The approach to outcomes at PR19
Summary presentation for Thames Water

October 2016
Our report for Thames Water addresses the key questions faced by the industry regarding the evolution of the outcomes framework for PR19.

Thames Water commissioned Economic Insight to develop a report setting out an assessment of the issues relevant to the design of the outcomes framework at PR19.

Our report seeks to identify practical opportunities for building on the successes of the PR14 approach – and addresses a number of key questions.

1. What is the appropriate **balance between ‘bespoke’ and ‘common’ performance commitments?** (and further, when should common performance commitments be assessed on a ‘comparative’ basis)?

2. How does the outcomes framework **interact with cost assessment** – and how can consistency between the two be achieved?

3. What is the appropriate **measure of the benchmark** for comparative performance commitments? (where here, by measure, we are referring to the choice between: upper quartile, upper quintile, average of best three, or frontier, etc).

4. Should outcome **benchmarks** for comparative performance commitments be set on an **individual or aggregate basis?** (i.e. separate benchmarks for each comparative performance commitment, or one combined benchmark)?

5. What are the relative pros and cons of setting outcomes benchmarks on a **static or dynamic basis?**
The outcomes framework consists of a number of inter-related elements. It is therefore important to be clear about the meaning of various terms. Key definitions are provided for reference in the table below.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes</td>
<td>Refers to what companies propose to deliver for customers (and the environment) within their Business Plans.</td>
</tr>
<tr>
<td>Performance measures</td>
<td>Refers to the metric(s) used to measure performance in an outcome area (above).</td>
</tr>
<tr>
<td>Performance commitments (PCs)</td>
<td>Refers to outcomes for which companies, within their Business Plans, ‘commit’ to a specified level of performance. This is assessed via the relevant performance measure (above).</td>
</tr>
<tr>
<td>Comparative information</td>
<td>A subset of performance measures (with common definitions) that can be used - by Ofwat, Customer Challenge Groups (CCGs), customers and companies - to inform the comparative performance of companies across outcome areas. These need not have any associated PCs.</td>
</tr>
<tr>
<td>Bespoke PCs</td>
<td>Refers to outcomes, with an associated PC and ODI, that are unique to the company proposing them in their Business Plan.</td>
</tr>
<tr>
<td>Common PCs</td>
<td>Refers to outcomes for which all companies are required to have an associated PC, with a consistent definition of the measure, and an ODI within their Business Plan (although the level of the PC itself could vary across companies). These are a subset of comparative information (above).</td>
</tr>
<tr>
<td>(PCs subject to) Comparative assessment</td>
<td>Refers to common PCs whereby the level of the PC is determined through comparative (horizontal) analysis of performance across companies. Therefore, the level of the PC is the same across companies. These are a subset of common PCs (above).</td>
</tr>
<tr>
<td>Outcome Delivery Incentives (ODIs)</td>
<td>Refers to outcomes for which companies propose an ‘incentive’, based on their performance. The incentive can be reputational or financial. Where it is financial, the incentive can include penalties and / or rewards.</td>
</tr>
<tr>
<td>Willingness to Pay (WTP)</td>
<td>Refers to the ‘value’ that customers attach to receiving an outcome (typically defined as being the ‘maximum’ they will pay in exchange for the outcome).</td>
</tr>
</tbody>
</table>
Question 1: The choice between bespoke, common, and comparative PCs turns on identifying when efficient outcomes are ‘the same’ or ‘differ’ across firms.

We have developed a decision-making framework for guiding the choice between the appropriate approach to setting performance commitments (PCs).

At the heart of this is determining where the economically efficient outcome performance is ‘the same’ or ‘differs’ across companies.

Importantly, one would never expect this to be ‘exactly the same’ across firms.

So, the choice between a ‘common PC’ and a ‘common PC subject to comparative assessment’ (i.e. with a benchmark) is a matter of degree.

Figure 2: Our proposed decision tree for selecting the appropriate approach to setting Performance Commitments

*In principle a broadly defined outcome could have more than one ‘performance commitment’, where the appropriateness of whether each PC was ‘bespoke’, ‘common’ or ‘comparative’ would depend on the questions set out here.
Question 1: Critically, it is inherently ‘uncertain’ as to where the efficient level of provision ‘differs’, or is ‘similar’, across companies – and getting it wrong risks customer detriment.

The efficient level of outcome performance would differ across firms if the ‘marginal cost’ of delivering outcomes differed.

Equally, efficient outcome performance would differ if customers’ ‘marginal value’ of outcomes differed across firms.

However, observed variation in cost or value could be due to:

- **Genuine differences** in the economically efficient cost and / or genuine differences in how customers value outcomes (for example, due to differences in geography or historical investment).
- **Measurement error.** In relation to cost assessment, this could be because factors that drive differences in efficient costs across companies are omitted. In relation to outcomes, it could include differences in definitions, differences in data recording and so on.
- **Inefficiency.** That is, one company’s costs could be higher because it is ‘less efficient’ than another. Equally, a company might deliver a outcome performance below the efficient level.

**Failure to distinguish between the above factors risks real customer harm:** Ofwat might inappropriately ‘remove’ genuine variation (or vice versa). This would mean customers receiving higher / lower quality service and paying more / less than they should.

- Accommodate the uncertainty appropriately.
- Find ways to reduce the uncertainty.

There is no ‘perfect’ way of distinguishing between ‘genuine differences’; ‘measurement error’; and ‘inefficiency’. Some uncertainty is inherent. Therefore….
Question 1: Therefore – at PR19 an approach is needed that ‘accommodates’ the uncertainty appropriately and ‘mitigates’ it where possible.

- **We recommend a ‘tolerance band’ around any common PCs benchmarked using comparative assessment.**
- It would be for **companies to provide appropriate evidence** as to why their PC should vary from the benchmark.
- It would be for **Ofwat to evaluate that evidence.**

In the long-run a **‘common’ customer research method across the industry is needed** to determine where value varies across firms. For PR19, the existing evidence base could be used.

- Ofwat should identify comparator PCs, and publish benchmarking of them, sufficiently in advance of companies submitting their plans.
- **Companies should propose their own PCs** as part of their plans (including those where comparative assessment has been applied).
- From a process perspective, were Ofwat not to allow any deviation, **it would call into question why** (in relation to PCs subject to comparative assessment) **companies should develop their own research in the first place.**

Uncertainty relating to cost assessment is discussed subsequently.

*Figure 5: illustrating our proposed ‘tolerance range’ for PCs*
Question 2: As a matter of principle, a consistent and concurrent approach to outcomes and cost assessment is essential.

At the heart of this issue is that the ‘frontier’ (benchmark) is both technically and allocatively efficient. Therefore, the **appropriate benchmark for cost must also be the benchmark for outcomes** – and vice versa.

What this means is that, if the cost assessment analysis benchmarks firms against upper quartile, then firms’ allowed revenues are consistent with: ‘whatever outcomes performance an upper quartile cost efficient firm can deliver’. As highlighted by the CMA, and shown in the figure below, **there is no intrinsic reason ‘why’ this would coincide with upper quartile for outcomes performance**, if assessed separately (as at PR14).

The above explains why, in the real world, there are cost / quality trade-offs, even in highly competitive markets (this is supported by empirical evidence and precedent).

Therefore, unless outcomes and costs are ‘joined up’ there is a real risk that firms would not receive the appropriate amount of allowed revenues (i.e. allowed revenues could be ‘too high’ or, more likely, ‘too low’).

For this reason, whilst the UKWIR study suggested that such ‘join up’ is “ideal”, we think it is “essential”.

Figure 6: why the ‘benchmark’ UQ for cost is not the same as UQ for outcomes

Source: Economic Insight
Question 2: Our recommendation is that consistency is achieved through ex-post modelling adjustments during the PR19 process.

- **Achieving the consistency we describe is not straightforward.** Our initial work to modify the cost assessment econometric models from PR14 proved unreliable (in statistical terms).
- **We therefore advocate the use of ex-post modelling adjustments.** This could either be to: (i) reduce the benchmark (for common PCs subject to comparative assessment) to make it consistent with what an UQ cost efficient firm could deliver – as illustrated in the figure; or (ii) adjusting allowed costs to reflect the ‘incremental cost’ of firms delivering UQ outcome performance.

Either approach could be effective, but ideally **requires more robust data on how outcomes impact incremental cost.**

If such evidence cannot be developed for PR19, the first approach (adjusting the PC level) might be more straightforward.

In that event, given the uncertainty regarding the cost / outcome interaction, Ofwat might be concerned about PCs being ‘too lenient’. In which case, **Ofwat could apply the adjustment selectively**, only to those PCs is was ‘most confident’ that a cost relationship should exist (but the application should always be consistent across companies).
The choice of benchmark flows from a ‘measurement problem’. That is, regulators want to encourage firms to move towards frontier performance for costs and outcomes, but a benchmark for the frontier cannot be accurately identified.

Therefore, to mitigate the risk of underfunding companies, it is widely accepted that an ‘unadjusted frontier’ approach is not appropriate.*

Instead, regulators might use the frontier with adjusted residuals or use different benchmarks, such as upper quartile. The choice is a balance of ‘how challenging’ the regulator wishes to be, versus ‘how achievable’ the benchmark is (and therefore, how effective an incentive is provided).

Our analysis (see figure) suggests that ‘best’ or ‘average of best three’ benchmarks would not be appropriate.

The choice between upper quartile and upper quintile is more balanced, but upper quintile would still require already strongly performing firms to make material improvements. So, on balance, we recommend maintaining an upper quartile approach.

Our full report describes the additional evidence and analysis Ofwat could examine to further inform its choice.

*Note, for this reason, even where regulators do apply a ‘frontier’ approach to cost benchmarking (for example) the model ‘residuals’ are ‘adjusted down’ to reflect this. This was also true of Ofwat when it formerly applied a frontier approach.
Question 4: The evidence suggests that an ‘aggregate’ benchmark for comparative PCs should be considered – but this question cannot be answered definitively yet.

![Figure 9: Impact of choosing different benchmark measures](image)

- **At PR14**, for common PCs benchmarked using comparative assessment, **individual PCs were set for each area.** Alternatively, at PR19 Ofwat might consider applying an ‘aggregate’ PC over a basket of outcome areas.
- **The choice between the two turns on the extent to which there are ‘trade-offs’** (both in terms of cost and value) in performing strongly across the various outcomes.
- **Our analysis shows that water companies’ rankings varied considerably across the comparative PCs at PR14.** This would be **consistent with there being trade-offs.**

- **In addition**, we analysed quality of service performance in **other, competitive, industries.** Here we found that, even firms that performed at the ‘frontier’ overall, had varying performance levels at a less aggregate level – **consistent with trade-offs.**
- **For this reason**, we recommend Ofwat give consideration to applying an aggregate approach at PR19. However, this **requires further evidence** – in particular, on the extent to which customer valuations are ‘more similar’ across firms at an aggregate level than they are for individual PCs.
Question 5: The choice between static and dynamic approaches to PCs is akin to considering the length of price controls – it turns on a trade-off between technical and allocative efficiency.

Static benchmark = “a benchmark that results in annual PCs being set at the start of a price control period and remaining unchanged during it. “

Dynamic benchmark = “a benchmark that results in annual PCs being set at the start of a price control period, but that can change during the price control period.”

At PR14, the benchmark for common PCs subject to comparative assessment was set on a ‘static’ basis. One option for PR19 would be to set dynamic benchmarks.

This is analogous to the choice between ‘longer’ and ‘shorter’ price controls:
- Longer price controls provide stronger technical efficiency incentives, because companies retain outperformance gains for longer, until the regulator resets the control.
- Shorter price controls encourage allocative efficiency, because they allow costs and outcomes to align to the efficient level more quickly.

Accordingly, a ‘static’ approach to setting outcomes benchmarks should deliver better technical efficiency.

A ‘dynamic’ approach should deliver better allocative efficiency.

The choice between the two, therefore, turns on which form of efficiency one should put most weight on.

We have developed a decision framework for the choice, reflecting this, and other relevant factors (see figure).
Following from the previous slide, one would not characterise the water industry as being ‘fast moving’ either in terms of:

- changes to customer preferences; or
- changes in technology.

In turn, this would point to customers being better protected by having strong technical efficiency incentives – therefore a static approach.

In addition, the issue of ‘controllability’ is key. That is, is outcome performance might be more controllable by companies over a period of time, than is it year-to-year.

Therefore, the risk of a dynamic approach, is that the benchmark would reflect short-term variations that carry little incentive power.

Our analysis of the data (see table) reveals considerable volatility in the upper quartile level of performance (>10% is not uncommon) year-to-year across outcome areas.

One would need to be very sure that this volatility was not primarily caused by external factors outside of company control before a dynamic approach could be considered.

We therefore recommend retaining a static approach.

Also, consistency with cost assessment is essential. One should not have static cost targets and dynamic outcome targets.

Table 1: Changes in upper quartile performance over time

<table>
<thead>
<tr>
<th>Water quality contacts per '000 population</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper quartile performance level</td>
<td>1.27</td>
<td>1.32</td>
<td>1.18</td>
<td>1.23</td>
</tr>
<tr>
<td>Percentage change</td>
<td>4%</td>
<td>-11%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Properties subject to sewer flooding per '000 properties</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper quartile performance level</td>
<td>0.13</td>
<td>0.17</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td>Percentage change</td>
<td>25%</td>
<td>-15%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Water drinking quality (mean zonal compliance %)</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper quartile performance level</td>
<td>99.98</td>
<td>99.97</td>
<td>99.97</td>
<td>99.97</td>
</tr>
<tr>
<td>Percentage change</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pollution incidents per length of sewer km</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper quartile performance level</td>
<td>31.89</td>
<td>43.04</td>
<td>40.46</td>
<td>41.56</td>
</tr>
<tr>
<td>Percentage change</td>
<td>35%</td>
<td>-6%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Length of supply interruptions per property</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper quartile performance level</td>
<td>10.20</td>
<td>11.70</td>
<td>9.30</td>
<td>10.40</td>
</tr>
<tr>
<td>Percentage change</td>
<td>15%</td>
<td>-21%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Economic Insight analysis of Ofwat ‘upper quartile comparative assessments’ data.
Concluding points.

- **This presentation summarises the main headlines**, and some key evidence, arising from our report into the approach to outcomes at PR19.
- **Our full report, for Thames Water, contains further details** regarding ‘how’ we think our recommendations can be implemented and the underlying rationale and evidence.
- Clearly, not all stakeholders will agree with all of the points raised in our report. However, by providing clarity regarding the framework and evidence we have used, **the hope is to encourage debate** and robust policy making – as well as providing **clear, practical recommendations for PR19**.
- In a number of areas, our report also highlights areas where **additional evidence is required in order to develop an appropriate methodology**. We would emphasise the importance of this. The issues identified here can all be informed by high quality evidence – and so we would encourage the industry to progress work in this area where possible.