Specification for Sustained Poor Performance (SPP)

SPP Arrangements for CP5

Final Report

January 2013

Redacted Version

This report has been redacted for reasons of commercial confidentiality.

Prepared for:
Network Rail
Kings Place
90 York Way
London
N1 9AG

Prepared by:
Steer Davies Gleave
28-32 Upper Ground
London SE1 9PD

+44 (0)20 7910 5000
www.steerdaviesgleave.com

CONTENTS

FIGURES

TABLES

APPENDICES

Executive Summary

The Sustained Poor Performance mechanism

1. The Sustained Poor Performance (SPP) mechanism is designed to provide protection to Train Operating Companies (TOCs) when performance falls to such a level that compensation under the standard Schedule 8 arrangements is considered to be inadequate. The mechanism provides for a TOC to claim for relevant losses, to the extent that these are not already compensated under the standard Schedule 8 arrangements, and may be invoked when performance averaged over 4 quarters (13 periods) is worse than a defined threshold.
2. Currently the SPP threshold is set so that additional compensation could be claimed when Network Rail performance is at least 10% worse than benchmark over a period of 12 consecutive months.

Consultation on Schedules 4 and 8 possessions and performance regimes

3. The Office of Rail Regulation (ORR) has released an industry consultation on Schedule 4 and 8 possessions and performance regimes (26th November 2012). ORR has stated in this consultation that it considers that at the current threshold, SPP is too easily triggered and is therefore minded to increase the threshold. The consultation invites respondents to give a view as to the level at which the threshold should be set.

Recommendation for SPP threshold

4. Our recommendation is that the SPP threshold should be raised from 10% of benchmark to 30%.

Rationale for recommendation

5. This recommendation is based on:

- A consideration of the purpose of the SPP mechanism;
- A review of the evidence for the effect of poor performance;
- A review of the history of SPP claims in CP4;
- An assessment of the risk to TOCs associated with the threshold; and
- An assessment of the risk to Network Rail associated with the threshold.

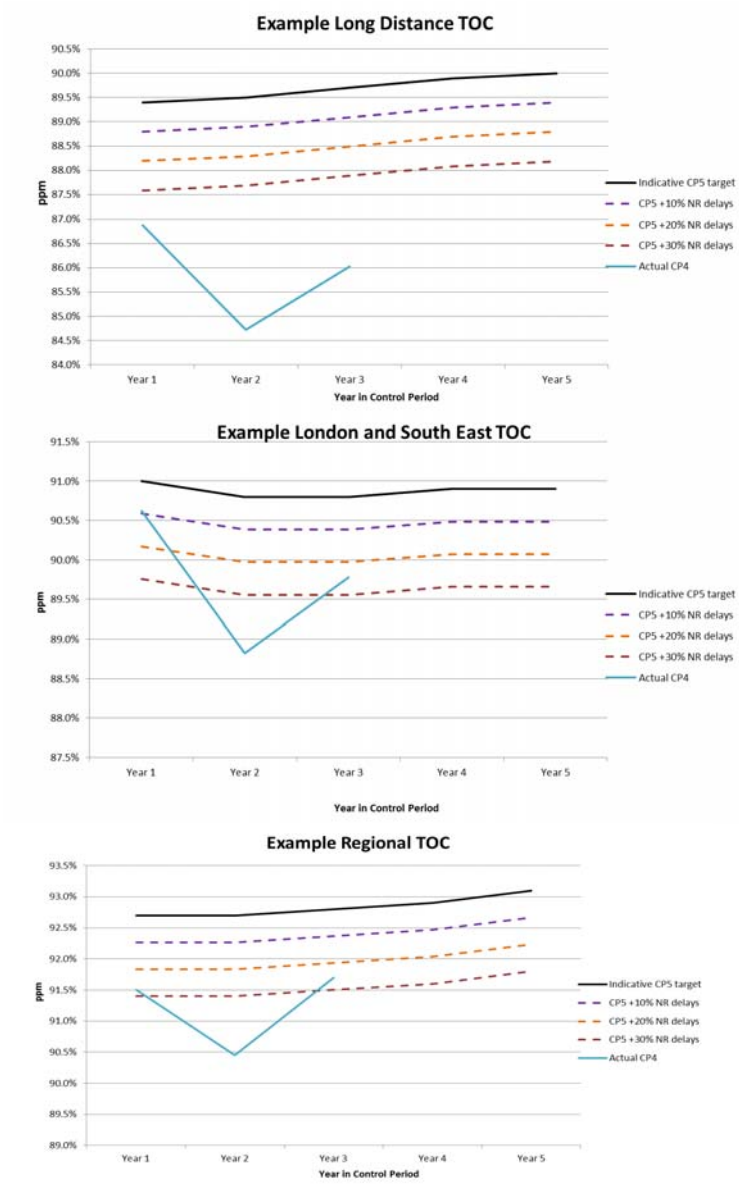
6. The justification for setting the benchmark at the 30% level is that:

- At this level, the risk to TOCs of suffering losses significantly in excess of the standard Schedule 8 compensation is small;
- Below this level there is a risk that the standard Schedule 8 regime would be undermined, with Network Rail facing a significant number of claims in instances where it is performing at benchmark in aggregate; and
- The level is broadly consistent, in terms of the performance levels at which SPP would be triggered, with previous threshold levels in CP3. It was therefore accepted at that time that standard Schedule 8 payments were adequate at these levels of performance.

Risk to TOCs

7. The SPP threshold is intended to represent the level of poor performance where compensation under the standard Schedule 8 arrangements is materially less than what is needed to reflect the actual impact on the train operator. This implies that there is a non-linearity between performance and revenue whereby, in theory at least, a worsening beyond a given threshold potentially affects passengers by a greater proportion than within that threshold, or that significant additional costs arise when performance falls below a certain threshold. A review of the Passenger Demand Forecasting Handbook found no evidence to indicate where, if at all, non-linearities might start to be experienced. Furthermore, no claims were made in CP4 based on the effects of such non-linearities on revenue, [REDACTED]. The evidence suggests that, unless there is a collapse in performance levels, the compensation under the standard Schedule 8 regime is adequate.
8. The graphs below illustrate that if performance in CP5 were below benchmark to the extent that SPP were triggered at a 30% threshold, PPM would still be of the same order as has been experienced during CP4 (the graphs show actual, rather than target, performance in CP4).
9. With a 30% threshold, sustained very poor performance on just one third of a TOC's service groups is likely to trigger SPP for the TOC.

FIGURE 1 PPM IN CP4 AND CP5



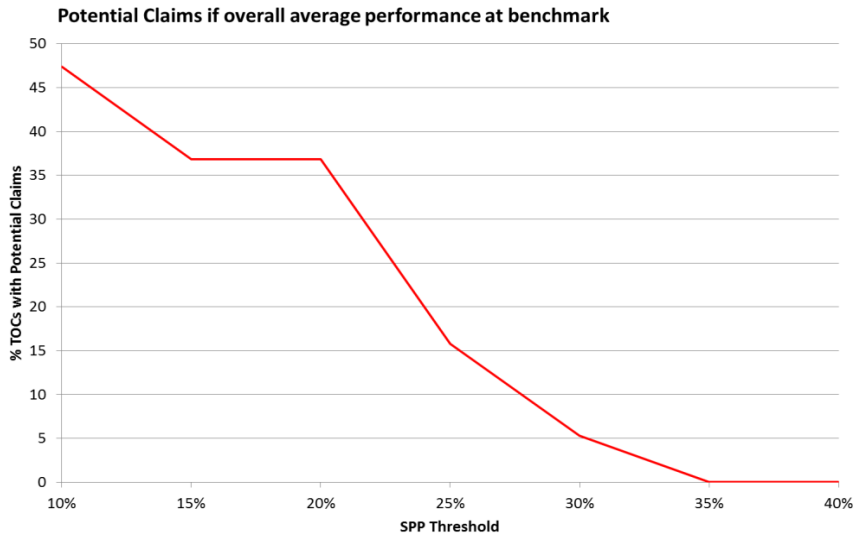
10. There is no evidence that performance at the levels currently experienced (and hence what might be experienced by TOCs just within the threshold were it set at 30% for CP5) causes an increase in TOC costs or a reduction in revenues beyond the levels compensated by the Schedule 8 formulae.

11. [REDACTED]

Risk associated with number of claims

12. A key strength of Schedule 8 is its liquidated sums nature which is simpler and less costly to administer than a bespoke claims process. This would be undermined if claims could too readily be triggered under SPP. The graph below shows the percentage of TOCs which would have been eligible to make a claim during 2011/12 if Network Rail were performing at benchmark levels in aggregate in 2010/11 and 2011/12 (but with the same degree of variability - as observed in those years - between service groups and from one period to another.)

FIGURE 2 POTENTIAL CLAIMS



13. The graph shows that where Network Rail is performing at benchmark in aggregate there would be a substantial risk of claims, with attendant significant costs in terms of management time, if the threshold were set at a level below 30%.

14. There is also a significant financial risk to Network Rail associated with SPP claims; this risk reduces as the SPP threshold is increased.

Consistency with former SPP threshold levels

15. During CP3 the SPP threshold was set between 25% and 20%.Performance benchmarks have become more demanding over time, therefore in terms of the level of performance which would trigger SPP, a level of 30% in CP5 would correspond to a threshold of less than 20% in CP3. This is illustrated by the fact that lateness at 30% above 2012/13 benchmarks is less, for all but one service groups, than lateness at 20% above 2009/10 benchmarks.

1 Introduction

Purpose and conclusion of the report

1.1 The Sustained Poor Performance (SPP) mechanism is designed to provide protection to Train Operating Companies (TOCs) when performance falls to such a level that compensation under the standard Schedule 8 arrangements is considered to be materially less than what is needed to reflect the actual impact on the train operator. The mechanism provides for a TOC to claim for relevant losses, to the extent that these are not already compensated under the standard Schedule 8 arrangements, and may be invoked when performance averaged over 4 quarters (13 periods) falls below a defined threshold.

1.2 The SPP threshold is currently set at 10% above (worse than) the Network Rail performance benchmarks defined in each TOC's Track Access Agreement (TAA). At the current threshold and current performance levels the majority of TOCs could make a claim under the SPP mechanism. Even if the effect of deviations from regulatory targets is accounted for, a substantial proportion of operators remain above the SPP threshold. As such Network Rail considers that the threshold is therefore set too low.

1.3 The Office of Rail Regulation (ORR) has released an industry consultation on Schedule 4 and 8 possessions and performance regimes (26th November 2012). ORR has stated by means of this consultation that it too considers that the threshold is too low and is therefore minded to increase it. The purpose of this report is to recommend an appropriate evidence-based level for the threshold, which Network Rail will propose as part of its response to the consultation.

1.4 Our conclusion is that the most appropriate level of the threshold is 30%. This level provides a balance between protecting TOCs against sustaining revenue losses for which they are not able to claim compensation and protecting Network Rail from the costs associated with a large number of claims which could be triggered through variability in performance, even when performance is at benchmark in aggregate.

1.5 The report is structured as follows:

- Section 1 describes the current SPP arrangements;
- Section 2 reviews the SPP claim history in CP4;
- Section 3 examines evidence in PDFH for the effects of poor performance which the SPP mechanism is designed to address;
- Section 4 considers the impact on the potential number of claims of varying the threshold, and examines the risk to Network Rail;
- Section 5 addresses the effect on TOCs of changing the SPP threshold, and
- Section 6 provides a comparison with other regimes.

A summary of the key findings is provided at the beginning of each chapter.

The current performance compensation regime

Schedule 8

1.6 Schedule 8 performance payments are designed to compensate TOCs for the loss in fare revenue in the event of unplanned disruption, and to incentivise Network Rail (and TOCs) to improve performance. Performance is measured against agreed benchmarks based on the forecast or 'expected' performance, which for CP4 is set out in Network Rail's CP4 Delivery Plan. These benchmarks are calculated for each TOC for each period, each quarter and each year.

1.7 Schedule 8 payments are based on a liquidated sums regime, whereby compensation payment rates are determined in advance using pre-determined formulae. The payment regime is re-calibrated periodically, typically prior to each Control Period, so that payments reflect revenues foregone as accurately as possible[1].

1.8 The calculation of the revenue impact of performance changes that underpins Schedule 8 Payments is as follows:

MRE (Marginal Revenue Effect per passenger minute for a specific flow) = RDE / T, where:

- R is revenue per journey
- D = delay multiplier
- E = elasticity
- T = Generalised Journey Time (GJT)

1.9 Under this formula, the revenue per journey is based on the calibrated payment rates, while the elasticity applied varies by ticket type (seasons and non-seasons) and type of journey (by geography and direction) reflecting the responsiveness of different passengers to changes in generalised cost. The value of the delay multiplier depends on the journey type and ticket type, and is based on the values set out in the Passenger Demand Forecasting Handbook (PDFH). A delay multiplier of X means that a minute's delay is valued by passengers (and weighted in demand forecasting) at X times the level of a minute of scheduled journey time. The effect of the formula overall, and the delay multiplier specifically, is that Schedule 8 payments vary proportionately with the change in lateness.

1.10 ORR has indicated in its Consultation on Schedules 4 and 8 that it is minded to continue to set Schedule 4 and 8 payment rates so that it compensates train operators for the full financial impact of service disruption due to Network Rail and other operators, where it does so currently. The analysis in this report is based on the assumption that this will be the outcome in CP5; it should be noted that the conclusions would be different were Schedule 8 rates to be set at a lower value.

Sustained Poor Performance mechanism

1.11 The SPP mechanism is intended to compensate TOCs (and incentivise Network Rail) where performance falls below a threshold which:

[1] A re-calibration exercise was undertaken in 2005 by AEA Technology (and took effect on 1 April 2006) to inform CP4, and Halcrow has been commissioned to update the Schedule 4 & 8 payment rates to inform CP5.

"should represent the level of poor performance where compensation under the standard schedule 8 arrangements is materially less than what is needed to reflect the actual impact on the train operator." (PR08 ORR Determination, p358)

1.12 Currently the threshold is set at 10%, so that a claim may be made if the average Schedule 8 payment made by Network Rail over 13 periods exceeds that which would be paid if performance were consistently 10% worse than benchmark. In principle, the threshold is intended to represent the point at which, in theory at least, a worsening of performance potentially affects passengers' travel behaviour to a greater extent than is assumed for the purposes of the Schedule 8 regime.

1.13 The standard TAA (Passenger Services), p182 states that:

"Network Rail shall indemnify the Train Operator against all Relevant Losses in accordance with this paragraph 18 [paragraph 18 describes the Schedule 8 payments] if, and to the extent that, the Average Periodic Liability shows Network Rail has exceeded (that is, equalled or been worse than) the relevant SPP Threshold."

1.14 Relevant losses are defined as:

"all costs, losses (including loss of profit and loss of revenue), expenses, payments, damages, liabilities, interest and the amounts by which rights or entitlements to amounts have been reduced, in each case incurred or occasioned as a result of or by such breach".

1.15 The 'average periodic liability' represents the Network Rail Performance Sums (NRPS) which provides the benchmark level against which the threshold is compared. The demonstration of SPP is based on the performance over the preceding 4 quarters (one year), measured over the preceding year from periods 1, 4, 7 and 11, i.e. in quarters ending in periods 3, 6,10 and 13.

1.16 The onus is on the TOC to demonstrate losses above this threshold, based on revenue foregone and / or net increase in operating costs.

Basis for claiming

1.17 The ORR consultation on Schedule 8 states that the Schedule 4 and 8 Industry Group has suggested that there is not enough clarity in the contractual wording as to what 'relevant losses' can be claimed[2]. Specifically, it has been argued that it is not clear whether claims can be made for losses incurred due management decisions arising from SPP. Examples include the delayed introduction of ticket restrictions, or the deferral of marketing campaigns, both of which could be triggered by poor performance and result in revenue foregone.

1.18 The ORR's consultation paper[3] points to the lack of agreement over what should be eligible for SPP claims, and its emerging view is that SPP should only be claimable for losses as a direct result of SPP[4].

[2] The Track Access Contract does not provide much guidance other than to state instances where TOCs are not entitled to claim. This essentially states that claims can only be made for losses not already covered by Schedule 8, and only under the current Track Access Agreement (i.e. it does not cover previous Track Access Agreements). P 183, TAA (Passenger Services), November 2011.

[3] Periodic Review 2013, Consultation on Schedules 4 and 8 possessions and performance regimes, ORR, November 2012.

Setting the threshold

1.19 The case for an SPP regime rests on the implicit assumption that there is a non-linearity between performance and revenue foregone once performance deteriorates beyond a given 'tipping point'. However, it is clear from our review of research that there is little in the way of firm evidence for the existence of such non-linearities, and hence no strong evidential basis for the level at which the threshold should be set (or arguably for the SPP regime itself).

1.20 In practice, the setting of the threshold has been based on a pragmatic assessment of the appropriate performance level reflecting the broad principles and balance of incentives that the SPP regime aims to achieve. There is no single statement of aims for SPP in any of the information reviewed, but based on available information and discussions with both the ORR and Network Rail, our interpretation of the aims of SPP are that:

- It should act as an appropriate incentive mechanism for Network Rail to avoid sustained periods of poor performance;
- It should compensate TOCs for genuine losses over and above those not captured by Schedule 8 payments, underpinned by evidence;
- SPP claims should not be triggered at a level where the transaction costs of claiming / determining claims are substantial relative to relevant losses;
- The mechanism should not be triggered by fluctuations that reflect the normal variability of performance; and
- The threshold should be informed by the "tolerance" of TOCs to performance that is materially worse than benchmark, demonstrated by the historic level of SPP claims.

1.21 In 2005 the SPP threshold was initially set at a level of 25% worse than benchmark performance for 2006/07, reducing to 22.5% in 2007/08 and 20% in 2008/09[5]. At this level the ORR deemed that too few TOCs were in SPP territory[6] (i.e. instances of apparently sustained poor performance were not necessarily resulting in SPP claims or payments), and the threshold has been reduced since to its current level of 10%.

1.22 At the 10% level, the majority of TOCs are now entering the SPP threshold (over 70% in 2011/12), even though PPM was only 0.3% away from its 92% target in the same year. Accordingly, Network Rail has argued that the current threshold is too readily triggered. The impact of revising the level of the threshold is considered further in Section 4.

[4] The ORR consultation suggests bespoke arrangements between Network Rail and TOCs could be put in place as appropriate to cover this.

[5] PR08 Determination, ORR, October 2008, p358

[6] Para 9.9 in the Periodic Review 2013 states: "In our PR08 determination, we set the SPP threshold so additional compensation could be claimed when performance is at least 10% worse than benchmark over a period of 12 consecutive months. This was to reflect the fact that the SPP threshold set for CP3 had not been reached at the time we produced our PR08 determination."

Claim history

1.23 [REDACTED]. The level of claims is therefore negligible in relation to the theoretical revenue impacts of poor performance experienced by TOCs eligible to claim under the SPP mechanism.

1.24 The ORR suggested[7] a number of possible reasons why SPP claims are so few in number. These are:

- compensation available through the liquidated sums regime is adequate;
- the extent to which costs and revenue losses have been above that covered in the liquidated sums regime is not high enough to justify the legal and administrative costs of making a claim;
- TOCs operating under 'cap and collar' provisions in existing franchise agreements and currently receiving revenue support would effectively have to give a proportion of the additional compensation to the franchise authority, further weakening the financial case for a TOC to make a claim; and
- It is difficult to determine which costs incurred arise from SPP, or it is difficult to identify long term revenue impacts arising specifically from a period of SPP.

1.25 These suggest specific instances where TOCs would be less minded to claim e.g. those receiving revenue support, or smaller TOCs where the legal and administrative costs would not justify a claim. However, there have also been a number of instances where larger TOCs with greater financial resources have experienced performance substantially worse than the SPP threshold, such that any associated 'material losses' should have been sufficient to encourage SPP claims to be made.

1.26 The fact that these have not been made either suggests that the Schedule 8 liquidated sums regime is adequate (i.e. there are no non-linearities), or that it is too difficult to identify long-term revenue impacts and hence to assemble the evidence in support of an SPP claim (i.e. there may be non-linearities, but assembling the data and evidence to prove this has not been possible). The risk of claims arising under the regime is considered further in Section 4.

[7] footnote p 62 from ORR Consultation

2 Review of SPP Claims

[REDACTED]

3 Review of Research Literature

Key findings

- We have reviewed available research and literature relating to the impacts of performance of demand. The main body of research is contained within PDFH.
- PDFH does not directly address Sustained Poor Performance as defined in the Schedule 8 regime. At a general level, the way in which demand reacts to poor performance in its different manifestations is not fully understood. More specifically, there is no evidence to suggest that of non linearities exist, and hence to support any given threshold level.
- As such, there is no empirical basis that would suggest standard Schedule 8 payments would be inadequate at levels of performance within an SPP threshold of 30%.

Introduction

3.1 We have undertaken a review of research into the evidence relating to the impact of performance on demand and revenue in order to identify evidence suggesting any non-linearity in the relationship between them. The primary focus of the review has been on the research underpinning the Passenger Demand Forecasting Handbook (PDFH), which provides recommended rail forecasting values and parameters based on a supporting compendium of research.

3.2 Our review covered:

- The discussion of reliability and recommended values as set out in the main section of PDFH; and
- The supporting research that underpins PDFH.

3.3 There is no direct reference in PDFH to Sustained Poor Performance as defined in the Schedule 8 regime, but there are some relevant themes from the PDFH research, which we identify in the discussion below together with the potential implications for SPP.

3.4 Appendix B contains an overview of reliability in PDFH.

The valuation of reliability

General findings

3.5 Most research has attempted to assess the value of reliability[8]. While in general the findings of individual studies tend to support the PDFH valuation of a 3 minute weighting of 'standard' journey time, disaggregate analysis suggests that this valuation varies significantly for different types of user and service. For example:

[8] The term 'Reliability' within a PDFH context is broadly defined, and covers all elements of unplanned disruption (lateness, delay, cancellations etc.) and the impacts these have on demand. Reliability in a train performance context has a narrower definition where the term reliability is associated with 'cancellations and scheduled lateness' (i.e. events that are very disruptive to passengers) whereas the term 'punctuality' is used to describe whether trains run on-time or late.

- A 1991 MVA study[9] found that the variability (standard deviation) of delay is valued highest by the longer distance travellers compared to shorter distance travellers.
- A study that reviewed existing reliability valuations[10] in 2000 found that while the then established value of 2.5 times mean delay was considered reasonable for frequent Connex Southeastern services, it undervalued the inconvenience due to unreliability in the case of longer distance trips (e.g. on Great Western) and higher value trips (work trips on Virgin, and trips to Manchester Airport). In each case the reliability value was between 3.9 and 5.5.
- A review by KPMG[11] of the above study suggested changing the Multiplier for Unscheduled Journey Time (MUJT), equivalent to the then reliability weighting of 2.5, for the purposes of re-calibrating performance regimes to reflect differences between service groups (with associated MUJT values between 0.78 and 6.52).

Implications for SPP

3.6 The research suggests that different types of passenger will place different 'valuations' on reliability. It is intuitive that these would be higher for longer-distance services (where frequencies are lower and ticket restrictions greater) and for higher value users (e.g. work trips and Airport trips, where the time at destination is important) and that the composition of demand for different services would mean that different 'average' reliability values would exist at the service group or TOC level.

3.7 However, for any given Schedule 8 rate there would be no implication for SPP of different valuations by different types of passenger unless there was evidence that there were differential reliability weightings at different levels of performance, and specifically that these were higher where performance was significantly worse. There is no such evidence available from the literature.

Evidence on non-linearities

3.8 Only one study among the PDFH research studies has investigated non-linearities related to poor performance.

3.9 The research into 'big delays'[12] sought to assess whether the values associated with these differed from those for ordinary delays. The study was largely based on a stated preference (SP) survey of passengers. The central finding was that the values attached to big delays were between 1.5 and 4.5, and that the average was around 3 (now the recommended PDFH value).

3.10 This result suggests that the value of big delays is similar to that for normal delays, and that there is no non-linearity in the performance-revenue relationship. However, there was no valuation of 'small delays' as part of the SP experiment, and therefore no direct comparator based on the same sample of respondents. A number of issues were identified in developing the framework for the SP research

[9] Market Research For ORCATS Model, MVA September 1991

[10] The Investigation of Punctuality and Reliability, John Bates, Peter Jones, John Polak, Andrew Cook February 2000

[11] Recalibration Of Rail Operational Performance Regimes, KPMG for the Office of the Rail Regulator July 2000

[12] Big Delays, SDG Research July 1995

and in the interpretation of results; these stemmed from the combinations of probabilities and delays put to respondents, which often elicited a 'protest' response that yielded unrealistic values. Again, this reinforces the more general limitations of SP research in this area.

3.11 The only other study giving an indication of where non-linearities might exist is the 1996 Rail OR study[13] which is reported below.

Implications for SPP

3.12 There is no direct evidence for the existence of non-linearities based on the only study to investigate this issue explicitly. However, limitations of the research also mean that the firm conclusions cannot be drawn to support the alternative hypothesis that non-linearities do not exist.

Types of lateness - reliability vs. punctuality

3.13 Some PDFH research has attempted to disentangle the different elements of unreliability to assess the relative importance of different aspects of reliability.

3.14 There is insufficient breadth of evidence in this area to form any general conclusions. The 1996 Rail OR Study investigated the relative importance of reliability (taken to be cancellations) and punctuality (average lateness), based on revealed preference research of North London Railways for three separate runs of poor performance ("events"). The impact on non-season ticket revenues (and compared to a control of other South East TOCS) was estimated.

3.15 Based on this research, the reliability factor (f) was calculated and compared to the (then) PDFH lateness weighting of 2.5. The results are re-presented below.

TABLE 3.1 EFFECTS OF POOR PERFORMANCE ON RELIABILITY FACTOR

	Event 1	Event 2	Event 3
Periods	9306-9313	9406-9501	9503-9602
Reliability Minutes	1.86	2.28	6.71
Punctuality Minutes	1.67	6.15	10.28
Implied factor f	3.7	1.5	1.9

3.16 The results suggest that the overall reliability factor (f) is higher when the reliability minutes make up a larger proportion of the total average lateness minutes (i.e. in Event 1 reliability minutes comprise more than punctuality minutes, whereas they are closer to one-third and two-thirds of the total for the Events 2 & 3).

[13] Performance, Perceptions And Revenue, Rail OR Report Number OHP014/07, March 1996

Implications for SPP

3.17 The finding of the 1996 study, that reliability is more important than punctuality, might suggest non-linearities exist if the relative ratio of these elements changed as performance worsened. Reliability and punctuality are train performance terms that are analogous to, respectively, schedule delay (expected delay that passengers build in a contingency for) and variability (unexpected delay that is, by definition, more disruptive).

3.18 While the study suggests that f is higher when the reliability minutes make up a larger proportion of the total average lateness minutes (although the PDFH commentary notes that an alternative interpretation is that if f is the same for reliability and punctuality, then the value of a cancellation is more than the value of the service interval) it also found lower values in the periods when performance was worse, so overall the results are inconclusive for SPP.

Awareness of delay

3.19 There is evidence from several studies to suggest that passengers place an additional value on the cost to them of delays that are not known in advance. The 1995 study by Steer Davies Gleave cited above compared the value of advertised delays associated with engineering works to unexpected delays, and found that the value attached to unexpected delays was three times that attached to anticipated delay communicated to passengers in advance. This finding is supported by more recent work[14] [15].

Implications for SPP

3.20 It is intuitive that passengers place a much higher value on unexpected unreliability. The question is whether there may be differences in the way information can be provided, or is sought, in advance to prospective passengers under SPP conditions. As research suggests 'planned unreliability' is valued lower than 'unplanned' then the ability to provide information in advance is a relevant consideration.

3.21 There is no evidence in this area, but it is possible that when operators experience sustained poor performance, passengers are more inclined to seek out information prior to travel, and hence to mitigate the impact of unreliability by delaying travel, re-routing, changing mode or not travelling at all. At the same time, from an operator perspective, it may be that it is harder to obtain an accurate picture of likely delays when performance 'breaks down', such that it is more difficult to provide useful advice to passengers.

3.22 This issue is likely to be of increasing relevance as technology improves rail industry performance information and communication, and passengers receive more accurate travel information and alerts in real-time, allowing greater flexibility in mitigating the impacts of unreliability. However, there is no basis for concluding what the implications of increasing awareness / information would be

[14] Demand effect of possessions, Steer Davies Gleave, Prepared for: ATOC, May 2006

[15] Demand Effect of Possessions, SDG, Prepared for: Passenger Demand Forecasting Executive, May 2006

in an SPP context – i.e. whether it has any impact on possible non linearities, or in which direction.

Demand and revenue impact

3.23 Most research has focused on valuing reliability from a passenger perspective. As the results are used as part of the generalised cost formulation there is an implicit assumption of a direct relationship between performance and revenue, albeit with an assumed lag as described in the Appendix (and reported in PDFH).

3.24 Some studies have attempted to examine the actual impacts of changes in reliability, although this is inherently challenging given the multiplicity of other service related and endogenous factors that can also affect demand and complicate the analysis. The evidence available is as follows:

- The Rail OR report found that a short-term lag exists in tracing the impacts of reliability through to revenue (this applied to both deterioration and recovery of performance), but the research was unable to distinguish any long-term effects on revenue from reliability.
- A study by Oxera[16] found that the market takes one to five years to achieve equilibrium in response to changes in fares, GJT and delay. This range is in line with the recommendations set out in PDFH (which suggest around 2 years).

Implications for SPP

3.25 The evidence above is piecemeal, and does not enable any broader conclusions about potential non-linearities to be drawn. The effect on demand of big delays does not necessarily translate into a non –linearity in the response to average lateness as measured by Schedule 8, unless worsening of average performance (as measured by Schedule 8) is associated with higher variability of individual delays.

3.26 While there is acceptance of a lag effect for changes in performance / reliability, there remains a degree of uncertainty over the period over which this occurs. This clearly makes providing the evidence for an SPP claim more difficult, as uncertainty about the rate and scale of the lag makes it harder to demonstrate cause and effect (reliability causing revenue loss).

[16] How do passengers respond to change?, Oxera, Prepared for: ATOC, March 2005

4 Impact on Network Rail of SPP Threshold

Key findings

- At current levels of performance, the majority of TOCs are eligible to make SPP claims.
- If Network Rail performance during 2010/11 and 2011/12 were at benchmark levels in aggregate, the variability in performance between TOCs and from one period to another would result in a significant risk of claims at thresholds below 30%. Therefore, in order to fulfil the aim that SPP should not be triggered, to a significant degree, when Network Rail is performing at benchmark levels in aggregate, it would be necessary to raise the threshold to a level of 30% or higher.
- The size of the financial risk to Network Rail depends upon the propensity of TOCs to make a claim, given the of excess Schedule 8 payments over the SPP threshold, the probability of success of the claim, and the size of the claim. It is not possible to quantify this risk accurately, but it does become considerably more manageable as the SPP threshold is increased.

The relationship between threshold and number of claims

4.1 The potential for a significant number of SPP claims would undermine the credibility of Schedule 8 as a liquidated sums regime, and pose a significant financial risk to Network Rail. It is therefore desirable that the threshold is set such that the number of potential claims is small, provided this can be done while providing protection for TOCs against the genuine effects of poor performance. In this section we examine the relationship between the threshold and the potential number of claims.

4.2 Table 4.1 shows the percentage of TOCs which could potentially have made a claim during 2011/12 at various levels of SPP threshold, and given the **actual** levels of performance achieved by Network Rail.

TABLE 4.1 PERCENTAGE OF TOCS IN SPP 2011/12

SPP threshold	% of TOCs in SPP during 2011/12
10%	74
15%	74
20%	68
25%	63
30%	53
35%	42
40%	37
45%	32
50%	16
55%	5
60% and above	0

4.3 To a certain extent, the number of potential claims arises from the fact that Network Rail fell short of its regulatory targets in 2011/12. However, even if performance were at target levels overall, the underlying variation in performance between TOCs and from one period to another would cause a significant number of TOCs to fall within the SPP threshold at current threshold levels.

The effect of removing variability

4.4 ORR has stated[17] that it is of the view that Network Rail should not face a large risk in respect of SPP claims in instances where it is performing at benchmark in aggregate. We have therefore extended the modelling previously carried out by Network Rail, and have tested the extent to which variability in performance between TOCs and over time could cause TOCs to fall above the SPP threshold when Network Rail performs at benchmark in aggregate. We have calculated the number of TOCs which would fall within the SPP threshold during 2011/12 if Network Rail performance (measured by performance minutes) were improved evenly (i.e. by a fixed percentage) across all service groups and periods within each of 2010/11 and 2011/12[18], such that the total Schedule 8 payment by Network Rail in each year fell to zero. Table 4.2 shows the percentage of TOCs which could potentially have made a claim during 2011/12 at various levels of SPP threshold, in the case that there is no overall Schedule 8 payment by Network Rail in either 2010/11 or 2011/12.

[17] Consultation on Schedules4 and 8 possessions and performance regimes Para 5.33

[18] Both these years are considered as they contribute to the calculation of the thresholds in 2011/12

4.5 If the variability in performance between TOCs is eliminated, the variability in performance from one period to another for a single TOC will still lead to a significant number of TOCs falling above the SPP threshold at current threshold levels. Table 4.3 shows the percentage of TOCs which could potentially have made a claim during 2011/12 at various levels of SPP threshold, if the benchmark were reset to be equal to the moving average of the payment levels over the previous two years. (This differs from the previous Network Rail analysis, which used the average of the moving averages calculated for 2011/12, and hence took account of 23 periods' data, but with greater weighting to the periods in the middle of that time span. The two year moving averages tend to lead to lower benchmarks than average of the moving averages, hence a larger number of potential claims.)

TABLE 4.2 PERCENTAGE OF TOCS IN SPP 2011/12 IF NETWORK RAIL OVERALL PERFORMANCE IS AT BENCHMARK LEVELS

SPP threshold	% of TOCs in SPP during 2011/12
10%	47
15%	37
20%	37
25%	16
30%	5
35% and above	0

TABLE 4.3 PERCENTAGE OF TOCS IN SPP 2011/12 IF THRESHOLD FOR EACH TOC IS SET TO AVERAGE PERFORMANCE OVER PREVIOUS TWO YEARS

SPP threshold	% of TOCs in SPP during 2011/12
10%	68
15%	53
20%	32
25%	11
30%	5
35% and above	0

4.6 These tables demonstrate that at the current level of the SPP threshold, a significant number of TOCs would be in SPP territory even if Network Rail performance overall were at benchmark. In order to fulfil the aim that SPP should not be triggered, to a significant degree, by fluctuations that reflect routine variability, it would be necessary to raise the threshold to a level of 30% or higher.

The effect of removing a single poorly performing period

4.7 While the SPP regime is, by definition, designed to be triggered by sustained levels of performance significantly worse than benchmark, the thresholds could be breached largely as a result of a single period in which performance was very poor. If this effect were significant, there might be a case for considering a revision to the SPP mechanism, whereby the single worst period (in terms of TOC Schedule 8 payment) over the previous 13 periods is excluded from the calculation of the threshold.

4.8 We have examined the effect of recalculating the moving average by eliminating the single worst period over the previous 13 periods. Table 4.4 shows the percentage of TOCs which could potentially have made a claim during 2011/12 at various levels of SPP threshold, if the performance were calculated on this basis.

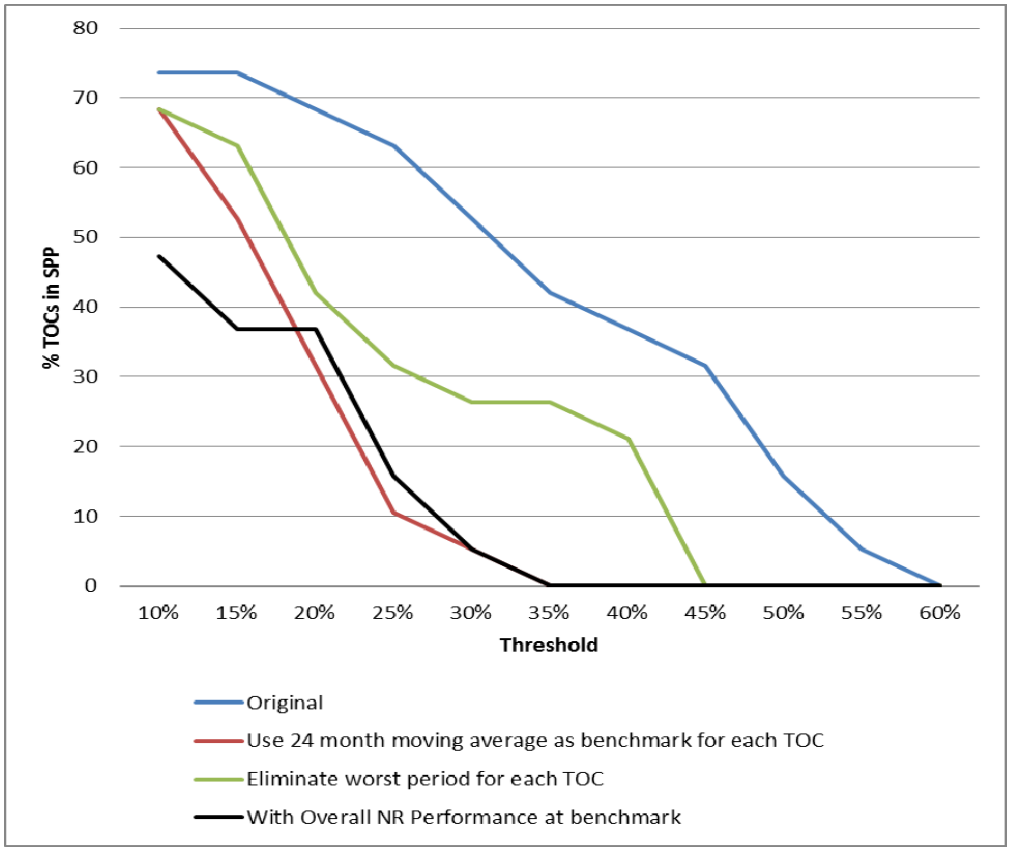
TABLE 4.4 PERCENTAGE OF TOCS IN SPP 2011/12 IF WORST PERFORMING PERIOD IS EXCLUDED

SPP threshold	% of TOCs in SPP during 2011/12
10%	68
15%	63
20%	42
25%	32
30%	26
35%	26
40%	21
45% and above	0

4.9 The table shows that while adding complexity, the change in definition would still leave a significant number of TOCs in SPP at low and medium values of the threshold.

4.10 The figures in tables 4.1 to 4.4 are illustrated in the graph below.

FIGURE 4.1 PERCENTAGE OF TOCS IN SPP IN 2011/12



Risk to Network Rail

4.11 We have also considered the risk to Network Rail of an increase in the level of claims under SPP relative to the low volume observed during CP4. In principle, this is a significant risk with the threshold set at the current level in view of the number of TOCs eligible to make claims, as reported above. As noted by the ORR in its recent consultation, it is not clear why the volume of claims has been low, although it is possible that the cost of claiming in terms of administrative and management time, and the effect of TOCs being in revenue support, have tended to discourage use of SPP.

4.12 There is, however, a risk that the number of claims could increase in the future, as fewer TOCs are in revenue support, and if, for example, the gap between the administrative cost and the potential gain becomes wider.

4.13 In order to inform an understanding of the claims risk, we have investigated the extent to which TOCs have experienced sustained poor performance potentially triggering claims under SPP. We have also drawn on the review of claims history summarised above and identified those TOCs benefitting from revenue support in 2011/12 with a view to inferring why claims may not have been made in particular cases.

4.14 [REDACTED]

4.15 [REDACTED]

FIGURE 4.2 TOCS ELIGIBLE TO CLAIM UNDER SPP [REDACTED]

4.16 [REDACTED]

4.17 In the light of the evidence, we have undertaken some analysis of historical performance to determine the probability of Schedule 8 payments exceeding the threshold by a substantial amount. We have considered a range of potential values of between £200,000 and £600,000 for the excess (measured in terms of a per period average over a year) required to trigger a claim, and estimated the proportion of TOCs with performance deteriorating to this level by analysing Schedule 8 payments by period between 2010 and 2012 (measured in terms of the 13-period moving average payment in period 13 of each year). The results of this analysis are shown in the table below.

TABLE 4.5 PROPORTION OF TOCS REACHING OR EXCEEDING TRIGGER SCHEDULE 8 PAYMENTS BETWEEN 2010 AND 2012

	SPP Threshold								
Trigger for claim	10%	15%	20%	25%	30%	35%	40%	45%	50%
£200,000	21%	14%	12%	11%	9%	5%	2%	2%	0%
£300,000	16%	11%	11%	7%	4%	2%	2%	0%	0%
£400,000	12%	11%	5%	2%	2%	2%	0%	0%	0%
£500,000	9%	5%	2%	2%	2%	0%	0%	0%	0%
£600,000	4%	2%	2%	2%	2%	0%	0%	0%	0%

4.18 The table shows how the risk of a TOC potentially making a claim falls with the level of the SPP threshold (specified in the TAA) and the trigger level of excess Schedule 8 payments (the level being determined by the TOC's own assessment of the loss required to make a claim worth serious consideration). Hence, for a trigger level of £400,000 excess over the SPP threshold, the risk of a TOC considering a claim in any given period falls by more than a half if the SPP threshold is increased from 10% to 20%, and by over 80% if the threshold is increased from 10% to 30%.

4.19 Even if a TOC considers a claim to be worthwhile in principle, based on the estimated value of the associated loss due to poor performance, it will nevertheless need to consider whether the claim is likely to succeed in practice, taking account of the argument on which it is based and the circumstances prevailing while the poor performance was experienced. Hence, notwithstanding Schedule 8 payments above the trigger level, it is possible that a claim will not be pursued or, where it is, that it will not be successful. The data provided by the claims history, reported earlier, is insufficient to estimate the likelihood of a potential claim proving successful, and we have therefore undertaken scenario analysis to illustrate the effect of changing the SPP threshold without seeking to estimate likely payment outcomes.

4.20 For the purposes of this analysis, we have considered a range for the probability of a claim being made and being successful of between 10% and 50%. We have also assumed an average value for a successful claim of £15 million for one year's poor performance. The resulting outcomes for each trigger level and SPP threshold value are shown in the tables below.

TABLE 4.6 ILLUSTRATIVE AVERAGE CLAIMS PER YEAR – 10% CHANCE OF CLAIM BEING MADE AND SUCCESSFUL

	SPP Threshold								
Trigger for claim	**10%**	**15%**	**20%**	**25%**	**30%**	**35%**	**40%**	**45%**	**50%**
£200,000	£ 6,000,000	£ 4,000,000	£ 3,500,000	£ 3,000,000	£ 2,500,000	£ 1,500,000	£ 500,000	£ 500,000	£ -
£300,000	£ 4,500,000	£ 3,000,000	£ 3,000,000	£ 2,000,000	£ 1,000,000	£ 500,000	£ 500,000	£ -	£ -
£400,000	£ 3,500,000	£ 3,000,000	£ 1,500,000	£ 500,000	£ 500,000	£ 500,000	£ -	£ -	£ -
£500,000	£ 2,500,000	£ 1,500,000	£ 500,000	£ 500,000	£ 500,000	£ -	£ -	£ -	£ -
£600,000	£ 1,000,000	£ 500,000	£ 500,000	£ 500,000	£ 500,000	£ -	£ -	£ -	£ -

TABLE 4.7 ILLUSTRATIVE AVERAGE CLAIMS PER YEAR – 50% CHANCE OF CLAIM BEING MADE AND SUCCESSFUL

	SPP Threshold								
Trigger for claim	**10%**	**15%**	**20%**	**25%**	**30%**	**35%**	**40%**	**45%**	**50%**
£200,000	£ 30,000,000	£ 20,000,000	£ 17,500,000	£ 15,000,000	£ 12,500,000	£ 7,500,000	£ 2,500,000	£ 2,500,000	£ -
£300,000	£ 22,500,000	£ 15,000,000	£ 15,000,000	£ 10,000,000	£ 5,000,000	£ 2,500,000	£ 2,500,000	£ -	£ -
£400,000	£ 17,500,000	£ 15,000,000	£ 7,500,000	£ 2,500,000	£ 2,500,000	£ 2,500,000	£ -	£ -	£ -
£500,000	£ 12,500,000	£ 7,500,000	£ 2,500,000	£ 2,500,000	£ 2,500,000	£ -	£ -	£ -	£ -
£600,000	£ 5,000,000	£ 2,500,000	£ 2,500,000	£ 2,500,000	£ 2,500,000	£ -	£ -	£ -	£ -

4.21 The tables suggest that the average claim value could be significant under the existing threshold, depending on the probability of a successful claim. Increasing the threshold to 30% reduces value substantially, although we note that if the probability of a successful claim is relatively high, annual expected payments could exceed £10 million if the assumption of an average claim value of £15 million were broadly correct. While it is not possible to value Network Rail's risk exposure on the basis of this analysis, it does indicate that the risk becomes considerably more manageable as the SPP threshold is increased.

5 Impact on Operators of SPP Threshold

Key findings

- At a threshold of 30%, SPP would be triggered at levels of performance similar to that experienced during CP4, during which there is no evidence that the standard Schedule 8 arrangements are providing inadequate compensation to TOCs.
- For TOCs on which there is sustained poor performance on individual service groups, SPP would usually be triggered at a 30% threshold.
- We did not find evidence that as average lateness gets worse, there is a significant increase in the variability of lateness.
- There is no evidence that performance at the levels currently experienced (and hence what might be experienced by TOCs just within the threshold were it set at 30% for CP5) causes an increase in TOC costs.

Impact on TOC revenue

5.1 The SPP threshold should be set at such a level that a TOC is able to make a claim if the level of Network Rail performance is such that the TOC is likely to suffer financial loss in excess of the amount compensated under the standard Schedule 8 arrangements.

5.2 In order to assess the extent of the risk to TOCs losses, we have examined a number of questions:

- **What is the relationship between minutes lateness and PPM?** If the SPP threshold were raised would this imply that the level at which TOCs moved into SPP territory represented such a low value of PPM that Schedule 8 was not providing full compensation?
- **Are there elements of the Network Rail effect on TOC performance which could be very poor, while overall TOC performance is at a level such that SPP is not triggered?** In particular:
  - Given that SPP operates on a TOC level, could a failure to reach the SPP threshold mask sustained poor performance on individual service groups?
  - Is there a reasonable degree of correspondence between the SPP measure and the number of cancellations - could a failure to reach the SPP threshold mask a significant number of cancellations?
- **Is there evidence that as average lateness gets worse, there is a significant increase in the variability of lateness?**

Relationship between minutes lateness and PPM

5.3 For each TOC we have estimated the change in PPM which would be caused by an increase in Network Rail performance minutes. This was done by establishing a relationship for each individual service group in the TOC between PPM and both Network Rail and TOC performance minutes, on the basis of 13 periods of data covering the period from 2011/12 period 9 to 2012/13 period 8. The results are summarised in table 5.1 below, showing the decrease in PPM for a 10% increase in

Network Rail performance minutes. TOCs have been grouped by the predominant market sector. Results for individual TOCs are shown in Appendix C.

TABLE 5.1 RELATIONSHIP BETWEEN MINUTES LATENESS AND PPM

Market sector	Reduction in PPM associated with 10% increase in Network Rail performance minutes
Long distance	0.6%
London and South East	0.4%
Regional	0.4%

5.4 The table shows that, typically, a 10% increase in Network Rail performance minutes would imply a reduction in PPM of around one half a percentage point.

5.5 The graphs below indicate the implication for PPM in CP5 of increases in 10%, 20% and 30% in Network Rail delay minutes, shown with CP4 PPM for comparison. They illustrate that with SPP thresholds up to at least 30%, PPM for TOCs within the threshold would be in line with PPM levels experienced in CP4.

FIGURE 5.1 PPM IN CP4 AND CP5 LONG DISTANCE SERVICES

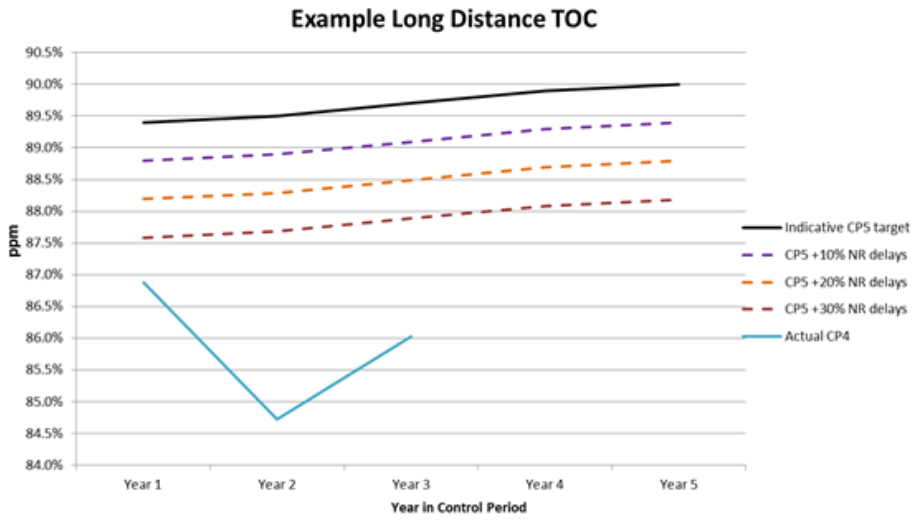


FIGURE 5.2 PPM IN CP4 AND CP5 LONDON AND SOUTH EAST SERVICES

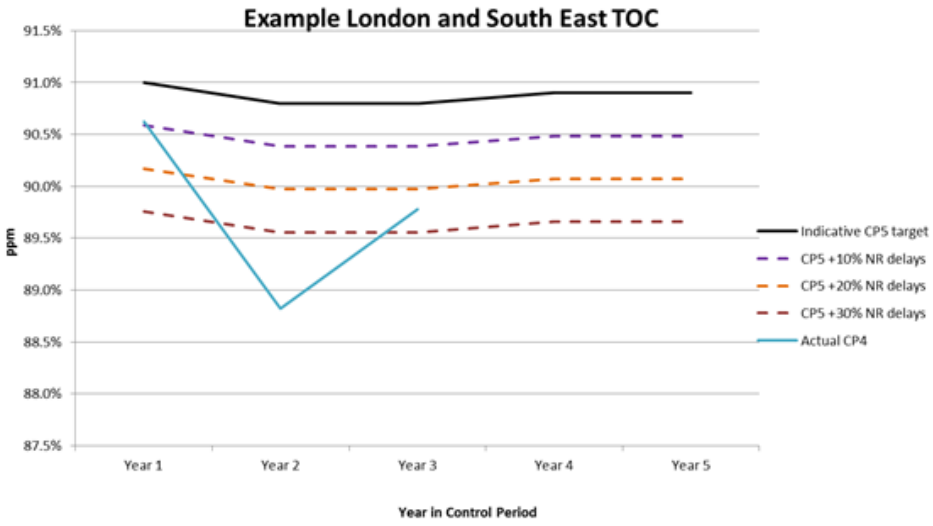
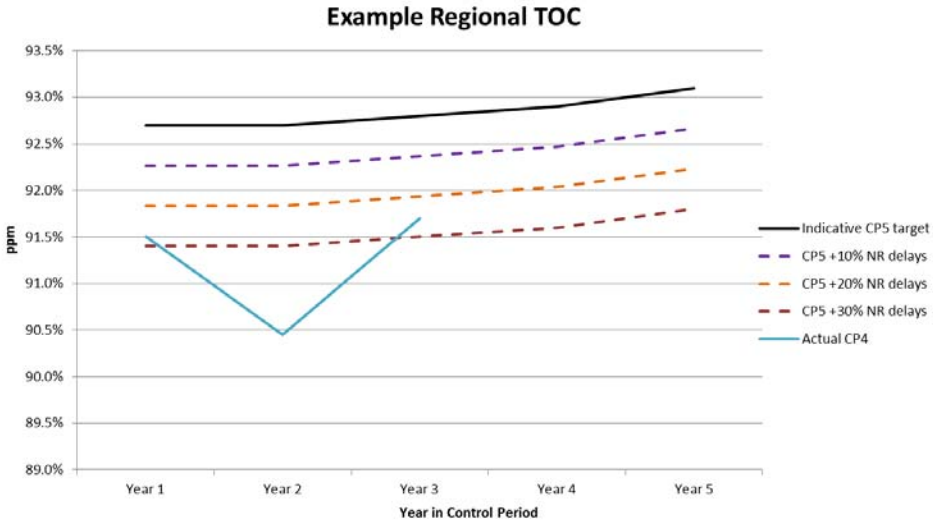


FIGURE 5.3 PPM IN CP4 AND CP5 REGIONAL SERVICES



Poor performance on individual service groups

5.6 It is conceivable that a TOC operating within the SPP threshold might be experiencing sustained poor performance on a significant number of individual service groups, which could give rise to uncompensated financial effects on those service groups.

5.7 [REDACTED]

Table 5.2 Individual service group performance vs SPP [REDACTED]

5.8 In the majority of cases, TOCs with a significant number of service groups experiencing very poor performance for three or more periods fall into SPP

territory at the 20% threshold, and in all cases where more than 30% of service groups experienced very poor performance for three or more periods the TOC fell into SPP territory at the 30% threshold.

High level of cancellations

5.9 It is conceivable that a TOC operating within the SPP threshold might be experiencing a high level of cancellations. This could be significant given that TOCs could incur costs specifically in order to reduce the level of cancellations. The potential impact on costs is considered below.

5.10 [REDACTED]

Table 5.3 Cancellations vs SPP [REDACTED]

5.11 [REDACTED]

Variability of lateness

5.12 A potential cause of non-linearity in the relationship between average lateness and revenue loss would be a greater variability in lateness as the average lateness gets worse, with the effect that passengers have a disproportionately greater chance of a very large delay.

5.13 A possible interpretation of the factor of 3 which is used to calculate the value of lateness is that passengers allow a contingency in planning a journey such that the lateness exceeds the contingency only on a certain percentage of occasions. This percentage is such that the contingency which needs to be allowed is 3 times the average lateness. If the ratio of

lateness at an appropriate percentile of the distribution of lateness: average lateness

were to rise as average lateness rises, this would provide some evidence of the possible existence of a non-linearity.

5.14 In order to investigate whether there is any evidence of this effect, we examined punctuality data provided by Network Rail for long distance services for the period from 11/11/12 to 06/12/12. For each day and for each TOC, we calculated the average lateness of trains on the TOC, and also the lateness of the worst 5% and 7.5% of trains (i.e.at the 95th and 92.5th percentile of lateness). We calculated the ratio of lateness at the 95th and 92.5th percentile, to the average lateness, in order to examine whether there is any evidence that this ratio rises as average lateness increases. These percentiles would seem to be appropriate to examine, as the ratio associated with them is close to 3.

5.15 Figures 5.4 and 5.5 below show the relationship between the ratio and the average lateness at the 92.5th and 95th percentile, each point representing one TOC on one day, and a trend line imposed. While this analysis is based on a relatively small

sample of trains, it indicates that, for long distance services at least, the ratio does not rise as lateness increases[19].

FIGURE 5.4 RATIO OF LATENESS AT 95TH PERCENTILE TO AVERAGE LATENESS

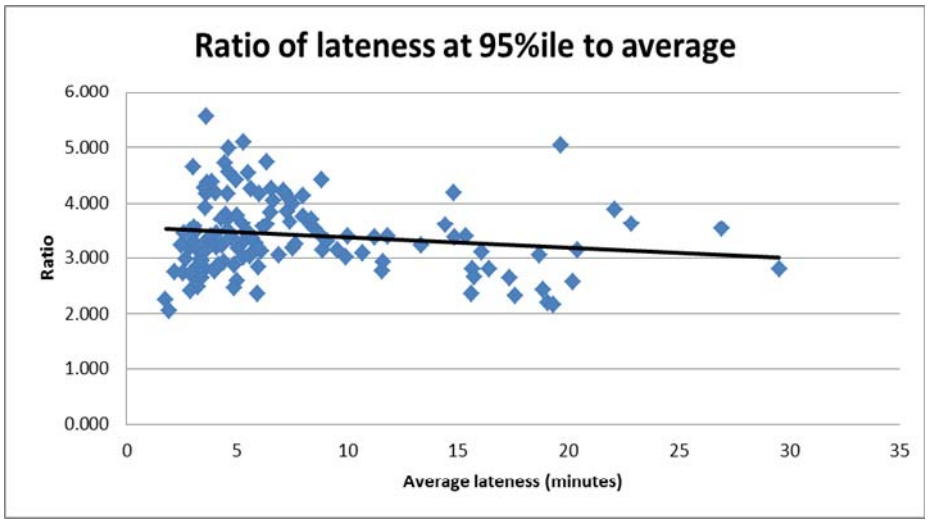
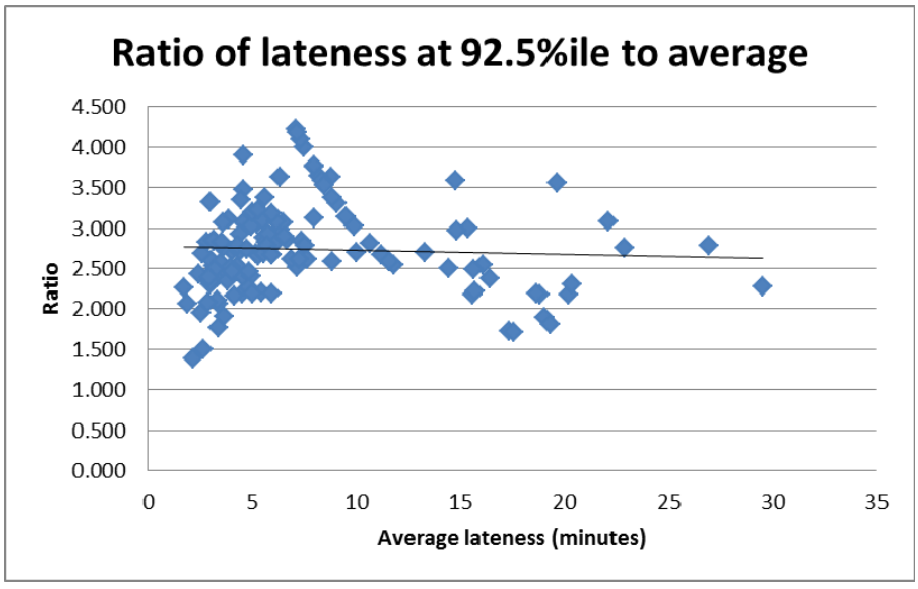


FIGURE 5.5 RATIO OF LATENESS AT 92.5TH PERCENTILE TO AVERAGE LATENESS



[19] The trendlines are slightly downward sloping, but this is not statistically significant

Impact on TOC costs

5.16 Individual incidents causing disruption to train services will have an impact on TOC costs, for example staff overtime where rostered hours are exceeded, or the provision of alternative transport for passengers. Where poor performance is sustained, there may be other costs which arise by virtue of the frequency of incidents. In the absence of SPP claims in CP4, we do not have examples of costs incurred by train operators, and further, this lack of claims suggests that there is no evidence that, at the levels of performance experienced, TOCs have incurred significant costs either as a result of, or in an attempt to mitigate the effects of, poor performance.

5.17 Based on our knowledge of train operations, we have identified that a potential consequence of sustained poor performance would be that train operators deployed additional resources in order to reduce the delay caused by Network Rail incidents. In particular, if the level of cancellations or late starts due to lack of rolling stock or train crew – because they were delayed on the inward journey – rose to a critical level, a TOC might need to deploy additional resources. The general point about the lack of claims made above applied in particular to additional rolling stock and train crew. No TOC has made a claim on the basis of identifying the need to deploy additional resources. This implies that current levels of resource do not include resources that were required to be provided purely as a result of poor performance.

5.18 An analysis made for the ORR consultation on Schedule 4 and 8 (on the question of whether payment rates should be set below 100% of the financial impact to train operators of service disruption) indicates that there is no case in which the level of cancellations is such that it would be worthwhile for a TOC to deploy additional resources given 2011/12 performance levels. This indicates that current poor performance is not likely to cause an increase in resources above current levels.

6 Comparison with other regimes

Key findings

- The SPD mechanism which applies to losses associated with planned disruption is similar to SPP. The threshold for that regime was designed to apply only to a given percentage (set at 1%) of all possessions. It may therefore be appropriate to inform the SPP threshold by reference to the percentage of occasions on which it is triggered.
- We have found no analogous mechanisms in other industries.

Schedule 4

6.1 Schedule 4 is analogous to Schedule 8 in that is compensates TOCs for revenue losses associated with planned disruption (and accounts for cost of replacement buses, and cost savings from reduced rail mileage among other things). Moreover, it also has a Sustained Planned Disruption (SPD) mechanism designed to protect operators from revenue losses over a sustained period, by compensating them at a level above the liquidated sums payable under Schedule 4.

6.2 The SPD is triggered when the revenue or increased cost associated with disruption crosses a pre-defined level. Revenue or cost can be used as the basis for the SPD threshold as this reflects the sustained nature of the impact, whereas individual Schedule 4 payments are based on hours of disruption (Type 1 is less than 60 hours, Type 2 60 – 120, and type 3 longer than 120 hours).

6.3 The SPD claim can be agreed by the relevant parties within a given timeframe or, if they are unable to agree, a dispute resolution process is applied. In either event, the calculation of changes in rail and bus replacement operating costs and of revenue losses are based on a tiered structure and agreed formula set out in the TAA.

6.4 The SPD thresholds are set out in **TABLE 6.1**.

TABLE 6.1 SUSTAINED PLANNED DISRUPTION THRESHOLDS

SPD Category	Threshold	Duration
SPD Cost Threshold No.1	£552,000	3 consecutive periods
SPD Cost Threshold No.2	£1,104,000	7 consecutive periods
SPD Revenue Threshold No.1	20% of defined service group revenue	3 consecutive periods
SPD Revenue Threshold No.2	15% of defined service group revenue	7 consecutive periods

Note: Figures derived from TAA, 2011, and exclude any subsequent index linking. Threshold amount based on Schedule 4 contract, rather than actual revenue.

Basis for revenue threshold

6.5 The SPD revenue threshold was designed to apply only to the most disruptive possessions, defined as approximately 1% of all possessions[20].

6.6 Consultation responses were provided to the ORR that suggests that there are fewer than anticipated possessions that are categorised as SPD. Following further consultation the ORR has stated that it is minded not the change the SPD threshold[21].

Comparison with SPP

6.7 The SPD mechanism contains an indication of the proportion of possessions that should be covered (around 1%), which effectively defines the balance between claims being too easily triggered and not being triggered at all.

6.8 The arguments for changing the SPP threshold have largely revolved around the belief that SPP was triggered too infrequently (when it went down from 25% to 10% above benchmark), to the current emerging view that is may be triggered too easily.

6.9 While a target percentage of occasions for triggering SPP would not be appropriate as the sole criterion for setting the threshold, the SPD mechanism does suggest that it is valid to inform the threshold by reference to the percentage of occasions on which it is triggered. A level of 1% of the worst performance would imply an SPP threshold well in excess of 30%.

London Underground PPP Performance Regime

6.10 The London Underground upgrade programme was developed as a Public Private Partnership (PPP) initiative, under which infrastructure providers (infracos) were incentivised through a performance mechanism to deliver good performance.

6.11 A feature of the PPP performance mechanism is that it is non-linear, whereby the penalty for 'poor' and 'unacceptable' performance (defined against a benchmark level) was double and treble (£6 and £9 per passenger hour respectively) the incentive payment rate for above-threshold performance (£3). The regime operated on a liquidated sums basis, and in this sense was analogous to Schedule 8 rather than SPP.

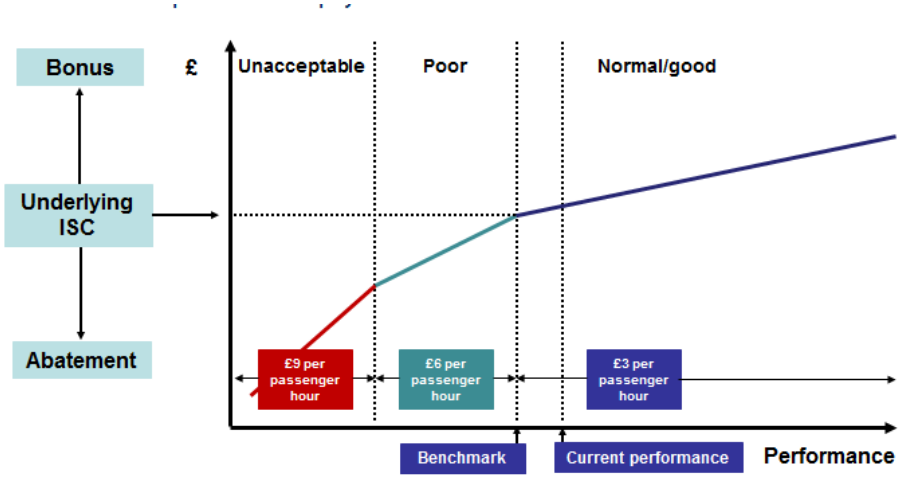
6.12 The PPP performance regime is illustrated in Figure 4.1.

[20] This is set out in the Industry Steering Group's (ISG) recommendation to ORR on changes to the regime for compensating disruptive possessions:

http://www.rail-reg.gov.uk/upload/pdf/pr08poss-recs_comp_regime_310108.pdf.

[21] Periodic Review 2013, Consultation on Schedules 4 and 8 possessions and performance regimes, ORR, November 2012.

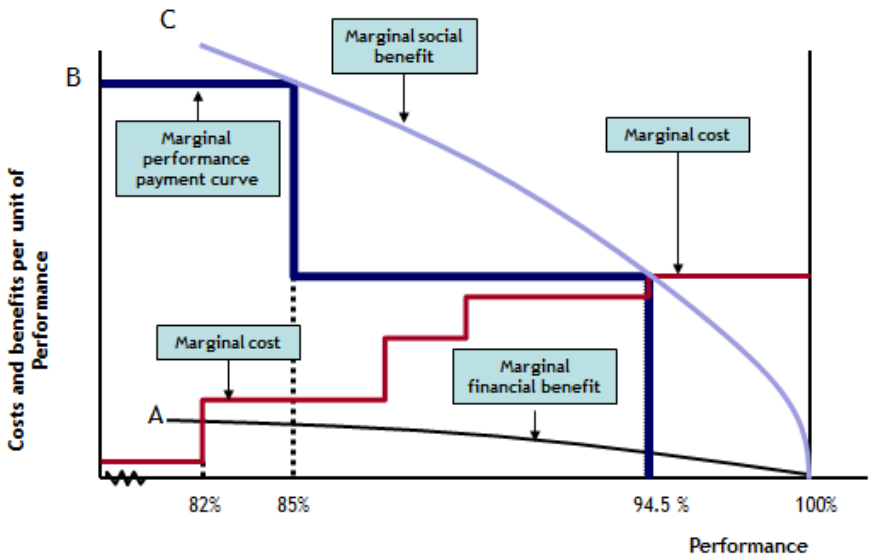
FIGURE 6.1 PPP PERFORMANCE PAYMENT REGIME



6.13 An importance point is that the liquidated sums are intended to compensate society for losses in economic welfare rather than to compensate the operator (London Underground) for loss of revenue alone. Hence, the incentive payments relate to passenger hours rather than revenue foregone.

6.14 The underlying premise for the non-linearity of payments is that the welfare effect at the margin is proportionally greater as performance deteriorates. This is shown in Figure 6.2 (Line C shows marginal social benefit).

FIGURE 6.2 PPP – BASIS FOR PERFORMANCE REGIME



6.15 While economic theory does support the premise of diminishing marginal benefit (at least in general terms applied to consumption of goods, if not specifically for performance), the payment regime adopted for PPP was established primarily as an incentive mechanism, and the differential valuations for 'poor' and 'unacceptable' performance have no evidential basis. Indeed, the stepped nature of the payments underscores its function as an incentive mechanism.

Comparison with SPP

6.16 The PPP regime is based on the principle that there are non-linearities related to performance, albeit expressed in terms of welfare rather than revenue. However, given individual 'welfare' is measured by generalised travel cost, we would expect non-linearities in passenger welfare to translate to non-linearity in revenue. The precise nature of this relationship would depend on the elasticity of demand.

6.17 However, the payment regime itself reflects its purpose as an incentive mechanism, and the valuations are not based on empirical evidence of the non-linear effects of performance on welfare.

6.18 In terms of developing a stronger conceptual underpinning for the SPP regime, the hypothesis that travellers experience diminishing marginal utility as performance improves would be worthy of further research.

Electricity regulation

6.19 We have investigated the regulatory regime administered by Ofgem covering the performance incentives for the quality of service received by customers from local electricity Distribution Network Operators (DNOs) – the companies that operate the local distribution networks that deliver electricity to homes and businesses.

6.20 The incentives address a range of issues such as how often customers have power cuts, how quickly electricity supplies are restored following such cuts and the quality of communication with customers and other parties. In terms of performance, there are two main categories of incentive. The first covers general performance (quality of service) against a benchmark / target level, with incentive payments (rewards and penalties) applied symmetrically and linearly.

"The interruption incentive scheme has symmetric annual rewards and penalties depending on each DNO's performance against their targets for the number of customers interrupted per 100 customers (CI) and the number of customer minutes lost (CML). The proportion of revenue exposed under the scheme is 1.2 per cent for CI and 1.8 per cent for CML." Ofgem website[22]

6.21 Second, there is a 'Guaranteed Standard' of performance which provides protection to individual customers, and must be met by each DNO. These standards have been set to guarantee a level of service that is reasonable to expect companies to deliver in all cases. If a company fails to meet a guaranteed standard of performance it must make a payment to the customers affected, subject to certain exemptions. The guaranteed standards cover 12 key service areas, including supply restoration, connections, and voltage quality.

6.22 There are a series of payments that accrue directly to customers (Guaranteed Standard payments). There are fixed penalties per incident, or per hour that supply is lost / interrupted[23], but there is no additional penalty for 'sustained' poor performance (i.e. the time-based penalties are linear).

[22] http://www.ofgem.gov.uk/Networks/ElecDist/QualofServ/Pages/QualofServ.aspx

[23] The schedule of payments is available here:

http://www.ofgem.gov.uk/Pages/MoreInformation.aspx?docid=41&refer=Networks/ElecDist/QualofServ/GuarStandds

Comparison with SPP

6.23 The regime appears broadly analogous with the regime that operates in the rail sector, though without the SPP dimension.

- The general performance reflects the quality of supply, measured against a benchmark, which is analogous to Schedule 8 payments.
- There are specific customer-focused performance incentives – a 'Guaranteed Standard' in the context of electricity and those contained in the Passengers' Charter for rail, where payments are incurred where the customer experience falls below a defined threshold level.
- There is no direct equivalent of SPP.

Other regimes considered

6.24 As part of this exercise we considered whether there were other regimes that provided a useful comparison with SPP. In particular, we looked at the Aviation sector (Airline and Airport regulation) and other UK utilities. Our scoping of these sectors, and dialogue with sector experts, suggested there we no regimes of equivalent relevance compared to those presented in this Chapter.

7 Conclusion

7.1 Our conclusion is that the most appropriate level of the SPP threshold is 30%. This level provides a balance between protecting TOCs against sustaining revenue losses for which they are not able to claim compensation and protecting Network Rail from the costs associated with a large number of claims which could be triggered through variability in performance, even when performance is at benchmark in aggregate.

7.2 At the 30% level, the risk to TOCs of suffering losses significantly in excess of the standard Schedule 8 compensation is small. If performance in CP5 were below benchmark to the extent that SPP were triggered at a 30% threshold, PPM would still be of the same order as has been experienced during CP4, during which there is no evidence that TOCs have been exposed to financial effects beyond what is compensated under standard Schedule 8 arrangements.

7.3 Below the 30% level there is a risk that the standard Schedule 8 regime would be undermined, with Network Rail facing a significant number of claims in instances where it is performing at benchmark in aggregate. There is also a significant financial risk to Network Rail associated with SPP claims; this risk reduces as the SPP threshold is increased.

7.4 The 30% level is broadly consistent, in terms of the performance levels at which SPP would be triggered, with previous threshold levels in CP3. It was therefore accepted at that time that standard Schedule 8 payments were adequate at these levels of performance.

APPENDIX

A

CLAIM HISTORY

[REDACTED]

Appendix

B

RELIABILITY IN PDFH

Summary of PDFH Commentary

B1.1 The main body of PDFH contains recommended values for all aspects of rail demand and revenue forecasting. These values are based on the overall supporting evidence in a particular field. The recommendations on Punctuality and Reliability are contained in Section B5 of PDFH. The key points are summarised below.

B1.2 *Reliability is fundamentally different from other elements of 'generalised travel cost /or time', as passengers have no accurate prior perception of reliability.* All other aspects of travel cost – fares, journey time, service level, interchange etc. can, in theory, be determined prior to travel, and built in to the passenger's decision. The nature of reliability is that it measures actual service quality (measured by deviation from the timetable) that cannot be fully accounted for in advance by passengers – any train could be delayed.

B1.3 *There are several key elements of reliability, and there is limited understanding of the relative importance of each for passengers.* Elements of reliability include average lateness, the variability of late arrivals and the frequency of big delays.

B1.4 A distinction is made in the literature between two elements of unreliability:

B1.5 'Schedule delay' is delay that is expected based on past information (measured by reference to observed average delay) and causes passengers to build-in a contingency or safety margin for their journey to allow for the consequence of anticipated additional travel time (e.g. arriving late at a destination). The evidence suggests that passengers expectations of this form of delay correlates reasonably well with actual delays.

B1.6 Delay variability (measured by the standard deviation of delay) makes delay to a particular journey inherently difficult to predict, with the result that passengers find it more difficult to 'build-in' contingency time. PDFH suggests that intuitively unpredictable delays (i.e. where there is greater variability of delay) are more important to passengers in terms of their impact on overall perceived journey time.

B1.7 When looking at rail performance data, 'schedule delay' corresponds more with train punctuality (probability of arriving on time or late), whereas variability corresponds more with the reliability element rail performance data (cancellations and significant lateness) – which capture instances of more variable and unexpected delay. If passengers do value variability more than average lateness, this raises the question of whether the relationship between reliability and performance changes as performance worsens.

B1.8 *Passenger perception is an issue.* For regular users, there is limited evidence on the extent to which passengers accurately perceive actual reliability, although the evidence that exists suggests that passengers have a reasonable perception of actual delay distribution. For less frequent users little is known, although their perception could be expected to be informed by their most recent journey experience (which may be atypical), or by the experience of others' journeys and

press reports. This suggests that in assessing non-linearities it is important to consider the impact on different types of user[24].

B1.9 *Passenger awareness is also important*. If passengers have advance knowledge of a reliability problem they can make more informed decisions about whether to continue travelling, or to use an alternative route or mode, to change the time / date of travel or not to travel at all. The role of prior awareness in affecting passenger behaviour will be more important now and in the future as technology should enable better rail industry information to be available, and for this information to be provided directly to passengers in real-time.

PDFH Recommended Valuations

B1.10 The valuation of reliability, for forecasting purposes, has focused on measuring reliability in conjunction with other elements of generalised journey time. The recommended valuation of reliability is 3.0, meaning that one minute of lateness is equivalent to 3 minutes of scheduled journey time. Use of this value in forecasting means that, for a given change in reliability, the impact on demand and revenue can be estimated.

B1.11 The valuation above represents the average (and hence recommended default) value of reliability. There is evidence that these values can vary by service type and passenger type – for example the value can be higher for longer-distance and airport services.

B1.12 PDFH also suggests there is a lag between the effects of a change in reliability and the resulting demand response. The recommended values for a lag associated with a deterioration in reliability or punctuality are consistent with a build-up of the demand response of 60% within 3 months, 85% within a year and 100% within 2 years[25].

B1.13 There is no guidance on whether this lag or build-up differs at different levels of deterioration (i.e. whether sustained poor performance might result in a faster response). We note the guidance suggests that the lag effect is greater for improvements in reliability (e.g. only 80% of the effect within a year), which suggests a slight asymmetry between improvement and deterioration.

Review of PDFH Supporting Evidence

B1.14 We have identified relevant research based on a three-stage process. First, we obtained summary information for all research under the 'topic area' reliability[26]. Second, we reviewed the research summaries of all the papers identified to assess the degree of relevance to the study (categorising them as no relevance, low, medium or high relevance). For selected papers of medium or high relevance,

[24] However, it is also noted in PDFH that nearly all research is focused on existing / regular rail users. For example, stated preference work would typically focus on regular users while any revealed preference work following 'poor performance' would normally focus on remaining users – any irregular users dissuaded from travelling would be out-with the sample.

[25] From PDFH v5 – August 2009, Table B12.1 'Recommended Values for Lags'.

[26] In extracting these, all other research papers were scanned tor potential relevance to this study (e.g. for other instances of non-linearities), and these were included in the body of work reviewed.

where the summary indicated that more information might be contained in the full paper, the latter was reviewed in detail.

General Overview and Summary

B1.15 The focus of the research in the area of 'reliability' has been to examine the valuation of reliability for the purposes of demand forecasting. The research is based primarily on stated preference studies, with some studies attempting to use revealed preference data to determine passengers' valuation and response to reliability.

B1.16 Revealed preference studies tend to be subject to data limitations, making it difficult to identify individual passenger responses to experiences of poor reliability (as opposed to other factors)[27]. Stated Preference (SP) research therefore comprises the bulk of research in this area. However, while SP is reasonable at eliciting responses around schedule delay / lateness (i.e. choices expressed in terms of the likelihood of being x minutes late), it deals poorly with unexpected variability (e.g. the nature of the SP experiment cannot really capture the impact of arriving at a station to discover the train is cancelled).

B1.17 While a substantial body of evidence therefore underpins the core valuation of reliability, the evidence on non-linearities is limited in volume and relatively weak. There is no direct evidence on the impact of sustained poor performance, and the research has therefore sought to identify the impact of relatively poor performance at any given point in time, rather than specifically over time. Again, the evidence is limited, there being no individual study offering any firm conclusions, or group of studies offering common findings that suggest any non-linearities.

B1.18 In summary, the evidence from the literature does not support the existence of non-linearities, and hence does not provide a basis for any particular SPP threshold. However, given the lack of available relevant research, neither is there clear evidence suggesting that non-linearities do not exist. The hypothesis underpinning SPP – that there may be non-linearities, is one that has largely gone untested.

[27] For example, the 2008 Revealed Preference Study by ITS Leeds for the DfT stated at the outset that 'Although the objective of the study was to gather RP evidence on the effect of reliability it was recognised at the outset that this would be difficult and this was the reason for inclusion of a parallel SP exercise'. The study duly found that the RP analysis could not overcome methodological inadequacies.

APPENDIX

C

EFFECT ON PPM OF ADDITIONAL MINUTES LATENESS

[REDACTED]

[REDACTED]

CONTROL SHEET

Project/Proposal Name	Specification for Sustained Poor Performance (SPP)
Document Title	SPP Arrangements for CP5
Client Contract/Project No.	Click here to enter text.
SDG Project/Proposal No.	22541401

ISSUE HISTORY

Issue No.	Date	Details
Draft v1	4/1/13	
Draft v2.1	17/1/13	
Final report	24/1/13	

REVIEW

Originator	Philip Whitehead	
Other Contributors	Simon Ellis Tom Higbee	
Review by:	Print	Simon Ellis
	Sign	

DISTRIBUTION

Client:	Network Rail
Steer Davies Gleave:	Simon Ellis Tom Higbee Alex Wood