

Caitlin Scarlett
Schedule 8 recalibration lead
Rail Delivery Group
Caitlin.Scarlett@raildeliverygroup.com

Deren Olgun
Principal Economist
Office of Rail and Road
deren.olgun@orr.gov.uk

2 February 2018

Dear Deren

Network Rail Payment Rates: Submission to ORR from the revenue sub-group of the Schedule 8 recalibration working group

The purpose of this letter is to request ORR approval and determination, where appropriate, on the two outstanding issues in the methodology of the recalibration of the Schedule 8 Network Rail Payment Rates for CP6. This letter is sent on behalf of the revenue sub-group of the Schedule 8 recalibration working group.¹

Context

At the RDG Schedule 8 recalibration working group on 15 January 2018, two outstanding areas in the methodology for the recalibration of the Schedule 8 Network Rail Payment Rates were discussed:

1. Whether to use the findings from the Oxera study “The impact of unplanned disruption on train operator revenue” for the recalibration of Marginal Revenue Effects (MREs) for London and South East commuter flows only; and
2. Which set of Delay Multipliers to use in the recalibration of the Network Rail Payment Rates, for all flows not affected by (1), above.

Noting the complexity of the issue, the RDG Schedule 8 recalibration working group agreed to delegate authority for the decision on these two outstanding areas to the revenue sub-group. The revenue sub-group met to discuss these issues on 25 January 2018, and all members of the Schedule 8 recalibration working group were invited to attend. The following companies were represented: Network Rail, Govia Thameslink Railway, Great Western

¹ The revenue sub-group was formed to discuss technical aspects of the Schedule 8 recalibration, in particular the recalibration of the Network Rail Payment Rates. The revenue sub-group consists of representatives from train operating companies, Network Rail and ORR.

Railway and South Western Railway (by correspondence). ORR also attended so that it could better understand the reasons behind the decisions made at the sub-group, and to provide advice on next steps.

This letter summarises the decision of the revenue sub-group on each of the two outstanding areas. Where the revenue sub-group has been unable to reach agreement on an area, representations from industry parties are included in the annexes to this letter for ORR's consideration.

Revenue sub-group decisions

The revenue sub-group came to the following conclusions on each issue:

1. Whether to use the findings from the Oxera study for the recalibration of MREs for London and South East commuter flows.

The revenue sub-group concluded that the findings from the Oxera study **should** be used in the recalibration of MREs for the London and South East commuter flows. The flows which should be recalibrated on this basis are the PDFH-defined flows "London TravelCard Area" and "South East to/from London". The Network Rail Payment Rates will therefore be calculated according to the following methodology.

Formula:

$$NRPR_{Service\ Group} = \sum MRE_{Service\ Code} \div Annualisation\ Factor_{Service\ Group}$$

where:

$$MRE_{Service\ Code} = \sum MRE_{Flow}$$

$$Annualisation\ Factor_{Service\ Group} = \sum Busyness\ Factors_{Service\ Group}$$

For the London and South East commuter flows (defined above), the flow MRE equals the Flow Revenue multiplied by the exponential of the Semi-Elasticity for that Sector and Ticket Category, multiplied by the incremental change in Performance Minutes.

$$MRE_{Flow} = Revenue_{Flow} \times [EXP (Semi-Elasticity_{Sector, Ticket\ Category} * -1) - 1]$$

The Oxera study provided Semi-Elasticities for both of the flows combined (i.e. one set of Semi-Elasticities for both the London TravelCard Area and South East to/from London). In calculating the MREs for these two flow types, we propose instead using Semi-Elasticities which have been calculated separately for the London TravelCard Area and South East to/from London. The SDG Peer Review of the Oxera study provides these disaggregated Semi-Elasticities.

The MREs for flows not captured by the above will be calculated according to the CP5 methodology, as per previously agreed at the Schedule 8 recalibration working group and agreed, in principal, by ORR.

A detailed methodology will be produced by SDG which will set out the Network Rail Payment Rate calculation in more detail, including the process by which revenue is assigned, and by which flows are assigned to the PDFH sectors. This note will also include a detailed explanation of the CP5 methodology, which will be used for all flows other than those falling into the categories of the London TravelCard Area and South East to/from London.

2. Which set of Delay Multipliers to use in the recalibration of the Network Rail Payment Rates, for all flows not affected by (1).

The revenue sub-group **did not agree** on which set of Delay Multipliers should be used in the recalibration of the Network Rail Payment Rates. Several options were discussed at the revenue sub-group meeting on 25 January 2018:

- a. Using the Delay Multipliers as set out in PDFH v5.1 (i.e. the Delay Multipliers used in the CP5 recalibration).

The meta-analysis which produced these Delay Multipliers ignored all studies which produced performance elasticities which were statistically insignificant.

- b. Using the Delay Multipliers which were originally issued alongside PDFH v6.0.

These Delay Multipliers are the unweighted average of the PDFH v5.1 Delay Multipliers, and a set of Delay Multipliers which takes evidence from the same studies as PDFH v5.1, but enters an elasticity of 0 for each study which produced a statistically insignificant result.

- c. Take an unweighted average of the Delay Multipliers in (a) and (b), above.

As agreement could not be reached at the revenue sub-group, it was agreed that each party that wishes to do so will make a submission to ORR detailing which set of Delay Multipliers should be used, and the reasons why. These submissions are provided within the annexes to this letter.

Next steps

The revenue sub-group asks that ORR:

- Approves its decision to use the findings from the Oxera study to recalibrate the MREs for London and South East commuter flows; and
- Considers each party's submission, set out in the annexes to this letter, and determines which set of Delay Multipliers should be used in the recalibration of all other MREs. We ask that ORR sets out the reasoning behind its determination.

As discussed at the revenue sub-group on 25 January 2018, **ORR's response to both of these issues is required by 16 February 2018**, so as not to delay the Schedule 8 recalibration work. As ORR is aware, the timescales for the Schedule 8 recalibration are incredible tight, and so any delays to ORR's response will have a material impact on the rest of the recalibration work. The revenue sub-group will therefore support ORR in its decision-making in any way that it can, for example by providing additional information if required.

Yours sincerely,

Caitlin Scarlett

ANNEX 1

GTR/GWR/SWR Joint Response to Delay Multiplier Decision

Background:

At the meeting of the Schedule 8 recalibration sub-group on 25th January 2018, a discussion of the choice of delay multipliers to use as inputs into the CP6 recalibration was unable to reach a consensus. This joint response from GTR (Go-Ahead/Keolis), GWR (FirstGroup) and SWR (FirstGroup/MTR Europe) represents our views as to the appropriate choice of delay multipliers. At this late stage, we understand there to be just two options, derived from studies all listed in the Wardman & Batley paper (2014):

1. Use the same Delay Multipliers as used during CP5, taken from PDFH 5.1, and based on a meta-analysis of those studies reporting statistically significant results; or
2. Use a straight average of two meta-analyses: one including only significant study results (as in Option 1) and one including both the significant and insignificant study results.

Other options were touched on – and in our opinion dismissed – at the meeting. Nevertheless, we wish to be clear that, to our mind, no option has been presented that is superior to those described above. To consider any further options at this stage (that have not been discussed more widely) would undermine the integrity of the process into which we have entered in good faith.

Summary:

We believe that the best and only justifiable option, without undertaking further research or analysis, is **Option 1**. In particular, we do not see any aspect in which Option 2 is preferable over Option 1. Our reasoning to support this conclusion is set out below under the following headings:

1. Lack of further evidence against current values;
2. Use of insignificant results;
3. Assumption of zero for insignificant results;
4. Averaging of the meta-analyses.

Given the timescales and dependencies inherent in the recalibration process, we do not believe that there is time to perform new research or undertake further analysis of existing studies in a robust and meaningful manner. Existing research must therefore be used, which – given our concerns regarding Option 2 – leaves Option 1 as the only acceptable position.

Supporting Comments:

1. Lack of further evidence against current values

Following the meta-analysis that informed the PDFH 5.1 Delay Multipliers used at PR13, the only new evidence on the demand effect of unplanned disruption is the OXERA study proposed for use for LSE flows. Indeed, no party involved in the recalibration sub-group has been able to supply further studies/evidence. The Wardman & Batley report that also produces Option 2 does not include any new evidence.

While it has been asserted in some quarters that Delay Multipliers are too high, there is no robust evidence to support this. In particular, there is no evidence to suggest that the CP5 Delay Multipliers are no longer fit for purpose. Although it could be inferred that this is a key conclusion of the Wardman & Batley paper, on page 1067 of that report they state that the use of the insignificant values “adds a significant element of uncertainty into [their] review”. We agree and believe the only defensible position – particularly statistically (see below) – is to reject the use of the Wardman & Batley paper. Consequently, there is no evidence to depart from the current set of parameters, which should remain as previously (i.e. adopting Option 1).

2. Use of insignificant results

Given the existence of numerous studies that do have significant results, we do not understand the basis for also using insignificant ones. There is substantial supporting evidence (e.g. NRPS and numerous disruption studies) indicating that train passengers treat unplanned delays with greater importance than planned journey time. Thus, we believe that studies with insignificant results are likely to be of an inferior quality: with poorer data sources and preparation, and less robust analysis. Consequently, these studies should not be used in any analysis, including the calculation of the demand multiplier: Option 2 must therefore be rejected.

3. Assumption of zero for insignificant results

In the Wardman & Batley meta-analysis, the insignificant results are included by taking them to be zero. Even allowing for our point 2 above, we disagree with this use of zero to “proxy” as the output of an insignificant study for use in the wider meta-analysis. Although the misconception is widespread, an insignificant result is not the same as a zero result. Insignificance simply implies that we cannot be confident that the estimated result did not occur by chance (and this is typically demonstrated by zero falling within the stated confidence interval). Clearly, an estimated value could be both ‘large’ and insignificant; zero would naturally be a very poor proxy in this case. Furthermore, such a choice of zero is completely arbitrary: even if zero falls within the confidence interval, any other number within the same interval could have been chosen with no less justification.

Moreover, not only is this use of zero mathematically wrong, the implication of such a finding to the Delay Multiplier itself would be that passengers are indifferent to arriving on time against arriving infinitely late.² This implication is clearly nonsensical.

If insignificant results *were* to be included in a meta-analysis (with which we disagree anyway), then we believe that mathematically there are only two acceptable options:

- a. Use each study’s central estimate – even though it is not statistically significant – since this is the best estimate the study can supply; or
- b. Use a value equivalent to a Delay Multiplier of 1, i.e. assuming passengers value unplanned journey time exactly the same as planned journey time.³ (We assert that statistically this should be the null hypothesis against which the studies were testing, thus being used when the result was found to be insignificant.)

However, neither of these approaches were adopted by the Wardman & Batley meta-analysis, meaning that Option 2 must be rejected, leading back to our conclusion of the use of Option 1.

4. Averaging of the meta-analyses

We do not understand how or why the creation of these figures is any more robust, justifiable or beneficial than those derived by Wardman & Batley’s meta-analysis including both significant and insignificant studies. We would therefore question why these “all studies” meta-analysis results should carry equal weight with Option 1 in this average. In effect, this weights the significant results twice against the insignificant results. However, a host of other relative weightings could be used with at least equal and possibly greater merit. It appears that this average has been created arbitrarily to provide ‘acceptable’ values due to the incredulous results generated by the “all studies” meta-analysis. Clearly Option 2 must be rejected. Moreover, since any choice of non-zero weighting to average the meta-analyses is necessarily

² Note that the Delay Multiplier describes the passenger’s relative weight of delay time to normal journey time. A weight of zero would therefore suggest that passengers place no value on delays, and would trade one minute of planned journey time for any amount of unplanned delay to their journey.

³ Note that a lateness elasticity of zero will always correspond to a Delay Multiplier of zero.

arbitrary, the arguments above apply no matter the weighting, and so any such average (e.g. an average of Options 1 and 2) must also be rejected in favour of the figures given by Option 1.

Concluding Comments:

In order to deviate from the accepted Delay Multipliers used at PR13, there must be significant cause and substantial new evidence; a subjective marginal “improvement” between two approaches does not warrant change. Moreover, it appears that Wardman & Batley’s research was approached under a given preconception, which seems to have influenced both their analysis and their conclusions. Indeed, the creation of the figures underpinning Option 2 is apparently principally due to the counter-intuitive results generated by Wardman & Batley’s new meta-analysis. Adopting such an approach instead of recommending further robust research does not benefit the industry, nor will it stand up in the long-term. It seems clear to us that such subjectivity would skew revenue analyses, as well as Schedule 8, preventing the regime from fulfilling its intended purpose. Therefore Option 2 should be rejected, in favour of Option 1.

Rob Moss
Head of Strategic Risk
Govia Thameslink Railway

Peter de Boeck
Finance Analyst
Great Western Railway

Verity Hinde
Senior Revenue Manager
South Western Railway

Please note: the views in this paper are endorsed by the respective franchise owning groups, as well as the TOCs listed above. Together these owning groups are responsible for the Greater Western, South Eastern, South Western, TSGN and Trans Pennine Express franchises, which together receive approximately 50% of GB Passenger Income (based on 2015-16 figures published by the ORR).

ANNEX 2

Network Rail's recommendation for Delay Multipliers in Schedule 8

Author: David Harding, Head of Analysis and Economics, Network Rail

Network Rail considers that the Delay Multipliers that should be used in the recalibration of the Schedule 8 Network Rail Payment Rates for CP6 are those that correspond with the elasticities in the latest version of the Passenger Demand Forecasting Handbook (PDFH v6.0). This will ensure that the most up-to-date, independent evidence is used. This is consistent with option (b), as set out in the main body of this paper.

To summarise, the Delay Multipliers presented alongside PDFH v6.0 are the unweighted average of the PDFH v5.1 Delay Multipliers (i.e. those which only include the figures from statistically significant results), and the Delay Multipliers which include the statistically insignificant results, assigned a value of 0.

The following sets out our rationale for not supporting PDFH 5.1 delay multipliers:

1. PDFH v5.1 is an upper bound of what the Delay Multipliers should be, because:
 - a. In updating the evidence base that underpins the Passenger Demand Forecasting Handbook, Wardman and Batley recognised that the recommendations they made in PDFH v5.1 for Delay Multipliers were likely to overstate the 'true' impact of delay on demand, and they have since revised them in their updated work for PDFH v6.0. The following is a direct quote from their independent study (Wardman and Batley 2014):

"summary measures based solely on the statistically significant evidence can be expected to lead to inflated mean values to the extent that insignificant elasticities even though, not necessarily zero in reality, will generally be lower than significant results."

There are other reasons why the delay multipliers are expected to be lower than in PDFH 5.1, other than those explicitly recognised by Wardman and Batley:

- b. The main study with a specific purpose of investigating performance impacts was Batley (2011) and this study returns some of the lowest elasticities across all studies in the meta-study. Other studies only included performance among many other variables that were investigated.
 - c. Evidence within the 'Revisiting the elasticity based framework' study, which was part of the meta-analysis, was largely ignored for PDFH updates across other areas of the handbook. This suggests that there is a lack of trust in this particular study. The meta-analysis itself notes that removing the 'Revisiting the elasticity based framework' study and a single outlier value would reduce the estimated elasticity from -0.13 to -0.10 for significant observations and from -0.09 to -0.08 for all observations. This study also estimated the elasticities using annual data when the

4-weekly data used in other studies would seem to provide a superior dataset for detecting performance impacts on demand as there will be greater variability in the performance variable. By using annual data you remove the sensitivity of the elasticities to within year fluctuations in performance.

- d. With the exception of this study and a single SDG observation, all AML elasticities in the meta-analysis imply responses between 0 and -0.06 (much lower than the average of all studies). These four studies all used 4-weekly or quarterly data.
- e. The median elasticity value of significant estimates is significantly lower than the mean, demonstrating that the proposed value is dragged upwards significantly by a small number of observations of higher elasticities. The median value may be a safer representation of the average effect where we are not confident in the extreme values. In this case, many of the highest values are from a particular study which uses annual PPM data, the limitations of which have been discussed above. Delay Multipliers can also be inferred from the PDFC-sponsored Oxera study into the revenue impact of performance on London commuter flows. These are also lower compared to those presented in PDFH v5.1, for the relevant sectors.

Network Rail considers that it is inappropriate to use Delay Multipliers which are shown to be an upper band of their 'true' values. Using Delay Multipliers which are too high would result in Network Rail Payment Rates which would create perverse incentives through Schedule 8, whereby Train Operators could be considered to financially 'prefer' Network Rail to cause disruption to their services. We recognise that Train Operators, of course, face other incentives to reduce delay outside of Schedule 8 (for example, reputation), however it is important that the Schedule 8 regime does not actively create these perverse financial incentives.

The following is our rationale for, instead, adopting our recommendation of using PDFH v6.0 Delay Multipliers:

1. Some concerns were raised at the revenue sub-group about the allocation of an elasticity of 0 to those studies which found statistically insignificant results. We do not consider that the allocation of a 0 value for insignificant results is a cause for concern. Typical interpretation of a statistically insignificant result is that the data does not provide us with reassurance, beyond reasonable doubt, that the elasticity is not 0. Put another way, a statistically insignificant result signifies that the evidence does not support the hypothesis that a relationship between the two variables in question exists (in this case, a relationship between performance and demand). Therefore, we consider that it is entirely appropriate in most cases to assume an elasticity of 0 for a statistically insignificant result –to use anything but 0 would mean that we are artificially imposing a relationship, when the evidence suggests there is none. Furthermore, it may even have been the case that initial estimates of performance elasticities that have been dropped from regression specifications were of the 'wrong sign', which strengthens the case for attributing an elasticity of 0 in the absence of further information.

2. Having said this, we recognise that in this case, the estimates including all the insignificant results are likely to be a lower bound, because there may also be other reasons that there is not enough evidence to demonstrate a relationship, such as a lack of variation in the independent variable.

The best estimate of the Delay Multipliers implied by the elasticities will lie somewhere between the two. The following is a further excerpt from the independent Wardman and Batley study, where “all observations” refers to the results which include the 0 values for all statistically insignificant studies:

“...the ‘true’ elasticity might be bounded at the upper end by the elasticity from Significant observations and at the lower end by the elasticity from All observations, given that non-significant elasticities can be expected to be less than significant elasticities but possible greater than zero.”

3. Therefore, we consider that the Delay Multipliers that correspond to the elasticities to be published in PDFH v6.0, which are half way between the PDFH v5.1 results and those which include a value of 0 for all insignificant results, provide suitable approximation of the ‘true’ Delay Multipliers which recognise all of the issues noted above, and address some of the concerns raised at the revenue sub-group. We also note that these results were produced by an independent study for the passenger demand forecasting council, which recognised the issues with the PDFHv5.1 Delay Multipliers. The study concluded that it is not appropriate to ignore the insignificant results, and that a lower value should be attributed to insignificant elasticities when compared to their full, reported values:

“It would be prudent to allow for these insignificant elasticities being lower rather than treating them as implicitly the same as significant elasticities.”

The Delay Multipliers reported in PDFH v6.0 achieve this, since they are an average of PDFH v5.1 (significant results only) and the results established from entering a value of 0 for all insignificant results. Effectively, this can be thought of as increasing the value of insignificant results above 0, to somewhere between 0 and the PDFH 5.1 value (thereby resulting in a lower value for the insignificant results, consistent with the recommendations of the independent study).

4. We consider that, while this approach may not result in Delay Multipliers which are precisely correct, it produces results which are in-between what we recognise to be the upper bound (PDFH v5.1) and the lower bound (those which include a 0 value for all statistically insignificant results).

Summary

Network Rail considers that it is inappropriate to use Delay Multipliers which are shown to be an upper band of their ‘true’ values. Using Delay Multipliers which are too high would result in Network Rail Payment Rates which would create perverse incentives through Schedule 8, whereby Train Operators could be considered to financially ‘prefer’ Network Rail to cause disruption to their

services. We recognise that Train Operators, of course, face other incentives to reduce delay outside of Schedule 8 (for example, reputation), however it is important that the Schedule 8 regime does not actively create these perverse financial incentives.

We consider that if the Delay Multipliers are too high, it will have the effect of unduly financially gearing the system. This would be perverse and ironic given that Schedule 8 is designed to de-risk franchises.

On this basis, we believe that there is strong evidence that the true elasticities and corresponding delay multipliers are lower than in PDFH 5.1 and the CP5 Schedule 8 regime, and that taking the recommendations of the most recent available independent evidence is the best and most defensible course of action.