionate way to incorporate a wider range of data in the CBA.

3.1.2 Improving the quality of cost data

As with customer valuations, the quality and availability of the cost data used in CBA is very important. Ofwat should set expectations on how companies develop cost data to be used in CBA at PR19. We have identified a couple of improvements that should be made at PR19.

- **Common costs** - As identified in section 2, companies faced challenges in allocating common costs across PCs at PR14. As companies often relied on implicit assumptions to split common costs, it was sometimes unclear what approach had been used, and whether approaches were consistent across companies. Ofwat might choose to set clearer expectations on how companies should deal with this issue at PR19 to improve consistency. This could include an explicit expectation that companies should find ways of allocating common costs across PCs, and ensuring that the allocation of common costs does not unduly affect the prioritisation of PCs and the level of stretch proposed in companies' targets in some areas. The way that companies allocate common costs should ideally be based on sound,

engineering analysis of the way that costs relate to measures. However, this may not be possible, in which case companies could split the common costs equally across PCs. If Ofwat was explicit in its expectations in this area, this should help to make the approach more pragmatic for companies and more consistently followed. Ofwat could consider engaging with companies on these explicit expectations, as the approach could then be based closely on companies' experiences.

- **Uncertainty around costs** - Ofwat should set an explicit expectation that companies should use proportionate sensitivity tests to check how the results of their CBA change when alternative cost assumptions are used. This could be incorporated into multi-input CBA if that is the triangulation approach that is adopted at PR19. If the results of these tests show that the outcome of the CBA is heavily dependent on the cost data, this would suggest that companies should apply more resources to estimating the cost information. It may also suggest that there is some uncertainty around the appropriate target to set. If however the sensitivity tests show that changes in the cost levels do not materially affect the outcome of the CBA, companies would not need to spend more time in considering how to reduce the uncertainty around the cost level. This would be a proportionate approach to dealing with this uncertainty.
- **Additional evidence** - We understand that Ofwat will expect companies to provide more detailed evidence on the costs they have used in CBA. In particular, Ofwat will expect companies to:
 - show that they have taken account of innovations that have already been identified, and reflected in the cost levels that they have used to conduct CBA and set targets; and
 - provide detailed evidence on how they have estimated their costs more generally.
- **Comparative information on costs** – We understand that Ofwat might compare information on marginal costs, for at least the common PCs, as a means to inform its checks and discussions with companies during PR19.

As we highlight below, Ofwat is likely to use this information when reviewing companies' business plans to assess how stretching the proposed targets are.

3.1.3 Affordability

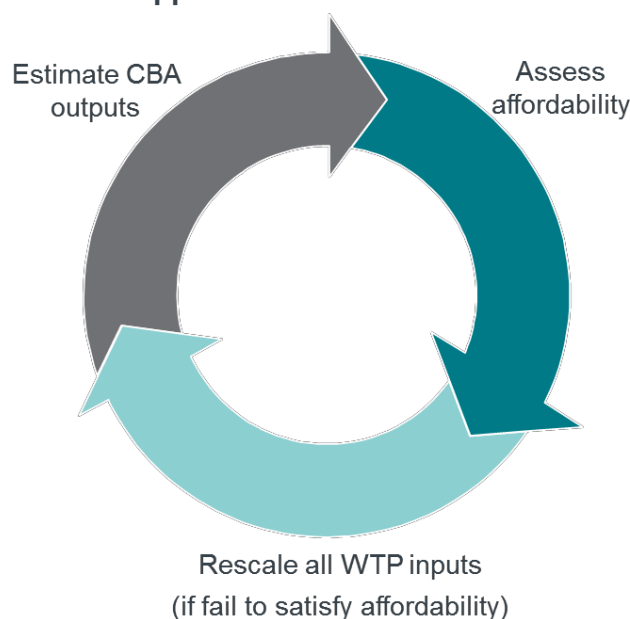
At PR14, companies used CBA to determine the target for each PC, but generally also used top-down cross-checks to check the overall acceptability of their proposed investment plans. As highlighted in section 2, it is possible that the overall WTP does not equal the sum of the individual WTP on PCs. If companies are faced with this issue, where their overall acceptability testing calls into question their disaggregated business plan, there is a question as to how companies should respond. This is an area of the outcomes framework that could be improved at PR19.

To meet the affordability constraint, commitment levels would need to be set below the optimal level, and at levels below the optimal level, marginal customer valuations would be higher than the marginal costs. To meet the affordability

constraint in the most allocatively efficient way (i.e. to make best use of the company's resources) commitment levels should be set so that the difference between marginal customer valuations and marginal costs (i.e. $WTP-MC$) is the same across all PCs. This would be challenging to achieve in practice though, as companies do not have sufficient data on customer valuations and marginal costs at all possible service levels. However, companies could use the logic behind this theoretical solution to set commitment levels under an affordability constraint. They could either reduce customer valuations, or they could increase marginal costs and then set commitment levels. We would suggest that amending customer valuations is the most appropriate option. This is largely because overall WTP may be lower than the sum of individual WTP due to the nature of customer preferences (i.e. substitutability across services) or because only a subset of measures are included in questions on individual WTP, as identified in section 2. Since these explanations relate to customer valuations, and not marginal costs, it seems most appropriate to make amendments to customer valuations. However, we note that it would also be possible to reflect affordability by making amendments to marginal costs.

In making changes to customer valuations, the key point is to maintain relative customer preferences across measures, and commit to setting stretching PCs within the affordability constraint. This will ensure that companies continue to use their resources efficiently, and to allocate them to PCs in a way that will deliver what customers want.

One simple approach that would retain relative preferences across PCs would be to re-scale the customer valuations that are used in each CBA by a common factor. We illustrate this approach in the figure below. We note that this approach assumes that the customer preferences illustrated in the individual WTP are "accurate" in reflecting customers' preferences/relative ordering across PCs, and that this is a rough approximation to the more theoretical, but impractical, solution identified above.

Figure 14 An iterative approach to CBA with an affordability constraint

Source: *Frontier Economics*

This approach would involve the following iterative process:

- The starting point would be the CBA tables that include customer valuations for individual PCs. This would give an initial set of targets.
- Next, companies would need to assess whether the total cost of delivering this set of targets exceeds the total bill increase that customers are willing to pay (or whatever the affordable level is deemed to be). If the affordability constraint was satisfied, this would be the final plan.
- If the implied bill level was too high (and if companies did not have any other evidence on customer preferences), then companies should reduce all customer valuations by 5%, or some other common scaling factor, and re-do the CBA for each PC. This would give a new set of targets. Companies may also have access to other evidence on customer preferences across PCs, including qualitative evidence, which may help them to allocate spending. If companies have access to such additional information, they could use this to justify movements away from the common factor approach to re-scaling customer valuations.

The total cost of the new plan would need to be calculated and assessed to see if it is acceptable and affordable overall. If the amended plan was found to be affordable, the process would end. If however it was found to exceed the affordable bill level, the iterative process would continue, by further re-scaling the customer valuations (again by a common factor). Using qualitative evidence is more likely to be appropriate for companies with relatively few PCs.

The iterative approach, with small adjustments made each round, should ensure that targets are still set in a stretching way and not too far away from those that would be set if there was no affordability concern.

This approach would have to be used more pragmatically if some PCs were not set using CBA, or the results of CBA were not used directly to set all PCs. In those cases, companies could apply the same principles which would involve judgements about how targets that are not based on CBA should be adjusted. This judgement would need to reflect the key concept of maintaining the relative preference across measures, and to ratchet down the value placed on them by the same factor.

3.1.4 Distributional concerns

Distributional concerns are an important policy issue that affects affordability for certain groups. It could be incorporated into the PR19 outcomes framework in the following two ways.

- When companies use multi-input CBA to test the impact of different customer valuations on targets, companies could specifically test how targets would change if the lowest customer valuation was used in each case (instead of the average customer valuations). This would provide insights on whether the difference in customer valuations across customer groups would materially affect the way that targets were set.
- Similarly, companies could explore how incentive rates and the maximum possible reward depend on the scale of customer valuations. In particular, companies could test whether ODIs would be materially different if the valuations of low income groups were used instead of the average.

It is important to recognise that the outcomes framework is not the primary tool for addressing distributional concerns. As this is a wider policy issue, that is not the result of the introduction of the outcomes framework²², there are also broader ways in which it can be addressed within the wider regulatory context. For instance, the government has required all companies to introduce social tariffs and many companies have explicit PCs on the number of people on social tariffs. Social tariffs are intended to protect the most vulnerable customers.

3.2 Increasing the effectiveness of Ofwat's role

One of Ofwat's stated goals in the outcomes consultation is to increase the level of stretch in outcome commitments. At the extreme, Ofwat could achieve this by shifting the framework so that it sets targets for all PCs, rather than companies. But such a change would be hard to reconcile with the over-arching principles of the outcomes framework such as companies owning their business plans and light-touch regulation. We have therefore considered options for Ofwat to increase the effectiveness of its challenge, while companies maintain ownership of their plans.

²² i.e. customer bills do not differentiate between service levels and are based on regional averaging, so although PCs may exacerbate this issue, the current way of addressing this issue is via the charging principles.

Process benchmarking

Process benchmarking involves assessing and comparing the systems and processes that companies use to develop their PCs. At PR14, Ofwat assessed the targets set by companies and the evidence provided in business plans qualitatively. At PR19, Ofwat could carry out process benchmarking to increase the quality of the data that is used in the underlying CBA to set targets, and to improve the robustness of the approach:

- Ofwat will have a better sense of how to make comparisons across companies' plans and to consistently apply its assessment. For instance, the assessment of plans could include a comparison of how companies have made use of the available data in their CBA, and the evidence that they have provided to justify their chosen targets.
- Where companies are found to have relatively poor processes to set PC levels, Ofwat could use best-practice examples to set expectations on how companies could improve post submitting their original business plans.
- We also note that, if there is an explicit expectation that companies should provide more detailed evidence on how they have estimated marginal costs, and how they have factored potential innovations into those costs, this should result in a more informed judgement from Ofwat. The level of information that is provided on cost data will be an important part of this process benchmarking.

Deep dives

As process benchmarking will focus on the systems and processes that companies have used to develop their PCs, it does not provide a detailed assessment of each PC. In addition to process benchmarking, at PR19, Ofwat could be explicit in that it would review companies' plans overall, and would commit to undertaking in-depth assessments (or deep dives) of companies' evidence on a subset of PCs. This approach would be proportionate, and light-touch, but would also bring additional confidence on the level of stretch in some areas. As part of these deep dives, Ofwat would look at the evidence provided on marginal costs, and may compare those marginal costs across companies. This may provide an indication of how stretching companies have been in the way that they set targets. To the extent that Ofwat chooses to carry out such comparisons of marginal costs, it would need to carry out deep-dives on the same PCs across companies.

As it would be an explicit commitment, companies would know that at least some element of their plans would be examined in detail. However, as companies would not know in advance which PCs would be subject to the deep dive assessments, they may consider it safest to provide a high level of stretch in all areas. That being said, deep dives are likely to be applied to high cost areas, or to PCs that are particularly important to customers. This may mean that companies have a reasonable idea of which elements of their plans are more likely to be assessed using deep dives. However, as this could simply shift companies' focus onto PCs with high cost implications and/or PCs that are of particular importance, this is unlikely to have unhelpful consequences.

RoRE range

While the RoRE range provides a useful tool for comparing the scale of incentives, we note that the scale of the RoRE range is outside the scope for this study. However, we have identified a number of improvements to the way the RoRE range is calculated:

- At PR14, the RoRE range was based on companies' performance at the P10 and P90 probability levels. However, Ofwat did not provide any guidance on how companies should estimate P10 and P90 performance. As a result, it is not clear how comparable the scenarios were between companies. More guidance and clearer expectations on the evidence that underpins P10 and P90 performance levels would lead to a higher degree of confidence in the comparative RoRE analysis.
- At PR14, the RoRE range was calculated without taking interdependencies into account. Interdependencies within the ODI framework (e.g. performance on one PC may influence performance on another PC) were not reflected and similarly the interaction between totex and ODI performance was not included in the modelling. As a result, the RoRE range did not provide the "true" P10 and P90 upside and downside but was a tool for comparing the size of different incentives on an individual basis. At PR19, Ofwat should require a more robust RoRE analysis that reflects interactions, or be clearer about the purpose of the analysis, if the impact of different incentives continues to be analysed in isolation.

3.3 Increasing the effectiveness of the CCG challenge

Ofwat has already made clear that CCGs will continue to provide challenge to companies' plans.²³ At PR14, CCGs were expected to challenge companies' proposed PC levels but had limited access to comparative information, reducing the effectiveness of their challenge to some extent. There is clearly an opportunity to learn from CCGs' experiences at PR14. We provide below our view on how this could work in practice, but note that this is not an exhaustive list of the possible improvements that could be made.

- CCGs, particularly new members but also existing members, could be offered the opportunity to attend workshops in advance of PR19. These workshops could clearly set out the role that the CCGs are expected to play, and explanations could be given on the way that CCGs can provide challenge to companies. This should mean that CCGs are clearer on the sorts of questions that they could present to companies. These could supplement the quarterly CCG chairs meeting which Ofwat organises. We also understand that CCWater and Ofwat are planning some workshops for new CCG members.
- More comparative information on companies' performance should be provided to CCGs in an easily accessible format, to help them challenge companies in areas where they are not performing as well as the rest of the

²³ Ofwat (2016), Ofwat's customer engagement policy statement and expectations for PR19, p. 4

industry. This could be done in several ways, including development of the Discover Water dashboard.

- Sharing of best practice on how other CCGs are providing challenge, including the questions they are asking and the evidence they are seeking from the companies. Some of this is being provided via the CCG chairs' quarterly meeting, could be provided in the introductory workshops, and if CCGs communicate during PR19 they could help to share best practice amongst each other.
 - This could include examples of how CCGs successfully provided challenge during PR14.
 - It could also include more engagement across CCGs during PR19 to share best practice.

All of these suggestions should improve the tools and information CCGs have available to effectively play their role. We also note that these will support the measures that Ofwat has already put in place to improve the effectiveness of the CCGs, including:

- the clarity on the CCGs' role it provided in the Customer engagement policy statement for PR19;
- the CCG chairs' quarterly meetings it organises; and
- the increased availability of comparative information.

3.4 Summary

We have identified a number of changes that should be included in the PR19 outcomes framework, regardless of what other choices are made. The box below summarises the recommendations that we make for the PR19 framework.

RECOMMENDATIONS FOR PR19: GENERAL IMPROVEMENTS

- Set higher expectations on the quality of the data and processes around Cost Benefit Analysis (CBA) to set PCs levels (i.e. both customer valuations and marginal cost data). This includes expectations for companies to carry out sensitivity tests to assess the impact of different levels of customer valuations and marginal costs. This should improve the robustness of the CBA, and result in more informed choices on where to set targets.
- Improve the effectiveness of Ofwat's role in assessing business plans by:
 - drawing on the lessons learned from PR14 to improve the effectiveness of process benchmarking of companies' plans;
 - committing to undertake deep-dives (i.e. more in-depth assessments) on a subset of PCs (that is selected after business plan submission); and
 - setting clearer expectations on how P10 and P90 performance levels should be calculated as part of the Return on Regulated Equity (RoRE) range, and be clearer on how interactions between PCs should be reflected in the range.
- Improve the tools CCGs have available to better enable them to play their role and challenge companies as effectively as possible:
 - workshops could be offered to new CCG members, and existing members that are interested, to explain the role the CCGs are expected to play and how they can best challenge / ask companies questions;
 - the availability of comparative information on companies' performance could be improved, including through the development of the Discover Water dashboard; and
 - best-practice from PR14, and updated during PR19, on how CCGs applied challenge, including how the questions asked could be shared between all CCGs, through the quarterly CCG chairs meetings and other means.

4 OPTIONS FOR KEY ELEMENTS OF PR19 FRAMEWORK

The previous section provides an overview of general improvements to the framework that we are recommending for PR19. In addition to these improvements, there are a number of fundamental choices that Ofwat needs to make for the PR19 framework on PCs and ODIs. In this section we describe the elements of the framework that are more flexible, that is, where Ofwat has a choice over several options and there is not an unambiguously better way forward. We explain the different elements of the framework and the options available, with their advantages and drawbacks. In section 5 we then combine different options in a coherent way to enable Ofwat to make a decision on the overall policy direction.

To identify options for the way forward in PR19, we have identified the following categories of options:

- Options for the high level instruments;
- Options for detailed methodology for PCs ;
- Options for detailed methodology for ODIs;
- Additional option: a measure-specific approach to PCs and ODIs.

4.1 High level instruments

We have defined high level instruments as elements of the outcomes framework that apply to companies at a high-level and are relevant to a company's overall outcomes proposal, rather than more detailed elements of the outcomes framework that are quite often linked to specific measures or specific components of the framework. High level instruments are designed to help Ofwat address the issue of information asymmetry and encourage companies to reveal truthful information on costs and costumer valuations. The key question is how Ofwat can create effective overall incentives for companies to challenge themselves and set stretching PCs.

Risk based review (RBR) incentives

Ofwat could state upfront that there will be a specific element of the RBR that considers how stretching companies' PCs are. There are a number of options for how this could work in practice.

- The type of incentive could be:
 - **Reputational** – Reputational incentives as part of the PR14 RBR appeared to be relatively strong, as companies were keen to receive a good performance rating as part of the RBR (for both individual aspects of the RBR and for the overall review).
 - **Financial** – Financial incentives would reinforce the reputational incentives and encourage companies to set higher targets for themselves,

trading off future rewards for the immediate financial gain from the RBR “prize”.

- **Procedural** – companies that have a clearly challenging plan could benefit from an early determination with no further scrutiny. The procedural incentive could also be two-sided, Ofwat could carry out more deep-dives on a company’s outcome plan when their PCs are amongst the least stretching of all companies (see next point on deep dives).
- The incentive could be **for specific elements or for the overall framework**. One option would be to have one overall score that summarised the level of challenge in the whole outcomes framework. Alternatively, Ofwat could provide multiple prizes for different elements of the outcomes framework. One overall prize will enable companies to choose how to focus efforts, and may be simpler to implement. However, multiple prizes for different elements of the framework would ensure that companies that show particular strength in certain areas are rewarded for their efforts.
- Ofwat would need to decide **how many companies** should be able to achieve the reward. On the one hand, Ofwat could set a very high standard by letting only one company receive the reward. This would encourage stronger competition between the companies at the top which were confident they could come first. However, the companies that were struggling would likely give up on trying to get the reward. As Ofwat increased the number of companies that could get the reward, or equivalently, lowered the standard for achieving it, more companies would be incentivised to challenge themselves and add more stretch to their plans, however, it would remove the incentives for companies that were anyway comfortably in a position to meet the required standard.
- Finally, the incentive could be **one-sided or two-sided**. One-sided incentives create the dilemma explained in the preceding paragraph. An alternative way of incentivising both the companies at the top and those at the bottom would be to combine “prizes” for the best plans with some sort of penalty for the plans that were least stretching. The penalty can be reputational (name and shame), financial (a deduction from the cost allowance) or procedural (additional scrutiny or the requirement to resubmit an improved plan).

Menu incentive for costs and outcomes

Ofwat has in the past used menu incentives to address the information asymmetry problem – companies have better information about their costs and about the uncertainties that will affect their future costs. A menu incentive mechanism “provides incentives for companies to be efficient and to reveal accurate information about their expectations of future costs”.

While, to date, this has been used exclusively as part of the cost assessment, Ofwat could design a menu for an integrated assessment of costs and outcomes. The menu incentive would need to be informed by an integrated modelling approach that combines costs and outcomes. It is therefore not compatible with the current separate assessment of costs and outcomes. The idea is that companies would therefore make their menu choice on outcomes and costs simultaneously. While there are significant challenges in designing such a menu,

this approach could be used to incentivise truthful revelation. While this approach may appear optimal in theory, the disadvantages are that it requires substantial resources and is not likely to be transparent as the modelling is complex.

4.2 Detailed methodology for the implementation of PCs

The aim in developing the methodology for setting PCs is to ensure PC levels are stretching and robust to any issues in the data. There are two broad approaches for achieving this:

- An approach built on CBA for each measure could be supplemented with additional challenges. This is equivalent to challenging the marginal cost that different companies face as additional challenges based on, e.g. comparative information, effectively question whether a company could achieve a particular level of service at lower marginal costs.
- Alternatively, the cost and outcomes assessments could be integrated into a single assessment. While this would require a fundamental change in the methodology, it is clearly an option that needs to be considered in the long-run.
- Alternative approaches could be used for some particular PCs: engineering based models; and basing targets on statutory targets (this links with the measure-specific approach discussed in section 4.4).

4.2.1 Complementing CBA with alternative approaches

The PR14 methodology required companies to undertake CBA for all measures to determine PCs. In our view the principle of CBA is sound, and should remain the basis of the methodology to set PCs. However, as identified in section 2, there were clear implementation issue with this approach at PR14. Data quality and data availability issues meant that companies were not always able to set PCs at the optimal level and the PCs were not always consistent with the data they presented in their data tables.

There is therefore a question as to whether Ofwat requires that companies use CBA to directly set PCs at PR19. Ofwat could either:

- maintain the PR14 methodology and require CBA to be used in all cases; or
- explicitly allow companies to rely on other methods to set PCs on some measures.

We have identified several options for how companies could complement the CBA approach with additional information. Ofwat could set the expectation that companies should use one or several of these alternative approaches wherever it is possible and would produce more robust PCs.

Using comparative information

Aside from the six measures with common commitment levels set by Ofwat, there will be many measures for which all or at least several companies monitor performance or set PCs or for which comparative data are available e.g. via the Discover Water dashboard. Ofwat could set the expectation that companies must use comparative information from PR14 commitment levels or actual performance levels, in order to challenge the new PC levels. There are several ways that this could work:

- This could be a prescriptive rule, for example, requiring that for the first year of PR19 all companies set PCs that are at least as high as the average or the upper-quartile for PR14. A prescriptive rule may introduce more challenge but it may not be appropriate where the quality of the data is not good enough, for example because only a small number of companies track that measure or because they measure it in slightly different ways that mean it is not perfectly comparable.
- A more flexible alternative is for Ofwat to change the reference point to the upper quartile level of performance from PR14, and require companies to justify where they set a PC below that. As this is only a shift in the reference point, companies would still be able to set a less challenging PC (if necessary) but would need to explain why they are not able to deliver the PR14 upper quartile level of performance. The goal would be to use information to provide challenge and context for the choice of PC without unduly restricting companies' discretion.

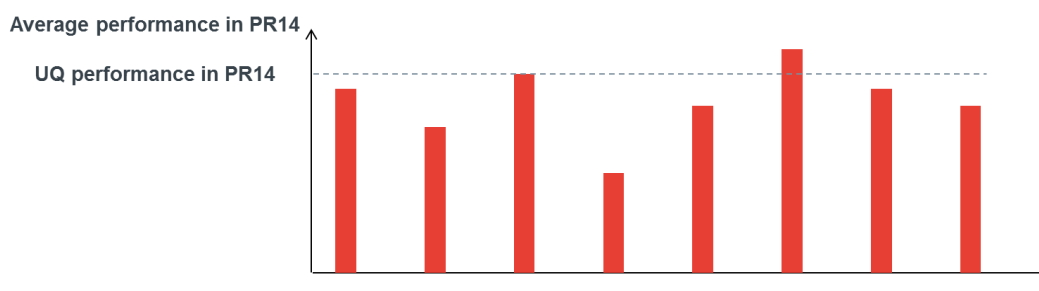
We provide a worked example of how this could work in practice in the box below.



Worked example: changing the reference point

The steps for using comparative information to set a more challenging reference point would be the following.

- Convert the units of the measure so that, wherever possible, they are the same for all companies. This may require some rescaling in some cases. For example, the number of properties at risk of persistent low pressure could be converted to the percentage of properties that are at risk of persistent low pressure.
- Choose a reasonable time horizon. For example, this could be the last year of the AMP, or the average over 5 years.
- Compute the upper quartile service level of performance across that time horizon.
- Set reference point, for example, at the average or the upper quartile of past performance.
- This reference point may be a required minimum level or may simply be a reference point, where companies who set PCs below that level would have to provide evidence for why their optimal PC was lower.



Using historical information

These are some ways in which historical information could be used to further inform the level at which PCs should be set:

- Ofwat could set the expectation that companies should calculate the average improvement that they achieved during PR14, and aim to deliver the same level of improvement during PR19. For example, if a company has historically been able to achieve a 5% year-on-year improvement on these measures, Ofwat could expect that companies include at least a 5% year-on-year improvement in the PCs that they propose for PR19. The following worked example, summarises how historical data can be used to set minimum improvement requirements.
- Where historical information is available for all companies, this could alternatively be based on rates of change in industry performance.
- Finally, Ofwat could define a minimum improvement for companies itself, based on PR14 historical information, potentially combined with some regulatory judgement for how circumstances may have changed since the last AMP.

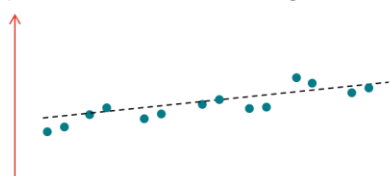
- In all cases, this would not necessarily be a requirement, but companies may instead be expected to carefully consider this reference point and fully justify cases when they have not included this level of stretch in their proposed PC targets.



Worked example: minimum improvement

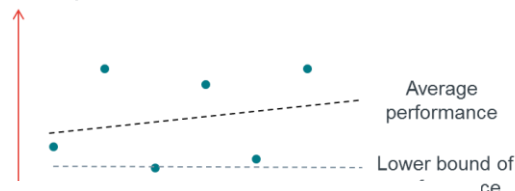
Ofwat could set expectations on the level of stretch that is included in PCs on fully bespoke measures. This could help CCGs to provide challenge to companies on these PCs.

The use of historical information to set expectations on minimum improvements will depend on the richness of the historical data and the variance in performance. The diagrams below provide two possible scenarios.



- Large number of observations (measured frequently)
- Low variance around the trend line
- Statistically significant average improvement

In a case like this, the expectation could be that companies set a minimum improvement based on average past improvement.



- Small number of observations (measured frequently)
- High variance around the trend line
- Statistically insignificant average improvement

In these cases, it may be more appropriate to use the lower bound of past performance to set a requirement on minimum levels.

Defining the maximum

Ofwat could set the expectation that companies should, where appropriate, define the maximum possible level for a given measure, and then work backwards from that level, rather than using their current level of performance as the starting point for analysis.

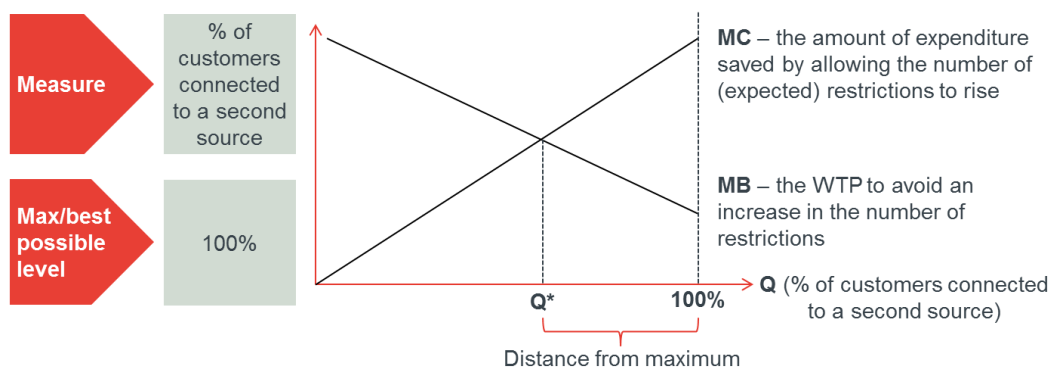
In some cases, thinking about the maximum may not be a helpful exercise because it may require the engineers considering a hypothetical network designed from scratch. But in other cases, this change in the reference point could make companies more ambitious in how they think about what is possible. The following worked example shows how this could be applied to decisions over investment projects to improve resilience in the network.



Worked example: defining the maximum

The steps for using the maximum performance level as the reference point when setting the PC for water restrictions could be the following:

- Start from what would be required in order to ensure no customer would ever face a water restriction (e.g. all customers connected to multiple water sources).
- Compare this to the level that is included in the business plan.
- Justify why the maximum possible level is not achievable.

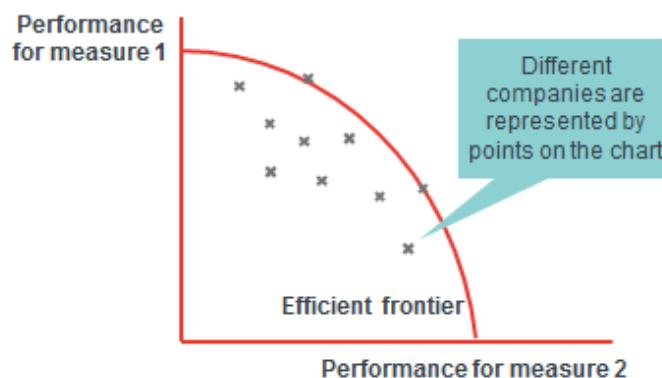


4.2.2 Integrated cost and outcomes benchmarking

Consistent with the menu incentive for costs and outcomes discussed in section 4.1, this option is a more radical change to the framework than the previous approaches we have discussed. It would involve moving away from the current approach: company-led led plans, based on CBA for individual measures, carried out independently of the cost assessment exercise. Instead this approach recognises that PCs are not independent of costs. Therefore, in order to determine whether PCs are stretching, ideally there should be an integrated assessment of outcomes and costs within a single framework.

This alternative framework would integrate the assessment of companies' costs and outputs, and would use benchmarking techniques to determine the efficient frontier: all the possible combinations of outcomes that could be delivered by a hypothetically fully efficient company, for a given cost level. Based on this, inefficient companies would be required to deliver more stretching PCs and more efficient cost levels.

The diagram below illustrates the principle of such a framework, with two service levels (Q1 and Q2) and a cost level (C).

Figure 15 Illustration of integrated approach: the efficient frontier

Source: *Frontier Economics*

While this framework would identify the efficient levels of cost and service performance, companies would still be required to carry out customer research to determine what weighting they should place on different aspects of customer service. This is akin to companies choosing their place on the line in the above illustration. This would mean that customers' preferences are still reflected in the way that PCs are set. In addition, if the framework was to deliver improved customer impacts in terms of better customer service at efficient cost levels (and potentially to lower bills), this could lead to higher trust and confidence in the sector.

However, as the framework would assess relative performance on service metrics at one time, it would be best suited to common measures. As a result, it would be difficult to incorporate bespoke measures into this framework. The nature of the framework would also mean that the overall approach is more interventionist than PR14, as it is likely that Ofwat would define the measures that are included in the framework (although it is also possible that the common set of measures could be selected by all the CCGs working jointly).

In practice, the implementation of such a framework would require new modelling techniques and significant research efforts to develop a practical approach. In Annex A we describe one possible technique that could allow for an integrated assessment of costs and outcomes – Data Envelopment Analysis (DEA) – but it is likely that a combination of techniques would need to be applied.

4.2.3 Alternative approaches to setting PCs in some cases

There are two other ways that could be used to set PCs, which may be particularly suited to certain PCs.

- **Engineering based models** - Some PCs, for example asset health measures, may be best set using engineering models that assess the quality of the network and where investments need to be made. These approaches are more focused on what inputs are required, to deliver certain long-term outcomes, and are therefore not appropriate to use to set targets in the majority of cases. But for PCs on asset health, they may be more pragmatic

than CBA, as they may provide the most informed view on what the target should be.

- **Statutory targets** - Companies face statutory targets for some PCs. Companies should be required to set their own targets in line with the statutory obligations, unless they have evidence (e.g. from CBA) to show that customers would prefer a more stretching target than the statutory obligations would imply.

4.3 Detailed methodology for the implementation of ODIs

As discussed in section 2, the outcomes consultation indicates that Ofwat may move towards a regime with more powerful ODIs. There are three main rationales for more powerful rewards:

- **Regulatory judgement** – Ofwat may consider that in PR14 the incentives for improvement were too weak to change company-behaviour significantly;
- **Innovation** – innovation by one company is likely to spill over to the rest of the industry extending the benefits beyond the set of consumers of a single company, since innovation requires risky investments it may be socially beneficial to offer more powerful rewards to stimulate innovation (see box below);
- **Offsetting changes to the WACC or other incentives** – Ofwat may choose to reduce the allowed return in the cost allowance or reduce the return available from cost reduction incentives and shift this return to the outcomes framework.

Further, at PR14 the ODI incentives were asymmetric with the overall potential for penalties outweighing rewards. We therefore focus in much of this section on “uplifts” to rewards, although we recognise that it is possible to apply similar uplifts to penalties.

POSITIVE EXTERNALITIES FROM INNOVATION

When one company achieves a large improvement in performance due to a significant innovation in its technology or managerial approach, this is likely to affect customers of other companies. Technological innovations are soon adopted or imitated by others in the industry so that one innovator can expand the efficient frontier for the whole industry. In addition, by raising the bar in any regulatory benchmarking between companies, the benefit of the innovation can be quickly passed onto customers of all companies.

This positive externality that one company could have on the consumers served by other companies could be reflected through an “innovation prize” – a special uplift to rewards for companies that reached exceptional performance on one or several quality measures. This uplift could be calculated to allow the innovator to share in the additional benefits created for other consumers.

4.3.1 Incentive rates and structure of ODIs

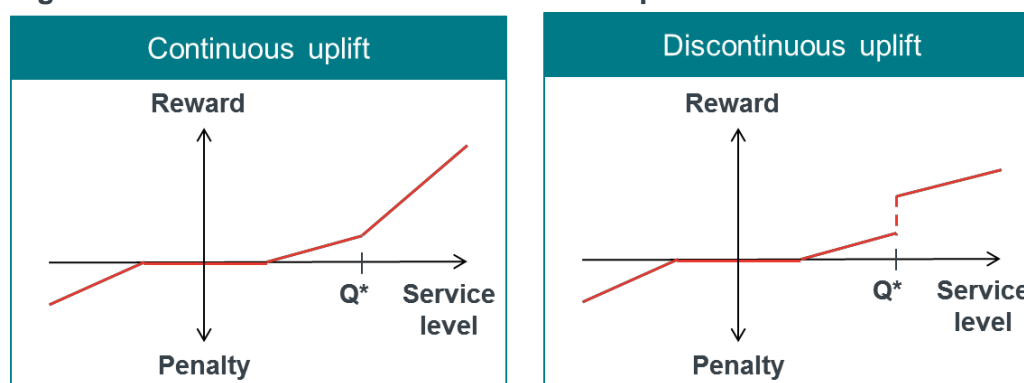
Setting incentive rates

We have identified a range of options for how the reward and penalty rates could be set.

- **Penalty rates** could be set according to:
 - the PR14 formula²⁴;
 - an adjusted PR14 formula – in principle uplifts could be applied to penalties although, as mentioned above, the rationale for penalties beyond the level given by the PR14 formula is not clear; or
 - marginal cost – this would be simpler to implement and wouldn't rely on the robustness of the WTP data.
- **Reward rates** could be set according to the PR14 formula²⁵ with or without additional uplifts. In principle, uplifts could be applied with a smooth profile or with a discontinuous prize for exceeding an “innovation threshold”. This is illustrated in the figure below.
 - In the first diagram we see that at the threshold level service level Q^* rewards receive an innovation uplift applied continuously, that is, the reward rate increases but there is no jump in the total amount of reward.
 - In the second diagram we illustrate the case of a discontinuous uplift, it is a lump-sum reward that is received for reaching service level Q^* . Beyond that, additional improvements earn the same incentive rate as those below Q^* (although the two approaches could be combined to have a lump-sum prize and a higher incentive rate above Q^*).

Discontinuities in the incentive profile can make incentives sharper; in the figure below, the lump sum reward at Q^* creates very strong incentives to reach that service level. However, discontinuities also increase the risk of unintended consequences compared to smoother incentive profiles.

Figure 16 Continuous and discontinuous uplifts to rewards



Source: Frontier Economics

²⁴ The PR14 formula for ODI penalties is $ODI_{penalty} = incremental\ WTP - (incremental\ cost \times p)$, where p is the customer share of expenditure performance.

²⁵ The PR14 formula for ODI rewards is $ODI_{reward} = incremental\ WTP \times (1 - p)$, where p is the customer share of expenditure performance.

The magnitude of uplifts – top-down versus bottom-up

If uplifts are applied to reward rates, there is then a question around how the magnitude of these uplifts should be determined. There are multiple ways that this could be done; we have considered two options in detail below.

- **Top-down scaling of rewards** - this is consistent with uplifts based on regulatory judgement or offsetting changes in the WACC as discussed above. Reward rates could initially be calculated using a formula akin to the PR14 formula, and could then be increased using a common uplift factor. If Ofwat sets a RoRE range for PR19, then the uplift factor could be based on the scale of the RoRE range. Using a common uplift factor across all measures would mean that customers' relative preferences across measures would be maintained in the final reward rates (assuming that the initial calculation of reward rates was based on customer valuations). However, the absolute magnitude of rewards would not be linked directly to customer valuations.
- **Uplifts based on innovation externality** - this is consistent with innovation as the rationale for uplifts. In this case, the rewards would be calculated as the additional benefit that innovative performance from one company would bring to all customers. These uplifts would therefore reflect the positive externalities that would be generated if one company develops new technologies and/or management approaches that all companies could adopt in future. The uplifts could be available to all companies, but limited to a specific number of companies who were successful first as they would be intended to reward innovation rather than general improvements. However, this would reduce the incentive to innovate as companies would need to factor in if they could, for example, be the first company to achieve the stretching performance level. The other downside of this approach is that companies may achieve improved performance by innovating in different ways (e.g. organisational structure, technology, better systems, etc) so this approach would limit the incentives to continuously find better ways of delivering the services.

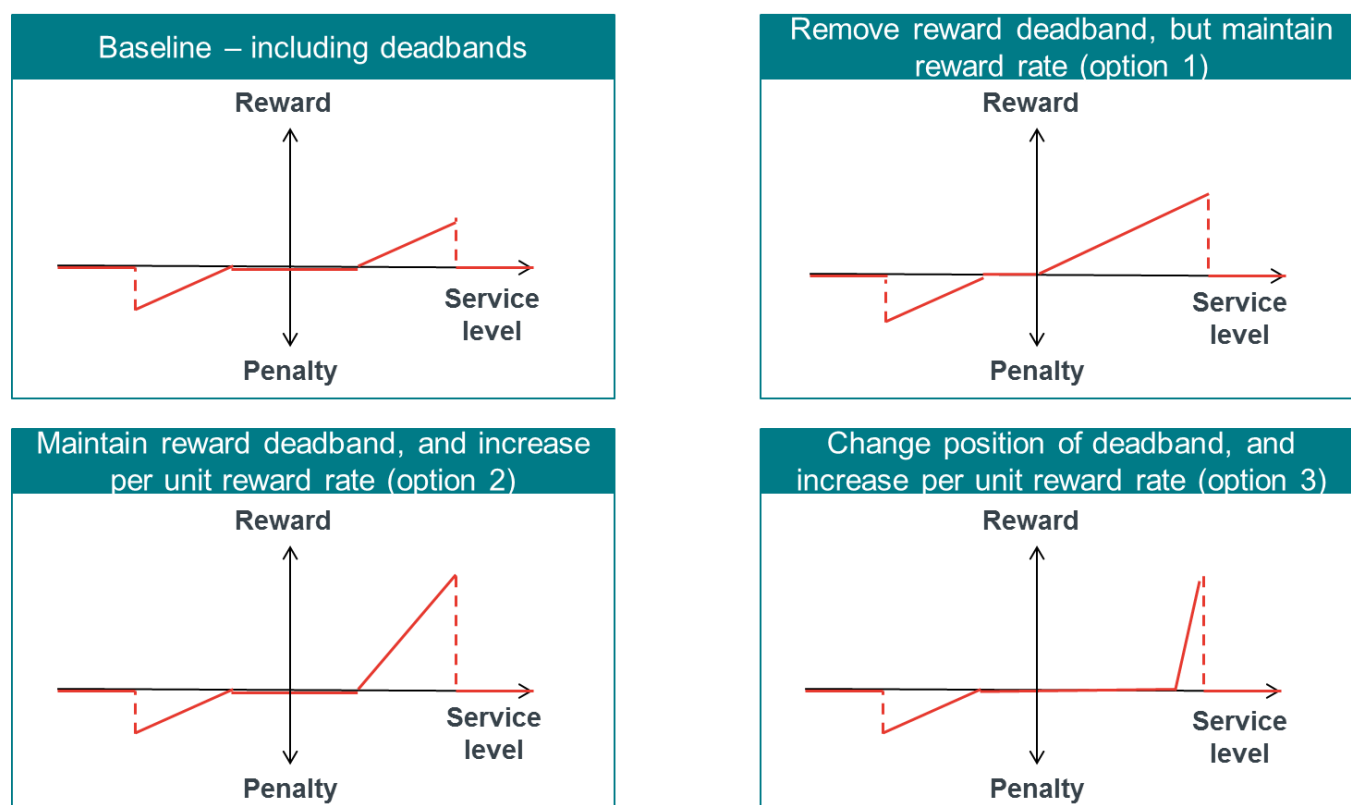
The structure of ODIs – deadbands, caps and collars

In addition to the question on the incentive rate formula, the structure of the incentives must also be addressed. For example, the following changes to the structure of ODIs would have the effect of increasing the power on rewards:

- Converting penalty-only ODIs into two-sided ODIs. This may not be appropriate in some cases, for example, where there is considerable uncertainty over what the company can achieve, say, because it is a new measure that was not previously monitored. It may be preferable to add power elsewhere in the framework.
- Removing or reducing deadbands so that rewards apply at a lower level of performance.
- Extending the cap so rewards accrue up to higher levels of performance.

The figure below illustrates different structures for rewards and is followed by a discussion of how these lead to different incentives for the companies.

Figure 17 Options for increasing the power of rewards



Source: Frontier Economics

- **Baseline.** The top left hand chart within the diagram depicts a baseline incentive rate structure. This could be considered to represent the PR14 incentive structure, as it includes both reward and penalties, and deadbands.
- **Remove reward deadband (option 1).** The top right hand chart presents one possible way in which the scale of rewards could increase. As the reward deadband has been removed, this means that any performance above the commitment level would be rewarded. This increases the incentive to deliver slightly better performance than the commitment level, as this would now receive a reward. However, as the unit incentive rate is the same as in the baseline case, the incremental incentive to deliver service improvements above the deadband level is the same as in the baseline case.
- **Increase reward rate (option 2).** The bottom left hand chart presents an alternative option, where the reward deadband has been retained, but the reward rate has been increased. Relative to the baseline, this would increase a company’s incentive to deliver performance above the deadband, which could be considered moderate sized improvements.
- **Move the deadband and increase reward rate (option 3).** The bottom right hand chart presents another option, where the deadband is moved to a level that constitutes “genuinely innovative performance”. The reward rate below

this level is equal to zero, or very low, but a steep reward rate is applied beyond the challenging level of performance. This incentive structure would mean that companies have no, or limited, incentives to deliver marginal improvements in performance, but increased incentives to deliver innovative performance.

This analysis shows that different incentive structures achieve different company behaviour. Note that the decisions on the incentive rate and the incentive structure are linked. If the rationale for uplifts is to reflect the externality benefits of innovation, then the structure should reflect this too – uplifts should be applied for step-change improvements as in option 3 above. In contrast, if uplifts are based on regulatory judgement (for example, if the regulator considers that the simple formula is not delivering sufficient power to change company behaviour), then these uplifts should apply for moderate improvement too, as in option 2. As such, the way that power is added to ODIs should reflect Ofwat's policy goals for PR19.

Reflecting substitutability / complementarity between measures

In PR14 preferences for different dimensions of quality were essentially treated as independent from each other.²⁶ In reality, these aspects of quality have interactions. For example, the better the taste of water, the more people will drink their tap water and the higher the average willingness to pay to avoid service interruptions. In other cases, measures may be substitutes if customers value different measures but have declining willingness to pay so that, if bill levels go up because one measure improves, they may become less willing to pay for further improvements in other measures.

If there are substitution or complementarity effects between measures then the WTP for individual measures cannot be added it up. For example, for substitutes this could lead to larger bill increases than consumers were willing to pay, at the PC levels of performance. A more sophisticated mapping of preferences would be required, reflecting how WTP changed on each measure as other measures improved.

²⁶ There was scope for grouped PCs at PR14 to reflect complementarities between PCs, but these were not used apart from a very small number of cases.



The customer valuation matrix

The following “customer valuation” matrix is an illustration of how willingness to pay may change with different combinations of two measures that are substitutes. The matrix maps individual’s monetised value from different combinations of performance across two measures.

		Quality measure 1				
		95	96	97	98	99
Quality measure 2	50	0.1	3.0	4.0	4.7	5.0
	51	3.0	4.5	5.4	5.9	6.1
	52	4.0	5.4	6.1	6.4	6.5
	53	4.7	5.9	6.4	6.6	6.6
	54	5.0	6.1	6.5	6.6	6.6

When quality measure 2 is at the level 51%, the average consumer is willing to pay £1.5 to increase the performance in quality measure 1 from 95% to 96%; however, if measure 2 were at a higher level, say 53% then the willingness to pay to increase performance in quality measure 1 from 95% to 96% would be lower, £1.2. For N measures this would be an N dimensional matrix of valuations, and given the number of parameters to be estimated, a large number of observations would be required. The industry, possibly a body like UKWIR or Water UK, would have to run comprehensive surveys in order to get to this information.

An alternative approach that would be less data intensive would be to apply judgement to estimate these interactions, which could be informed by qualitative customer research but without estimating every parameter directly from the data. For example, some interactions could be set to zero if two measures were unlikely to have any significant substitution or complementarity effects.

4.3.2 Gated ODIs

A system of gated ODIs could be used to safeguard against a situation where a company performs very poorly on one (or more) important dimensions of service, but still receives significant rewards because it outperformed its PCs on other measures. Gating would mean that companies would face a minimum standards requirement on a set of measures – if they failed to reach these minimum standards on a single measure (or more), they would not qualify for any rewards, regardless of how well they performed in other measures.

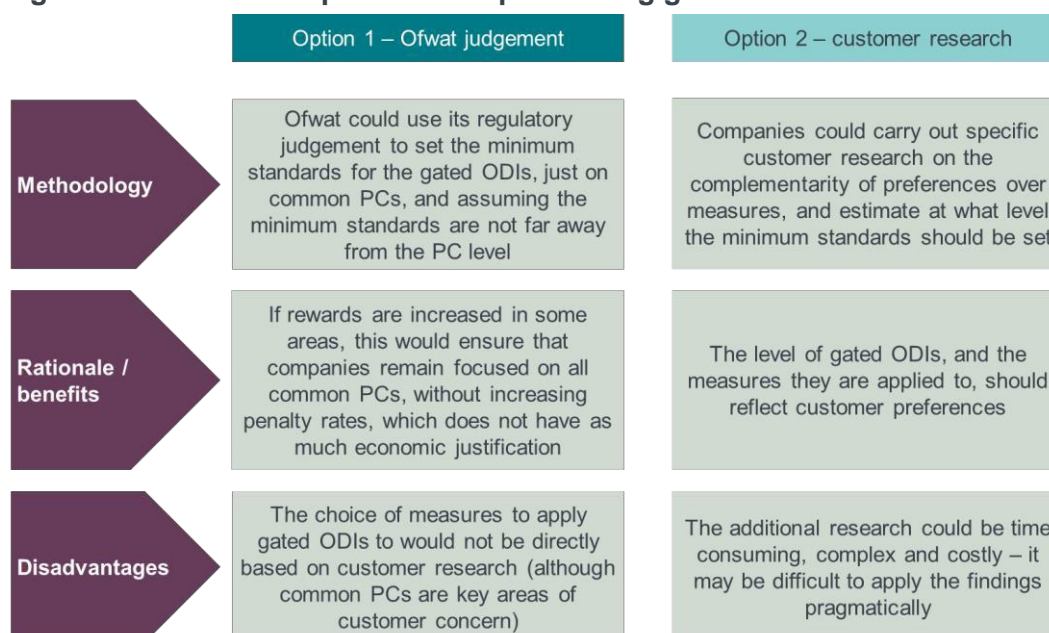
However, introducing gated ODIs would also have drawbacks. Companies might spend significant resources improving their service and have achieved it

relatively consistently but fail to earn rewards if they had one measure that didn't meet the gate. This may be perceived as unfair by companies, and, more importantly, it may reduce the incentive for companies to chase rewards undermining the entire incentive scheme. Gated ODIs also add complexity to the ODI structure, which might reduce incentives to pursue rewards. There are a number of detailed design choices within gated ODIs.

What information should the gates be based on?

The diagram below presents two possible options. In the first option, Ofwat would use its regulatory judgement to set the system of gated ODIs, i.e. which measures have minimum standards, and which measures have rewards potentially gated as well as what the level of the minimum standards should be. In the second option, companies are expected to carry out customer research to develop the system of gated ODIs.

Figure 18 Possible options for implementing gated ODIs



Source: Frontier Economics

An alternative to these two options would be for the CCGs to be involved in setting levels. This would be somewhere in between these two options, as it would be based on judgement at some level, but it is also likely to be based on customer research to some extent.

What measures should be included?

If the framework has very few measures with gated ODIs, the impact of introducing gated ODIs is likely to be reduced. However, if many gated ODIs are introduced, this would materially increase the impact of performing below PC levels and may disincentivise companies from pursuing reward. It may be particularly difficult to apply gated ODIs to new measures, as there may be a lack of robust performance data on these measures. This could mean there might be additional uncertainty around what is the appropriate PC level for these

measures. If gates were applied to these measures, companies may have a disincentive to develop and introduce new measures.

What level should the gates be set at?

If the levels of the gates are set at current levels (or even the lower quartile), the impact of them may be relatively small, although they would still provide an extra challenge as dropping just below the PC level on any gated measure would eliminate all potential rewards on gated measures. This is desirable if the gates are seen as a safeguard.

Given the scale of the potential impact of gated ODIs, setting the levels of the gates considerably above the PC may have a very material impact the incentives associated with the outcome framework. If there is a significant probability that companies may miss the gate on at least one measure, such that the chance of capturing any rewards is low, this could materially reduce the incentives to innovate and deliver beyond the PC on all other measures that are included in the gate.

4.3.3 Common versus company-specific ODIs (for measures with common PC levels)

In PR14, five measures had common performance commitment levels across all companies but different financial incentives. At PR19, Ofwat has the option of introducing common ODIs for common PCs. This would require Ofwat to determine the structure of these incentives centrally.

Matching common ODIs with common PCs is a more consistent approach and would prevent a situation where companies have the same target but very different incentives to meet that target, which could undermine the case for having common PCs at all. On the other hand, company specific ODIs would ensure that the incentives of each company reflect its actual cost structure as well as the value of improvement to its own customers. In addition, allowing for companies to have different incentives would be a less interventionist approach. This is particularly true as the common PCs are likely to be those with the largest incremental totex so common ODIs may shift the balance from companies to Ofwat with regard to designing the incentives. This is not entirely aligned with the overall methodology.

Which option is most appropriate depends on how much genuine heterogeneity between companies (and their customers) there is: if Ofwat expects that the costs and the value of improvement for these measures is not very different across companies, then opting for common ODIs may be preferable because it would prevent large differences in incentives that are driven by data issues for example. If Ofwat expects there to be large genuine differences between companies then company specific ODIs are more suitable.

4.4 Implementing a measure-specific approach

We have set out a large menu of options for the detailed methodology for setting PCs and ODIs. An overarching question regarding the detailed methodology is whether the exact same methodology is applied to all measures, or whether companies may adapt the methodology to the specific characteristics each measure (this approach is discussed in more detail in Annex B).

One obvious example is the application of CBA. The issues discussed in section 2 highlighted the difficulty companies had in implementing CBA. However, this difficulty varied depending on the measure; some issues were general, but some measures had specific barriers to implementing CBA.

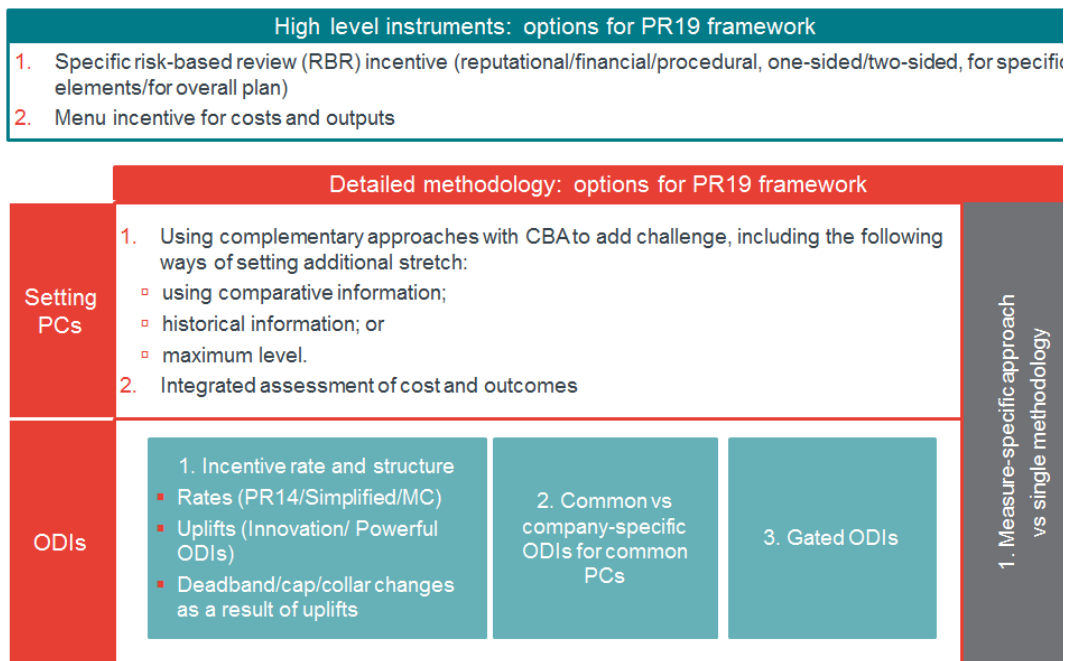
A measure-specific approach might be a pragmatic way of capturing how to set PCs and ODIs for different types of measures. In some cases, a straightforward CBA may lead to a sensible PC level which is robust to measurement error in the costs and valuation data. In others, CBA may not be the best tool as statutory targets require 100% compliance. Similarly, there may be some measures where it is difficult to estimate customers' valuation such as asset health. The measure-specific approach therefore reflects the reality of different types of measures and identifies the most appropriate method for setting PCs and ODIs.

It is worth noting however, that a measure-specific approach could have some drawbacks. As the approach would be more pragmatic and less mechanistic (i.e. the same approach would not be used for all measures), there is a risk that this may lead to some variations in the way that companies approach setting PCs, which could lead to some inconsistencies in the end results. However, overall a flexible measure-specific approach should make it easier for companies to develop meaningful and robust PCs and ODIs, especially if Ofwat provides clear expectations of how different measures should be considered. Annex B develops the thinking for the measure-specific approach in more detail and considers which approaches would work best for each category of measures.

4.5 Summary

As becomes clear from the discussion in this section, there are many elements of the methodology that need to be considered and several options available for each of them. The following figure summarises the decisions Ofwat will have to make in setting the expectations for companies in PR19.

Figure 19 Summary of decisions for Ofwat on the detailed methodology



Source: Frontier Economics

5 OVERVIEW OF PACKAGES

As there are many different choices to be made, we have developed four plausible packages (in addition to a base case – the PR14 framework). The rationale for the packages is to manage this complexity and facilitate the evaluation of Ofwat’s choices. Each package represents a different combination of options, and is focused on achieving a specific policy objective.

Figure 20 provides a summary of our four packages, and the base case scenario. The base case scenario reflects the approach that was used at PR14, and the other four packages will be assessed relative to the base case. Packages 1-4 all contain the general improvements that were outlined in section 3:

- clearer expectations on CBA (i.e. including multiple data sources for customer valuations, triangulation, and the use of sensitivity tests);
- increasing the effectiveness of the Ofwat challenge (i.e. including process benchmarking and a comparison of marginal costs, deep-dives on a subset of measures, and clearer expectations on P10 and P90 levels); and
- increasing the effectiveness of the CCG challenge, by improving the tools available to them to better carry out their role.

Figure 20 therefore identifies the aspects of the packages that are additional to the general improvements.

As noted above, there remains some flexibility within each of the packages, such that other options could be selected while retaining the policy focus of the package. In respect of this flexibility, we have developed the packages so that they have the “boldest” options within them. We have also identified ways in which the packages could be amended to reduce the risk of unintended consequences. For example, in package 1 we have suggested that it include uplifts for common PCs and for some bespoke PCs. However, as Ofwat may have less confidence in the level of stretch in the targets on bespoke PCs (due to a lack of comparative information), it may decide not to apply uplifts to bespoke PCs. In the remainder of this subsection, we provide a more detailed description of packages 1-4, and highlight the possible ways in which those packages could be amended to reduce the risk of unintended consequences.

Figure 20 Summary of packages

Base case

	High level instruments	Methodology
PCs	Rating of PCs and ODIs as part of overall RBR, high level review of all ODIs	PR14 (CBA for all bespoke, CCG provides challenge)
ODIs		PR14 (Company-specific ODIs on common PCs) PR14 (ODIs for all measures based on PR14 formula)

1. Focus on PR14 improvements

	High level instruments	Methodology
PCs	Same as base case (with general improvements)	CBA for all bespoke, more expectations set for CBA approach, rely on comparative info given to CCGs for extra challenge
ODIs		Common ODIs incl. uplifts for common PCs Based on PR14 formula for bespoke PCs with uplifts only on some measures

2. Focus on innovation

	High level instruments	Methodology
PCs	RBR incentive	Measure specific approach, proportionate application of CBA, incl. additional challenge
ODIs		Common ODIs with innovation uplift for common PCs Based on PR14 formula for bespoke PCs with innovation uplifts

3. Focus on tailored rewards

	High level instruments	Methodology
PCs	RBR incentive	Measure specific approach, proportionate application of CBA, incl. additional challenge
ODIs		Company specific ODIs for common PCs with uplifts Uplifts for ODIs on bespoke PCs where evidence is strong Gated ODIs

4. Focus on integrated assessment

	High level instruments	Methodology
PCs	Menu incentive for costs and outcomes	Assess costs and outcomes together combining results from multiple methodological approaches (e.g. SFA, DEA)
ODIs		Companies to collect data on customer valuations to apply to results from integrated benchmarking Rewards on each measure depend on performance on multiple measures, to reflect customer preferences across measures Innovation uplifts available for some measures

Source: Frontier Economics

5.1 Package 1 – Focus on PR14 improvement

This package is designed to focus on addressing the lessons learned from PR14, particularly with regard to the way that the PR14 framework was implemented. It therefore retains the overall policy direction from PR14, where companies are encouraged to own their business plan development and focus on designing outcomes that meet their customers’ preferences.

High level instruments

There are no additional high level instruments in this package, although there will be the general improvements that we identified in section 3. This package therefore does not provide any upfront incentives for companies to be ambitious in their plans and targets beyond the incentives provided in PR14, such as the financial, reputational and procedural incentives to achieve enhanced status.

Detailed methodology

The main elements of the methodology are:

- PR14 approach used to set PCs, with the general improvements on improving the way that CBA is done;
- common ODIs on PCs with common commitment levels, including top-down uplifts;
- top-down uplifts applied to the ODIs for all (or the vast majority of) remaining PCs.

As at PR14, CBA would be used to determine all bespoke PCs, possibly with a small number of exceptions (for example, for asset health measures). The general improvements to CBA would be applied, including: using multiple data sources on customer valuations; triangulation; and the use of sensitivity tests. There would be no explicit mechanism for adding more challenge to PCs beyond what CBA prescribes.

Where Ofwat sets common commitment levels, it would set common ODIs. Penalties could be set at the level of the highest marginal cost to ensure all companies are incentivised to meet the target. Rewards would also have a common reward rate, and Ofwat could apply uplifts to provide more powerful incentives for improved performance. The common PCs have the strongest rationale for applying uplifts: as comparative information is likely to be used directly to set PC levels, Ofwat should have confidence in the level of stretch in the PC levels; and these PCs are important to all customers, which means there are positive externalities from one company innovating and pushing out the efficient frontier for the whole industry.

For bespoke PCs, companies would be expected to apply the PR 14 formula to set ODIs, and where they have strong evidence that their PC level is stretching, they may also apply uplifts to rewards.

Potential changes to reduce the risk of unintended consequences

The following elements could be modified to reduce the risk of unintended consequences:

- add extra challenge using available information to set PC levels, for example, Ofwat could expect companies to deliver a minimum level of improvement, and provide a detailed justification when they cannot; and/or
- remove the uplifts on rewards for bespoke PCs.

5.2 Package 2 – Focus on innovation

The focus of this package is on delivering innovation. It should incentivise companies to deliver high quality plans, including innovative measures and using innovative customer engagement, and also sets significant rewards for companies that deliver innovation in service delivery.

High level instruments

In addition to the high level instruments included in our general improvements, this package includes a RBR incentive that is focused on innovation. The incentive would be used to reward companies that deliver high quality plans, including stretching introduction of innovative measures, new customer engagement approaches, etc. This upfront prize for companies that are innovative should provide strong incentives for companies to challenge themselves and set stretching targets.

Detailed methodology

The main elements of the methodology are:

- a measure-specific approach to setting PC levels for bespoke PCs;
- common ODIs on PCs with common commitment levels, including innovation uplifts; and
- innovation uplifts applied to ODIs on some bespoke PCs.

PC levels for bespoke PCs would be set using a measure-specific approach. This means that CBA would be used in most cases. Ofwat would set expectations for the level of improvement / target levels that it feels companies could achieve, and companies would need to provide detailed justifications wherever the CBA suggested they could not achieve those levels. This should re-frame the way that companies carry out CBA, and should lead to more stretching targets. However, CBA would not be required in all cases, and companies could use other approaches for some measures. In particular, companies could base asset health targets on engineering models and statutory PCs should be set according to statutory obligations (CBA could be used to justify performance above the statutory obligations). We provide more detail on the measure-specific approach in Annex B.

The main feature of the ODI methodology is the innovation uplift. The rationale for it is to reflect the externality benefit of innovation, where one company's innovation can spill over and lead to improvements across the whole industry. Since small improvements in performance are unlikely to be due to genuine innovation, these uplifts would only be applied for "exceptional" performance, with reward rates remaining moderate (i.e. PR14 levels of rewards) for smaller improvements in performance. While the rewards would be harder to achieve, the per unit rate would be significantly larger than in the other packages, such that companies would be incentivised to innovate, rather than focusing on gradual improvement. We note that it may be challenging for Ofwat to assess what level of performance is genuinely innovative, and therefore when companies should be allowed to earn these innovation uplifts.

For PCs where Ofwat sets common commitment levels, Ofwat would set common ODIs, with an innovation uplift. There are a number of ways in which the innovation uplifts could be calculated. One option would be to set the uplift with reference to the industry-wide customer valuation for the innovative performance. However the full valuation would represent an absolute upper limit, as all of the benefit of the innovation would be received by the company. An alternative option would be to use regulatory judgement to set the innovation uplifts, which could be informed by an analysis of the potential benefits of the innovation in future (i.e. how the innovation may change the way that benchmarks are set in future). One element of the regulatory judgement would be whether the innovation uplift should be set as a fixed amount (i.e. to reflect an estimate of the benefit) or set as a fixed multiple of the common ODI (i.e. so that the financial impact of the innovation across companies is similar relative to the scale of company). Where companies could justify that there could be positive externalities from innovation on bespoke PCs, they could also apply innovation uplifts to the ODIs on bespoke PCs. These ODIs should also reflect the benefits that all customers would receive from one company innovating on that bespoke PC, recognising that the positive externalities may be lower than on common PCs.

Potential changes to reduce the risk of unintended consequences

The following elements could be modified to reduce the risk of unintended consequences:

- remove innovation uplifts on bespoke PCs;
- reduce the scale of the RBR incentive; and/or
- add a system of gates to the innovation uplifts.

5.3 Package 3 – Focus on tailored rewards

The focus on this package is increasing the level of stretch in the PCs and increasing the power of the outcomes framework. This is therefore most in-line with the outcomes consultation. It should bring about benefits to customers by increasing the level of service that companies commit to delivering, and providing stronger incentives for companies to improve performance (in a proportionate way).

High level instruments

In addition to the high level instruments included in our general improvements, this package includes a RBR incentive and a system of gated ODIs. The RBR incentive would reward companies that set challenging targets and provide robust evidence in the business plan.

Gated ODIs would set minimum standards for a subset of measures, meaning that companies would only receive rewards if they meet the gates on all PCs. This would mean that companies would not achieve high rewards on some PCs, while performing very poorly on another measure. This would therefore encourage companies to focus more on overall performance.

Detailed methodology

The main elements of the methodology are:

- a measure-specific approach to setting commitment levels for bespoke PCs;
- company-specific ODIs for PCs with common commitment levels; and
- possible uplifts to ODI rates.

As in package 2, a measure-specific approach would be used to set bespoke PCs. More detailed on this approach is provided in Annex B.

This package differs from packages 1 and 2 as it includes company-specific ODIs for all measures, including the ones with common commitment levels. While matching common commitment levels with common ODIs may be more consistent, it makes it harder to set appropriate incentives where companies have different cost structures and customer valuations are not consistent.

Incentive rates would be based on the PR14 formula but depending on the overall RoRE range set by Ofwat, companies may be expected to apply uplifts to rewards. These would be more moderate uplifts than the innovation uplifts and would apply at all performance levels until the cap. Ofwat could set the expectation that larger uplifts should be applied to PCs where the evidence base for setting the target is strongest. These tailored rewards would increase Ofwat's confidence that power had been applied in a proportionate way.

Potential changes to reduce the risk of unintended consequences

The following elements could be modified to reduce the risk of unintended consequences:

- set minimum standards on more measures within the gated ODI system;
- make the level of the gates more challenging; and/or
- remove or reduce uplifts for some or all measures.

5.4 Package 4 – Focus on integrated assessment

The focus on this package is developing an integrated framework, where costs and outcomes could be assessed together. It would be based on a range of regulatory tools, using multiple benchmarking techniques to set allowances.

High level instruments

The high level instrument included in this package is a menu incentive for costs and outcomes. Companies would be required to submit plans on costs and outcomes, and a single menu would provide incentives for truthful revelation of information. The principles would be the same as those of the menu incentive that is currently used, but further research by Ofwat would be required to adapt the mechanism to deal with costs and outcomes simultaneously.

Detailed methodology

The main elements of the methodology are:

- a combination of benchmarking techniques to assess costs and outcomes together;

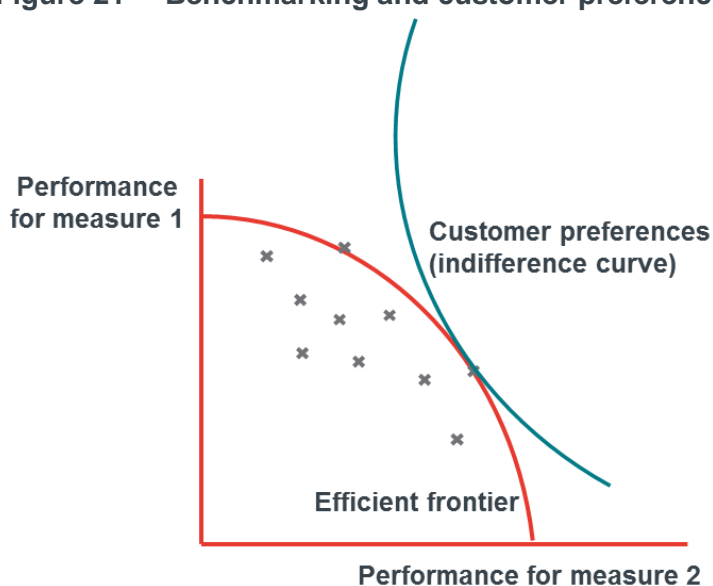
- only common PCs;
- incentive rates set to reflect preferences over combinations of measures; and
- innovation uplifts on PCs with clear externality benefits.

A combination of benchmarking techniques would be used to assess cost and service quality simultaneously. These techniques could include Data Envelopment Analysis, Stochastic Frontier Analysis, and other econometric approaches. These models would estimate the efficient frontier: the combinations of service performance across all PCs that could be delivered, for a given cost allowance.

In practice this benchmarking exercise would have to be carried out by Ofwat, and would be based only on common PCs. Triangulation of the benchmarking results would help to ensure the framework is robust, providing a richer set of information and avoiding over-reliance on a single set of benchmarking results.

Companies would choose a point on the efficient frontier, according to their customers' preferences. Therefore, while the efficient frontier is common to all companies, the outcomes targets will not generally be (unless companies' customers have identical preferences). Companies would be able to trade-off different aspects of service quality, settling for lower targets on measures that are less important to their customers, and aiming for higher performance on service measures that matter particularly customers. Companies would still have to carry out customer engagement and would be required to map out customer preferences over *combinations* of measures, rather than single measures, reflecting substitution and complementarity effects. CCGs could be involved in challenging companies on how they carry out customer engagement, and how they use that research to determine their outcomes targets. Figure 21 below illustrates how customer preferences would be used to determine the company's place on the efficient frontier.

Figure 21 Benchmarking and customer preferences



Source: Frontier Economics

Companies would be expected to set incentive rates based on performance on multiple PCs, rather than just one PC. This means that each incentive rate would reflect customers' preferences across a combination of PCs. Finally, companies would be incentivised to innovate by applying larger uplifts to reward rates for exceptional levels of performance.

Potential changes to reduce the risk of unintended consequences

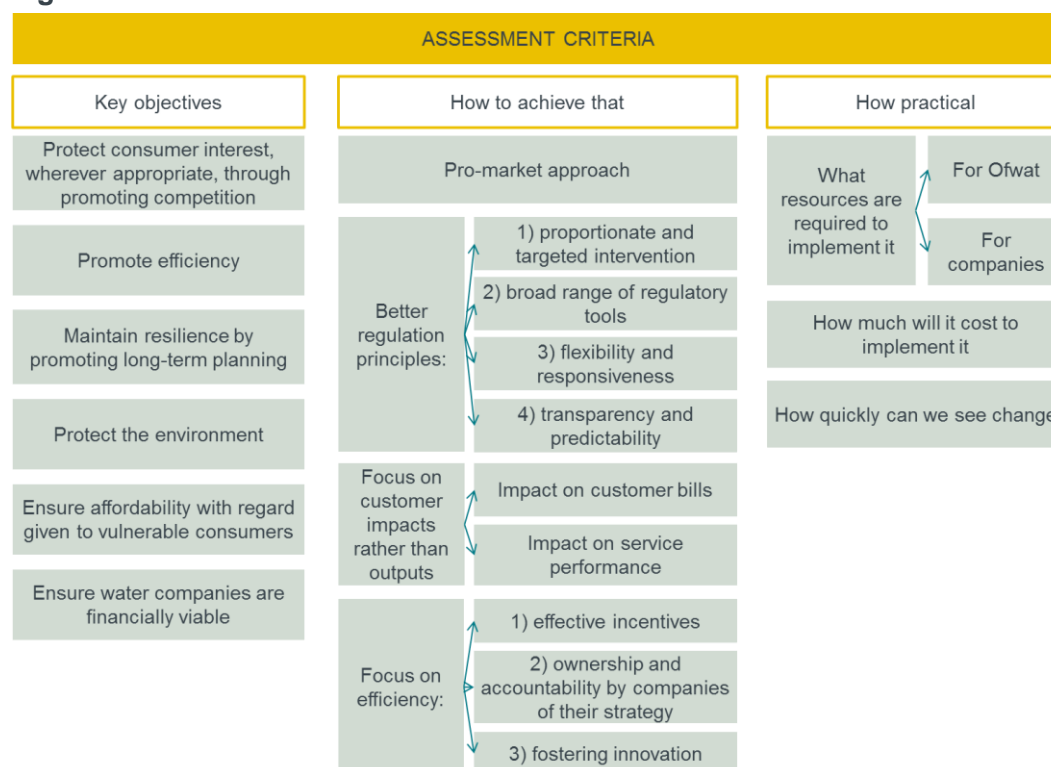
The following elements could be modified to reduce the risk of unintended consequences:

- add gated ODIs; and/or
- reduce or remove the innovation uplifts.

6 ASSESSMENT OF OPTIONS

We applied Ofwat’s assessment framework, summarised in Figure 22 below.

Figure 22 Ofwat’s assessment framework



Source: Ofwat appraisal framework

6.1 Assessment against criteria

Figure 23 below summarises how the different packages perform against the assessment criteria. Annex C provides a more detailed discussion of the rationale for each of the ratings. Note that the scoring is relative to the base case (PR14 framework) so that the traffic light scores can be interpreted as:

- Red – no significant improvement compared to PR14, or worse than PR14;
- Amber – moderate improvement compared to PR14; and
- Green – large improvement compared to PR14.

Figure 23 Summary of the options assessment

	Package 1 – Focus on PR14 improvement	Package 2 – Focus on innovation	Package 3 – Focus on tailored rewards	Package 4 – Focus on integrated assessment
Protect customer interest				
Promote efficiency				
Maintain resilience	n/a	n/a	n/a	n/a
Protect the environment	n/a	n/a	n/a	n/a
Ensure affordability	n/a	n/a	n/a	n/a
Ensure financial viability	n/a	n/a	n/a	n/a
Pro-market approach	n/a	n/a	n/a	n/a
Better regulation principles:				
1) Proportionate and targeted				
2) Broader range of regulatory tools				
3) Flexibility and responsiveness				
4) Transparency and predictability				
Focus on customer impacts:				
1) Impact on customer bills				
2) Impact on service performance				
Focus on efficiency:				
1) Effective incentives				
2) Ownership and accountability				
3) Fostering innovation				
What resources are required to implement it:				
1) For Ofwat				
2) For companies				
How much will it cost	n/a	n/a	n/a	n/a
How quickly will we see change				

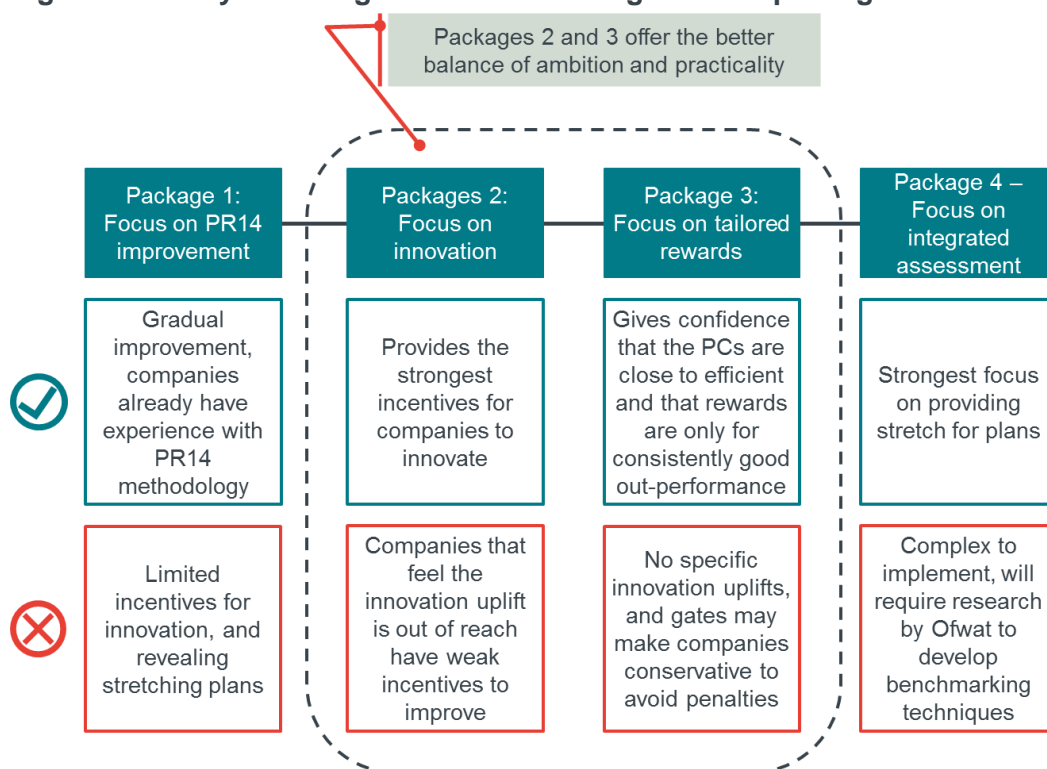
The packages have not been scored against some criteria that are not directly applicable to the outcomes framework. More detail on why these criteria are not scored is provided in Annex C.

6.2 Ofwat's key choices for PR19

The assessment above demonstrates that there are important trade-offs between the packages and there is no package that delivers every possible objective as there are trade-offs between the different objectives.

The following figure summarises the key advantages and disadvantages of each package:

Figure 24 Key advantages and disadvantages of the packages



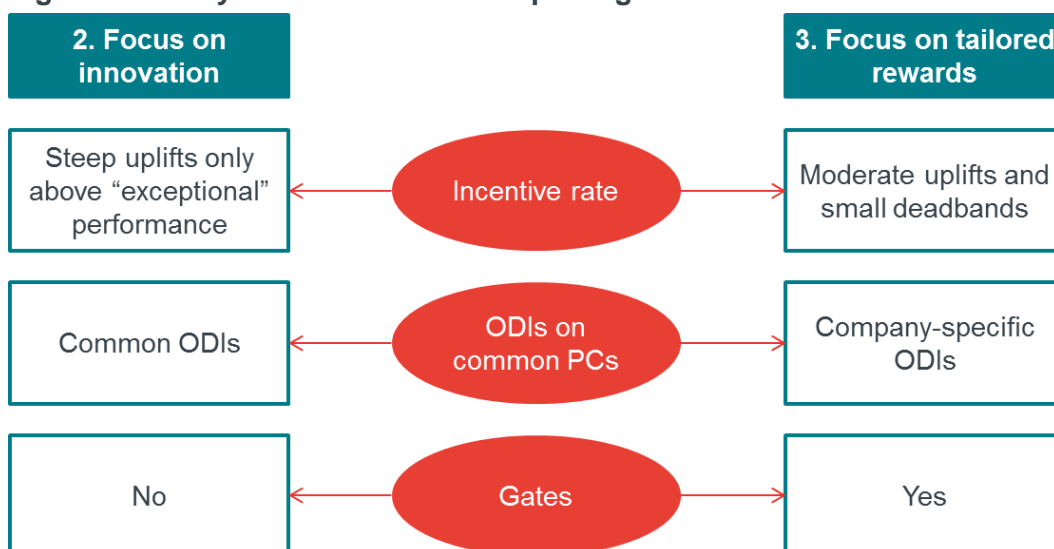
Package 1 is a conservative package. It would build on PR14 by addressing the most immediate issues with the methodology and companies would have clearer expectations of how to apply CBA. However, it is unlikely to achieve Ofwat's goal of more stretching PCs. In contrast, package 4 is highly ambitious and would use advanced benchmarking techniques to set the most stretching targets for companies. However, it is an entirely new framework and would require substantial research on integrated benchmarking of costs and outcomes.

Our assessment suggests that packages 2 and 3 are the better options for PR19, as they offer a good balance between ambition and practicality. In the remainder of this section we provide a more detailed assessment of the pros and cons of these two packages, in order to highlight the key differences.

Key differences between packages 2 and 3

The following diagram highlights the key differences between packages 2 (Focus on innovation) and 3 (Focus on tailored rewards).

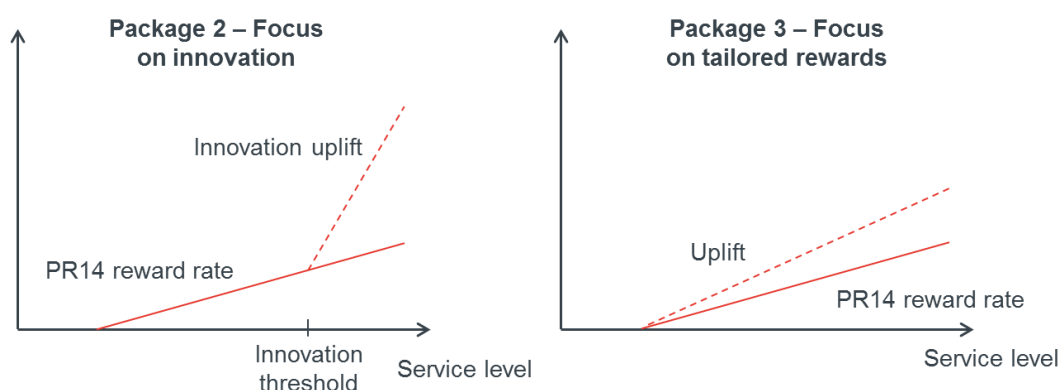
Figure 25 Key differences between packages 2 and 3



Both packages rely on the same approach to setting the PCs: companies should use a measure-specific approach that complements CBA with other approaches where these would add robustness and challenge. The differences between the two packages therefore lie in how the ODIs are set.

- Package 2 is focused on delivering innovation.
 - The reward rates would be relatively small (similar to PR14 level or, at the extreme, set at zero) for small improvements in performance, but there would be material rewards for truly innovative performance.
 - Common PCs with common commitment levels would have common ODIs. These common ODIs would incorporate innovation uplifts, to reflect the positive externality of innovation.
 - No gates would be required, because the larger rewards would only be awarded to companies that brought true innovation to the sector.
- Package 3 is focussed on delivering improvement across the board with tailored rewards.
 - Moderate uplifts would be applied to increase the incentives for companies to improve, relative to PR14, and these rewards would be attainable for companies that made any significant improvement (that is, the deadband would have the purpose of avoiding rewards for small exogenous variation in performance) beyond stretching PC levels.
 - The approach would be fully tailored, and therefore the ODIs for common PCs with common commitment levels would be specific to each company, given its costs and customer preferences.
 - Gates would be used to ensure that companies did not chase rewards at the cost of reducing performance in some areas.

The following diagram illustrates the structure of rewards under each of these packages.

Figure 26 Structure of rewards under packages 2 and 3

How would the packages affect company behaviour?

Company behaviour is likely to depend on how well the package is implemented:

- Package 2 (Focus on Innovation) would be successful if it set the innovation threshold at a level of performance that could only be reached if the company found an innovative solution.
 - Companies would have incentives to chase innovation rewards by researching new technologies and management methods.
 - It is likely that only a few companies, if that, would succeed in innovating and pushing out the efficient frontier.
 - Some companies may not chase rewards at all, especially if they feel the innovative level of performance is too difficult for them within a five-year period, which is more likely to be the case if the PC levels themselves are truly stretching.

But the crucial implementation question is where to set the threshold for innovation uplifts. If these are set too low lots of companies may obtain innovation rewards which could undermine confidence. If these are set too high, there may not be any impact on industry behaviour.

There is also a question as to how to set the innovation incentive. One option would be to set the innovation reward as the industry-wide customer valuation. However, this would be the maximum possible level that could be used, and it may be challenging to set in some circumstances if there is uncertainty around the data. It may be considered that industry-wide customer valuation is too high, as multiple companies could potentially receive this reward. Although there are multiple options for how to set the reward rate, we consider that, if anything, this is a less challenging issue than where the innovation threshold should be set. At the same time there is an interaction between the innovation threshold and the magnitude of the uplift. For example, if the innovation threshold is too challenging, setting the incentive rate too high could mitigate the impact of setting the innovation threshold too high. But equally, if the innovation threshold is set too low, and the innovation uplift is set at a high level, this would increase the impact of setting the innovation threshold too low.

- Package 3 (Focus on tailored rewards) would be successful if the PCs were set at a challenging level that companies could achieve, but not easily exceed.
 - Companies would have stronger incentives to improve, relative to PR14, but would not have as strong incentives to deliver an innovative frontier-shift performance.
 - Companies may spread their efforts across PCs, rather than focusing efforts in one area.
 - More companies are likely to deliver gradual levels of improvement than under package 2, but fewer companies are likely to deliver substantial frontier-shifting levels of improvement.

The differing behaviours across the two packages may affect the way that the industry as a whole develops. Under package 2, innovation may be led by a small set of companies, and would only affect the improvements required from other companies at price control reviews, when the improvements made by the small set of innovative companies are used to set more challenging targets for the whole industry. Whereas under package 3, companies may develop at a more similar rate to each other, as it is unlikely that any companies would “give up” in delivering service improvements and also unlikely that any companies would deliver genuine innovation.

What are the biggest risks of each package?

Our discussion so far has focused on what these packages could achieve if they were implemented successfully, but both packages also present risks.

The main risks we have identified for package 2 (Focus on innovation) are the following.

- The innovation threshold is too challenging – Because it will be necessary to set the innovation threshold at a high level of performance to drive companies to innovate, there is a risk that the threshold is set so high as to fail to attract any companies into chasing it. If all companies considered the innovation uplift was infeasible, or too unlikely to achieve in five years, they would not invest in new solutions and technology. The result would be very similar to the status quo since the rewards at lower levels of improvement would be the same as in PR14.
- Ofwat underestimates the innovation threshold – this would result in many companies getting material rewards. This risk could be partially mitigated by limiting the rewards to only the first two or three companies to meet the innovation threshold. However, if this limit is applied, and the threshold is actually set too high, this would only exacerbate the first risk we identified.

There are similar issues with setting the uplift rate, as the incentive rate could over or under-stimulate innovation efforts. However, as explained above, we expect this incentive rate to be large in any case and company behaviour is likely to be more sensitive to the threshold which creates very sharp incentives at a particular performance level. As noted above, this issue is likely to relate to the risks around the setting of the innovation threshold. The main risk we have identified for package 3 (Focus on tailored rewards) is the following.

- PC levels are set too low – due to the asymmetry of information between Ofwat and companies, it is possible that companies set PC levels that are not sufficiently challenging. If this were the case, companies would receive even higher rewards for delivering improvements in service performance, even if they were not challenging to achieve. We note that the uplifts themselves present an incentive for companies to set less stretching PC levels, such as the benefit to companies from doing so has increased. Given this risk, it is of particular importance that uplifts are only applied in cases where there is more confidence in the level of stretch in outcomes targets.

Making the choice between packages 2 and 3

The choice between packages 2 and 3 depends on the policy direction that Ofwat would like to take: Package 2 is focused on innovation; while package 3 should deliver more tailored, gradual improvements in performance. It also depends on the risks that Ofwat is most concerned about, as each option presents different risks. The following table provides a guide for thinking through the factors that will determine the conditions under which each package is more appropriate.

Figure 27 Key choice between packages 2 and 3

Package 2 (focus on innovation) is more appropriate when:	Package 3 (focus on tailored rewards) is more appropriate when:
<ul style="list-style-type: none"> ■ The objective is to shift the efficient frontier through innovation 	<ul style="list-style-type: none"> ■ The objective is to improve performance across the board
<ul style="list-style-type: none"> ■ There are concerns about the ability to test the level of stretch in commitment levels, but relative comfort in assessing the level at which companies would receive innovative rewards 	<ul style="list-style-type: none"> ■ There is comfort in being able to test the level of stretch in PC targets
<ul style="list-style-type: none"> ■ It is felt the level of PR14 rewards are appropriate in general, but would not incentivise genuine innovation / reward for the risks involved in innovation 	<ul style="list-style-type: none"> ■ It is felt that the level of rewards at PR14 are not sufficient to incentivise gradual improvement beyond PC levels
<ul style="list-style-type: none"> ■ Increasing the ODI proportion of the overall return is not a priority so there is no need for an approach that is consistent with that 	<ul style="list-style-type: none"> ■ Ofwat would like to increase the ODI proportion of the overall return
<ul style="list-style-type: none"> ■ Ofwat is comfortable with multiple companies delivering some improvement but only a few companies delivering innovative levels of performance 	<ul style="list-style-type: none"> ■ Ofwat is comfortable with multiple companies delivering gradual improvement beyond stretching PC levels, but not necessarily innovative levels of performance
<ul style="list-style-type: none"> ■ Ofwat is comfortable with some companies receiving material rewards for innovative performance 	<ul style="list-style-type: none"> ■ Ofwat is comfortable with multiple companies receiving rewards greater than at PR14 for gradual improvement beyond stretching PC levels
<ul style="list-style-type: none"> ■ Ofwat considers companies are likely to respond to innovation rewards 	<ul style="list-style-type: none"> ■ Ofwat considers firms likely to respond to more gradual rewards

Source: *Frontier Economics*

Finally, we note that Ofwat could choose to implement one of packages 2 or 3, with some modifications. This is because although we have defined coherent packages, there are still options within the packages and changes could be made to one or more of the detailed aspects of the packages, without losing the policy focus of the package. For example, Ofwat could choose to implement package 2 with company-specific ODIs, or it could choose to implement package 3 without gated ODIs. In addition, the package design also depends on the way it is implemented so the detailed approach may require some modifications. For example, if package 3 was implemented with relatively small uplifts, this could reduce the need for gated ODIs.

ANNEX A DATA ENVELOPMENT ANALYSIS

One of the modelling approaches that could be used for an integrated assessment of costs and outcomes is Data Envelopment Analysis (DEA).

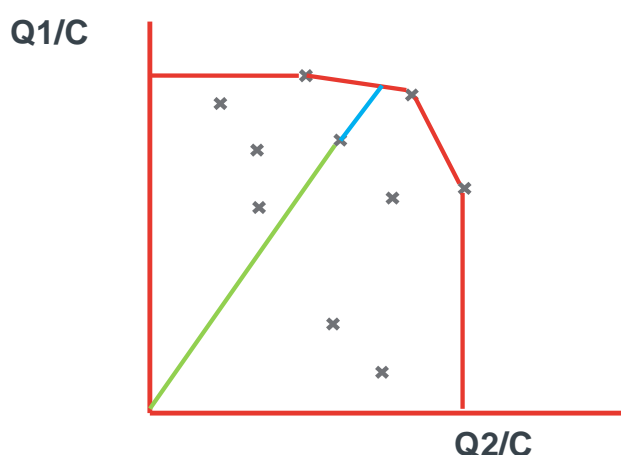


What is DEA?

DEA is a form of ratio analysis that can take account of multiple dimensions, and explicitly reflects the mix of different dimensions that are being offered. When assessing the efficiency of a given offering, the point of comparison is a specific set of companies, as opposed to the average offering (e.g. in regression analysis).

When setting up DEA, costs need to be defined and also outputs need to be defined. One of these two elements needs to be fixed and the other can be determined. Typically, it is assumed that outputs are fixed and the cost is the element to be determined. Ratios should then be calculated between cost and each output. The diagram below illustrates this for two outputs, Q1 and Q2.

The first step is to determine which company has the highest ratio (amount of output per unit of cost) for Q1, and then which company has the highest ratio for Q2. The second step is to identify the companies that have delivered a different mix of Q1 and Q2, such that they do not have the highest ratio, but still have presented an offering that is efficient across both outputs. In this case, there is a third company that has provided such an offering. The final step is to calculate efficiency scores. These are calculated relative to the origin, and relative to a notional company that has the same mix of outputs, but would be on the efficient frontier. The diagram below shows this calculation as the green and blue lines. The inefficiency would be calculated as the blue line divided by the sum of the blue and green line.



There are however a number of implementation issues that would need to be considered if DEA was to be used.

- When DEA is used in the energy industry (for example in Germany and Austria), the outputs are defined as measures such as peak demand and number of customers, that are generally considered to be outside of the companies' control. As a result, the outputs are defined as fixed elements of service that companies have to deliver, and the cost is the item which is considered to be determined using the DEA. However, in the water sector, the framework would be aiming to identify the efficient mix of costs and outcome targets, meaning that both costs and outcomes are controllable. This is a technical issue that would require further research. For example, it may be possible to use alternative techniques such as regression analysis to assess cost levels, and then to define costs as fixed in the DEA assessment of outcomes.
- The more measures that are included in the framework, the more likely it is that any given company has the best ratio for at least one measure. This means that the more measures that are included, the more likely it is that a company is classified as being on the frontier. This would mean that there would be a restriction on how many measures could feasibly be included in the framework. It also means that, if taken to the extreme, if all companies are the best at something, all companies would be classified as efficient under this framework (unless there was a way to assess cost and outcomes directly together in the DEA).
- The following general issues with DEA would need to be considered.
 - Missing data is only beneficial to companies, unlike regression analysis. If there is a missing data point that is within the frontier, including the data point would not affect the efficiency performance of companies. If there is a missing data point that is outside the frontier, then including this would shift the frontier out, and would make some companies appear less efficient.
 - A decision would need to be made about what assumption is made on returns to scale.

CASE STUDY: USE OF DEA IN ENERGY SECTOR IN GERMANY

In Germany, the energy regulator (BNetzA) caps the revenue of the electricity and gas distribution network operators according to a RPI-X adjustment factor. This factor reflects revenue changes due to inflation and (required) improvements in efficiency over time. The X-factor contains a general and an individual firm factor. The individual X-factor reflects how much the company would reduce its totex if it were to become “fully efficient”. It is estimated using DEA and Stochastic Frontier Analysis (SFA). The two methods are applied to two cost bases (book value and standardised capital costs) and then the regulator applies a “best of four” principle. This means that it selects the efficiency score from the four estimates available that is most favourable to the company (with a minimum efficiency factor set at 60%).

The German Ordinance is very explicit about the technical details for how to apply the benchmarking, for example, specifying a minimum set of cost drivers to be included, which estimation techniques must be used, what assumptions to make about returns to scale and how to treat outliers (and the criteria to exclude them). Despite that, there is still discretion for the regulators in how they specify the models, for example, they may choose to include additional cost drivers beyond the minimum set required by the guidelines. The different choices often face trade-offs and there is not a clear “better” option. This creates room for the operators to challenge the methodology and reduces its acceptability in the industry.

In practice, applying DEA has been challenging in Germany. The lack of transparency from the benchmarking and sensitivity of the models, for example to parameter choice and the presence of outliers, has led to many operators taking legal steps to challenge the regulator.

We provide below a summary of the practical issues that have arisen in the regulatory framework in Germany, and how they might be affected if this approach was applied to the Ofwat outcomes framework.

- It is a data intensive exercise. If service quality measures like those in the Ofwat outcomes framework were to be included in this exercise, it may become even more challenging. There would also be a potential collinearity problem between the fixed outputs and outcome measures.
- The current approach seems relatively sensitive to the assumptions that are made. Introducing service quality outcomes may add to this volatility. Note that combining DEA and SFA approaches is one way in which the German regulator has reduced the sensitivity of results to data and model specification issues.
- The model is not particularly transparent and is therefore more likely to result in legal challenge from the regulated companies than simpler methodologies. This risk may be exacerbated if the framework included service level / quality measures which companies considered were not comparable.

Given these challenges with DEA, the new framework may need to draw on range of different techniques, and combine the results from that analysis to make its overall conclusion.

ANNEX B IMPLEMENTING A MEASURE-SPECIFIC APPROACH

As noted in section 2, we identified some inconsistencies in the way that data was presented for some measures in the PR14 data tables. This may suggest that companies experienced particular challenges when setting PCs and ODIs for some measures. There is an important choice for Ofwat in whether it expects companies to use the same approach to set PCs and ODIs for all measures, or whether it designs a measure-specific approach.

In this annex we explore how a measure-specific approach would work in practice. We consider how the methodology would be adjusted for each category of measures. The parts of the methodology that would vary by measure include:

- how PCs could be set; in line with the discussion above, we use CBA as a starting point and consider alternative approaches where required;
- The type of expectations on how challenging PCs should be (e.g. at least at the upper quartile, or at least as much improvement in PR14), or whether Ofwat should expect companies to justify their PC levels in the context of additional information (e.g. historical information, comparative information);
- The structure of the ODIs; and
- The method for determining incentive rates.

This section is split by the following categories of measures:

- Common PCs – we consider how to set ODIs only (as the commitment level for common PCs is outside the scope of this project);
- Measures for which there is comparative information (all or several companies set PCs for these measures, or comparative information is available despite no other company setting a PC) but the commitment level is company-specific;
- statutory measures that do not fall within common PCs;
- bespoke (additional) asset health measures;
- bespoke (additional) resilience measures; and
- fully bespoke measures.

Having too many different categories each with their own methodology could result in confusion among companies as to what they should do. We have therefore tried to identify the smallest set of categories that will capture the most crucial differences in their characteristics and necessarily require a measure-specific approach.

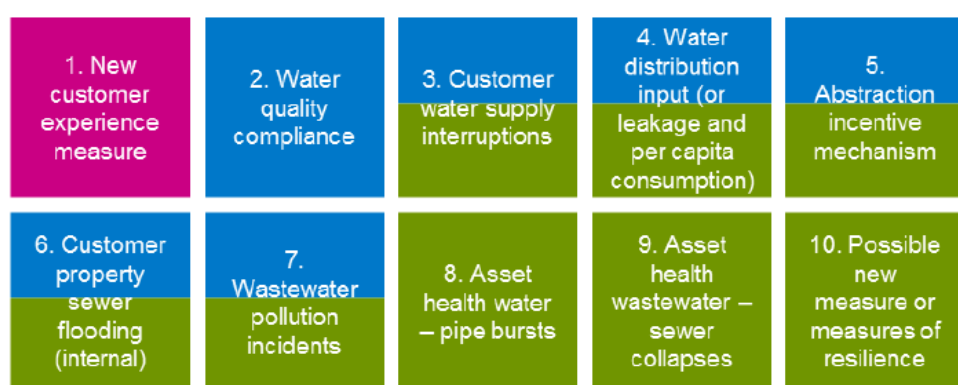
We consider each of these categories of measures in the remainder of this section and provide our recommendations for the most suitably methodology for each.

B.1 Common measures: measures where Ofwat is likely to set PCs (common PCs)

These are measures for which Ofwat is likely to set common targets at PR19. The only element of the methodology that is within the scope of this project is how ODIs are set.

The measures that Ofwat listed as likely candidate for a common PC in the recent consultation were:

Figure 28 Ofwat’s proposed measures for common PCs in PR19



Area of focus

■ Customer satisfaction ■ Current performance ■ Resilience/future performance

Source: Ofwat (2016), *A consultation on the outcomes framework for PR19*, Figure 6

These measures can be described as:

- **High confidence in PC levels.** Since will be measured by all companies, and in many cases were already measured at PR14, comparative data is available and can be used to ensure that PCs are stretching, although Ofwat has recognised that common commitment levels might only be appropriate for six of the measures above (water quality compliance, customer water supply interruptions, sewer flooding, pollution incidents, bursts and collapses). As a result, the level of confidence in the PC level is likely to be relatively high, which means that there is a relatively low risk of increasing the power of incentives. This is because, if a company outperforms its PC level, it is likely to have had to challenge itself to find more cost effective ways of delivering that improved performance level.
- **Strong rationale for externalities.** As these are measured by all companies, there is strong potential that positive externalities may result from one company innovating, and spreading best practice across the industry.

These two features suggest that, if Ofwat would like to encourage innovation and increase the potential upside in the outcomes framework, putting more power into the rewards for these measures could be based on a clear rationale.

As a result, ODIs for these measures could be based on:

■ **Option 1 – Common two-sided ODIs (same incentive rate per unit of improvement)**

This option has the following feature:

- Common incentive rates would be set for all companies. This would mean that ODIs would not directly reflect companies' costs, or their customers' valuations of service improvement. In this option, the structure of ODIs would be flat, such that the same incentive rate would be used for each unit of improvement (or decline).
 - Penalties are based on the highest marginal cost. Setting penalties according to marginal cost only is simpler than the PR14 formula, as it doesn't require information on customer valuations. Setting the penalty at the highest marginal cost ensures that all companies have an incentive to meet the PC.
 - Rewards are based on the average customer valuation across the industry (with a possible uplift). Ofwat would need to calculate average customer valuations for each measure. This could be based on an industry-wide survey conducted by Ofwat itself or it could be derived from the company-specific values obtained in companies' own surveys.
 - Uplifts could be applied to the average customer valuation to create stronger incentives for companies to innovate in service delivery.
 - Deadbands are applied. This would reduce the impact of relatively small fluctuations in performance. Deadbands could be set closer to the PC level on measures that are more within companies' control.
- **Option 2 – Common two-sided ODIs with a discontinuous prize for innovation**

As in option 1, option 2 would have common, two-sided ODIs. Penalties would be set in the same way, but the way that rewards are set would be different. In this option, the reward rate would be low or zero at performance levels just above the PC. The structure of rewards would be discontinuous, such that at some level of performance which is considered to be truly innovative, the company would receive an innovation prize (i.e. the reward rate would be significantly higher for these innovative levels of performance).

■ **Option 3 – Company-specific two-sided ODIs (with uplifts)**

This option would have the following features:

- ODIs are set by each company, based on costs and customer valuations.
- Penalties could be based on PR14 formula or marginal cost. This approach could be run with either the PR14 formula for penalties, or with penalty rates set using only marginal cost. Both of these two options create the incentive for companies to meet the PC, however using only marginal cost data could be easier to implement.
- **Rewards based on PR14 formula with uplifts.** Rewards would be set based on the PR14 formula (i.e. both customer valuations and marginal costs). The scale of the rewards could be uplifted, for instance to fit within

a RoRE range. This uplift could be applied using a common scaling factor for all measures within this category, such that rewards still reflect customers' relative preferences.

This approach may lead to a higher probability of moderate rewards than option 2, as the uplifts apply to rewards at all levels of outperformance, not just once the company's performance goes above some innovation threshold. This approach may therefore incentivise companies to aim for more moderate improvement across the board, rather than focusing their efforts on genuine innovation in one area of service.

B.2 Common measures: measures where companies set PCs (other common measures)

There are some measures for which several companies measure performance, but do not have common commitment levels (e.g. size of the carbon footprint, river quality improved or properties at risk of persistent low pressure). This comparative information exists, could be used to challenge and inform what levels of performance are possible for companies. In addition to comparative information, each company may have historical information on these measures. This historical information could also be used to put companies' proposed PCs into context.

For these measures, we have identified the following approach:

- **PCs:** CBA would provide a good starting point to set PCs. However, Ofwat could require or expect companies to set PCs that achieve certain levels of improvement. This additional information would frame the way that PCs are set, and should result in more stretching PCs as it changes would require companies to justify cases where they were setting less ambitious PC levels than other companies. Ofwat could, for example, set the following expectations on how PCs should be set for these measures.
 - Ofwat could change the reference point to the upper quartile level of performance from PR14;
 - Ofwat could define a minimum improvement, based on PR14 historical information.
- **ODIs:** these could be two-sided financial incentives wherever the evidence base for setting the PC was sufficiently strong (given the availability of comparative information).
 - Penalties could be based on the PR14 formula, or only on marginal cost. The rationale for including customer valuations in the penalty formula could depend on how directly customers are affected by the measure, how important customers generally find this measure, and/or the quality of the available data on customer valuations.
 - Rewards could be based on the PR14 formula. Ofwat could allow companies to apply uplifts to these reward rates where companies are able to provide sufficient justification that their PCs are challenging.

- Deadbands should be based on historical information wherever possible to ensure that small fluctuations in performance that occur fairly regularly (and are unlikely to be driven by significant changes to technology or management practices) do not result in financial penalties or rewards.

B.3 Statutory measures (that are not within common PCs)

These are measures that are monitored by government departments or other regulators, for example the Environment Agency. Some statutory measures (e.g. drinking water quality) are included in the set of common PCs. Companies often choose PCs for other statutory measures where this reflects their customers' preferences.

These other statutory measures often reflect aspects of service quality that tend to have more public exposure than some other measures within the outcomes framework, and as such companies are likely to face strong reputational incentives to meet any statutory obligations. In addition, companies may face fines if they fail to meet any statutory obligations on these measures.

We have developed the following options:

- **PCs:** Where statutory obligations are set by an external agency, the PC level should be no lower than this statutory requirement. However, a company may choose to set the PC at a more stretching level (where the statutory target is not at 100%). If a company chooses to do this, it would need to use CBA to justify the higher PC. Ofwat could set the expectation that companies consider the following.
 - Companies could first identify the maximum possible level of performance, and work backwards from this point, rather than starting with the statutory requirement and work from there (assuming that the statutory requirement is not already the maximum possible level of performance).
 - For measures that have been monitored for a long period of time, Ofwat could use historical information to set expectations on what it considers would be a reasonable level of performance improvement. If this would take companies beyond the statutory minimum, companies could use this tougher PC level as the baseline when considering its CBA.
- **ODIs:** If Ofwat considers that the potential fines are unlikely to reflect the costs that companies would save if they failed to meet targets, it could set the expectation that companies also set penalties as part of the outcomes framework. The scale of penalties should reflect any fines that companies would pay, if they failed to meet the statutory target, to avoid any double counting of penalties/fines. The scale of any financial penalties should also reflect the significant reputational impact of any poor performance. Any financial penalties should therefore be based on avoided costs, minus the value of any fines; but we would not necessarily expect all companies to set financial penalties, due to the existence of fines and the fact that reputational incentives are very powerful on these PCs.

If companies deliver improved performance above the PC, it seems reasonable for rewards to be paid. These rewards could be based on the PR14 formula.

B.4 Additional asset health measures (that are not within common PCs)

These are indicators of long-term asset health that are not part of the common PCs (i.e. mains bursts and sewer collapses fall outside of this category). For example, at PR14 most companies had a PC on the RAG asset health indicator for infrastructure. An example of a more bespoke measure was the population in centres >25,000 at risk from asset failure.

These measure intermediate indicators of likely service levels. They are not true outcomes, but neither are they inputs. While customers benefit indirectly from good asset management (i.e. through service quality that is captured in other measures), they are likely to be unfamiliar with the measures, as the measures do not reflect service elements that customers experience directly. Additionally, the impact of poor asset health may not be realised for a long time and customers may have myopic preferences. It is therefore challenging to get meaningful customer valuations for improvements in asset health.

We have developed the following options:

- **PCs:** Commitment levels could be informed by engineering and asset management expertise. The levels should be based on what companies want to achieve on other measures over the longer term (i.e. could be based on other longer term plans, and indicative plans for PCs over the longer term) and what that implies for companies' required asset health (i.e. what health a company needs to deliver those long term outcomes in the least cost way). We understand that companies have systems in place to set asset health targets, and we feel it makes sense for companies to use those, rather than being required to use CBA. Therefore, companies would not necessarily be expected to use CBA to set PCs, or base PCs directly on customer valuations. Commitment levels could also be informed by historical and comparative information where these are available.
- **ODIs:** Generally these will not be two-sided incentives.
 - Penalties should be based on costs. They should be set in a way that reduces the risk that companies focus on short term outcomes and do not invest in asset health, i.e. reflect at least the marginal cost of not investing in asset health. Given there may be uncertainty over incremental costs, it may be preferable to set an expectation that companies should err on the side of caution, providing strong incentives to meet the PC.
 - Rewards may not be appropriate. If companies found cost-saving ways to deliver improvements in asset health, they would still be incentivised through the cost sharing mechanism to use those new methods. Companies would also indirectly be incentivised by the rewards that are applied to measures in other areas, as improved asset health may increase the chance that companies are able to improve their performance

on other measures. As a result, additional rewards on asset health may have unintended consequences. The exceptions to this would be where an asset health PC is a customer priority e.g. reducing low pressure or where a company is seeking to deliver a step change in the level of asset health and could earn rewards for delivering the change earlier than planned for in its PC level.

B.5 Additional resilience measures (that are not within common PCs)

As identified in section 2, resilience is now one of Ofwat's primary duties and is one of the 4 key policy areas for PR19. Ofwat has stated that it will introduce a resilience measure at PR19, and that it is likely to be a common measure. Companies are also likely to set PCs for other resilience measures, which is the focus of this section (i.e. the common resilience measure falls within common PCs). It is possible that a common PC level is not appropriate for some of these more bespoke resilience measures, due to the differing nature of the operating environments that companies face or legacy network issues. Further, the approach to monitoring resilience is still evolving and there is likely to be a benefit to encouraging companies to develop new resilience measures.

Resilience measures assess companies' abilities to withstand external shocks. For example, how a company's assets could withstand a flood, a drought or unexpected population growth. A specific example is the proportion of households that are supplied by one source. This is a resilience measure, as those households would be particularly affected if there are any issues with supply sources, as there is no alternative supply source for them.

We have developed the following options:

- **PCs:** CBA would be used to set PC. Since customers struggle to place value on low probability / high impact events, CBA should be adapted so that WTP was only measured for realised events (the willingness to pay to avoid a two day interruption, rather than to reduce the probability of a two day interruption by 0.1%).

Some resilience measures may be common to several companies, and Ofwat could set expectations on how to use comparative information to add challenge to the PCs. Comparative information may not be available for all these measures, or it may not be appropriate to use as companies may face different resilience challenges. However, we consider that Ofwat could use the following aspects to re-frame the way that PCs are set.

- Companies could be asked to define the maximum performance level first (for example, 100% of customers connected to two sources), and then justify why their proposed PC level is below that. Companies would be expected to provide detailed evidence on why that level of service is not cost beneficial. We provide a worked example of how this could work in practice below.
- Ofwat could define a minimum improvement, based on PR14 historical information. For example, if a company has historically been able to

achieve a 5% year-on-year improvement on these measures, Ofwat could expect that companies include at least a 5% year-on-year improvement in the PCs that they propose for PR19. Again, this would not necessarily be a requirement, but companies would be expected to carefully consider this reference point and fully justify cases when they have not included this level of stretch in their proposed PC levels.

- **ODIs:** companies would be expected to set two-sided financial incentives.
 - Penalties could be based on marginal cost, as this would provide sufficient incentives to meet the target. The rationale for including customer valuations in the penalty formula could depend on how directly customers are affected by the measure, how important customers generally find this measure, and/or the quality of the available data on customer valuations. In the case of resilience metrics customer valuations can be large. For example, work led by Water UK on the water resources long-term planning framework identified a central estimate of household willingness to pay to avoid emergency drought orders of £80 per year per avoided day of interruption per year from emergency drought orders.
 - Rewards could be based on the PR14 formula, with uplifts, or even with discontinuous rewards for exceptional performance. The structure of the reward rate would depend on Ofwat's ambition, and whether it wished to incentivise gradual improvement or step-wise innovation.
 - Deadbands should be based on historical information wherever possible.

B.6 Fully bespoke measures

These are measures that only one company, or a small subset of companies, use. They are therefore often based on local circumstances or a company's specific customer preferences. Since only a small set of companies monitor performance on these measures, very limited comparative information is likely to be available, and also by definition the measure is unlikely to be comparable across companies.

While these measures are relatively narrow, there is value in companies having the freedom to develop new, innovative measures. These fully bespoke measures may capture important needs of a particular community, or they may bring to light new aspects of service quality (or new ways of measuring existing aspects of service quality) that other companies would benefit from monitoring in future. In defining these new measures and committing to monitoring them, companies bring forward new information that could be used in future to expect other companies to be monitoring these additional measures. Assuming these measures are of material importance to customers in other areas, this could result in positive externalities in future.

Some examples of fully bespoke measures from PR14 include measures such as "Partnership working (community/partnerships)" or "Sites with eel protection at intakes".

Ofwat indicated in its November outcomes consultation that there are several areas in which it would welcome the development of innovative bespoke

measures, such as those capturing the outcomes for vulnerable consumers. For example, some measures that companies introduced in PR14 included the reduction in the proportion of customers falling into arrears due to their water bills, or the percentage of eligible households that actually receive a social tariff.

We have developed the following options:

- **PCs:** CBA should be used as the foundation for setting fully bespoke PCs. While no comparative information is likely to be available on these measures, companies may have historical information on their own performance, which could be used to increase the challenge in PCs. For example, Ofwat could state that it expects to see a minimum level of improvement in companies fully bespoke measures, which would be based on average historical improvements on these types of measures.
- **ODIs:** where a measure is entirely new, there may be some uncertainty around the level of performance that could be achieved. In those cases, it may not be appropriate to set financial incentives. If companies were required to apply financial incentives to new measures, there may be a disincentive to define innovative measures.

However, where the evidence on potential performance is more robust, financial incentives could be applied in the following way.

- Penalties could be based on marginal cost, or the PR14 formula.

Rewards could be based on the PR14 formula. We do not consider that it would be appropriate to apply uplifts to rewards, as Ofwat is unlikely to have particular confidence in the level of stretch included in the PCs, due to limited information on performance levels. Also, there is less rationale for uplifts, as not all other companies currently measure these aspects of service. To the extent that there is a positive externality from these measures, it comes about from companies defining these measures in the first place, which could be rewarded through the risk based review incentive.

ANNEX C DETAILED ASSESSMENT OF PACKAGES

In section 6.1 we listed the full set of criteria in Ofwat’s assessment framework. We consider that some criteria are not applicable to the assessment of packages in this instance for the following reasons:

- Maintain resilience – resilience is now one of Ofwat’s primary duties, and the common PCs will include a performance measure on resilience. This will be true for all packages, and one can imagine that, even in the case of package 4, Ofwat would include a measure of resilience in the benchmarking. Therefore, the packages do not vary substantially in this respect.
- Protect the environment – all packages would allow companies to develop measures of quality reflecting their impact on the environment, and again there are a number of common PCs that reflect environmental concerns. Therefore, the packages do not vary substantially in this respect either.
- Ensure affordability – while the different packages will affect the bill levels that are likely to arise, we have considered this within the criterion on customer impacts. Wider affordability concerns would be reflected in the iterative approach to re-scaling incentive rates that we included in our general improvements, and would therefore be incorporated into all packages.
- Ensure financial viability – whether the framework creates a risk for companies’ financial viability depends on the overall scale of the risk and reward balance, which falls outside of the scope of this project. We have not therefore assessed this aspect.
- Pro-market approach – this is outside the scope of this project and framework.
- How much will it cost to implement – we consider that this question is largely equivalent to the question on resourcing (since the majority of the cost will be resource costs). We have therefore not duplicated the answers, but we recognise that the different options will have different cost implications.

The following table provides the justification for the scoring of packages for all remaining criteria.

Figure 29 Options assessment with explanatory notes

	Package 1 – Focus on PR14 improvement	Package 2 – Focus on innovation + measure- specific	Package 3 – Focus on tailored rewards	Package 4 – Focus on integrated assessment
Protect customer interest	Not much improvement in the level of stretch or in companies’ incentives to innovate.	Measure-specific approach should result in more challenging PCs; innovation uplifts encourage innovative improvements and mean rewards only given out for exceptionally high performance.	Measure-specific approach should result in more challenging PCs; gated ODIs provide further safeguards so that only all-round good performers receive rewards.	Integrated cost and quality benchmarking should result in the most stretching targets, achieving both lower bills and better service outcomes.
Promote efficiency	Improvements to CBA mean PCs should be set closer to the efficient level.	The measure-specific approach should make best use of information to produce PCs at or close to the efficient level.	The measure-specific approach should make best use of information to produce PCs at or close to the efficient level.	Integrated benchmarking would allow Ofwat to identify the efficient combinations of service delivery levels – the efficient frontier.
Better regulation principles:				
1) Proportionate and targeted	Minimal change with respect to the base case.	The innovation uplifts can be targeted to measures with larger externality benefits.	More moderate uplifts and the measure-specific approach mean that intervention is targeted and proportionate.	Targeted on set of common PCs, with proportionate uplifts for over performers.
2) Broader range of regulatory tools	Incorporates different valuation techniques and triangulation.	Use a larger set of techniques within the measure-specific approach, for example comparative data, historical data etc.	Use a larger set of techniques within the measure-specific approach, for example comparative data, historical data etc.	Expands the toolkit used, adding sophisticated benchmarking and survey techniques.
3) Flexibility and responsiveness	Minimal change with respect to the base case.	Measure-specific approach is flexible and adaptable.	Measure-specific approach is flexible and adaptable.	Responds to issues but only common measures are captured to it is less flexible.

	Package 1 – Focus on PR14 improvement	Package 2 – Focus on innovation + measure- specific	Package 3 – Focus on tailored rewards	Package 4 – Focus on integrated assessment
4) Transparency and predictability	Minimal change with respect to the base case.	Measure-specific approach makes the framework more transparent.	Measure-specific approach makes the framework more transparent.	Benchmarking is more of a black box approach that makes it harder to companies to know what to expect and makes it less transparent for the purpose of engaging with stakeholders.
Focus on customer impacts:				
1) Impact on customer bills	Minimal change with respect to the base case.	Higher bill initially (one company’s customers pay for the externality value) but lower bills for all customers in the long run.	Moderate uplifts should be offset by more challenge to PCs so that overall customer bills should only increase by less than the benefit to consumers from higher service levels.	Integrated cost and quality benchmarking should drive bill levels down for any inefficient companies.
2) Impact on service performance	Minimal change with respect to the base case.	Innovation uplifts should lead to service innovation, driving up performance across the industry.	Moderate uplifts to rewards would incentivise more effort to improve than the base case.	Benchmarking should push companies to sit on the efficient frontier wherever they didn’t, achieving more for the same cost allowance. Innovation uplifts would ensure companies continued to expand the efficient frontier.
Focus on efficiency:				

	Package 1 – Focus on PR14 improvement	Package 2 – Focus on innovation + measure- specific	Package 3 – Focus on tailored rewards	Package 4 – Focus on integrated assessment
1) Effective incentives	Improvements to the CBA approach should mean the PCs are set closer to the efficient level.	Measure-specific approach should deliver PCs that are at or close to the efficient level.	Measure-specific approach should deliver PCs that are at or close to the efficient level.	Benchmarking should identify the efficient performance targets; the approach to preferences over combined measures should ensure the incentives are appropriate and reflect interactions between measures.
2) Ownership and accountability	No change with respect to the base case.	Measure-specific approach puts onus on the companies to provide the most robust plans possible based on the best evidence available for each measure.	Measure-specific approach puts onus on the companies to provide the most robust plans possible based on the best evidence available for each measure.	Benchmarking would be done centrally by Ofwat, taking away part of the ownership of the plans from companies.
3) Fostering innovation	No change with respect to the base case.	Innovation uplifts that only apply for large improvements will encourage companies to seek step level change through innovation.	Weaker incentive to innovate but uplifts still incentivise companies to outperform their targets (and rewards are within easier reach so likely to try to achieve them across several measures).	Innovation uplifts that only apply for large improvements will encourage companies to seek step level change through innovation.
What resources would be required to implement it:				
1) For Ofwat	Clearer expectations and experience from PR14 should make it easier for companies to implement the framework.	Measure-specific approach means there will be more information and different techniques for Ofwat to inspect but should be easier for Ofwat to interpret.	Measure-specific approach means there will be more information and different techniques for Ofwat to inspect but should be easier for Ofwat to interpret.	Potentially very resource intensive for Ofwat to implement the benchmarking with triangulation of multiple estimation techniques.

	Package 1 – Focus on PR14 improvement	Package 2 – Focus on innovation + measure- specific	Package 3 – Focus on tailored rewards	Package 4 – Focus on integrated assessment
2) For companies	Clearer expectations and experience from PR14 should make it easier for companies to implement the framework.	Flexibility in the measure-specific should mean companies find it easier and more intuitive to implement the framework.	Flexibility in the measure-specific should mean companies find it easier and more intuitive to implement the framework.	Potentially less resource-intensive for companies if much of the analysis is done centrally by Ofwat (although companies would still have to measure their customers' preferences to set targets).
How quickly will we see change	Minimal change with respect to the base case.	Depends on the level of stretch achieved and on the level of uplifts and deadbands – higher potential rewards combined with stretching targets is more likely to result in quick and significant improvement.	Depends on the level of stretch achieved and on the level of uplifts and deadbands – higher potential rewards combined with stretching targets is more likely to result in quick and significant improvement.	Should deliver strong improvements, once in place (but does require considerable research to design the new methodology).

