

Office of Rail Regulation and
Network Rail

**AO/039: Review of Performance
Measures**

Interim Report

Issue | 2 July 2013

This report takes into account the particular instructions and requirements of our client.

It is not intended for and should not be relied upon by any third party and no responsibility is undertaken to any third party.

Job number 223767-21

Ove Arup & Partners Ltd
13 Fitzroy Street
London
W1T 4BQ
United Kingdom
www.arup.com

ARUP

Contents

	Page	
1	Executive Summary	1
2	Introduction	4
	2.1 Background	4
	2.2 Objectives and Scope of Review	4
	2.3 Audit Methodology	5
	2.4 Structure of Report	5
3	Progress on Recommendations	7
	3.1 April 2011 Review (Q4 2010/11)	7
	3.2 Mandate AO/033 – Right Time Performance Review (December 2012)	13
4	Performance KPIs	16
	4.1 Introduction	16
	4.2 Data Management Processes & KPI Production	16
	4.3 Data Assurance in the Context of the Right Time Performance Report – PPM & CaSL	22
	4.4 Confidence Ratings	28
5	Delay Attribution Review	30
	5.1 Introduction	30
	5.2 Methodology	30
	5.3 The use of FO/TO Delay Cause Codes	31
	5.4 The use of Z* Delay Codes	33
	5.5 Rates of Commercial Takeback	34
	5.6 Recoded Delays	36
	5.7 Delays in Dispute on Day 2	37
	5.8 The level of Roll Up Delays	38
	5.9 Management Incident Levels	39
	5.10 Overview	41
	5.11 Recommendations for Phase Two	42
	5.12 Other possible reviews	42

Appendices

Appendix A

Mandate of Review

Appendix B

Spot Checks of Sample TOC PPM & CaSL Data

Appendix C

Data Assurance in the Context of Right Time Performance Matrix

1 Executive Summary

This commission requires the Reporter to review a series of KPIs and measures produced by Network Rail for the ORR, to ensure their correctness. Also, the Right Time Performance Measure study (RTP) in December 2012 raised concerns about the reliability of an undocumented process for manually recorded data. The report stated that there could be implications for the reliability of PPM and CaSL data, and this review seeks to establish whether this is the case. ORR also requested that the Reporter team carry out an investigation into whether the quality of delay attribution had deteriorated during CP4.

Recommendations from the April 2011 Performance report, and the December 2012 Right Time Performance Measure report, have been reviewed and are substantially complete.

For cancellations data, NR has implemented a revised reconciliation procedure for TOC-provided data, to ensure that differences between NR and TOC data are always positively investigated.

In respect of both PPM and CaSL, the Reporter team have reviewed the end-to-end calculation process through the spreadsheets provided by Network Rail. All linkages remain formulae based, with the only area of hard coding remaining as the input of TOC supplied cancellation data. A series of spot checks, based on a sample of TOCs, have been carried out on the data flow from extraction from PSS through to final figures, to ensure this is accurate. These spot checks are summarised in Appendix B and show no concerns. On the basis of these checks, the Reporter team are content that the data is still flowing through correctly.

Following the publication of Arup's Right Time review¹, we have assessed the implications of the identified timing errors on PPM and CaSL. This follow-up review uses the key findings from the Right Time review and performs a statistical analysis to measure the impact on PPM and CaSL. The results from the statistical analysis show that the expected impact is negligible. The conclusion is that the identified errors in the timing recordings have an immaterial impact on the national PPM and CaSL metrics.

The ratings for the performance KPIs are outlined in the table below, demonstrating any changes since the last review presented in the 2011 Quarter 4 report.

KPI	Original Score	New Score	Comments
5a: PPM	Reliability A Accuracy 1	Reliability A Accuracy 1	No Change. As noted previously, with the exception of the use of TOC data, the process is fully automated. The findings of the Right Time study have negligible impact on PPM, and so provide no reason to downgrade this rating.
5b: CaSL	Reliability A Accuracy 2	Reliability A Accuracy 2	No Change. Network Rail are still reliant on TOC

¹ 223767-07 Mandate AO/033 - Right Time Performance Measure Final Report (Complete)

KPI	Original Score	New Score	Comments
			cancellation data for reporting purposes. The provision of TOC data provides confidence in terms of enabling a reconciliation exercise between data held within PSS and by TOCs, but there do remain differences in the datasets. NR only investigate and reconcile differences of over 0.1%. The findings of the Right Time study have negligible impact on CaSL.
5c: Network Rail Delay Minutes to TOCs	Reliability A Accuracy 1	Reliability A Accuracy 1	No Change. The impact of the Adjusted Data Series calculation has added greater confidence in the accuracy of these figures (see section 4.2.2 of this report).
5d: Network Rail Delay Minutes to FOCs per 100 train kms	Reliability A Accuracy 3	Reliability A Accuracy 3	No Change. The accuracy rating reflects the on-going issues with the collation of the freight mileage data used as the normaliser. NR have now devised a more accurate methodology for calculating freight mileages but have taken the decision to introduce it in April 2014 (see update on recommendation 2010.5.4a in Section 3.1). Once this is fully implemented the measure should be able to rise to A1.
6a/6b: Asset Management (Track / Non Track Delay Minutes)	Reliability A Accuracy 1	Reliability A Accuracy 1	No Change. This dataset is a direct derivative of Network Rail delay minutes, and so is reflective of the KPI score for 5c.

In regard to Delay Attribution, ORR requested that the Reporter team carry out an investigation into whether the quality of delay attribution has deteriorated during CP4. The request arose from concerns that reductions in staffing levels or other causes, such as the impact of Alliances between NR and TOCs, may have led to deterioration in the quality of attribution. Data analysis was undertaken for a period before changes were perceived to have occurred (2009/10) and the data from corresponding periods in 2012/13.

The view of the Reporter Team is that the overall levels of delay attribution show *an improvement* between the two review periods. There are no obvious national trends within the data that highlight significant country wide issues of concern. Most of the measures show a positive improvement since 2009/10. This reflects well on the greater focus on data quality now carried out by NR and increased levels of data integrity checks being carried out both at Route and through the central team as defined in the Performance Measurement Manual. The use of periodic data quality checks has assisted the Routes in improving data quality, and the figures shared with the Reporter Team show an improving trend across the country with areas of concern being tackled.

However, the Reporter team have identified certain issues affecting individual routes and operators that are worth further examination. In general these should

be investigated by NR and reported back to ORR where they have not already given details noted within this report.

Given the overall improvements in the areas investigated the Reporter team do not feel it would be sensible use of resources to undertake a detailed follow up (Phase 2) to the specifics highlighted within this report, as originally envisaged at the initial launch meeting.

Although the overall quality of delay attribution has improved, we have identified variations in practice between the Routes. We therefore suggest that the Reporter team explore the emerging arrangements across three Routes alongside their lead TOCs, to understand the drivers behind the variations being observed. It would also give us the opportunity to explore some of the specific changes on these routes observed during CP4 that we have identified in our analysis. The suggested routes for these discussions are:

- Scotland and ScotRail;
- Sussex and Southern; and
- Wessex and SWT *or* LNE and Northern.

2 Introduction

2.1 Background

Arup was appointed by the Office of Rail Regulation (ORR) and Network Rail in 2009 to undertake the role of Independent Reporter (Part 'A'). This commission requires the Reporter to review a series of KPIs and measures produced by Network Rail for the ORR, to ensure their correctness. These reviews are undertaken as part of a rolling programme and are reported to the ORR in a series of Quarterly Reports. This report covers the Reporter's data assurance review of Performance Measures, the mandate for which is included in Appendix A.

The last Reporter review of NR's performance data took place in March/April 2011. On Public Performance Measure (PPM), and Cancellations and Significant Lateness (CaSL), the review assessed the confidence ratings as (respectively) A1 and A2. Since that review, the Reporter has undertaken a separate review of Right Time Performance data (RTP) in December 2012. This study raised concerns about the reliability of an undocumented process for manually recorded data. The report stated that there could be implications for the reliability of PPM and CaSL data, and this review seeks to establish whether this is the case.

ORR also requested that the Reporter team carry out an investigation into whether the quality of delay attribution had deteriorated during CP4. The initial request was due to concerns that reductions in staffing levels or other causes, such as the impact of Alliances between NR and TOCs, may have led to deterioration in data quality. At the initial project tripartite meeting it was agreed that this review should be addressed in two phases. The first phase would carry out empirical comparisons between selected data from early in CP4 and recent periods, to test if there are any identifiable trends either at national or Route level that indicated a deterioration in data quality.

The Reporter team will then identify potential hypotheses if the data shows any worsening trends and recommend further, more detailed analysis in a second phase of investigation.

2.2 Objectives and Scope of Review

This review is required to include:

1. PPM – the implications of the findings of Arup's Right Time review² (e.g. berthing offsets, manual reporting) on PPM;
2. CaSL – the implications of the findings of the Right Time review (as above) as well as implications on missed intermediate stations;
3. Delay Attribution – ORR would like assurance on whether an apparent reduction in resources available to 'level one' delay attribution (undertaken on shift, in real time) has impacted the quality of delay minute data. To that end, ORR would like a better understanding of the delay datasets (adjusted and unadjusted) and confidence that the data received from Network Rail is of a high standard.

² AO/033 – Review of Right Time Performance (RTP) Data

The Reporter should briefly review PPM, CaSL and delay minutes KPIs and:

- comment on the reliability, quality, consistency, completeness and accuracy of the reported data;
- present a confidence grade for each KPI and comment on the change since last reviewed in April 2011; and
- report on progress against recommendations made in both the April 2011 review and the December 2012 Right Time Performance Measures review, and make appropriate recommendations where necessary.

This report covers all of the above objectives with the exception of Delay Attribution, where we report on phase one and recommend work for phase two.

2.3 Audit Methodology

Following the tripartite inception meeting to agree scope and timescales for this review, a number of meetings were established, and various NR train performance datasets were downloaded:

Date	Network Rail Attendees	Purpose	Location
25 March	Performance Support Analyst Performance Process & Controls Manager	Review of Right Time Performance recommendations & progress	Milton Keynes
27 March	Performance Analyst	Core KPI review	Milton Keynes
23 April	Performance Support Analyst Performance Process & Controls Manager	Delay Attribution Findings review	Milton Keynes
20 May	Performance Support Analyst Performance Analysis Manager	Delay Attribution Findings review	Milton Keynes

At these meetings, the specific areas of review were checked and then data was collected to check that the evidence presented could be verified. Analysis of the data provided was undertaken after the meeting, with any clarifications raised directly with the provider. The findings are set out in the following sections.

2.4 Structure of Report

To enable easy comprehension, the Report is structured as follows:

Section 3 - Progress on Recommendations

- April 2011 Review of Performance KPIs
- December 2012 Right Time Performance Measures Review

Section 4 - Performance KPI Review

- Data Management Processes and KPI Production
- PPM, CaSL & Delay Minutes Data Assurance in the Context of the Right Time Performance Measures Review

Section 5 - Delay Attribution Review

In accordance with Reporter review protocol, confidence ratings have been awarded in the Performance KPI Review section of the Report.

There is no evidence that there is any reduction in quality which could be attributed to a reduction in staffing levels engaged in TRUST DA activities. In some cases it appears there are slightly more staff involved in this area now than in 2009/10.

However there are some specific areas of concerns on Routes as highlighted in the individual sections and that variability is highlighted above in Table 5.12.

One factor that is apparent from discussions with NR is that at the margins of delay attribution there are varying practices within the industry on how delay attribution is handled. There are differences between Routes and TOCs on how delay attribution is managed and the impact of devolution and the creation of Alliances might cause these to widen further.

The analysis in this section also highlights some wide variations in the numbers of incidents being allocated to certain codes between the Routes. Although we stated at the outset that we would not cross-compare Routes, this is perhaps an area worthy of further investigation. The scale of these variations is not as well understood as they could be and there are no clear boundaries as to how far they will widen. In our opinion, the pressure to save costs within the industry means that this will be an area where NR and TOCs will jointly explore cost reductions. This may well mean agreements being reached to avoid costly investigation of minor or difficult areas such as small, difficult-to-investigate delays. The purist DAG view will state that this is not acceptable but a more pragmatic view is likely to challenge this.

5.11 Recommendations for Phase Two

Within this section, the Reporter team have highlighted certain issues affecting individual routes and operators that are worth further examination. In general these should be investigated by NR and reported back to ORR where they have not already given details noted within this report.

Given the overall improvements in the areas investigated the Reporter team do not feel it would be sensible use of resources to undertake a detailed follow up to the specifics highlighted within this report as originally envisaged at the initial launch meeting.

5.12 Other possible reviews

As noted above, we have observed some wide variations in practice between the Routes. In discussion with NR, it is clear they are looking for guidance on what is acceptable in allowing some degree of variation in attribution practices across the country.

We therefore suggest that the Reporter team explore the emerging arrangements across three Routes alongside their lead TOCs, to understand the drivers behind the variations being observed. It will also give us the opportunity to explore some of the specific changes on these routes observed during CP4 that we have identified in our analysis. The suggested routes for these discussions are:

- Scotland and ScotRail;
- Sussex and Southern; and

- Wessex and SWT *or* LNE and Northern.

To carry out the review, the Reporter team would meet jointly with the Route and TOC Performance Managers to understand the reasons behind changes, what direction they see their arrangements taking, and discuss issues picked up in this report. It would also be appropriate to hold discussions with the Chair of the DAB and with representatives from the ORR performance team, to understand their perspectives.

The deliverable would be a report highlighting the reasoning behind the variations and how these affect, if at all, the focus on understanding the causes of delay and delivering improvements. It would set out potential future minimum standards that should be applied in managing delay data and set boundaries within which any future alliances should operate.

Appendix A

Mandate of Review

Mandate for Independent Reporter Part A – Review of performance data

Audit Title:	Review of performance measures
Mandate Ref:	AO/039
Document version:	Final
Date:	17 January 2013
Draft prepared by:	Andy Lewis
Remit prepared by:	Chris Fieldsend
Network Rail reviewer:	John Thompson

Authorisation to proceed

ORR	John Larkinson	
Network Rail	Angelique Tjen	

1 Purpose

This mandate sets out the scope of work for the Part A Independent Reporter (Arup) to review Network Rail's (NR) performance data. As a regulated target, it is critical that ORR has assurance of the quality of this data which offers stakeholders key headlines on industry performance.

2 Background

Arup last reviewed NR's performance data in April 2011. On Public Performance Measure (PPM) and Cancellations and Significant Lateness (CaSL), the review assessed the confidence ratings as (respectively) A1 and A2 (see Annex A below). Since that review (in December 2012) Arup carried out a separate review of Right Time Performance data (RTP). This study raised concerns about the reliability of an undocumented process for manually recorded data. The report stated that there could be implications for the reliability of PPM and CaSL data. ORR would like to know whether this is the case

Additionally, ORR would like assurance on whether an apparent reduction in resources available to 'level one' delay attribution (undertaken on shift, in real time) has impacted the quality of delay minute data. To that end, ORR would like a better understanding of the delay datasets (adjusted and unadjusted) and confidence that the data received from Network Rail is of a high standard. This will enable stakeholders and the public to make better informed decisions.

3 Scope

The review should include:

1. PPM – the implications of the findings of ARUP's right time review¹ (e.g. berthing offsets, manual reporting) on PPM;
2. CaSL – the implications of the findings of the right time review (as above) as well as implications on missed intermediate stations;
3. Delay minutes – level one attributions. ORR is conscious of an apparent reduction in resources committed to level one delay attribution for both Network Rail and Train Operating Companies and requires assurance that this is not impacting data quality. On that basis ORR would like ARUP to review data from areas that have recently changed their delay attribution complement (e.g. York). ORR would like to know if these resource changes could account for an increase in unexplained delay on the Sussex route.

The reporter should briefly review PPM, CaSL and delay minutes KPIs and :

- comment on the reliability, quality, consistency, completeness and accuracy of the reported data;

¹ AO/033 – Review of Right Time Performance (RTP) Data

- present a confidence grade for each KPI and comment on the change since last reviewed in April 2011; and
- report on progress against recommendations made in the April 2011 review and make appropriate recommendations where necessary.

On delay attribution, there are several ways in which a deterioration in quality or accuracy might become evident after a reduction in the amount of resource devoted to level one delay attribution, especially by a TOC. The reporter should therefore comment on:

- whether there has been an increase in the number of occasions when the delay attribution system has become overwhelmed where incidents have to be initially coded into a management incident and subsequently re-coded to their correct incident in the following days;
- whether there has been an increase in unexplained or un-investigated delays, and what the causes of this are;
- whether there has been an increase in delays coded finally simply to TO or FO (TOC and FOC unexplained) with no information available from operator – i.e. the operator does not contest the delay but has not managed to specifically allocate it to fleet, driver, station or whatever, thereby losing the opportunity to identify root cause and take corrective action);
- whether there has been an increase in ‘commercial takeback’ due to delay incidents being incorrectly coded to Train Operating Companies and subsequently re-coded to Network Rail or vice versa;
- whether there has been an increase in delays coded to Network Rail and then subsequently, and without sufficient time for a reasonable investigation to have taken place, re-coded to a Train Operating Company. Special attention should be paid to incidents which are eventually resolved back to Network Rail or subject of resolution by means of a ‘commercial split’;

Answers to these questions should include a review of data “before and after” backed up by a few interviews (which could be by telephone rather than site visits).

4 Methodology

The Reporter should meet with relevant Network Rail employees to understand any procedural changes to the processes used to report the above KPIs since the April 2011 review. The Reporter should also review all relevant documentation and systems, and comment on their quality and fitness for purpose. The Reporter should draw on (and not duplicate) work previously undertaken in their review of RTP. The Reporter’s proposal should articulate clearly how they will address each element of the above scope.

5 Deliverables

The Reporter should provide a publishable report, including findings, conclusions and recommendations (explaining the reasons for them and the benefits if implemented along with timescales for completion). The report should be prepared in draft form and sent electronically to Network Rail and ORR, at the same time. The Reporter should facilitate feedback (via a tripartite feedback session if appropriate) and provide a revised report with track changes. This should be followed by a final report for publication on ORR’s website.

6 Timescales

A fully costed proposal for this work is required by 31 January 2013. Work is expected to commence shortly after following approval by NR and ORR. A draft report is required by 29 March 2013 and a final report is required by 30 April 2013.

7 Independent Reporter remit proposal

The Independent Reporter shall prepare a fully costed proposal for review and approval by NR and ORR on the basis of this mandate. The approved remit will form part of the mandate and shall be attached to this document.

The proposal will detail methodology, tasks, programme, deliverables, resources (including consideration of the mix of seniority and skills required) and costs.

8 Confidence grades

The Independent Reporter shall provide a confidence grade for the PPM and CaSL measures as well as for NR delay minutes to TOCs and NR delay minutes to FOCs per 100 train km. The confidence grading system in Annex A should be used. For each measure, the Independent Reporter should include the:

- confidence grade for this review;
- commentary on the grade against ORR's benchmark; and
- an indication of the highest achievable grade at each level.

Confidence grades should be provided for:

- PPM (% MAA) England & Wales long distance
- PPM (% MAA) England & Wales London & South East
- PPM (% MAA) England & Wales Regional
- PPM (% MAA) England & Wales Total
- PPM (% MAA) Scotland (ScotRail)
- CaSL (%MAA) England & Wales long distance
- CaSL (%MAA) London & South East
- CaSL (%MAA) Regional
- Delay mins – passenger (000's) England & Wales
- Delay mins – passenger (000's) Scotland (ScotRail)
- Delay mins per 100 train km – freight

9 Conflict of interest

The Reporter should explicitly highlight any conflicts of interest.

10 ARUP quality assurance

The Reporter should describe the internal processes in place to quality assure the work delivered under this mandate.

11 Annex A: Confidence grading system

System reliability grading system

System Reliability Band	Description
A	<p>Appropriate, auditable, properly documented, well-defined and written records, reporting arrangements, procedures, investigations and analysis shall be maintained, and consistently applied across Network Rail. Where appropriate the systems used to collect and analyse the data will be automated. The system is regularly reviewed and updated by Network Rail's senior management so that it remains fit for purpose. This includes identifying potential risks that could materially affect the reliability of the system or the accuracy of the data and identifying ways that these risks can be mitigated.</p> <p>The system that is used is recognised as representing best practice and is an effective method of data collation and analysis. If necessary, it also uses appropriate algorithms.</p> <p>The system is resourced by appropriate numbers of effective people who have been appropriately trained. Appropriate contingency plans will also be in place to ensure that if the system fails there is an alternative way of sourcing and processing data to produce appropriate outputs.</p> <p>Appropriate internal verification of the data and the data processing system is carried out and appropriate control systems and governance arrangements are in place.</p> <p>The outputs and any analysis produced by the system are subject to management analysis and challenge. This includes being able to adequately explain variances between expected and actual results, time-series data, targets etc.</p> <p>There may be some negligible shortcomings in the system that would only have a negligible effect on the reliability of the system.</p>
B	<p>As A, but with minor shortcomings in the system.</p> <p>The minor shortcomings would only have a minor effect on the reliability of the system.</p>
C	<p>As A, but with some significant shortcomings in the system.</p> <p>The significant shortcomings would have a significant effect on the reliability of the system.</p>
D	<p>As A, but with some highly significant shortcomings in the system.</p> <p>The highly significant shortcomings would have a highly significant effect on the reliability of the system.</p>

Notes:

1. System reliability is a measure of the overall reliability, quality, robustness and integrity of the system that produces the data.
2. Some examples of the potential shortcomings include old assessment, missing documentation, insufficient internal verification and undocumented reliance on third-party data.

Accuracy grading system

Accuracy Band	Description
1*	Data used to calculate the measure is accurate to within 0.1%
1	Data used to calculate the measure is accurate to within 1%
2	Data used to calculate the measure is accurate to within 5%
3	Data used to calculate the measure is accurate to within 10%
4	Data used to calculate the measure is accurate to within 25%
5	Data used to calculate the measure is accurate to within 50%
6	Data used to calculate the measure is inaccurate by more than 50%
X	Data accuracy cannot be measured

Notes:

1. Accuracy is a measure of the closeness of the data used in the system to the true values.
2. Accuracy is defined at the 95% confidence level - i.e. the true value of 95% of the data points will be in the accuracy bands defined above.

Benchmark grades

As agreed with Network Rail, from Q3 2011-2012 data assurance reviews have been using this new confidence grading system. A characteristic of the new system is the introduction of a benchmark grade; the grade at which ORR believes the measure should be, given what we know about the processes and level of subjectivity in deriving it. It should be noted that the derivation and application of benchmark grades has recently been introduced, and all parties should decide how useful this element is throughout the review. The table below provides ORR's benchmark grades for the measures under review.

Measure	Benchmark grade
PPM	A1*
CaSL	A1
Delay Minutes - Level One attribution	A1

Appendix B

Spot Checks of Sample TOC
PPM & CaSL Data

This Appendix provides examples of the checks of data flow carried out by the Reporter team on Network Rail's performance data for calculating PPM and CaSL. These checks are based on 2012/13 Period 11, and are broad repeats of prior checks of the process. A sample of 5 TOCs was selected at random, although ensuring at least one TOC from each sector was included, and at least one TOC operating across more than one sector was included.

Flow of Data from 'PSS Data Export Spreadsheet' to 'TOC Reconciliation Spreadsheet'

The Network Rail spreadsheet '*PPM & CaSL (1) TOC-PSS.xls*' contains separate worksheets for individual TOCs and Sectors. The PSS cancellations data and the number of trains run are linked by formulae from the PSS export ('*PPM & CaSL (0) All Day P1311 20130311.xls*') to this reconciliation spreadsheet. Checks carried out have shown that the formulae are accurately set up, as shown in the table below.

TOC	PPM & CaSL (0) All Day P1311 20130311.xls			PPM & CaSL (1) TOC-PSS.xls			Difference	Comment
	Total Cancellation	Part Cancellation ⁵	Trains Run	Total Cancellation	Part Cancellation	Trains Run		
Northern Rail	730	566	62,191	730	566	62,191	0	OK
First Great Western (LSE)	307	210	20,066	307	210	20,066	0	OK
First Great Western (Regional)	181	205	13,107	181	205	13,107	0	OK
First ScotRail	364	341	57,248	364	341	57,248	0	OK
East Coast	84	70	3,709	84	70	3,709	0	OK

Source: '*PPM & CaSL (0) All Day P1311 20130311.xls*', '*PPM & CaSL (1) TOC-PSS.xls*'

⁵ In all tables, part cancellation includes trains which were over 120 minutes late

TOCs Cancellation Data

Spot checks on the data supplied by TOCs and that which has been reported by Network Rail were carried out for the TOCs shown in the table below. The data supplied for this review covered 2013/P11. The spot checks carried out, as listed below, indicate the TOC cancellation data has been accurately inputted into the spreadsheet.

TOC	TOC Provided		Network Rail Reported		Difference	Comment	TOC Data Source
	Total Cancellation	Part Cancellation	Total Cancellation	Part Cancellation			
Northern Rail	731	553	731	553	0	OK	_Sch 7 1 Benchmarks DfT - Northern 2012 05
First Great Western (LSE)	306	211	306	211	0	OK	PHIS Export PPM Figures-2012-13 By Day.xls
First Great Western (Regional)	184	203	184	203	0	OK	
First ScotRail	358	335	358	335	0	OK	Bwop1213 1.pdf
East Coast	84	66	84	66	0	OK	Network Rail Performance Sheet.xls

Source: 'PPM & CaSL (1) TOC-PSS.xls', TOC data as indicated above

Note that this table shows, for these TOCs, the magnitude of difference in TOC reported figures (as above) with the initial figures extracted from PSS (in table in previous subsection). This is summarised in the table below for quick reference. It is the TOC provided figures which are used for reporting, and the responsibility of the Route to review any significant differences in these figures.

TOC	PSS Extracted		TOC Provided		Difference	
	Total Cancellation	Part Cancellation	Total Cancellation	Part Cancellation	Total Cancellation	Part Cancellation
Northern Rail	730	566	731	553	1	13
First Great Western (LSE)	307	210	306	211	1	1
First Great Western (Regional)	181	205	184	203	3	2
First ScotRail	364	341	358	335	6	6
East Coast	84	70	84	66	0	4

-

-

-

-

-

Calculation of PPM and CaSL

Values from the PSS Data Export spreadsheet are also linked by formulae to the calculation spreadsheet (PPM & CaSL (2)). The values in the table below have been checked to ensure accuracy and have been found to be robust. Checks have also been carried out to confirm the TOC cancellation data has been correctly linked to this spreadsheet from the 'PPM & CaSL (1)' spreadsheet, and found to be consistent (although details not shown in this table).

Train Numbers	PSS Export					PPM & CaSL Calculation				
	('PPM & CaSL (0) All Day P1311 20130311.xls')					('PPM & CaSL (2) Consolidation.xls')				
	Northern	FGW (LSE)	FGW (Reg)	ScotRail	East Coast	Northern	FGW (LSE)	FGW (Reg)	ScotRail	East Coast
PPM Passes	56,602	18,411	12,077	53,747	2,760	56,602	18,411	12,077	53,747	2,760
Within 15	61,097	19,779	12,881	56,966	3,023	61,097	19,779	12,881	56,966	3,023
15-20 Late	457	126	98	151	203	457	126	98	151	203
20-30 Late	425	92	78	82	227	425	92	78	82	227
30-61 Late	190	57	45	42	205	190	57	45	42	205
61-120 Late	22	12	5	7	51	22	12	5	7	51

Source: 'PPM & CaSL (0) All Day P1311 20130311.xls', 'PPM & CaSL (2) Consolidation.xls'

Train Numbers	Difference (from above table)				
	Northern	FGW (LSE)	FGW (Reg)	ScotRail	East Coast
PPM Passes	0	0	0	0	0
Within 15	0	0	0	0	0
15-20 Late	0	0	0	0	0
20-30 Late	0	0	0	0	0
30-61 Late	0	0	0	0	0
61-120 Late	0	0	0	0	0

Calculation of Sector Level PPM/CaSL

The calculation spreadsheet is set up with a reference worksheet (*'Master'*) which contains a list of all TOCs and the relevant Sector they operate within (and for those TOCs which operate in more than one sector, these are explicitly separated, e.g. First Great Western LD, First Great Western LSE and First Great Western Regional).

These values are then transposed onto a worksheet called *'Template'*. The TOC values for each of the measures, as summarised in each of the individual measures calculation sheets (e.g. "Within 5"), are then multiplied with the values on the *'Template'* sheet to obtain the aggregated sector values. The calculation also includes input from the sheet called *'Timelines'* which takes into account the validity of franchise dates.

Auditing of the process of amalgamating TOC figures to produce sector results focussed on two sample measures, *'PPM'* and *'CaSL'*.

The values from individual TOC sheets were checked to ensure they had accurately been fed through to the *'individual measures'* worksheets. These checks are summarised in the table below for the same four sample TOCs, which showed no concerns.

TOC	From Individual TOC Worksheet		To Measures Worksheet		Difference	Comment
	PPM Passes	CaSL Trains	PPM Passes	CaSL Trains		
Northern Rail	56,602	1,496	56,602	1,496	0	OK
First Great Western (LSE)	18,411	586	18,411	586	0	OK
First Great Western (Regional)	12,077	437	12,077	437	0	OK
First ScotRail	53,747	742	53,747	742	0	OK
East Coast	2,760	406	2,760	406	0	OK

Source: *'PPM & CaSL (2) Consolidation.xls'*

These figures are then fed into the Sector level figures as reported by Network Rail.

Appendix C

Data Assurance in the Context of Right Time Performance Matrix

C1 Detailed Methodology

The methodology used for the PPM and CaSL calculations is essentially the same although different datasets have been utilised. In both cases the data has been aggregated by reporting method, operator and sector across all periods. The error band of +/- 13 seconds has been applied against those services measured by automatic reporting, while the error band of +/- 35 seconds has been applied against those services measured by manual reporting. In this analysis it is assumed that the error distribution is symmetric around a mean of zero. Furthermore, the error distribution is assumed to be uniform. If the distribution was normal then the impact would be significantly less than calculated here since the errors would be clustered close to the mean of zero.

As the error bands extend less than one minute on either side of the recorded arrival times, the one minute period either side of the target (or threshold) time has been considered. With a PPM target of 5 minutes for the London & South East, Scotland and Regional services, the area of interest is arrivals between 4 minutes late and 5 minutes and 59 seconds late. For Long Distance services where the PPM target is 10 minutes, the area of interest is between 9 minutes late and 10 minutes and 59 seconds late. In the case of CaSL where the threshold is 30 minutes for all services, the area of interest is between 29 minutes late and 30 minutes and 59 seconds late.

Given the respective error bands there is a chance that a train arriving in the minute prior to the threshold was in fact late (a false positive). Conversely, there is a chance that a train recorded as arriving in the minute after the threshold was in fact early (a false negative). The false positives and the false negatives will tend to cancel each other out, so the net potential value is required to estimate the impact on PPM or CaSL.

A three-step process was used:

1. Estimate net potential errors;
2. Calculate expected net errors; and
- 3a. RT error percentage = expected net errors / number of trains run;
- 3b. PPM error percentage = expected net errors / number of trains run; or
- 3c. CaSL error percentage = expected net errors / number of trains planned.

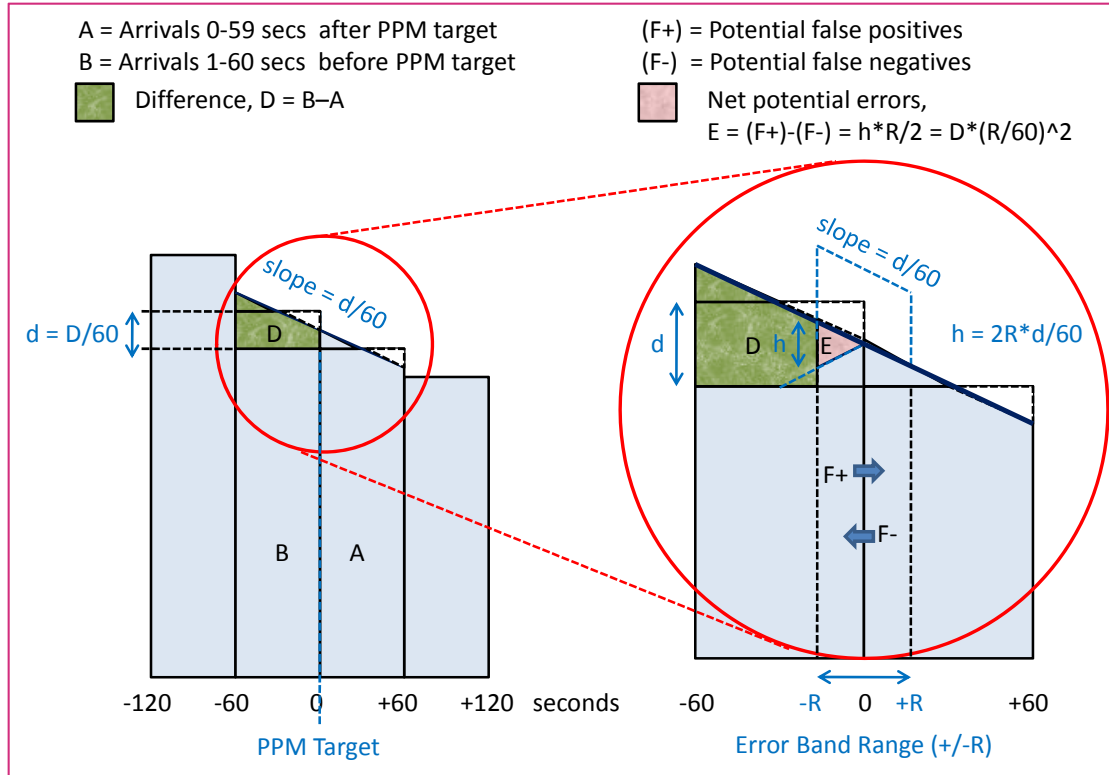
In Step 1, as illustrated below, given the shape of the arrival distribution and the timings of interest, more trains will generally arrive in the minute before the threshold than the minute after. Thus, area B is greater than area A in the illustration on the left. Area D represents the difference. Assuming the arrival rate drops gradually over the 120 second period of interest, the slope of the line is $d/60$, where $d=D/60$

In the magnified illustration on the right, the error band range of +/- R provides a bound to the potential false positives (F+) and false negatives (F-). As area B is greater than area A, more false positives are expected than false negatives and area F+ exceeds area F-. Under the assumption that the error distribution is symmetric most of the errors in recording will cancel each other out.

The net potential errors (represented by the triangular area denoted E) is equal to the number of false positives minus the number of false negatives. The error

triangle has a height (h) of $2R*d/60$ and a base of R . The area of the triangle is calculated as half the base times the height or $R/2*h$. Using substitution, area $E = D*(R/60)^2 = D*R^2/3600$.

Step 1: Estimate net potential errors - PPM



In step 2, as described below, numerical integration is used to weight the potential errors according to the probability that the error for an individual train x is large enough to require reclassification. In other words, a probability density function is used to calculate the expected number of errors. For convenience, arrival times between $-R$ and 0 are relabelled from 0 to R to avoid an excess of minus signs in the algebra.

Step 2: Expected net errors (using numerical integration)i) Probability density, P

$$= p(\text{error exceeds gap between arrival and target})$$

$$= x/2R \text{ (assuming uniform distribution of errors)}$$

ii) Net potential errors, e

$$= h * (1 - x/R)$$

$$= 2R * (D/3600) * (1 - x/R)$$

$$= 2D/3600 * (R-x)$$

iii) Expected net errors = Integral {Net potential errors * Probability density}, $x = 0$ to R

$$= \text{Integral} \{ (2D/3600) * (R-x) * x/2R \}, x = 0 \text{ to } R$$

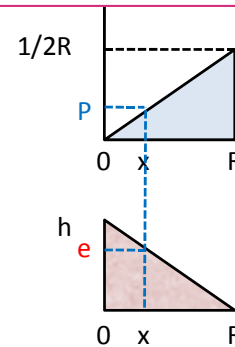
$$= (D/3600) * \text{Integral} \{ x - (x^2/R) \}, x = 0 \text{ to } R$$

$$= (D/3600) * \{ x^2/2 - (x^3/3R) \}, x = 0 \text{ to } R$$

$$= (D/3600) * \{ R^2/2 - (R^2/3) \} - \{ 0 - 0 \}$$

$$= (D/3600) * (R^2/6)$$

$$= E/6$$



Assuming the error distribution is symmetric, at time 0 (relabelled R) there is a 50% chance of an error either way, hence a 50% chance of reclassification. An error in the opposite direction will not generate a false positive so half of the density function can be discarded as irrelevant. Assuming the distribution of errors is uniform, the probability that an error warrants reclassification will increase linearly between 0 and R seconds, reaching a maximum of $1/2R$ as illustrated by the upper triangle in step 2. Thus, the density $p = x/2R$ and the area of this triangle is 0.5.

The number of net potential errors (e) as derived in step 1, is represented by the lower triangle in step 2 and although it now has a different shape with the height adjusted to start from zero, it has an equivalent area. Assuming a simple linear fit to the arrivals distribution, the number of potential errors will decrease linearly from h to zero over R seconds.

Multiplying the values of p and e gives a peak at $R/2$ and zero probability at 0 and R . The expression is a quadratic curve (parabola) and the area under the curve which represents the expected net error, can be calculated using numeric integration. The result is to reduce E by a factor of 6, implying that 1 in 6 potential errors are expected to be large enough to require reclassification, thus producing the expected net errors.

The worst case estimates have been obtained by ignoring the possibility of false negatives and assuming all arrivals within range R of the target are false positives. In this case the worst case errors are $D \cdot R/60$.