Europe Economics

Recommendations on Mid-Control Period Updates to Schedule 8

November 2022

Europe Economics
4th Floor
5 Chancery Lane
London
EC4A 1BL

Tel: (0) 20 3862 9252

www.europe-economics.com

Contents

Executive Summary

In this report, we assess reform options relating to mid-control period updates to benchmarks and payment rates in the Schedule 8 performance regime for passenger operators.

The reform options assessed for benchmarks in this report are:

- fixed benchmarks based on a forward-looking trajectory that incorporates traffic recovery;
- fixed benchmarks based on historical performance;
- annual updates using rolling historical performance;
- annual updates using a modelled approach;
- annual updates using Network Rail scorecard performance targets; and
- mid-period recalibration of benchmarks.

For payment rates, we have assessed:

- mid-period recalibration of TOC payment rates; and
- mid-period recalibration of Network Rail payment rates.

We envisage these options for payment rates being linked to a mid-period recalibration of benchmarks.

We have carried out illustrative modelling of the impact that the benchmark reform options would have on Schedule 8 financial flows for Network Rail and operators.[1] As part of this, we have modelled the impact of there being no Schedule 8 payments between GBR and its operators for all of the reform options. Due to the confidentiality of the data that we used, the results of this modelling have been provided to ORR in a separate note, but where relevant we refer to our findings in this report (without referencing specific figures).

Design of mid-period recalibration mechanism for benchmarks

We have designed a policy option for a mechanism to recalibrate benchmarks during CP7, following analysis of a number of dimensions over which the policy design could vary. Our selected policy design is for a triggered recalibration process, with a trigger based on deviations in outturn traffic away from the traffic level observed in the data period used for PR23 benchmark recalibration. Under this policy option, when the trigger is met ORR would consider the nature of the shock to traffic and determine if a mid-period recalibration is appropriate. (We understand that ORR has not yet made a decision on the design of any triggers.)

Each year during CP7 (barring the first year), the traffic observed during the PR23 recalibration period (which ORR indicated to us is likely to be Period 8 2021/22 to Period 7 2022/23) would be compared against outturn traffic on the network (probably with a one-year lag due to the time necessary for traffic data to become available). If outturn traffic in the most recent thirteen rail periods available deviates from the traffic level in the PR23 recalibration period by more than a specified percentage,[2] consideration by ORR of a recalibration of Schedule 8 benchmarks would be triggered. A key issue for ORR to consider would be whether the deviation in traffic was due to a temporary or permanent shock. If a mid-period recalibration was initiated it

[1] We did not model annual updates using Network Rail scorecard performance targets. It is difficult to assess what benchmarks this option might result in, as it would depend on the outcome of discussions between Network Rail and operators.

[2] In this report we do not make a recommendation on the amount of deviation from PR23 projections that is acceptable. The amount of deviation from projected traffic levels that triggers consideration of a recalibration is a policy decision for ORR to make, and would need to balance the benefits of more accurate benchmarks (e.g. avoiding large financial flows that do not reflect the underlying performance of parties) against the costs of recalibration.

would be likely to involve using a different data period to set benchmarks from the initial PR23 recalibration. The data period would be decided at the time of the recalibration based on an assessment of which years appear likely to be the most representative of future performance levels.

Recommendations

Following our assessment of all policy options, we would recommend against basing benchmarks solely on historical performance and against both forms of annually updated benchmarks. Setting benchmarks solely on historical performance could lead to unrealistic benchmarks if performance in the recalibration period is not reflective of expected performance in CP7 (e.g. due to an expected recovery in traffic levels), resulting in large financial flows for Network Rail. Benchmarks with rolling annual updates would suffer from COVID-19 distortions to performance in the early years, as well as weakening Network Rail's incentives and increasing the administrative costs of the regime. Modelled annual updates suffer from practicality issues – the lag in traffic data availability means that the benchmarks set by this method may be unrealistic during the COVID-19 recovery period. It would also be the most burdensome option administratively due to the modelling requirements.

We would not recommend implementing Network Rail's new proposal to update benchmarks annually in line with scorecard targets, given our understanding of the current governance process for agreeing those targets. This would be a costly approach as it involves recalibrating Network Rail's benchmarks every year, and it is not clear that this could be done mechanistically as the measure used for performance can vary in scorecards. Increasing the financial importance of the scorecard targets that are agreed between operators and Network Rail could lead to protracted negotiations and introduce perverse incentives. Moreover, Network Rail may not face the same challenge from operators to its proposed scorecard performance targets under the GBR model.

Benchmarks for Network Rail based on a forward-looking trajectory that incorporates traffic recovery has merit as a policy option.[3] Network Rail's trajectory would be based on the funding Network Rail receives for CP7, forecasts of traffic volumes in CP7 and a performance stretch target set by ORR, and therefore the overall shape of Network Rail's trajectory is an unknown at the time of this report. Assuming that the trajectory is set using all available information at the time, then incorporating the trajectory into Network Rail's benchmarks will mean that benchmarks are neutral on expectation, consistent with its funding and more likely to be accurate than benchmarks that do not factor in anticipated traffic recovery in CP7. If traffic forecasts are accurate, this policy option would lead to reduced financial flows compared to Network Rail benchmarks based solely on historical performance or annually updated benchmarks. However, if traffic projections for CP7 are inaccurate then benchmarks would be set at an incorrect level, potentially leading to large financial flows between Network Rail and operators. Given the uncertainty over future traffic levels, this is a significant concern.

As a result, alongside use of a forward-looking trajectory for Network Rail, **our recommendation is to reform CP7 benchmarks for Schedule 8 by introducing a mid-period recalibration trigger mechanism.** We suggest a trigger based on the traffic level on the network, which we consider to be more appropriate than a trigger based on outturn performance as it is largely exogenous to Network Rail and operators, and more appropriate than a trigger based on CP7 traffic projections as these would not be incorporated into TOC benchmarks. This policy option introduces flexibility to the regime at a time of uncertainty, without significantly hampering the incentive properties of the regime. Our illustrative modelling found that this option is effective at preventing substantial and recurring financial flows between Network Rail and operators. When initial PR23 benchmarks are set at a level which turns out to be unrealistically high,

[3] TOC benchmarks based on a forward-looking trajectory that incorporated expected traffic recovery would also have merit, as benchmarks would be more likely to be accurate. However, we understand that incorporating a forward-looking trajectory into TOC benchmarks is not currently under consideration. We assess this hypothetical policy option in Section 4.

the trigger activates to "correct" benchmarks with more up-to-date data and thus to reduce financial flows in subsequent years.

We note that there is also the possibility of performance trajectories changing during CP7 through an updated Managing Change process. We would recommend that any decision to change performance trajectories should also be a trigger for ORR to consider recalibrating benchmarks. If benchmarks are recalibrated in such a scenario, they should incorporate the new performance trajectories to reflect ORR's updated position on the outputs Network Rail is expected to deliver in terms of train performance.

In a scenario in which benchmark recalibration is carried out but performance trajectories are left unchanged, ORR would need to consider at the time whether or not the existing performance trajectories were too outdated to incorporate into the new benchmarks.

We also recommend that ORR introduce the option of recalibrating TOC and Network Rail's payment rates alongside a mid-period recalibration of benchmarks.

Updating the TOC Responsibility Matrix (used in calculating TOC payment rates) alongside a benchmark recalibration may be justified if the incremental cost of doing so is relatively low. This is because it would sharpen TOC incentives, provide improved information on the cost of delay and increase Schedule 8's flexibility to changing circumstances.

In deciding whether to update Network Rail's payment rates, ORR would need to think about whether the change in industry circumstances is likely to have significantly changed the Marginal Revenue Effect (MRE) of poor performance. Examples of circumstances in which the MRE of delay might change significantly during CP7 could include the following:

- Traffic recovery post-COVID could increase the amount of revenue that there is to lose when passengers are deterred by delays.
- The shift to home-working / hybrid working arrangements is likely to have made commuter traffic more sensitive to delay than previously, potentially increasing the MRE of delay. This effect may reverse if there is a shift back to office-based working during CP7.
- The effect of delays on revenue may be non-linear. For example, there may be little effect for delays up to a certain threshold, but large effects once delays go beyond this point as passengers become "fed up" with using trains. If increased traffic on the network during CP7 leads to increased delays, then it could move the industry beyond the threshold at which the previously-estimated MRE of delay is applicable.

In circumstances in which the MRE of delay may have changed significantly, ORR would need to reach a judgment on whether the potential benefits of updating Network Rail payment rates to reflect this change justify the additional costs of including Network Rail payment rates in the recalibration. These costs would include the costs of an update to the semi-elasticities that feed into the MRE calculation. We understand from ORR that the feasibility of such an update would depend on the availability of robust survey evidence from the Passenger Demand Forecasting Council. Network Rail's payment rate could also be updated based on new revenue data, even if the elasticities are not changed.

We also recommend that if ORR decides to recalibrate Network Rail payment rates alongside a mid-period recalibration of benchmarks, it should also recalibrate TOC payment rates (including the TOC Responsibility Matrix) at the same time. We do not think it would be feasible to update Network Rail's payment rate and not update TOC payment rates without risking imbalance in the star model.

Finally, our illustrative modelling of the impact of an opt-out for GBR operators on Schedule 8 financial flows shows that the scale of many of the impacts of Schedule 8 reform are greatly reduced following an opt-out from financial payments for these operators, as financial flows are reduced by an order of magnitude.

1 Introduction

This report has been carried out by Europe Economics for the Office of Rail and Road (ORR).

ORR is reviewing the policy approach to take at the 2023 periodic review (PR23). Part of this involves determining the performance incentive regime that will apply in control period 7 (CP7).

Schedule 8 has been adjusted to various degrees since it was introduced following railway privatisation. Despite adjustment, there remain issues that might be addressed. Some of these issues have been further highlighted by the onset of the COVID-19 pandemic and its impact on railway traffic.

The pandemic has also accelerated broader rail sector reform that has implications for the incentive effects of Schedule 8. In May 2021, the government published the Williams-Shapps Plan for Rail which envisages the creation of Great British Railways (GBR) to integrate track and train. One implication is the proportionality of making significant changes when a smaller share of train operators is expected to face revenue risk – and hence the full incentive effect of Schedule 8 payments – during CP7. We note that there is uncertainty regarding the timing of the implementation of the Plan for Rail.

This report qualitatively evaluates some key options to reform Schedule 8 in areas relating to benchmarks and payment rates, including possible mechanisms for updating those parameters within a control period.[4] We have carried out multi-criteria analysis to identify how well each option is expected to perform, relative to the counterfactual, against a set of objectives agreed with ORR.

The document is structured as follows:

- Section 2 outlines the issues and rationale for reform, including some background on Schedule 8, and presents the objectives that reform of Schedule 8 should aim to meet.
- Section 3 introduces the set of specific reform options for benchmarks that are assessed.
- Section 4 presents our assessment of the reform options for benchmarks against the reform objectives.
- Section 5 sets out our assessment of a mechanism for the mid-period recalibration of payment rates.
- Section 6 summarises our recommendations.
- Appendix 1 describes the PR18 framework for mid-period recalibrations.
- Appendix 2 summarises responses to ORR's April 2022 consultation on Schedules 4 and 8, focusing on points relevant to this report on mid-control period updates to Schedule 8 parameters.

2 Issues and Policy Objectives

This section sets out an overview of Schedule 8, discusses the problem of a lack of flexibility to adjust to changing circumstances, and presents some Schedule 8 policy objectives that are used later in the report to assess reform options.

2.1 An overview of Schedule 8 and the wider train performance incentive framework

Schedule 8 is the performance regime set out in track access contracts that compensates train operators for delays caused by Network Rail and by other train operators. In doing so, it aims to fulfil three core functions:

1. To reduce train operators' exposure to the revenue losses that arise due to delay and cancellations they have no control over.
2. To provide train operators with financial incentives to limit the delay they cause to other operators.
3. To provide Network Rail with incentives to improve train performance.

The Schedule 8 performance regime is complemented by a wider set of factors that influence train performance. This includes Schedule 4 incentives on Network Rail to minimise the impact of rail possessions on train operators, commercial mechanisms and reputational incentives.

The reduced effectiveness of certain financial incentives on Network Rail was recognised at the start of CP5 when the company was confirmed as a public sector body. Since CP6, reputation has been used alongside financial incentives to encourage good train performance. Management incentives continue to apply to Network Rail and operators, such as senior manager bonuses tied to good performance.

The reputational incentive framework involves Network Rail Scorecards, which display Network Rail's performance against 21 measures in categories including customer satisfaction, safety, train performance and financial performance. Scorecards are set for specific regions and show performance against annual targets and end-of-year forecast performance. Stakeholder research for ORR carried out by Systra finds that the scorecards are only perceived to positively incentivise Network Rail's decision-making among strategic stakeholders at Network Rail and in local government, with train operators generally expressing concern that their incentive effects are limited. Some Network Rail representatives view them as secondary influences on performance decisions compared with individuals' pride in their work.

2.2 The main benefits of Schedule 8

Research for ORR, carried out prior to the COVID-19 pandemic as well as through more recent stakeholder engagement, provides evidence that certain aspects of Schedule 8 are viewed positively.

Stakeholders highly value the revenue loss compensation provided by Schedule 8 when performance is poor due to delay caused by other parties. This has been noted as a key benefit during the franchise era, and stakeholders that will continue to be exposed to revenue risk (e.g. freight and charter operators) highly value the protection this affords them.

Schedule 8 also has a positive role in incentivising investments in train performance improvement, providing information and a basis for supporting investment decisions. The stakeholder research carried out by Systra identified examples of Schedule 8 data being relied upon to justify internal and external business cases.

2.3 The lack of flexibility to adjust to changing circumstances

A potential problem with the Schedule 8 regime is that it may not respond well to changing circumstances. Benchmarks and payment rates are fixed at the start of each control period and do not adjust during the period to reflect external changes. For example, benchmarks do not adjust within the control period to reflect the inverse relationship that exists between traffic levels and train performance (when routes are more congested, the knock-on effects of specific performance incidents on the performance of other trains can be far greater).

The COVID-19 pandemic has exposed the lack of flexibility of Schedule 8. CP6 benchmarks and payment rates were set when passenger numbers were reaching unprecedented highs. Substantial drops in passenger volumes and passenger-related disruptions during the pandemic have enabled train performance targets to be outperformed, causing large Schedule 8 bonus payments to be paid by operators to Network Rail. Lower passenger numbers per train also means that the revenue impact of delays is smaller, so payment rates set at PR18 may currently be too high for passenger operators. Significant changes to timetables will also impact on the payment rates that are appropriate to reflect the revenue impact that TOCs face.

At PR23 there could be problems with benchmarks and payment rates in the reverse direction. Recent data available for Schedule 8 recalibration at PR23 will reflect the lower traffic levels that have been observed since the COVID-19 crisis due to increased levels of home-working. If rail traffic recovers over the coming years, benchmarks and payments rates calibrated using such data may not be realistic for the new, higher level of traffic. For example, benchmarks may end up being unrealistic due to the higher levels of performance observed in the recalibration data period becoming unachievable as the network becomes more congested again. This could lead to large financial flows under Schedule 8 driven by external changes rather than how well companies are managing their performance.

As summarised in Appendix 1, at PR18 ORR did set out a framework for mid-period recalibrations of Schedule 8 parameters for passenger operators. However, this framework did not get used to adjust Schedule 8 parameters for the effects of the COVID-19 crisis. The PR18 framework had a number of limitations, including the following:

- It only covered passenger operators, and hence did not allow for any adjustment to Schedule 8 parameters for freight operators.
- It required either Network Rail or a passenger operator to request a change to Schedule 8 parameters, with no scope for ORR to initiate a recalibration. Parties will have no incentive to request a recalibration if the external shock has worked in their favour, and the former franchise passenger operators also had little incentive to request a recalibration once they were placed on concession contracts that passed Schedule 8 payments through to their funders.
- The framework implied that the evidence base used for the initial PR18 recalibration would also be used for the within-period recalibration, unless there was compelling reason to do otherwise. Using the same evidence base as at PR18 would limit the value of a mid-period recalibration in circumstances in which there has been a material change in circumstances.

2.4 Policy objectives for Schedule 8

The reform options are evaluated against a number of objectives for Schedule 8 which have been developed collaboratively with ORR:

1. To provide train operators with appropriate protection from losses arising from delays and cancellations outside their control.
2. To provide incentives for Network Rail to improve performance on the network for the benefit of customers.

3. To provide incentives for train operators and freight operators to improve train performance for the benefit of customers.
4. To provide information on the costs of delays to enable efficient allocation of resources.
5. To avoid undue discrimination between different services.
6. To avoid perverse incentives.
7. To be simple, predictable and practicable.
8. To be resilient to changing circumstances.
9. To provide consistent performance incentives across the industry.

3 Policy Options for Benchmarks

This section describes the policy options for setting Network Rail benchmarks that will be considered in this report. The policy options we consider are:

- Fixed benchmarks based on a forward-looking trajectory that incorporates traffic recovery
- Fixed benchmarks based on historical performance
- Annual updates using rolling historical performance
- Annual updates using a modelled approach
- Annual updates using Network Rail scorecard performance targets
- Mid-period recalibration of benchmarks if a trigger is met

3.1 Fixed benchmarks based on a forward-looking trajectory that incorporates traffic recovery

This policy option, designed by Europe Economics, is an alternative to the current approach for setting Network Rail and TOC benchmarks. It involves setting benchmarks for both Network Rail and TOCs based on a forward-looking trajectory that accounts for the expected recovery of network traffic following the COVID-19 crisis. The benchmarks would be set during the PR23 recalibration and would remain fixed for the duration of CP7. Network Rail's benchmarks would also be consistent with its PR23 funding settlement. We understand that setting TOC benchmarks based on a forward-looking trajectory is not an approach under consideration because ORR does not have the role of holding operators to account in the way that it does for Network Rail, and as a result this policy option is hypothetical.

Under this option, benchmarks would be calculated by first making a forecast of how traffic will evolve over CP7. There would then be a forward-looking looking adjustment applied to Network Rail and TOC benchmarks that is consistent with this projected level of traffic, with Network Rail's benchmarks also adjusted for the renewals and enhancement funding it receives for CP7 and potentially assuming an improvement in its operational efficiency. Higher levels of traffic on the network increase delay, *ceteris paribus*, and we note that under this option there is a distinct possibility that benchmarks will imply higher levels of delay during the course of CP7. In other words, the forward-looking performance trajectory may be a negative trajectory for TOCs, and the same may also be true for Network Rail if anticipated traffic recovery on the network is only partially offset by anticipated efficiency improvements and enhancement funding.

3.2 Fixed benchmarks based on historical performance

This option refers to Proposal A from ORR's initial consultation. The proposal is to base Network Rail's performance benchmarks on historical data alone, without adjustments for performance trajectories. Different benchmarks for different passenger service groups would remain. This would align the approach to setting Network Rail's benchmarks with that currently used to set the benchmarks of passenger operators. This option would require consideration of how to use past performance data when one-off events in the past could mean that unadjusted historical data produces inappropriate forward-looking benchmarks.

3.3 Annual updates using rolling historical performance

This option refers to Proposal B1 from ORR's initial consultation. This would involve updating benchmarks each year using average outturn performance during a period that is a set amount of time in the past (e.g. the

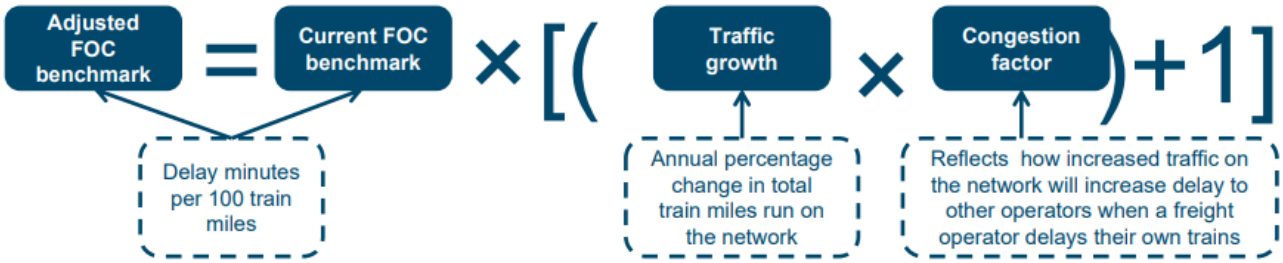
most recent five years for which data are available). The data period would thus move forward by a year each year when the benchmarks are recalibrated.

There are a number of different forms this option could take, varying by the length of period used to construct the rolling average and how adjustments (if any) are made to account for exogenous shocks to traffic levels in the rolling period. For the purpose of this report, we assume that the approach would be implemented through the use of a five-year rolling average of historical performance, with outliers removed by striking out the years with the highest and lowest performance levels.

3.4 Annual updates using a modelled approach

This option refers to Proposal B2 from ORR's initial consultation. This would involve mechanistically adjusting benchmarks each year based on changes in network traffic levels observed during the past year. A precedent for this option would be the Schedule 8 freight and charter regime in CP5, under which the benchmark was adjusted each year to reflect changes in traffic growth. The figure below illustrates the adjustment calculation that was used.

Figure 1: The formula for adjusting freight operator (FOC) benchmarks during CP5.



Source: ORR (2017) "Slides on Schedule 8 recalibration benchmarks in the freight regime 28 April 2017", p.13 [online].

The main input to the benchmark update, annual traffic growth, would reflect total mileage on the network (as in Figure 1). This would mean that the benchmark level of delay would be adjusted in proportion to the degree of traffic growth observed over the relevant recent period (i.e. the last year for which data are available). The congestion factor measures the extent to which delay caused by operators results in more delay to third party trains as traffic on the network increases. The congestion factor was fixed for the duration of a control period when this approach was implemented (as set out in ORR's letter on the mechanism, during CP5 it was fixed at 1.044). The congestion factor was calculated in the capacity charge model. Capacity charges were removed at PR18, meaning implementing this option would require consideration of whether the model was still appropriate to calculate the congestion factor or whether another method of calculating this factor should be used.

For the purpose of this report, we assume that this approach is applied without any deadbands. However, we note that deadbands could in theory be used to reduce the administrative cost of this approach by limiting changes to the benchmarks to those years in which there has been a significant change in traffic levels (or a significant change since the benchmarks were previously recalibrated).

3.5 Annual updates using Network Rail scorecard performance targets

In Network Rail's response to ORR's April 2022 consultation, it suggests that "Network Rail benchmarks could be set annually, using the annual performance trajectories that Network Rail and operators work together to set.". Network Rail does not make explicit what the "annual performance trajectories" it refers to are, but our understanding is that it is a reference to the performance targets that are reported in Network Rail's scorecards.

To be able to assess the impact that this policy option would have, it is critical to understand exactly what the scorecard performance targets are, and the governance process in place for setting these targets. We set out our understanding of how scorecard train performance targets were set in the next sub-section. In summary, we understand that that the annual scorecard targets agreed between Network Rail and passenger operators are not subject to ORR approval and note that basing Schedule 8 benchmarks on the scorecard targets could have a significant impact on the incentives to set and agree the level of those targets.

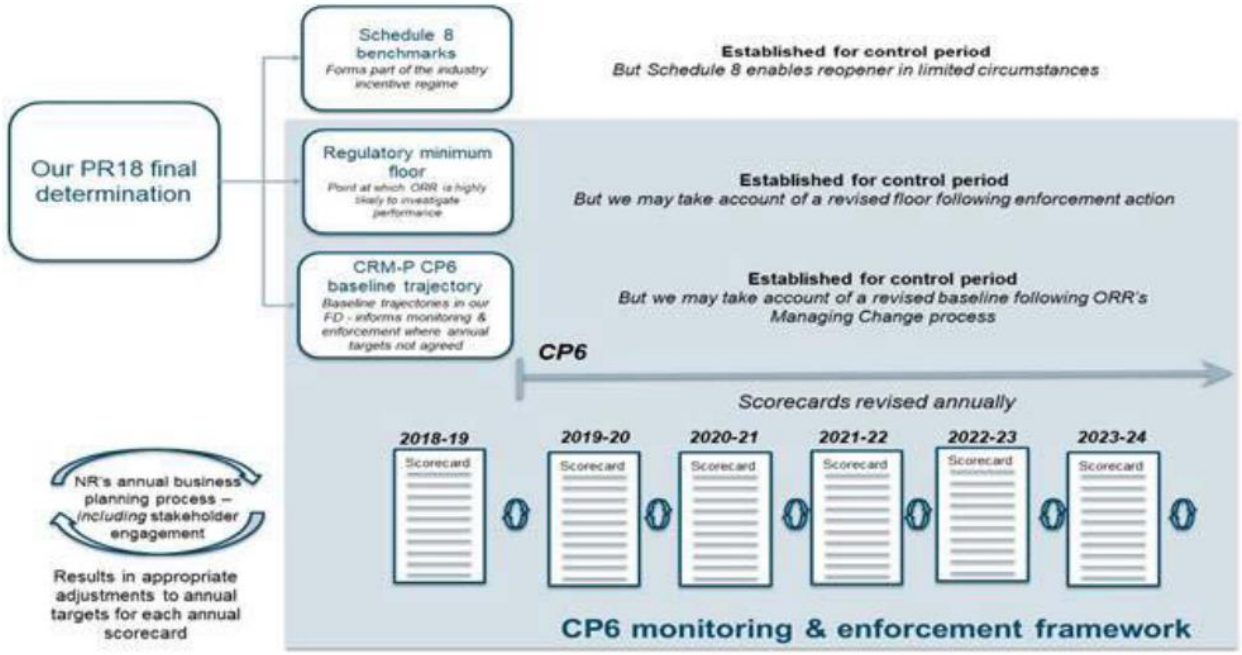
3.5.1 Governance process for Network Rail's scorecard performance targets

At PR18, ORR decided to make more use of reputational incentives in its regulation of Network Rail in CP6, given the likely reduced effectiveness of financial incentives following Network Rail becoming a public sector body. As explained in a supplementary document to ORR's PR18 final determination, one of the approaches ORR took was to implement scorecards for each of Network Rail's routes and the System Operator function. Network Rail's strategic business plan (SBP) included scorecard measures and proposed trajectories for the routes and SO over the five years of CP6.

The route scorecards contain a set of measures across Network Rail's activities, including stretch performance targets. Route-level management pay for Network Rail is affected by performance against the measures in its annual scorecards. During PR18, ORR asked the routes to engage with operators and agree scorecard targets for CP6. At the time of ORR's PR18 final determination, the majority of routes and operators were not able to reach agreement on scorecard targets. Where targets were agreed, they were reflected in route settlement documents. ORR stated that where agreement could not be reached with operators, Network Rail must continue to ensure that each route has a "stretching but realistic" target in each year of CP6.

Network Rail was required to commit to a number of processes for how scorecards would be updated, including a commitment to update scorecards annually to reflect external events or changes in customer priorities, a commitment to explain changes to scorecard trajectories with reference to its previous plan and a commitment to explain how its plan had changed. Generally, changes to scorecard targets are outside the scope of ORR's Managing Change Policy.

ORR stated in its final determination that over the course of CP6 it would increase the weight it puts on annual targets in its monitoring, where these have been explicitly agreed with operators (with sufficient governance). The figure below, taken from ORR's PR18 supplementary documents, sets out the monitoring regime for CP6.



3.5.2 Summary of Network Rail's proposal

Network Rail has suggested that its Schedule 8 benchmarks be updated annually in line with the annual performance targets included in its scorecards, rather than basing Schedule 8 benchmarks on the CRM-P based performance trajectories set in ORR's PR23 final determination. The annual targets are agreed by Network Rail routes with their operators each year, and are often not targets based on CRM-P but alternative measures such as "on time", meaning that there would need to be some conversion of this measure to delay minutes to be used for Schedule 8 benchmarks. ORR requires the annual scorecard targets to be agreed by both the route and the operator in question, but we understand that ORR does not subject these to formal approval in the same way that periodic review trajectories are agreed and approved.

3.6 Benchmarks featuring a mid-period recalibration mechanism

This policy option involves a mechanism for a mid-period recalibration of Schedule 8 benchmarks. There are a number of approaches that could be taken for a mid-period recalibration, with several dimensions along which the approach could vary. In order to arrive at a single policy option to assess, we have first assessed what might work best along different dimensions. We have grouped the dimensions into two categories; *when* a recalibration is triggered and *how* a recalibration would work. We assess these below before setting out our proposed design for policy option.

3.6.1 When a recalibration is triggered

The first consideration is how a recalibration of Schedule 8 parameters is initiated, for which there are two main options. The first is to have a trigger, where the deviation of a selected metric beyond an allowed threshold activates an ORR review of whether a recalibration is appropriate. The second option is to require a formal request for recalibration to be made by one or more parties to ORR, which ORR can then decide to accept or reject. (The latter is the approach used at PR18, which we summarise in Appendix 1.)

We prefer the first option, a trigger to consider a recalibration. It represents a more systematic approach to mid-period recalibrations than requiring a request from Network Rail or operators, which may choose not

to request a recalibration when it might be appropriate to do so because it is not in their own financial interests. It is notable that there have been no requests for a system-wide recalibration of Schedule 8 during CP6, despite the unforeseen and very sizeable impact that COVID-19 has had on network traffic (and consequently performance). This may be because the reduced traffic on the network meant that Network Rail benefitted from being able to comfortably outperform the benchmarks that were set for it at PR18, and passenger operators were transferred to concessions contracts that held them financially neutral to Schedule 8 payment flows. While a recalibration was not in the interests of Network Rail or operators, it may have been appropriate to recalibrate benchmarks to better reflect expected performance at the reduced level of network traffic during the COVID-affected years. A trigger, whether based on traffic levels or performance, would probably have been activated during that period and would have allowed benchmarks to be "corrected" to better align with the actual level of traffic on the network. Hence, we prefer an approach that uses a trigger to identify when ORR should consider a recalibration.

While the trigger itself would be mechanical, ORR would have the discretion not to carry out a mid-period recalibration if the trigger is met. There may be circumstances in which the trigger is met but a recalibration would not be deemed appropriate, such as if there was a temporary shock to the rail sector. It is important that the trigger does not *automatically* lead to a recalibration, to avoid unnecessary or inappropriate recalibrations. Instead, the trigger should always initiate *consideration* of a mid-period recalibration, with the final decision left to ORR.

With the above in mind, we then need to consider what measure to use to set a trigger. The two obvious candidate measures are traffic levels on the network and parties' outturn performance against their PR23 benchmarks. We strongly prefer using network traffic as the basis for a trigger, because it is less endogenous to the performance of Network Rail and the operators. If consideration of a recalibration were to be triggered when a party's outturn performance deviates from its benchmark by a given amount, it could create perverse incentive effects. An underperforming party that is close to triggering consideration of a recalibration would have an incentive to intentionally worsen its performance in order to trigger a potential recalibration that would lead to an adjustment to its benchmarks. There is less scope for these perverse incentives if traffic levels are used as the basis for a trigger, although GBR would have some ability to affect traffic levels through its timetabling and fare decisions. Our recommendation, therefore, is that a trigger is set on the basis of network traffic so that when outturn traffic deviates outside a set range, consideration by ORR of a recalibration of Schedule 8 is activated.[5]

This leads to a further consideration; how should deviations in traffic be measured? The observed traffic levels in CP7 could be compared to the traffic levels in the base period used to set benchmarks, or to a set of forecasts made at PR23. Our preference here is dependent upon how the initial CP7 benchmarks are set for Network Rail and for TOCs. If benchmarks are set using historical performance data, then the trigger should be based on the traffic levels in the historical data period used to set benchmarks. If, instead, benchmarks take account of a projected recovery in traffic levels throughout CP7, then the recalibration trigger should be based on deviations in traffic levels away from these projections.

We would also recommend that the trigger be based upon the divergence between the average traffic level over the course of the PR23 recalibration period and the average outturn traffic level in each year of CP7. One drawback of this approach is that consideration by ORR of a recalibration could be triggered by temporary shocks to the traffic level in a single year. However, in these circumstances ORR could use its discretion not to carry out a recalibration despite the trigger being met.

[5] An argument against using traffic levels as the trigger is that it would not pick up situations in which factors other than traffic have a significant impact on train performance. For example, in 2000/01 train performance reduced significantly due to the identification of gauge corner cracking across the network following the Hatfield accident and the widespread imposition of speed restrictions. This increased Schedule 8 performance payments and was one of the key factors leading to Railtrack being put in Railway Administration.

A final issue to consider is the times at which recalibrations can happen (recognising the time and resources involved in undertaking a recalibration). Again, there are two main possibilities. There could be one specified year in CP7 during which a recalibration could be triggered if actual traffic levels are outside a specified range, probably the middle (third) year. Alternatively, the framework could allow scope for a recalibration in any year of CP7 in which the trigger is met (possibly with the exception of the first year, so that the benchmarks derived in the PR23 recalibration exercise apply for at least one year). We prefer this second, more flexible, option. It allows for multiple shocks occurring over the five years of CP7, or for a shock that occurs in a year other than the middle year of the price control period.

3.6.2 How a recalibration would work

When thinking about how a mid-period recalibration triggered by a deviation in outturn network traffic would work, the main consideration is what data would be used to calibrate the updated parameters. As discussed in Appendix 1, the PR18 guidance on mid-period recalibrations suggested the same base data period used for the initial calibration is used for the recalibration, unless there is a clear and compelling reason not to do so. We do not think this is a sensible approach for PR23. The rationale for a recalibration would be that existing parameters, based on the best data available at PR23, are not representative of the actual state of the rail network in CP7, due to unforeseen evolution of network traffic or an unexpected shock to the industry. It would therefore be optimal to base a recalibration on more recent data than that which were available at PR23, so that parameters are more reflective of performance expectations at the time of the recalibration. The best data period for recalibration would need to be decided upon at the time of recalibration rather than specified in advance, as the optimal choice would be dependent on which years of data are least distorted by atypical events.

3.6.3 Interaction with updates to performance trajectories during CP7

ORR's recent consultation on PR23 Policy Framework included proposals for reforms to the Managing Change process. This included scope for the CP7 baseline performance trajectories ORR determines for Network Rail to be changed during CP7, if Network Rail demonstrated evidence of a substantial change in circumstances that could not reasonably be expected in its Annual Business Plan.

Network Rail's performance trajectories fed into its Schedule 8 benchmarks in CP6 and may also do so in CP7. The performance trajectory was converted into Average Minutes Lateness (AML) for the purposes of Schedule 8 benchmarks. The AML calculated from the baseline recalibration period was then adjusted for each year of the control period in line with the performance trajectories.

The "substantial change in circumstances" that prompted Network Rail to request new trajectories could be related to a shock to traffic levels that would simultaneously lead to the trigger being met for ORR to consider a benchmark recalibration, but that may not necessarily be the case. There are three possible scenarios to consider, discussed below.

Scenario 1: Performance trajectories change during CP7, but consideration of a recalibration of benchmarks is not triggered by a change in traffic

If ORR agreed to change Network Rail's performance trajectories, we would recommend that this is taken as an alternative "trigger" for considering a benchmark recalibration if traffic levels had not triggered a recalibration anyway. While traffic is a good indicator of an external shock to expected delay, it is not the only possible cause of a shock to performance. ORR would only change performance trajectories if there was clear evidence of a "change in circumstances that could not reasonably be expected" (i.e. an external shock). This is sufficient grounds to consider whether that change in circumstances also warrants a benchmark recalibration. If ORR decides that benchmarks should be recalibrated, the new benchmarks should be set using the new performance trajectories.

Scenario 2: Consideration of a benchmark recalibration is triggered by a change in traffic, but performance trajectories have not changed

If increased traffic led to consideration of a benchmark recalibration during CP7 but performance trajectories had not changed since PR23, we think ORR would need to come to a judgement on the relevance of the PR23 performance trajectories for Network Rail's Schedule 8 benchmarks given the change in traffic levels. An in-depth discussion of when a change in performance trajectories is warranted is beyond the scope of this paper, and we acknowledge that the primary purpose of the trajectories is not to set Schedule 8 benchmarks. If trajectories are not changed when consideration of a recalibration is triggered, then ORR would need to consider either ratcheting the new benchmarks (based on a more recent data period) using the old performance trajectories or not using performance trajectories at all. There would also be the option of deciding not to recalibrate existing benchmarks.

Scenario 3: Consideration of a benchmark recalibration is triggered and performance trajectories also change

Our proposed design for a recalibration trigger allows for the possibility of a recalibration in any year of CP7 (after Year 1). Similarly, the proposed changes to the Managing Change process indicate that it would be up to Network Rail in its Annual Business Plan to suggest changes to performance trajectories. This introduces the possibility that ORR might decide in favour of a benchmark recalibration following the traffic trigger being met at the same time in CP7 as trajectories are also being changed. This possibility is not insignificant, as the same external shock could cause both events to occur. In such a situation, the new trajectories should feature in new benchmarks, with the same ratcheting process as used previously. The new trajectories will reflect ORR's updated position on the outcomes Network Rail is expected to achieve in the remainder of CP7 regarding train performance.

3.6.4 Proposed design for benchmarks with a mid-period recalibration mechanism

On the basis of the above analysis of the different dimensions along which a mid-period recalibration mechanism could vary, the policy design for a mid-period recalibration mechanism for benchmarks that we are proposing, and that we will assess in Chapter 4, is set out below.

Benchmarks for Network Rail would be set at PR23 using ORR's preferred recalibration period, and would be set for the full duration of CP7. However, scope for recalibrating benchmarks within CP7 would be introduced in the form of a trigger that when activated would initiate consideration of a mid-period recalibration of benchmarks.

In each year during CP7 (barring the first year), the traffic observed during the PR23 recalibration period (Period 8 2021/22 to Period 7 2022/23) would be compared against outturn traffic on the network (probably with a one-year lag due to the time necessary for traffic data to become available). If outturn traffic in the most recent thirteen rail periods available[6] deviates from the traffic level in the PR23 recalibration data period beyond a certain threshold,[7] consideration by ORR of a potential recalibration of Schedule 8 benchmarks would be triggered. Following consideration of the causes of the shock to traffic levels, ORR would determine if a benchmark recalibration is appropriate. A mid-period recalibration would be likely to use a different data period from the initial PR23 recalibration, with the data period to be decided at the time of the recalibration based on an assessment of which years appear likely to be the most representative of future performance levels.

[6] There are 13 rail periods in a year.

[7] In this report we do not make a recommendation on the amount of deviation from PR23 projections that is acceptable. The amount of deviation from projected traffic levels that triggers consideration of a recalibration is a policy decision for ORR to make, and would need to balance the benefits of more accurate benchmarks (e.g. avoiding large financial flows that do not reflect the underlying performance of parties) against the costs of recalibration.

We would recommend that any decision to change Network Rail's performance trajectories should also be a trigger for ORR to consider recalibrating benchmarks. If benchmarks are recalibrated in such a scenario, they should incorporate the new performance trajectories to reflect ORR's updated position on the outputs Network Rail is expected to deliver in terms of train performance.

4 Assessment of Policy Options for Benchmarks

This section assesses each of the policy options for benchmarks outlined in Section 3. It sets out our multi-criteria analysis of each option, which involves giving scores based on the scoring approach in 4.1 and explaining the reasoning behind them. A comparative summary of the scores given to each policy option is provided at the end of this section.

4.1 Approach to assessment: multi-criteria analysis

We assess how each reform option performs against each objective relative to the counterfactual ("Do minimum") using the following two-dimensional scale relating to the direction and strength of the impact:

Direction of impact	Negative impact	Neutral impact	Positive impact	Depends
Strength of impact	Weak	Strong		

For example, if an option scores "neutral", it means that it is not discernibly better or worse than the counterfactual against that criterion.

Some of the scores against some criteria are influenced by the possible characteristics of the industry following rail reform. In such cases, the comments in the "Explanation of score" column already take into account the impact of rail reform. Where comments are given in square brackets in the same column, these explain how the score would be different if industry reform is delayed.

"Do minimum" departs from the standard "Do nothing" counterfactual to the extent that the existing performance regime would, as a minimum, require some calibration at PR23.

A number of policy options face disadvantages as a result of the endogeneity of the metric – for example, Network Rail or operators facing perverse incentives to reduce performance in order to secure easier future benchmarks under certain of the options. This could be mitigated by ORR being aware of this risk and retaining the power to take it into account in any update to Schedule 8 parameters, if it was suspected that a party had been attempting to game the system.

Illustrative modelling by Europe Economics

To inform our multi-criteria analysis of the reform options relating to benchmarks, we carried out illustrative modelling of Schedule 8 benchmarks and performance payments throughout CP7 for Network Rail and TOCs under most of our benchmark policy proposals. The modelling provided a forward-looking analysis of the potential impact of the policy options. It involved three different scenarios for how traffic on the rail network could recover from the low level seen during the COVID-19 crisis. Our modelling considered a scenario in which there are no Schedule 8 payments between GBR and its contracted operators in CP7.

Due to the confidentiality of the data used for our modelling, we do not report any quantitative results from the modelling in this report. Instead, we draw on the qualitative conclusions from the modelling in carrying

out our multi-criteria analysis of the policy options. (The exception to this is our modelling of the potential impact of different benchmark methodologies on the incentives of Network Rail and operators to carry out investments to improve performance, which does not use confidential data.)

4.2 Option 1: Fixed benchmarks based on a forward-looking trajectory that incorporates traffic recovery

To recap, this reform option involves the following:

- Setting benchmarks for Network Rail and TOCs for the duration of CP7 with a forward-looking adjustment that accounts for the expected recovery of network traffic following the COVID-19 crisis. In the case of Network Rail, benchmarks would also take account of any renewals and enhancement funding it receives for CP7 and an assumption for improved operational efficiency.
- The performance trajectory could be a negative trajectory, in which case TOC and Network Rail benchmarks would imply increased delays during CP7 as traffic recovers.

In summary, the main advantages of this option are:

- Benchmarks take account of the anticipated recovery of traffic from the COVID-19 crisis and are therefore likely to be more accurate than benchmarks that do not incorporate a forward-looking trajectory.
- Benchmarks are neutral in expectation.
- Network Rail's benchmarks would be consistent with the funding settlement it receives in terms of enhancement expenditure, maintenance and renewals, a key concern for many stakeholders.

The main disadvantages of this option are:

- If traffic projections for CP7 are inaccurate then benchmarks would be set at an incorrect level, potentially leading to large financial flows between Network Rail and operators.
- Incorrect traffic forecasts would also increase the likelihood of requests for recalibration of benchmarks, which creates administrative costs for the industry.
- Benchmark-setting is more complex and costlier as it requires the incorporation of traffic projections into the benchmark calculations.
- Network Rail and operators might have more incentive to game regulatory trajectories at price reviews.

The table below sets out our multi-criteria analysis for this policy option.

Table 1: Multi-criteria analysis of fixed benchmarks based on a forward-looking trajectory

	Objective	Direction of impact	Strength of impact	Explanation of score
1	Provide train operators with protection from losses due to delays outside their control	Positive	Weak	Network Rail's benchmarks would be consistent with its funding settlement for CP7 and operators would receive compensation for performance below this level by Network Rail throughout CP7. Incorporating traffic recovery into the benchmark reduces large financial flows between operators and Network Rail, as shown in our illustrative modelling.
2	Provide incentives for Network Rail to improve train performance	Neutral	n/a	As with the "Do minimum" option, the influence of past performance on benchmarks would be complemented by the use of forward-looking regulatory trajectories. This avoids the risk of endogenously determined benchmarks based solely on past performance weakening incentives on Network Rail to improve its performance.

	Objective	Direction of impact	Strength of impact	Explanation of score
3	Provide incentives for train operators and freight operators to improve train performance	Neutral	n/a	No incentive impacts envisaged.
4	Provide information on the costs of delays	Neutral	n/a	No change envisaged.
5	Avoid undue discrimination between different services.	Neutral	n/a	No change envisaged.
6	Avoid perverse incentives	Negative	Weak	Network Rail and operators might have an incentive to game traffic forecasts and regulatory trajectories at price reviews. Network Rail might try to secure less challenging trajectories for itself at price reviews and TOCs might try to secure more challenging trajectories for Network Rail.
7	Be simple, predictable and practicable.	Negative	Weak	There is increased complexity due to the need to produce traffic projections for CP7. Even if traffic projections were going to be produced regardless of Schedule 8, there is a cost involved in incorporating those projections into the Schedule 8 recalibration.
8	Be resilient to changing circumstances	Positive	Weak	Our illustrative modelling shows that this policy is resilient to changing circumstances as the economy recovers from the COVID-19 crisis. By factoring in traffic recovery, Schedule 8 payment flows are minimal for this option when traffic projections are accurate. Even if traffic forecasts are inaccurate, our modelling indicates payment flows are lower than they would be if benchmarks were based on historical performance alone. However, there still would be large financial flows under this option if PR23 traffic projections deviate from actual CP7 traffic.
9	Provide consistent performance incentives across the industry.	Depends	Strong	The approach would be consistent with the funding settlement Network Rail receives for enhancement expenditure, maintenance and renewals as it would reflect the improved performance expected from Network Rail. However, the approach to setting benchmarks for Network Rail would not be aligned with the approach used for operators.

4.3 Option 2: Fixed benchmarks based on historical performance

This approach involves the following:

- Removing the regulatory performance trajectories from the construction of Network Rail's train performance benchmarks, leaving historical performance as the key input.

In summary, the main advantages of this option are:

- Benchmarks may be less administratively burdensome to calibrate since regulatory trajectories would no longer be used as an input for benchmark calculations. We understand from ORR that trajectories would be developed regardless of whether this reform is implemented, but the cost of adjusting benchmarks in the light of these trajectories would be avoided by this policy option.
- Network Rail and operators would have less incentive to game regulatory trajectories at price reviews, as they would no longer influence Schedule 8 benchmarks.
- Network Rail and operator benchmarks would be set on a consistent basis.

The main disadvantages of this option are:

- Our illustrative modelling shows that this option would lead to substantial "windfall penalties" for Network Rail and for operators in CP7 because the high performance in COVID-19 affected years creates unrealistic benchmarks in CP7. These penalties would be particularly large for Network Rail.
- Operators would eventually receive less compensation if poor performance by Network Rail is sustained over multiple price reviews, and conversely Network Rail would receive lower rewards if good performance is sustained over multiple price reviews.
- The benchmarks do not take account of any information about the likely future path of performance that is not encapsulated in historical performance. Using historical performance as a forecast is more likely to be inaccurate than forecasts based on all available information. This may increase the likelihood of requests for mid-period recalibrations, creating administrative costs for the industry.
- Network Rail's benchmarks would be driven entirely by its own historical performance, which may (marginally) reduce its incentives to improve performance, as improved performance would lead to tougher benchmarks in the future.
- Network Rail benchmarks would not be aligned with the funding settlement (regarding enhancement expenditure, maintenance and renewals) as they would not reflect the improved performance that Network Rail would be expected to achieve from the funding it receives.
- The endogeneity of benchmarks would increase in a scenario in which GBR manages both trains and tracks, because network traffic levels would be partially dependent on the timetabling and fare decisions that GBR makes. There is a (small) risk this could impact negatively upon incentives for GBR. In particular, GBR may not take full account of the impact that increased traffic will have on delays to third party operators because it will know that any increase in such delays will lead to more relaxed Schedule 8 benchmarks in relation to those operators in the next price control period.

The table below sets out our multi-criteria analysis for this policy option.

Table 2: Multi-criteria analysis of setting Network Rail's benchmarks on historical performance

	Objective	Direction of impact	Strength of impact	Explanation of score
I	Provide train operators with protection from losses due to delays outside their control	Negative	Weak	The extent to which operators would continue to receive compensation for ongoing poor performance over the course of multiple price controls would diminish as Network Rail's benchmarks would become less challenging, relative to traffic levels, over time.

	Objective	Direction of impact	Strength of impact	Explanation of score
2	Provide incentives for Network Rail to improve train performance	Negative	Weak	There is a risk that endogenously determined benchmarks based on past performance could weaken incentives on Network Rail to improve its performance, since doing so could result in a more challenging benchmark in the next control period. This risk may be small given the time lag until benchmarks are adjusted, since Network Rail can still benefit from outperforming its benchmarks during the price control period. The perverse incentive may also be reduced by the fact that Network Rail cannot be certain that ORR will retain the same approach in future periodic reviews. If rail reform takes place as expected, GBR benchmarks would be more endogenous under this option, because the volume of traffic (which affects the amount of delay through congestion effects) will be, at least partially, dependent on the timetabling and fare decisions of GBR. This could impact negatively upon incentives for GBR; it may not take full account of the impact increased traffic will have on delays to third party operators because it will know that any increase in such delays will lead to more relaxed Schedule 8 benchmarks in relation to those operators in the next price control period. The above point (which also applies to other benchmark reform options analysed in this report) is a small risk because it is likely that the costs of increased Schedule 8 compensation from more congestion only would only have a small impact on GBR's timetabling and fare decisions. Other factors, such as expected additional fare revenue and the costs of running more trains, or the perceived social benefits of running a minimum level of train service, are likely to be stronger drivers of decisions on whether or not to add more services to the timetable.
3	Provide incentives for train operators and freight operators to improve train performance	Neutral	n/a	No incentive impacts envisaged.
4	Provide information on the costs of delays	Neutral	n/a	No change envisaged.
5	Avoid undue discrimination between different services.	Neutral	n/a	No change envisaged.
6	Avoid perverse incentives	Neutral	n/a	Removing the regulatory performance trajectories from the recalibration could reduce any perverse gaming incentives on Network Rail to try to secure less challenging Network Rail trajectories at price reviews, as well as the perverse gaming incentives on TOCs to try to secure more challenging Network Rail trajectories. However, this could simply be replaced by gaming around the new approach, such as when deciding which precise data period should be used to calculate benchmarks.

	Objective	Direction of impact	Strength of impact	Explanation of score
7	Be simple, predictable and practicable.	Depends	Weak	Setting Network Rail benchmarks on the basis of historical performance simplifies the recalibration at periodic reviews and increases its transparency, since performance trajectories do not need to be transformed into benchmarks. Reducing the cost of recalibration of the regime is arguably appropriate given uncertainty over whether Schedule 8 will continue to exist after industry reform and, if it does, what form it will take. However, the reduced scope of Schedule 8 under the updated counterfactual reduces these cost savings, as does the fact that performance trajectories will need to be calculated anyway. Therefore, the impact on this criterion is fairly small. On the other hand, this policy option would not score well in terms of predictability and practicality, as historical benchmarks could lead to large windfall payments during CP7, rather than being neutral on expectation.
8	Be resilient to changing circumstances	Negative	Strong	Our illustrative modelling shows that this policy is not resilient to changing circumstances as the economy recovers from the COVID-19 crisis. With no adjustments, this option would lead to substantial 'windfall penalties' for Network Rail and for operators in CP7, with especially large penalties for Network Rail. Including a forward-looking performance trajectory that factors in traffic recovery post-COVID is a far more resilient approach in our modelling than this option.
9	Provide consistent performance incentives across the industry.	Depends	Strong	The approach to setting benchmarks for Network Rail would be aligned with the approach used for operators. However, the approach would not be consistent with the funding settlement Network Rail receives for enhancement expenditure, maintenance and renewals as it would not reflect the improved performance expected from Network Rail.

4.4 Option 3: Annual updates using rolling historical performance

This approach involves the following:

- Benchmarks for Network Rail and operators are updated based on average outturn performance in a group of recent years. (Note that this policy inherently involves setting Network Rail's benchmarks solely on the basis of historical performance and therefore under this approach regulatory trajectories would no longer be used to set Network Rail's benchmarks.)
- This option could take a number of forms. Here we assume that the approach would be implemented through the use of a five-year rolling average of historical performance, with outliers removed by striking out the years with the highest and lowest performance levels.[8]

In summary, the main advantages of this option are:

[8] We note that there might need to be a lag between the group of five years used for the rolling calculation and the year in which the benchmark would apply, in order to allow time for the required data to become available and to feed into the calculations.

- It provides scope for Schedule 8 flexibility by more quickly reflecting changes in delays caused by permanent or long-lived external shocks. This advantage would particularly apply in later years of CP7 when a benchmark set on the basis of CP6 data may no longer be reflective of the latest reality.
- The benchmarks for Network Rail and operators would be set on a more consistent basis.

The main disadvantages of this option are:

- Benchmarks could be based on performance in previous years that have limited relevance to the setting of a challenging target for the forthcoming year.
- The benchmarks do not take account of any information about the likely future path of performance that is not encapsulated in historical performance.
- Operators would eventually receive less compensation if poor performance by Network Rail is sustained over a number of years and Network Rail would receive lower rewards if good performance is sustained over a number of years.
- Network Rail benchmarks based on rolling past performance could (marginally) blunt incentives to improve performance because improved performance in one year would contribute to more challenging benchmarks in future years.
- Benchmarks are more endogenous to Network Rail/GBR's own decisions on timetabling and fares, as these decisions will affect the volume of traffic and hence the amount of congestion and delays.
- Benchmarks may reflect temporary shocks in previous years that will not affect performance in the year to which the benchmark applies. The longer the rolling window is, the longer the distortions caused by a temporary shock will remain in the benchmark. Conversely, a shorter rolling window would reduce how long a temporary shock distorts benchmarks, but the extent of the distortion would be greater than with a longer window due to having less "normal" years in the window offsetting the distortion. A top-and-tail approach would not entirely address the volatility issue when a temporary shock affects multiple years of data, as the COVID-19 crisis has, and would delay the inclusion of impacts that systematically change performance over time (for example, changes to safety standards that impact on performance).
- It could reduce the incentive to undertake investment to improve performance as that would lead to more challenging benchmarks in future years. Our modelling shows that the rolling average approaches, both with and without top-and-tailing, reduce investment incentives. The shorter the period used to calculate rolling benchmarks, the greater the incentive problem.[9]
- There is more uncertainty as to what benchmarks will be over the course of CP7 which could reduce incentive for long-term investment, as it will be harder to factor Schedule 8 financial flows into business plans.

The table below provides our more detailed multi-criteria assessment against our policy objectives.

Table 3: Multi-criteria analysis of annual updates to benchmarks based on rolling historical performance

	Objective	Direction of impact	Strength of Impact	Explanation of score
I	Provide train operators with protection from losses due to delays outside their control	Negative	Strong	The extent to which operators would continue to receive compensation for ongoing poor performance outside their control would diminish over the course of CP7 as Network Rail's benchmarks would become less challenging, relative to traffic levels, over time.

[9] Our modelling indicates only investments with a payback period of 2 years or less are worthwhile under a 1-year or 2-year rolling average. This changes to investments with a 3-year payback period for 3-year or 4-year rolling benchmarks, with or without top-and-tailing. 5-year rolling top-and-tailing benchmarks allow for investments with a 4-year payback period. All of these approaches fare worse, in terms of incentivising investment, than historical approaches that use performance data from one or two years in the previous control period and remain fixed for 5 years.

	Objective	Direction of impact	Strength of Impact	Explanation of score
				Endogenously determined benchmarks based on past performance could weaken incentives on Network Rail to improve its performance, since doing so could result in a more challenging benchmark in future years (and similarly, poor performance would contribute to by easier targets in future years).
2	Provide incentives for Network Rail to improve train performance	Negative	Weak	Under this approach, benchmarks are more endogenous to Network Rail/GBR's own decisions on timetabling and fares, as these decisions will affect the volume of traffic and hence the amount of congestion and delays. There is a small risk this could impact negatively upon incentives for GBR; it may not take full account of the impact increased traffic will have on delays to third party operators because it will know that any increase in such delays will lead to more relaxed Schedule 8 benchmarks in relation to those operators in the future.
3	Provide incentives for train operators and freight operators to improve train performance	Negative	Weak	This option could reduce the incentive for operators to undertake investment to improve performance as that would lead to more challenging benchmarks in future years. This means operators are less able to recoup upfront investment costs through Schedule 8 payments. Our illustrative modelling shows that rolling updates (with or without a top-and-tail adjustment) reduce the incentive to invest in performance improvements relative to the current approach for operators of fixing benchmarks for the duration of each control period using historical performance data from the previous control period.
4	Provide information on the costs of delays	Neutral	n/a	No change envisaged.
5	Avoid undue discrimination between different services.	Negative	Weak	There is a small risk this option could create (weak) incentives for GBR to discriminate against non-GBR operators due to increased endogeneity of benchmarks. GBR may be less inclined to account for the impact of increased delays to non-GBR operators when making timetabling and fare decisions that increase traffic volumes on the network, because its benchmarks would get easier when traffic increases and performance deteriorates as a result.
6	Avoid perverse incentives	Depends	Weak	Removing the regulatory performance trajectories from the recalibration could reduce any perverse gaming incentives on Network Rail to try to secure less challenging Network Rail trajectories at price reviews, as well as the perverse gaming incentives on TOCs to try to secure more challenging Network Rail trajectories. However, this could simply be replaced by gaming around the new approach, such as when deciding which precise data period should be used to calculate benchmarks.
7	Be simple, predictable and practicable.	Neutral	n/a	While the option would involve updating benchmarks each year, the calculation involved is relatively straightforward.

	Objective	Direction of impact	Strength of Impact	Explanation of score
8	Be resilient to changing circumstances	Depends	Weak	Exogenous events that cause permanent changes in traffic levels would feed through into benchmarks, which in turn could make benchmarks more appropriate for contemporary levels of network activity. (While under the counterfactual, operator benchmarks would also eventually adjust to changes in traffic levels through recalibration at price reviews, such adjustment would happen on a more continuous basis under this policy option.) However, temporary shocks that cause large changes in traffic / performance followed by a rebound (e.g. the COVID-19 crisis) might lead to inappropriate benchmarks and large "windfall" financial flows in post-rebound periods. Our illustrative modelling found that this risk is only partially mitigated by a top-and-tail approach because multiple years of data are affected by COVID-19.
9	Provide consistent performance incentives across the industry.	Depends	Strong	The approach to setting benchmarks for Network Rail would be aligned with the approach used for operators. However, the approach would not be consistent with the funding settlement Network Rail receives for enhancement expenditure, maintenance and renewals as it would not reflect the improved performance expected from Network Rail.

4.5 Option 4: Annual updates using a modelled approach

This approach involves the following:

- Benchmarks are recalibrated at PR23 (as under the counterfactual) but are then adjusted each year of CP7 based on changes in rail traffic levels observed over the past year and the modelled relationship between traffic levels and delays. (Note that this option does not necessarily imply any move away from the use of regulatory trajectories to set Network Rail's initial benchmarks.)

In summary, the main advantage of this option is:

- Benchmarks adjust more quickly to reflect contemporary traffic levels.

The main disadvantages of this option are:

- There would be significant complexity and costs associated with determining the appropriate adjustment factor (which is likely to require the development of new modelling) and applying the annual updates.
- Practical issues around data lag mean that older than ideal traffic data would need to be used, probably two-year-old traffic data. The result of this may be unrealistic benchmarks during the period of recovery from the COVID-19 crisis, with our illustrative modelling showing this option would lead to substantial windfall penalties.
- Benchmarks are more endogenous to GBR's own decisions on timetabling and fares, as those decisions impact on traffic volumes which are then used to model delays.
- There is more uncertainty as to what benchmarks will be over the course of CP7 which could reduce incentive for long-term investment, as it will be harder to factor Schedule 8 financial flows into business plans

The table below shows our multi-criteria analysis of this policy option.

Table 4: Multi-criteria analysis of annual updates to benchmarks based on modelling

	Objective	Direction of impact	Strength of Impact	Explanation of score
1	Provide train operators with protection from losses due to delays outside their control	Negative	Weak	GBR would be protected from having to pay out as much compensation to devolved authority, open access and freight operators when it increases traffic on the network. This would have a negative effect on the protection provided to operators by Schedule 8.
2	Provide incentives for Network Rail to improve train performance	Negative	Weak	GBR would have reduced incentives to take account of the effect on delays of increases in traffic resulting from its timetabling and fares decisions, as its benchmarks would be relaxed when traffic increases.
3	Provide incentives for train operators and freight operators to improve train performance	Neutral	n/a	No incentive impacts envisaged, as the adjustment to an operator's benchmark is exogenous to the operator's performance.
4	Provide information on the costs of delays	Neutral	n/a	No impacts envisaged.
5	Avoid undue discrimination between different services.	Negative	Weak	This option could create incentives for GBR to discriminate against non-GBR operators due to increased endogeneity of benchmarks. In particular, GBR may be less inclined to account for the impact of increased delays to non-GBR operators when making timetabling and fare decisions that increase traffic volumes on the network, because its benchmarks would get easier when traffic increases. By contrast, GBR would be fully exposed to the revenue and cost impacts of delays to its own operators.
6	Avoid perverse incentives	Depends	Weak	Removing the regulatory performance trajectories from the recalibration could reduce any perverse gaming incentives on Network Rail to try to secure less challenging Network Rail trajectories at price reviews, as well as the perverse gaming incentives on TOCs to try to secure more challenging Network Rail trajectories. However, this could simply be replaced by gaming around the new approach, such as when deciding which precise data period should be used to calculate benchmarks and during the modelling work to determine traffic-related adjustments.

	Objective	Direction of impact	Strength of Impact	Explanation of score
7	Be simple, predictable and practicable.	Negative	Strong	Introduces complexity in recalculating all benchmarks each year, which could require auxiliary calculations. For example, the approach would require calculation of a congestion factor such as that used in CP5 when a similar adjustment mechanism applied in the freight and charter regimes.[10] This would be costly given that the capacity charge model (from which the congestion factor was taken) was not recalculated at PR18 due to the removal of the capacity charge. This approach scores poorly for practicality. It may not be possible to accurately estimate the relationship between traffic and performance. Moreover, it is unlikely that the value of this parameter is the same for all service groups, so the model would need to be able to deal with this through disaggregation. Data lags are also a problem. In practice, historical data would have to be used to calculate the annual update (probably from two years previously, given the need for the data to be available in time to calculate benchmarks in advance of each year).
8	Be resilient to changing circumstances	Depends	Weak	Exogenous events that cause permanent changes in rail traffic would be reflected in performance benchmarks more quickly than at each periodic review. In theory, this approach would automatically adjust for any difference in traffic levels as traffic recovers from the COVID-19 crisis. However, problems with data lags partially negate this benefit. In practice, historical data would have to be used to calibrate the model (probably from two years previously, given the need for the data to be available in time to calculate benchmarks in advance of each year). The resulting benchmarks may be unrealistic during the period of recovery from the COVID-19 crisis. Our illustrative modelling indicates that a modelled approach to annually updating benchmarks could lead to substantial windfall penalties. Further, temporary shocks that cause large changes in traffic followed by a restoration of previous traffic levels might lead to inappropriate benchmarks, as due to data lags the temporary change in traffic would affect benchmarks in subsequent years in which traffic has returned to normal.
9	Provide consistent performance incentives across the industry.	Neutral	n/a	Depending how it was implemented, the base benchmarks might still be set on a different basis for Network Rail and operators. For example, the base benchmarks for Network Rail could still be set based on regulatory trajectories, and the annual update applied for any change in traffic which differed from what was assumed in formulating those trajectories.

[10] The congestion factor was fixed for the duration of CP5.

4.6 Option 5: Annual updates using Network Rail scorecard performance targets

This option involves:

- Network Rail's benchmarks being updated annually in line with the annual performance targets it agrees with operators each year for its scorecards. (For the purpose of assessment, we assume that operator benchmarks are set in the same way as under the "Do minimum" option.)

The main advantages of this option are:

- It provides scope for flexibility in Network Rail's benchmarks by allowing Network Rail and operators to respond to changing circumstances or external shocks when agreeing annual targets.
- Benchmarks would be consistent with Network Rail's annual scorecard performance targets.

The main disadvantages of this option are:

- Network Rail benchmarks may not be aligned with the funding settlement (regarding enhancement expenditure, maintenance and renewals) as they may not reflect the improved performance that Network Rail would be expected to achieve from the funding it receives.
- The governance process for Network Rail and operators agreeing annual performance targets is not robust when used for this purpose and the infrastructure manager may not face the same challenge from operators to scorecard performance targets under the GBR model.
- It would be a costly approach as it involves recalibrating Network Rail's benchmarks every year, and it is not clear that this could be done mechanistically.
- There is more uncertainty as to what benchmarks will be over the course of CP7 which could reduce incentive for long-term investment, as it will be harder to factor Schedule 8 financial flows into business plans

The table below shows our multi-criteria analysis of this policy option.

Table 5: Multi-criteria analysis of annual updates to benchmarks based on scorecard targets

	Objective	Direction of impact	Strength of Impact	Explanation of score
I	Provide train operators with protection from losses due to delays outside their control	Depends	Weak	The extent to which operators are protected from losses outside their control would no longer be linked to Network Rail's funding settlement or its past performance. It would depend upon the ability and appetite of operators to negotiate challenging annual performance targets for Network Rail. If operators can get Network Rail to agree to challenging targets, then this would have a positive impact on this objective. Conversely, if Network Rail can get operators to agree to a less challenging target, this would have a negative impact on this objective.

	Objective	Direction of impact	Strength of Impact	Explanation of score
2	Provide incentives for Network Rail to improve train performance	Depends	Strong	Network Rail's benchmarks would no longer be directly linked to Network Rail's funding settlement or its past performance. The incentives provided for Network Rail under this option would depend entirely upon how challenging the annual performance targets in Network Rail's scorecards are. This in turn depends upon how effectively Network Rail and operators are able to negotiate those targets. Benchmarks could be set too high or too low and therefore provide sub-optimal incentives because they would not reflect anything other than the relative bargaining power of Network Rail and passenger operators and the ongoing commercial incentives that they face (e.g. from passenger revenue, franchise contracts or GBR relationships).
3	Provide incentives for train operators and freight operators to improve train performance	Neutral	n/a	No incentive impact envisaged.
4	Provide information on the costs of delays	Neutral	n/a	No change envisaged.
5	Avoid undue discrimination between different services.	Negative	Weak	The different bargaining power of different passenger, charter and freight operators may lead to inconsistent approaches to setting scorecard targets and therefore performance benchmarks.
6	Avoid perverse incentives	Negative	Strong	There is a clear perverse incentive for Network Rail to try and agree easy annual performance targets with operators to reduce its benchmarks. Conversely, operators have an incentive to try and agree overly difficult benchmarks for Network Rail to increase the Schedule 8 payments they receive.
7	Be simple, predictable and practicable.	Negative	Strong	Benchmarks would need to be recalibrated every year under this option to reflect the annual scorecard performance targets. There is further complexity due to the fact that scorecard targets agreed between routes and operators are not always based upon the same measures of performance. There would therefore be more complexity in transforming targets into benchmarks, and it is not clear if under this option the recalibration process could be automated. Finally, given the increased importance the annual scorecard targets would have under this option, the negotiation process would be likely to become longer and costlier as both Network Rail and operators place more weight on annual scorecard targets.

	Objective	Direction of impact	Strength of Impact	Explanation of score
8	Be resilient to changing circumstances	Positive	Weak	This option allows scope for benchmarks to be flexible to changing circumstances. In theory, the annual targets can adapt each year to external shocks to the rail sector although this depends on how far parties correctly believe the shocks to be short or long term.
9	Provide consistent performance incentives across the industry.	Negative	Strong	Network Rail's benchmarks would not be consistent with operators' benchmarks under this option. Benchmarks would also not be consistent with the CP7 funding settlement.

4.7 Option 6: Mid-period recalibration of benchmarks

This option involves the following:

- Each year, consideration by ORR of a recalibration is triggered if actual outturn traffic deviates from traffic levels in the data period used for PR23 benchmark recalibration by more than a pre-determined threshold. ORR would then determine if a recalibration is warranted on a case-by-case basis.
- The data period used for the mid-period recalibration would be decided at the time of recalibration, but would probably be different from the data period used for the initial PR23 recalibration.

The main advantages of this option are:

- If there is a significant recovery in traffic levels, benchmarks will be recalibrated in the light of this. Our illustrative modelling shows that when a recalibration is triggered, financial flows for both Network Rail and TOCs in subsequent years are significantly reduced. In this regard, this reform performs better than fixed benchmarks based solely on historical data or annually-updating benchmarks with either a rolling or modelled approach. Naturally, the impact of the mid-period recalibration on financial flows varies according to the level of divergence in traffic at which the trigger threshold is set (as well as the assumption made in the model about the evolution of traffic over CP7).
- Benchmarks are flexible to unforeseen shocks or circumstances.
- The variable used in the recalibration trigger, traffic level, is reasonably exogenous and therefore minimises scope for parties to game the trigger process.

The main disadvantages are:

- There is more uncertainty as to what benchmarks will be over the course of CP7 which could reduce incentive for long-term investment, as it will be harder to factor Schedule 8 financial flows into business plans.
- Temporary shocks to traffic could lead to a recalibration which would "correct" benchmarks in line with the temporary shock, only for the adjusted benchmarks to become inappropriate when the shock has passed. This drawback is mitigated by the fact that ORR would have discretion to not recalibrate benchmarks if it was aware that a shock to traffic was temporary.

The table below shows our multi-criteria analysis of this policy option.

Table 6: Multi-criteria analysis of mid-period recalibration of benchmarks

	Objective	Direction of impact	Strength of Impact	Explanation of score
1	Provide train operators with protection from losses due to delays outside their control	Positive	Weak	If traffic changes significantly, which could lead to operators receiving insufficient compensation, a recalibration will be triggered during CP7. Benchmarks would be adjusted and operators would regain protection from losses.
2	Provide incentives for Network Rail to improve train performance	Negative	Weak	There may be a weak negative affect on Network Rail's incentives in a situation where it becomes clear that a recalibration is inevitable due to unexpected outturn traffic levels. In the intervening period, there may be less incentive for Network Rail to improve its performance as this would translate into more challenging benchmarks following the recalibration.
3	Provide incentives for train operators and freight operators to improve train performance	Negative	Weak	There may be a weak negative affect on operators' incentives in a situation where it becomes clear that a recalibration is inevitable due to unexpected outturn traffic levels. In the intervening period, there may be less incentive for operators to improve their performance as this would translate into more challenging benchmarks following the recalibration.
4	Provide information on the costs of delays	Neutral	n/a	No change envisaged.
5	Avoid undue discrimination between different services.	Neutral	n/a	No change envisaged.
6	Avoid perverse incentives	Negative	Weak	To the extent that GBR can influence traffic levels, there may be gaming incentives – GBR could try to engineer a recalibration through its influence over traffic levels, if it thought doing so would leader to easier benchmarks following a recalibration. This impact is weak, because it would only come into force when the level of traffic is close to triggering a consideration by ORR of a recalibration. Further, there are other influences on GBR's timetabling and fare decisions that would be likely to outweigh Schedule 8 considerations.
7	Be simple, predictable and practicable.	Negative	Weak	There would also be administrative costs associated with any mid-period recalibration that is triggered. However, a recalibration would only be triggered when there is a significant deviation in outturn traffic from traffic levels in the data period used for PR23 benchmark recalibration. The predictability of benchmarks would be lower relative to a regime without a trigger for recalibration, as operators/Network Rail would be uncertain *a priori* when, or if, a mid-period recalibration would be triggered.

	Objective	Direction of impact	Strength of Impact	Explanation of score
8	Be resilient to changing circumstances	Positive	Strong	This option would be resilient to changing circumstances, as any significant, unforeseen shock to the traffic level would trigger a recalibration which would "correct" benchmarks in line with that change to traffic. Consequently, this would avoid the large financial flows that can occur due to forecast uncertainty when setting fixed benchmarks. Our illustrative modelling shows just this – in situations where traffic grows significantly from the initial recalibration period and the resulting financial flows are substantial, a mid-period recalibration improves the accuracy of benchmarks and reduces financial flows. However, this option could lead to benchmarks being recalibrated because of temporary shocks to traffic that occur during CP7, leading to inappropriate benchmarks when the temporary shock has passed. This drawback is mitigated by the fact that ORR would have discretion to not recalibrate benchmarks if it was aware that a shock to traffic was temporary.
9	Provide consistent performance incentives across the industry.	Neutral	n/a	Any triggering of a mid-period recalibration would apply to all operators and Network Rail consistently. However, only Network Rail's benchmarks would incorporate a forward-looking regulatory trajectory.

4.8 Summary of scores for benchmark options

The following table summarises the scores for options to reform performance benchmarks. In this table, we have replaced "Neutral" with dashes, to make it clearer at a glance in which areas the policies will have positive or negative effects.

Table 7: Summary of options to reform train performance benchmarks

Objective	Option 1: Fixed benchmarks with traffic projection	Option 2: Fixed historical benchmarks	Option 3: Annual rolling benchmarks	Option 4: Annual modelled benchmarks	Option 5: Annual benchmarks based on scorecards	Option 6: Mid-period recalibration trigger
Provide train operators with protection from losses due to delays outside their control	Weak Positive	Weak Negative	Strong Negative	Weak Negative	Weak Depends	Weak Positive
Provide incentives for Network Rail to improve train performance	-	Weak Negative	Weak Negative	Weak Negative	Strong Depends	Weak Negative

Objective	Option 1: Fixed benchmarks with traffic projection	Option 2: Fixed historical benchmarks	Option 3: Annual rolling benchmarks	Option 4: Annual modelled benchmarks	Option 5: Annual benchmarks based on scorecards	Option 6: Mid-period recalibration trigger
Provide incentives for train operators and freight operators to improve train performance	-	-	Weak Negative	-	-	Weak Negative
Provide information on the costs of delays	-	-	-	-	-	-
Avoid undue discrimination between different services.	-	-	Weak Negative	Weak Negative	Weak Negative	-
Avoid perverse incentives	Weak Negative	-	Weak Depends	Weak Depends	Strong Negative	Weak Negative
Be simple, predictable and practicable.	Weak Negative	Weak Depends	-	Strong Negative	Strong Negative	Weak Negative
Be resilient to changing circumstances	Weak Positive	Weak Negative	Weak Depends	Weak Depends	Weak Positive	Strong Positive
Provide consistent performance incentives across the industry.	Strong Depends	Strong Depends	Strong Depends	-	Strong Negative	-

As shown in the table, one advantage of having fixed forward-looking benchmarks that incorporate a traffic projection (Option 1) is that it provides operators with protection from losses caused by delays outside their control. This is because it sets benchmarks that reflect the expected level of future traffic, thus reducing the likelihood of large Schedule 8 payments driven by external factors. Another key benefit is that it maintains alignment between Network Rail's benchmarks and CP7 funding for Network Rail, a key priority raised by many stakeholders in consultation responses. The main drawback of this approach is it increases the complexity of the regime.

The downside of Options 2, 3 and 4 is that they could reduce compensation received by operators for prolonged poor performance by Network Rail and would dampen Network Rail's incentives to reduce delays. To different degrees, all three approaches increase the endogeneity of benchmarks, potentially causing perverse incentive effects. The increased endogeneity of benchmarks risks creating an incentive for undue discrimination between different services in the case of both approaches to annual updating benchmarks.

The table shows that historical benchmarks for Network Rail (Option 2) score negatively on resilience to changing circumstances because this approach is not resilient to COVID-19 affected data, causing unrealistic benchmarks as traffic levels recover and substantial windfall penalties for Network Rail and operators. The resilience of Options 3 and 4 depends on the nature of shocks to the sector – both approaches cope well with permanent or long-lived shocks but less well with temporary shocks. The table also shows that the

biggest downside of the modelled approach (Option 4) is its lack of simplicity and transparency, as it would require the development of potentially complex modelling to determine how traffic levels affect achievable levels of performance.

Network Rail's recent proposal to base benchmarks on the annual scorecard targets that it agrees with operators (Option 5) does not perform well in our multi-criteria analysis. Any potential positive impacts of this option are reliant upon there being a strong governance process for agreeing scorecard targets, something that would need to be implemented ahead of CP7. There are also strong negative impacts from increasing the financial importance of the annual scorecard targets – it increases perverse incentives for both Network Rail and operators and it increases the likelihood of long, drawn-out negotiations between parties, which increases the costs of the regime.

Option 6, fixed benchmarks with a trigger for a recalibration, scores similarly to Option 1 against several criteria. As shown in the table, it scores negatively in relation to Network Rail's and operators' incentives due to the possibility that Network Rail and/or operators may anticipate that a recalibration will be triggered and have less incentive to perform strongly in the periods that may subsequently be used to recalibrate benchmarks, although we think this is a weak negative effect. The other key disadvantage is that it increases the complexity of the regime and reduces the predictability of benchmarks for all parties. The key advantage of Option 6, and the reason it is our preferred option, is that it has a strong positive impact on the resilience of benchmarks to changing circumstances during CP7, as benchmarks will adapt to unforeseen changes in the level of traffic on the network.

5 Assessment of Mid-period Recalibration of Payment Rates

This section assesses two additional policy options relating to the recalibration of Schedule 8 payment rates for the passenger regime within CP7. First, the current "do nothing" policy for payment rate recalibration is briefly outlined, after which the new policy options are described. We then assess the new options using the same multi-criteria analysis as for the benchmark policy options.

5.1 Description of policy options

5.1.1 Do nothing option

The "do nothing" policy would involve calibrating the Network Rail and TOC payment rates at PR23 and fixing them for the duration of CP7.

Network Rail payment rates for the passenger Schedule 8 regime are based on the estimated marginal revenue effect (MRE) of poor performance on operators. MRE models the change in revenue for a given Service Group from a one-minute change in "Performance Minutes", which is a measure of lateness and cancellations. Calculating the MRE for a Service Group involves multiplying the total revenue for that Service Group in question over the recalibration period by the appropriate semi-elasticity (a measure of passenger sensitivity to disruption). The semi-elasticities are derived from econometric modelling of the relationship between passenger demand and delay.

Alongside Network Rail payment rates, a key input into the calculation of TOC payment rates is the TOC responsibility matrix (TRM). This sets out the relationship between the delays that operators cause to themselves and the amount of delay caused to other operators. Under the PR18 recalibration methodology, the TRM was calculated using data on delays taken from the calibration period of 2015/16 and 2016/17. The relationship observed in the calibration period was assumed to remain throughout the upcoming control period.

5.1.2 Mid-period recalibration of TOC payment rates alongside benchmark recalibration

One possible option to reform passenger Schedule 8 payment rates would be to add scope for recalibrating TOC payment rates within CP7. Under this option, payment rates would be calculated during PR23 in the same way that they were for PR18, making use of the TRM. However, a trigger based on outturn traffic levels compared to projected traffic levels in CP7 could be introduced. If the trigger was met, the TRM would be updated using the most appropriate period of delay data available at the time, and from the updated TRM new TOC payment rates would be calculated. This essentially mirrors Option 6 from our previous section on reforms to benchmarks.

However, as discussed in our assessment of updating TOC payment rates later in this section, updating the TRM is likely to have only small incentive benefits, and hence it would be unlikely that the time and effort of a stand-alone recalibration of TOC payment rates would be justified. However, in a scenario in which benchmarks are being recalibrated during CP7 anyway (i.e. because a trigger for benchmark recalibration has been activated) then the incremental cost of updating the TRM alongside benchmarks might be low. If this is the case, one process for mid-period recalibration of payment rates could simply be to incorporate it into

any benchmark recalibrations that occur during CP7. In other words, TOC payment rates are recalibrated during CP7 by updating the TRM if, and only if, benchmarks are recalibrated during CP7.

On this basis, the policy option for mid-period recalibration of TOC payment rates that we assess in this section is for recalibration of TOC payment rates to be linked to benchmark recalibration. When a benchmark recalibration is triggered within CP7, following the process set out in Section 3.6, a TOC payment rate recalibration, through an update of the TRM, is triggered alongside it. At all other times, payment rates are left unchanged.

5.1.3 Updating Network Rail payment rates alongside benchmark recalibration

As with TOC payment rates, one option for updating Network Rail payment rates is to add scope for triggering a recalibration in CP7. Again, divergence between outturn traffic in CP7 and traffic levels in the data period used for PR23 recalibration could be used as the basis for triggering a recalibration of Network Rail payment rates, mirroring Option 6 for benchmark reforms.

Updating the Network Rail payment rate would involve recalculating Service Group MREs. There would be two elements to updating MREs. Firstly, the revenue for that Service Group would be updated in line with the new recalibration data period. Secondly, the semi elasticities that feed into MREs might also be updated with more recent data.

This second element could be particularly important in the current rail environment. The COVID-19 pandemic forced a substantial increase in the number of workers that worked from home rather than in offices (or other business premises). Despite health restrictions on office-working no longer being in place, there have been significant legacy effects from the period of enforced home-working. Many employers are much more flexible in regards to their staff working from home than they were before COVID-19. The result of this is that rail commuters may be more sensitive to disruption to their commutes than they were before COVID-19, as there is greater scope for them to substitute away from commuting (by working from home). How this dynamic develops over CP7 cannot be predicted, but should be borne in mind by the rail industry. It may be that the shift to a greater amount of home-working is permanent, in which case there may have been a permanent change in commuters' elasticity of demand for rail. Conversely, there may be a gradual return to pre-COVID working practices, with employers requiring staff to come into the office as much as possible. Either way, if Network Rail's payment rates are being recalibrated it would be worth considering whether there is updated research evidence available that would change the elasticities that feed into the calculation of payment rates since the last time rates were calibrated.

Updating MREs could also be significant if the effect of delays on revenue is non-linear. For example, there may be little effect on revenue for delays up to a certain threshold, but large effects once delays go beyond this point as passengers become "fed up" with using trains. If increased traffic on the network during CP7 leads to increased delays, then it could move the industry beyond the threshold at which the previously-estimated MRE of delay is applicable.

5.2 Assessment of mid-period recalibration of TOC payment rates

The main advantages of the option are:

- If a mid-period recalibration was triggered, payment rates would more accurately reflect the latest performance of passenger operators in terms of the reactionary delays that they cause to other operators. This would provide more accurate incentives for TOCs.
- The policy might create a small incentive on TOCs to reduce delays. A TOC that has reduced the amount of reactionary delay it causes to other operators relative to the delay it causes to its own trains would see its payment rate fall as a result of the recalibration. This would benefit the TOC if it underperformed

its benchmark in the following periods (as it would pay less financial penalty), but would work to its financial disadvantage if it outperformed its benchmark in the following periods (as it would receive less financial reward). Overall, the TOC would be faced with lower financial risk associated with Schedule 8 payments as both rewards and penalties would be lower. This might act as a small incentive on TOCs to try to reduce reactionary delays that they cause to other operators in case a recalibration happens.

- If a mid-period recalibration was triggered, updated payment rates would reduce the risk of imbalance in the star model, lowering Network Rail's financial exposure to risks that it cannot control.
- A mid-period recalibration of the TRM would provide more information on the delays that each individual TOC was having on other passenger operators, a relationship which may evolve over CP7.

The main disadvantages of this option are:

- It could create a (small) risk of gaming of payment rates by operators, as they could allow their performance in terms of reactionary delay caused to other operators to deteriorate if they anticipate a mid-period recalibration, in order to gain higher payment rates in subsequent years in which they intend to invest heavily in outperforming their benchmark.
- There would be administrative costs associated with updating the TRM and recalibrating payment rates, potentially on multiple occasions.

We note that an opt-out mechanism that enabled DfT operators to remove themselves from Schedule 8 would reduce the incentive benefits of this option, which are already limited.

The table below provides a more detailed assessment against our policy objectives.

Table 8: Multi-criteria analysis of mid-period recalibration of TOC payment rates

	Objective	Direction of impact	Strength of impact	Explanation of score
1	Provide train operators with protection from losses due to delays outside their control	Neutral	n/a	If a mid-period recalibration is triggered, updating the TRM and payment rates would change payments made by perpetrator operators (i.e. those that cause delays), but would not change the compensation received by victim operators (i.e. those that experience delays caused by other operators).
2	Provide incentives for Network Rail to improve train performance	Neutral	n/a	Network Rail would remain incentivised to assist the system in recovering from reactionary delays it does not cause.

	Objective	Direction of impact	Strength of impact	Explanation of score
3	Provide incentives for train operators and freight operators to improve train performance	Positive	Weak	If a mid-period recalibration was triggered, payment rates would more accurately reflect the latest performance of passenger operators in terms of the reactionary delays that they cause to other operators. This would provide more accurate incentives for TOCs. The policy might create a small incentive on TOCs to reduce delays. A TOC that has reduced the amount of reactionary delay it causes to other operators relative to the delay it causes to its own trains would see its payment rate fall as a result of the recalibration. This would benefit the TOC if it underperformed its benchmark in the following periods (as it would pay less financial penalty), but would work to its financial disadvantage if it outperformed its benchmark in the following periods (as it would receive less financial reward). Overall, the TOC would be faced with lower financial risk associated with Schedule 8 payments as both rewards and penalties would be lower. This might act as a small incentive on TOCs to try to reduce reactionary delays that they cause to other operators in case a recalibration happens.
4	Provide information on the costs of delays	Positive	Strong	Mid-period recalibrations would enable payment sums to more closely represent the amount of delay caused to other operators by an operator delaying its own services.
5	Avoid undue discrimination between different services.	Neutral	n/a	No change envisaged.
6	Avoid perverse incentives	Negative	Weak	The policy option could in theory create a risk of train operators gaming the system for financial gain. An operator could have the perverse incentive to intentionally underperform in the run-up to an anticipated recalibration, knowing that this relationship would be reflected in higher payment rates after the recalibration. Subsequently, the operator could then commit significant resources to outperforming its benchmark so as to receive significant Schedule 8 payment sums based on its higher payment rates. In practice, this risk is mitigated by the fact that operators are likely to prefer to reduce their level of financial exposure – which would not be achieved by intentionally influencing their payment rates upwards. Such gaming would also would require an operator to have full control over whether it underperforms or outperforms. Further, there are other incentives for operators to run their trains on time, such as commercial pressures or contractual obligations, that are likely to outweigh the weak negative incentive effect of this policy option.

	Objective	Direction of impact	Strength of impact	Explanation of score
7	Be simple, predictable and practicable.	Depends	Weak	Multiple consultation responses indicated that TRM recalibration is a burdensome process. Repeating the calculations each time a recalibration is triggered during CP7 would increase administrative costs for Network Rail. On the other hand, if a mid-period recalibration is triggered, better alignment of the sums that Network Rail pays in relation to reactionary delay with the amounts received from operators would increase the predictability of financial flows for Network Rail.
8	Be resilient to changing circumstances	Positive	Strong	If there is a permanent change in the amount of reactionary delay that specific operators cause to other operators (relative to the delay they cause to their own trains), a recalibration would lead to payment rates being adjusted accordingly.
9	Provide consistent performance incentives across the industry.	Neutral	n/a	No change envisaged.

5.3 Assessment of mid-period recalibration of Network Rail payment rate

The main advantages of this option are:

- If a mid-period recalibration was triggered, payment rates would more accurately reflect the latest data on the financial impact of delays on operators. This would provide more accurate incentives on Network Rail to reduce reactionary delay. If TOC payment rates were also recalibrated to reflect the new Network Rail payment rates, then the incentives on TOCs would also become more accurate.
- A mid-period recalibration would provide more information on the long-run revenue impacts on passenger operators of delay, a relationship which may evolve over CP7. For example:
  - The shift to home-working / hybrid working arrangements is likely to have made commuter traffic more sensitive to delay than previously, potentially increasing the MRE of delay. This effect may reverse if there is a shift back to office-based working during CP7.
  - If traffic rises significantly during CP7, there could be an increase in the long-run revenue impact of delay. Rising traffic will directly increase the amount of revenue that can potentially be lost when delay occurs. Further, if congestion caused by rising traffic leads to delays rising to a high level, passenger dissatisfaction could reach a "tipping point" where there is a significant reduction in the use of the rail network.

The main disadvantages of this option are:

- There would be administrative costs associated with updating the MREs and recalibrating payment rates, potentially on multiple occasions. If this includes updating the semi elasticities that feed into MREs (for example, because of a significant shift in home-working patterns) the cost could be substantial.

The table below provides a more detailed assessment against our policy objectives.

Table 9: Multi-criteria analysis of mid-period recalibration of Network Rail payment rates

	Objective	Direction of impact	Strength of impact	Explanation of score
1	Provide train operators with protection from losses due to delays outside their control	Positive	Weak	If a mid-period recalibration is triggered, updating the MREs and payment rates would change payments made by Network Rail to victim operators for reactionary delays, providing more accurate protection against financial losses incurred by those operators.
2	Provide incentives for Network Rail to improve train performance	Positive	Weak	If a mid-period recalibration is triggered, the resulting payment rates would more accurately reflect the latest data on the financial impact that delays have on operators, thus improving the accuracy of Network Rail's incentives.
3	Provide incentives for train operators and freight operators to improve train performance	Neutral	n/a	A change in Network Rail payment rates on its own would not affect passenger operators' incentives to reduce reactionary delays that they cause to other operators. (That said, incentives on operators would become more accurate if TOC payment rates were updated using the new Network Rail payment rates.)
4	Provide information on the costs of delays	Positive	Strong	Mid-period recalibrations would provide information on the evolving long-run revenue impacts of delay upon passenger operators.
5	Avoid undue discrimination between different services.	Neutral	n/a	No change envisaged.
6	Avoid perverse incentives	Neutral	n/a	No change envisaged.
7	Be simple, predictable and practicable.	Negative	Strong	Recalibrating Network Rail payment rates would lead to administrative costs, which would potentially be significant if the recalibration includes the elasticities that feed into MREs.
8	Be resilient to changing circumstances	Positive	Strong	This option is likely to have a positive impact on the regime's resilience to changing circumstances. In particular, if there is a permanent change in the financial impact of reactionary delay (e.g. caused by significant changes to traffic levels or increased commuter sensitivity to delay), a recalibration could be triggered allowing payment rates to adjust to the change.
9	Provide consistent performance incentives across the industry.	Neutral	n/a	No change envisaged.

5.4 Interactions between the two policy options for payment rates

While updating TOC payment rates and updating Network Rail payment rates are two separate processes and two separate policy options, they are linked due to the need to keep the star model in balance. It is important to keep Network Rail financially neutral for disruption that one passenger operator causes another. If Network Rail's payment rates are updated then the payment sums Network Rail pays out to operators for reactionary delay will change. As a result, TOC payment rates (including the TRM) would also need to be

updated to maintain balance in the star model, so that TOCs' Schedule 8 payments continue to match compensation paid by Network Rail as closely as possible. If TOC payments rates (including the TRM) were not updated alongside the recalibration of Network Rail payment rates, Network Rail could be exposed to windfall financial gains or losses for the reactionary delays caused by operators.

The result of the interaction between these two policy options is that, while they are separate reforms with separate impacts against ORR's objectives, it would be difficult to update Network Rail's payments rates and not also update TOC payment rates, because of the risk of unbalancing the star model.

5.5 Summary of scores for payment rates options

The following table summarises the scores for options to reform payment rates. In this table, we have replaced "Neutral" with dashes, to make it clearer at a glance in which areas the policies will have positive or negative effects.

Table 10: Summary of multi-criteria analysis of payment rates options

Objective	Option 1: Mid-period recalibration of TOC payment rates	Option 2: Mid-period recalibration of Network Rail payment rates
Provide train operators with protection from losses due to delays outside their control	-	Weak Positive
Provide incentives for Network Rail to improve train performance	-	Weak Positive
Provide incentives for train operators and freight operators to improve train performance	Weak Positive	-
Provide information on the costs of delays	Strong Positive	Strong Positive
Avoid undue discrimination between different services.	-	-
Avoid perverse incentives	Weak Negative	-
Be simple, predictable and practicable.	Weak Depends	Strong Negative
Be resilient to changing circumstances	Strong Positive	Strong Positive
Provide consistent performance incentives across the industry.	-	-

As shown in the table above, the main strengths of introducing a mechanism for updating TOC payment rates are the information it provides on the cost of delay and its resilience to changing circumstances. It would also have a small positive impact on TOC incentives. The main disadvantage is that it would increase the administrative costs of Schedule 8 if any recalibrations were triggered during CP7. There would also be a small negative impact against the criterion of avoiding perverse incentives.

In a similar way, the main benefits of introducing a mechanism for updating Network Rail benchmarks are the information it provides on the cost of delay and its resilience to changing circumstances. It also has positive benefits for Network Rail incentives and helps to ensure operators are protected against losses for delays beyond their control. It has the same key disadvantage as updating TOC payment rates – administrative costs. This disadvantage is potentially greater for updating Network Rail payment rates as it might involve complex econometric modelling to update elasticities.

6 Recommendations

In this section we provide our recommendations for reform to Schedule 8 benchmarks and payment rates for the passenger regime. The recommendations are based on the preceding analysis.

6.1 Recommendation for benchmarks

In an ideal world, Schedule 8 benchmarks at PR23 for both Network Rail and operators would be based on a forward-looking trajectory that reflected traffic projections for CP7. This hypothetical option performs better than any alternative policy option for setting benchmarks in advance that we have assessed in this report, though the accuracy of benchmarks set with this approach depends on the accuracy of CP7 traffic projections. However, we understand that basing TOC benchmarks on a forward-looking trajectory is not being considered as an option for CP7. TOC benchmarks based solely on historical performance may lead to inaccurate benchmarks that lead to large financial flows between Network Rail and operators, because the level of traffic on the network during the data period used for recalibration may not reflect traffic levels during CP7.

Our recommendation is to introduce a system for a recalibration of benchmarks during CP7. Benchmarks for Network Rail would be set at PR23 using historical performance and the PR23 performance trajectories, and would be set for the full duration of CP7. TOC benchmarks would be set on historical performance. If outturn traffic in any year of CP7 deviates from traffic in the data period used for the PR23 benchmark recalibration by more than a specified percentage, this should trigger consideration by ORR of whether there should be a recalibration of Schedule 8 benchmarks. We do not make a recommendation as to what the threshold for triggering a recalibration should be – this would a policy question for ORR to determine. Once triggered, ORR would consider the nature of the shock to traffic (i.e. temporary or permanent) that caused the triggering and then determine if a mid-period recalibration is warranted.

This mid-period recalibration would be likely to use a different data period from the initial PR23 recalibration, with the data period to be decided at the time of the recalibration based on an assessment of which years appear likely to be the most representative of future performance levels.

We recommend this option because it allows benchmarks to be flexible to changing circumstances over CP7, while avoiding the negative impacts upon incentives and financial flows that would result from the options involving annual benchmark updates. By using an exogenous measure (traffic levels) as the trigger for a recalibration, it avoids perverse incentive affects that would result from having a trigger endogenous to the performance of Network Rail and/or operators.

We note that there is also the possibility of performance trajectories changing during CP7. We would recommend that any decision to change performance trajectories should be taken as a trigger for ORR to consider also recalibrating benchmarks. If benchmarks are recalibrated in such a scenario, they should incorporate the new performance trajectories to reflect ORR's updated position on the outputs Network Rail is expected to deliver in terms of train performance.

In a scenario in which benchmark recalibration is carried out but performance trajectories are left unchanged, ORR would need to consider at the time whether or not the existing performance trajectories were too outdated to incorporate into the new benchmarks.

6.2 Recommendation for payment rates

We recommend that an option to recalibrate TOC and Network Rail's payment rates alongside a mid-period benchmark recalibration is introduced.

If the incremental cost of updating the TOC Responsibility Matrix while recalibrating benchmarks is relatively low, this addition may be justified because it sharpens TOC incentives, provides information on the cost of delay and increases Schedule 8's flexibility to changing circumstances.

In deciding whether to update Network Rail's payment rates, ORR would need to think about whether the change in industry circumstances is likely to have significantly changed the MRE of poor performance. Examples of circumstances in which the MRE of delay might change significantly during CP7 could include the following:

- Traffic recovery post-COVID could increase the amount of revenue that there is to lose when passengers get put off by delays.
- The shift to home-working / hybrid working arrangements is likely to have made commuter traffic much more sensitive to delay than previously, potentially increasing the MRE of delay. If there is a shift back to office-based working during CP7 then this effect may reverse.
- The effect of delays on revenue may be non-linear. For example, there may be little effect for delays up to a certain threshold, but large effects once delays go beyond this point as passengers become "fed up" with using trains. If increased traffic on the network during CP7 leads to increased delays, then it could move the industry beyond the threshold at which the previously-estimated MRE of delay is applicable.

In circumstances in which the MRE of delay may have changed significantly, ORR would need to reach a judgment on whether the potential benefits of updating Network Rail payment rates to reflect this change justify the additional costs of including Network Rail payment rates in the recalibration. These costs could include the cost of updating the semi-elasticities that feed into the MRE calculation if new survey evidence from the Passenger Demand Forecasting Council is available. Network Rail's payment rate could also be updated based on new revenue data, even if the elasticities are not changed.

We also recommend that if ORR decides to recalibrate Network Rail payment rates alongside a mid-period recalibration of benchmarks, it should also recalibrate TOC payment rates (including the TRM) at the same time. We do not think it would be feasible to update Network Rail's payment rates and not update TOC payment rates without risking imbalance in the star model.

Appendix 1: Summary of PR18 Framework for Mid-period Recalibrations

This appendix sets out a summary of ORR's PR18 guidance on the circumstances under which it would consider a within-control period recalibration of Schedule 8 parameters for passenger operators for CP6.

Process for proposing recalibration of Schedule 8 during a control period

Schedule 8 of track access passenger contracts contains provision for either Network Rail or passenger operators to propose changes to Schedule 8 parameters during CP6. Any change in the arrangements between two parties requires sign off by ORR. The process begins with one party proposing a change to the other party. There are then two possibilities:

- The other party agrees with the change, in which case ORR approval can be sought; or,
- The other party disagrees, in which case either party can refer the matter to ORR to determine themselves or refer the matter for resolution under the Access Disputes Resolution Rules.

Types of within-period recalibration

In its guidance, ORR sets out three broad categories of within-period recalibrations, with the categories based on the scope of the recalibration and the reason the request has been made. The three recalibration types are:

- Type 1 "Basic" recalibrations
- Type 2 "Large-scale basic" recalibrations
- Type 3 "Forecast uncertainty" recalibrations

A Type 1 "Basic" recalibration is where there has been a material change in circumstances for an operator or operators, such as franchise remapping. The recalibration applies only to directly affected operators.

A Type 2 "Large-scale basic" recalibration is where there has been a material change in circumstances that affects all operators or impacts upon the star model. The recalibration would cover all operators. One circumstance that could lead to a Type 2 recalibration is a significant change to delay attribution practices that causes a significant reclassification of delay. This could warrant a within-period recalibration to ensure benchmarks are aligned with expected performance under the revised delay attribution approach. Another circumstance is a significant change in traffic on the network causing material imbalance in the star model. In either circumstance, the guidance indicates ORR would be minded to approve applications (agreed or disputed) for recalibration of affected Schedule 8 parameters that are proportionate and consistent with the principles of Schedule 8.

A Type 3 "forecast uncertainty" recalibration is intended to "correct" benchmarks during the control period to better reflect expected performance. An application for this type of recalibration needs to demonstrate that misalignment between benchmarks and outturn performance is a result of uncertainty in forecasting, rather than good or bad performance. Off-benchmark performance resulting from an acknowledged and accepted weakness of the recalibration methodology is not grounds for a Type 3 recalibration. Even if it is demonstrated that forecast uncertainty explains the difference between benchmarks and performance, ORR's guidance indicates that other considerations may lead it to decide not to approve a recalibration. In particular, the guidance states that consideration would be given to the impact a recalibration would have on any relevant franchise settlement.

Increases to scheduled journey times and/or introduction of timetable differentials

ORR's guidance document also acknowledges that within-period recalibration of Network Rail benchmarks may be appropriate if there are non-routine changes to scheduled journey times or significant timetable differentials.

A material increase in a scheduled journey time makes it easier for Network Rail to meet its performance target and thus increase the bonus payments a train operator has to pay to Network Rail (with the opposite true for a reduced journey time). Further, a timetable differential[11] could make it easier for both TOCs and Network Rail to achieve their targets. If differentials are significant, the guidance indicates that parties should consider a recalibration to ensure Schedule 8 benchmarks remain as challenging as intended.

A request for recalibration in these circumstances would need to demonstrate a change in the party's net financial position as a result of the change in scheduled journey times and/or introduction of timetabling differentials. In the case of an increase in scheduled journey times, it would also need to be demonstrated that previous journey times were not achievable.

Basis for within control period recalibrations

The guidance notes that for any within-CP6 recalibration, the assumption is that the evidence base used for the initial PR18 recalibration would serve as the basis for the within-period recalibration, unless there is compelling reason not to use this evidence base. It also notes that changes to the benchmarks should not be sought where an investment made by one of the parties has led to improvements in expected performance, as this would distort the incentives of the regime.

[11] A timetable differential is where there is a difference between the times shown in the working timetable and the times shown in the public timetable. Timetable differentials can be appropriate in certain circumstances, for example, where half minutes in the working timetable are rounded up to whole minutes in the public timetable.

Appendix 2: Summary of Consultation Responses

This appendix summarises the responses to ORR's April 2022 technical consultation on the Schedule 4 possessions regime and the Schedule 8 performance regime, focusing on the responses relevant to our work on mid-control period updates to Schedule 8 parameters. The note begins by recapping ORR's position before summarising stakeholder responses on each issue.

Network Rail gave a very detailed (and critical) response that addressed each policy proposal in turn, and an overall opinion on ORR's proposed approach to the recalibration of the passenger Schedule 8 regime. Other stakeholders typically focused their responses on the proposals that ORR was minded-to take forward, which in the case of Schedule 8 was only Proposal G (freight rate recalibration).

Recap of ORR's position

In its April 2022 consultation paper, ORR set out the following minded-to positions on Schedule 8:

- Schedule 8 will continue to apply between Network Rail/GBR and all operators in CP7. Following legal analysis, ORR was of the opinion that proposed alternatives to the current regime (an opt-out mechanism, setting zero payment rates or side-agreements between Network Rail/GBR and operators) were all incompatible with the current legal framework. ORR noted that if new legislation was passed that relaxed the requirement for a performance scheme, Schedule 8 payments may no longer need to be made between GBR and its operators, and potentially operators contracted to devolved bodies. ORR was also open to new proposals for alternative arrangements within the current legal framework, although it stated that these would need to be settled by autumn 2022 to be reflected in ORR's PR23 decision.
- Only incremental changes will be made to Schedule 8 at PR23**.** The possibility of legislative change affecting the application of Schedule 8 for some passenger operators lowered the expected benefit of any changes. Further, ORR considered that the impact of the COVID-19 pandemic on reference data was likely to make the Schedule 8 recalibration particularly challenging, such that there was an advantage in avoiding changes that might complicate the process.
- ORR would consider updating its guidance on mid-control period recalibrations**.** After PR18, ORR issued guidance on the circumstances under which it would consider a mid-control period recalibration of Schedule 8 for passenger operators, including a recalibration based on forecast uncertainty as a way to "correct" benchmarks during a control period to better align them with expected performance (see Appendix 1). ORR considered that uncertainty around traffic recovery from the COVID-19 pandemic might make an updated recalibration mechanism appropriate during CP7.

ORR then set out its updated positions on each of the proposals from its initial consultation. The positions relevant to updating Schedule 8 parameters are provided in the table below.

Table 11: Summary of ORR's initial Schedule 8 proposals and minded-to positions

Initial Proposal	Minded-to position
Proposal A Change the way Network Rail's benchmarks are set, basing them only on historical data	**Minded not to take forward** Keeping the current approach is proportionate and keeps consistency with the wider PR23 settlement. The current approach is contingent on production of performance trajectories.
Proposal B Update benchmarks annually to make them more flexible during control periods, either through: *Option B1*: average of rolling historical performance, or *Option B2*: modelled approach	**Minded not to take forward** Keeping the current approach keeps intact incentives to improve performance. While ORR did not envisage taking forward the proposal, due in part to concerns about its incentive effects, it stated that the recalibration working groups would consider the merits of more frequent updates to Schedule 8 parameters. ORR stated that it would take these discussions into account when making a decision on this proposal.
Proposal D Change how TOC-on-TOC delay is handled within Schedule 8 to address an existing gap in TOCs' incentives, either through: *Option D1:* full TOC-on-TOC delay measure, or *Option D2:* annual update of TOC responsibility matrix	**Minded not to take forward** Keeping the current approach keeps intact incentives to improve performance. While ORR did not envisage taking forward the proposal, due in part to concerns about its incentive effects, it stated that the recalibration working groups would consider the merits of more frequent updates to Schedule 8 parameters. ORR stated that it would take these discussions into account when making a decision on this proposal.

Network Rail's view on ORR's recalibration approach

Network Rail is concerned by ORR's proposed approach to the Schedule 8 recalibration for CP7. It strongly disagrees with ORR's decision to follow a "traditional" recalibration approach. Its concerns relate mostly to the passenger Schedule 8 regime, as it feels many of the issues are less significant for freight.

Network Rail stated that, with the exception of 2019/20, the remainder of CP6 performance and revenue data has been severely affected by the impacts of COVID-19. Network Rail added that the industry is still recovering from COVID-19 and is far from settled into a "new normal". It added that this issue is compounded by the most recent data, which is least COVID-affected, being impacted by industrial action that would leave several periods of this year's data unusable for recalibration. Network Rail is concerned that using CP6 data as a base for recalibrating Schedule 8 would significantly increase the risk of an inaccurate recalibration, which would in turn increase the likelihood of Network Rail or operators seeking mid-period recalibrations. Network Rail stated that mid-control period recalibrations of this sort are costly, time-consuming and often complex. This will reduce funding available to the industry for other, important projects, and distract management time. Network Rail acknowledged, however, that mid-period recalibrations can be entirely appropriate, and gave the upcoming delay attribution review as an example of a circumstance where a recalibration could be necessary.

Network Rail believes that ORR is wrong to rule out its initial proposals on the basis that they would overcomplicate the recalibration. It argues that that there are a number of proposals which would instead

simplify the recalibration and asks that ORR reconsiders these. It gives Proposals A, B and D as examples of ways to greatly improve recalibration outcomes for all parties. Network Rail encouraged ORR to take early decisions on these recalibration issues, after setting out various time constraints for the recalibration, so that there is sufficient time to implement decisions and "avoid unnecessary industry rancour".

Responses to specific issues

Below, we provide a summary of consultation responses to ORR's positions on the three proposals in the above table. In each case, Network Rail's response is summarised separately, given the high level of detail in its consultation response relative to responses from other stakeholders.

Responses to decision to not take forward Proposal A (historical benchmarks)

ORR is minded to not change the way Network Rail's benchmarks are set by basing them only on historical data.

Network Rail and GBRTT response

Network Rail disagrees with ORR's position. It supports Proposal A because it removes the need to forecast performance up to 7 years ahead of time, at a point where performance is very uncertain as the railway recovers from the impacts of the COVID-19 pandemic. Network Rail argued that it had demonstrated the inaccuracies of a forecasting approach with the evidence it supplied to ORR's initial consultation, and that it had also demonstrated how a historical approach could reduce large payment swings. It considers that an accurate prediction of future Network Rail performance is likely to be impossible to create, and that any performance trajectory would be likely to need to adapt to changing circumstances as the industry reaches a new normal post COVID-19. Network Rail believes that the current process of using the Final Determination trajectories results in inaccurate and inappropriate incentives. It argued that this can cause misalignment within the control period between performance levels that are actually being targeted (i.e. the annual performance targets), and those that are set through Schedule 8 (using the Final Determination trajectories). It stated that aligning the Schedule 8 benchmarks with the annual performance targets would ensure that incentives were aligned and that all industry parties knew what level of performance was being targeted.

Other responses

There was little in the way of direct responses from other stakeholders to ORR's minded-to position to not take forward this proposal, but many stakeholders expressed support with ORR's position to only take forward Proposal G.

Responses to decision to not take forward Proposal B (annual updates)

ORR is minded to not take forward annual updates of Network Rail benchmarks.

Network Rail and GBRTT response

Network Rail strongly disagrees with ORR's position. It argues that setting benchmarks for the entirety of CP7 will be incredibly difficult with historical data that is affected by the COVID-19 pandemic, and any attempt to do so would lead to unrepresentative parameters. It considers that taking a more flexible approach to benchmarks in CP7 is essential, and that any attempt to set benchmarks for the entire control period will lead to inaccurate results and therefore (potentially large) perverse payments.

It argues that Proposal B would help to minimise the importance of selecting an appropriate recalibration period, as the results from that recalibration period would only be in place for the first year of CP7. Network Rail recognises ORR's concerns over the incentive effects changing benchmarks annually might have, but does not think that setting benchmarks for the whole period would set better incentives due to the difficulty in

determining a base period for such a recalibration. Network Rail strongly considers that an annual, mechanistic change to the benchmarks would be the least-worst option for CP7.

Network Rail would like to work closely with ORR and industry to determine a "more suitable approach" over the coming months, preferably involving annual updates, but recognising that a one-off mid control period recalibration would be preferable to a fixed 5-year approach.

Other responses

There was little in the way of direct responses from other stakeholders to ORR's minded-to position to not take forward this proposal, but many stakeholders expressed support with ORR's position to only take forward Proposal G.

SouthEastern expressed disappointment that this proposal has not been taken forward, but acknowledged the financial and administrative burden this would have placed on the industry. It added that it would support use of mid-control period recalibrations in the event of volatility and forecast uncertainty.

TfL believes that it is important to be able to utilise existing mechanisms or reopeners to re-adjust Schedule 8 parameters to address emerging discrepancies, such as when performance sums are clearly disproportionate to revenue effects. It states that there is uncertainty inherent in forecasting substantial change with little historical data, and that Network Rail/GBR must be ready to act to correct any errors, particularly when bespoke approaches are adopted for regime calibration (necessary at times for substantial service changes such as full Elizabeth Line timetable implementation).

Responses to decision to not take forward Proposal D (change TOC-on-TOC delay)

ORR is minded to not update the TOC-on-TOC responsibility matrix annually. The other option for this proposal, a full TOC-on-TOC regime, was deemed infeasible.

Network Rail and GBRTT response

Network Rail disagrees with ORR's position. Network Rail agrees that the current PEARS system means that a full TOC-on-TOC regime is unfeasible, but supports more frequent updates of the responsibility matrix to recognise changes in the TOC-on-Self to TOC-on-TOC relationship as the rail sector recovers from the pandemic. Again, Network Rail thinks this proposal has the benefit of reducing the importance of the choice of recalibration period.

Network Rail suggested that ORR instructs its consultants to produce a model that can be mechanistically updated each year with the latest available data, an approach Network Rail considers would minimise the impact of the COVID-19 pandemic on the recalibration. It also sees this approach as pragmatic in regard to wider industry reform, which could lead to GBR operators opting-out of Schedule 8 just before, or during, CP7. This approach would mean that parameters for those operators would not need to be updated annually, reducing cost and complexity.

Other responses

There was little in the way of direct responses from other stakeholders to ORR's minded-to position to not take forward this proposal, but many stakeholders expressed support with ORR's position to only take forward Proposal G.