

McDonald's and the Art of Railway Performance: why it matters that trains in Scotland do actually run on time

Ross Clark, Rail Performance Manager, Transport Scotland
ross.clark@transportscotland.gsi.gov.uk

INTRODUCTION

Service industries are expected to deliver their services when they say they will, and to the standard they say they will. Passenger rail is absolutely no exception to this principle. However, rail and other service companies also need to know how well they are delivering their services – what you can't measure you can't change, never mind improve – and tendering authorities need to have appropriate measurements of and standards for performance, if they are to incentivise the operators to maintain the desired standards and improve on them.

This paper is a review of the situation and outlook for the 'hard performance' (that is, punctuality & reliability) of the Scottish railway franchise. It is structured in four parts. First, details are provided of the various ways of measuring "hard performance", and how they were applied during FirstGroup's tenure of the franchise. Second, the detail of how performance actually trended over this period is discussed. Third, the structure which will be used to monitor and incentivise Abellio's tenure of the main franchise and Serco's tenure of the sleepers franchise is laid out; as is some reasoning behind the change in approach. In conclusion, some reflections are shared along the lines of 'why, and where to from here'. Note also that the requirements in both franchises for improved 'soft performance' (e.g. train cleanliness, station services), are not discussed further in this paper.

The author has served as the Rail Performance Manager for Transport Scotland since August 2005, thus for most of the period of the Firstgroup franchise tenure.

1. How to measure performance

What might seem to be a reasonably straightforward question, is not so in practice, because of the use to which the performance data needs to be put. The following are the main ways to measure hard performance, and how these means were used during FirstGroup's tenure of the Scottish franchise.

Percent on time (punctuality) & Percent which ran (reliability)

These two measures are the basic measures of performance.

During the FirstGroup tenure, punctuality was measured to within five minutes for the East, Central, and Strathclyde areas; to within ten minutes for Express, South-west and Highland services; and to thirty minutes for sleeper services. This left the anomaly that Edinburgh & Glasgow line services, because they historically fell within the Express group, had a ten-minute 'window' to meet the punctuality target; despite having a market and journey time which for the most part is much more like a regional rail service like those serving Edinburgh or Glasgow proper.

These two measures were included and reported within the Passengers' Charter which First ScotRail had in place. The original Passengers' Charter arrangement was inherited from the dying days of British Rail. This specified a five percent discount on season tickets if the moving-annual-average for either punctuality or reliability fell below the specified targets (ten percent if both fell below the specified targets). In the event, FSR was in slight 'breach' on some service groups for the early part of their franchise. However, they came out of 'breach' on these standards in mid-2007 and did not get back thereafter to anywhere near those levels.

Public-performance-measure (PPM)

The PPM was developed as a composite measure of both punctuality and reliability, and is calculated on the basis of including all trains which run 'to time', with respect to the public timetable. Punctuality measures consistency with the public timetable, for the trains which did run – thus, excluding cancellations, which are reported separately – so it is higher than the PPM measure. It was developed in the 1990s, and data for the Scottish franchise is held back to April 1997. In this instance, though, the 'punctuality' component is measured to within five minutes of the timetabled time for all trains^{*}; which is why PPM and punctuality statistics can vary. Generally the data were reported on a moving-annual-average ("MAA") basis as well, because of the degree of seasonality in the process.

The importance of the measure is twofold. First, it lets us see performance from the point of view of the passenger. Second, it provides a way of comparing, as much as possible, 'like-with-like' performance across the industry in Great Britain. Data are provided for each of the franchises and open access operators and then a grouped result by sector (London & South-east, long-distance expresses and regional franchises) is provided as well. ScotRail's performance can be compared fairly with those of other regional franchises, notably Northern Rail and Arriva Trains Wales, as well as the calculated average for the regional sector as a whole. On the other hand, comparing ScotRail's performance with that of a London-area TOC is not a comparison of like-with-like. For FirstGroup's tenure of the franchise, this measure was included as a planning target, but not as an incentive measure, because this would have involved the operator being accountable and sometimes liable for the performance of the network infrastructure as well as other operators. An overall target for Scotland (92 percent for the twelve months ending March 2015) was set by the Office of Rail Regulation, and reported against as well.

Delay-minutes

This is perhaps the most straightforward of measures. A delay-minute measures the length of time for which a train is delayed, and the current systems also allow a cause to be attached for that delay. However, the measure places an equal weight on a delayed train that may be carrying relatively few passengers, as a delayed train that is fully loaded. This is what reduces its usefulness as a performance incentive, but as a way of reporting the overall picture it is very helpful. The measure can also be related logically to PPM.

This measure was not used as an incentive, but as a reporting and management measure it allowed us to see the proportions of delay caused to FSR by both the network and other operators. The delay that FSR has then caused to other operators is also reported through various systems.

Average minutes of delay

The final measure is for a value called the 'average minutes of delay' per train run, which includes an allowance for full- and part-cancellations. This was developed as part of the performance incentive structure within the Track Access Agreements between the various train and freight operating companies and Network Rail. It is often referred to as the Schedule 8 process, from where it fits in those Agreements. These assume that each party will deliver the other a given level of hard performance, as-reported against various benchmarks (one for the network, one for the operator). If the benchmark for performance is exceeded, then a proportionate payment is made; in either direction, depending on the measure. If the benchmark for performance is not met, then a proportional charge is made.

In Transport Scotland we did not have any direct involvement in the process by which reported delay was allocated to operators and the network. However, once the operators' delay was settled and confirmed, this value was then used for our own incentive structures. The major part of those incentive structures was the performance-incentive-payment, which 'piggybacked' off the results of the Network Rail/FSR performance process; the data for agreed operator delay was then applied against benchmarks (performance and financial) which were quite separate from those used in the 'Schedule 8' agreement. The original FirstGroup franchise in 2004 included a slow (2 percent per year) tightening-up of these benchmarks; the idea was to give the franchisee the incentive to lift its game,

^{*} More precisely, this is for all trains in Scotland – the PPM benchmark for the longer-distance franchises in Great Britain, such as Virgin East Coast and Virgin West Coast, is ten minutes.

although not with the particular expectation at the time that this could or would happen quickly. Importantly, the franchise was incentivised for that part of the performance system for which it was directly responsible – not the problems for performance arising from the network or other operators, which is why the allocation of delay between the network and the operation could often be quite a fraught process. Also, the franchising authorities did not at that stage have a way to incentivise the network's performance directly.

1.1 Other measures

The second part of the main process was the short-formation-incentive-payment (SFIP). This was a charge raised when FSR didn't provide the train seating capacity that it was meant to. The required seat capacity was provided through a Train Plan, which listed every single train run with its seat capacity, including where that capacity needed to be 'strengthened' (that is, more capacity provided than the 'base' provision of a single standard two- or three-car set).

The third part of the performance structure was the timetable-change-incentive-payment, which was intended to minimise unjustified tinkering with the on-the-day timetable. This did not generally have much of an effect, as the changes which were required on a day-to-day basis could generally be justified as being either about network disruptions, or alternatively, were about circumstances outwith the operator's control, such as weather disruptions. Finally, there were a series of 'enforcement' benchmarks. That is, if performance got bad enough, then the whole franchise could have ended up at risk of default. In the event, FirstGroup's performance was generally positive, so these were not an issue.

1.2 Right-time performance

An issue which has assumed more prominence in more recent years is that of right-time performance (that is, the number of trains which arrive within 59 seconds of the timetabled time). Right Time is not a target metric or regulated output in either the past or new franchises, or in Network Rail's High-Level-Output-Specification (HLOS) requirement for CP4. Therefore, there has not been a method for penalising poor right-time performance.

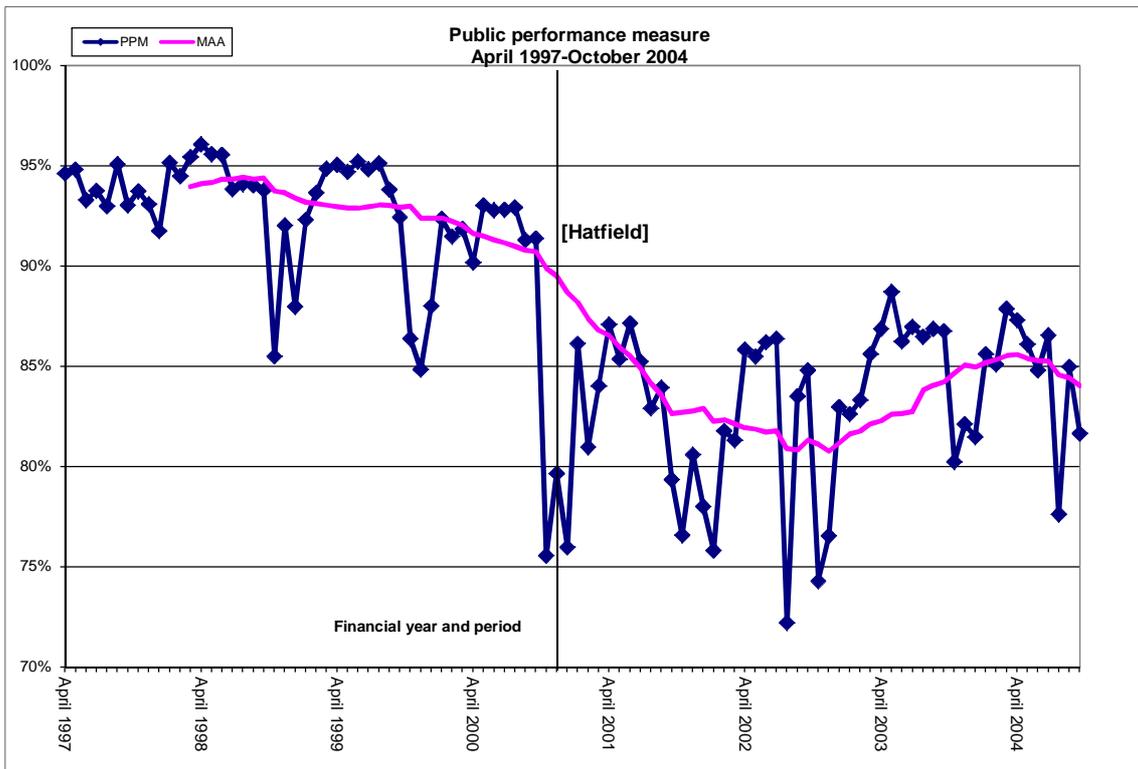
The HLOS for CP5 placed a requirement on the industry to develop and monitor a suite of key performance indicators (KPIs), including right-time performance. Accordingly the ORR and Network Rail now publishes right-time data each rail period, at a whole-TOC level, and the new franchise will be required, as a "committed obligation", to publish how right-time performance is trending. With right-time performance, two other aspects should be noted. First, differing conventions within TOCs as to how working and public timetables relate to each other means that making comparisons between TOCs as to their right-time performance is somewhat problematic – there is no standard convention, especially when a TOC's working timetable is tighter than its public timetable, although First ScotRail did not allow the two to diverge. Statistics also reveal that the gap between PPM and right-time widens during autumn and winter, thus showing the impact poor weather has on specific right-time performance as distinct from the 'normal' measure of punctuality.

2. The way we were: a brief history of performance

2.1 The initial situation

The track of performance in the period post- the Hatfield accident (October 2000) ¹ had been very poor, with at one point the PPM measure running at little better than 80 percent; it did improve after that, but by October 2004 one train in six was not running on time, assuming it ran at all. Chart 1 following illustrates the performance trends from 1997 through to the start of FirstGroup's tenure of the franchise. It is interesting that performance had been in a slow decline for about two years before Hatfield itself. The decline in the year *after* Hatfield was quite sharp, as can be seen. Note also the strong degree of natural seasonality in the data; generally about a ten percent range in PPM between a good summer period and a difficult winter one.

Chart 1 Performance prior to FirstGroup's tenure of the franchise



So, when ScotRail took over the franchise in October 2004, its hard performance was not at all in a good way. In contrast, Chart 2a shows the situation over the ten and a half years of the First ScotRail franchise, with Chart 2b showing the detail from April 2008 onwards.

Chart 2a ScotRail's performance during FirstGroup's tenure

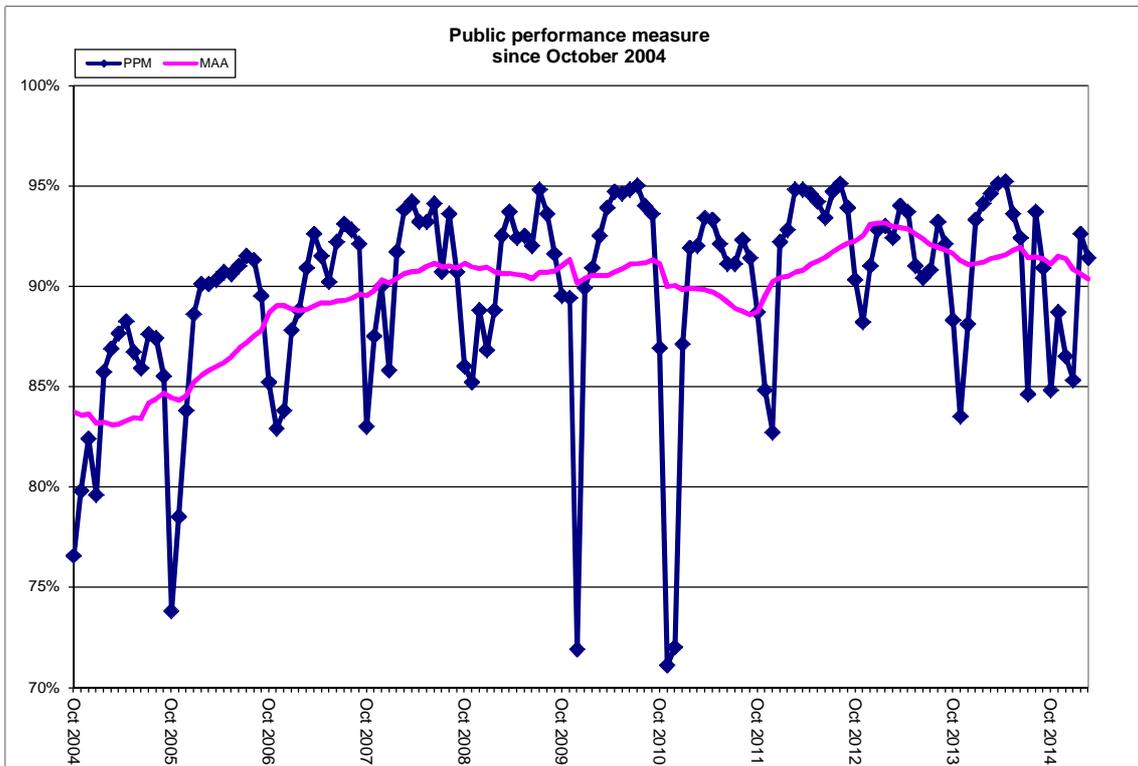
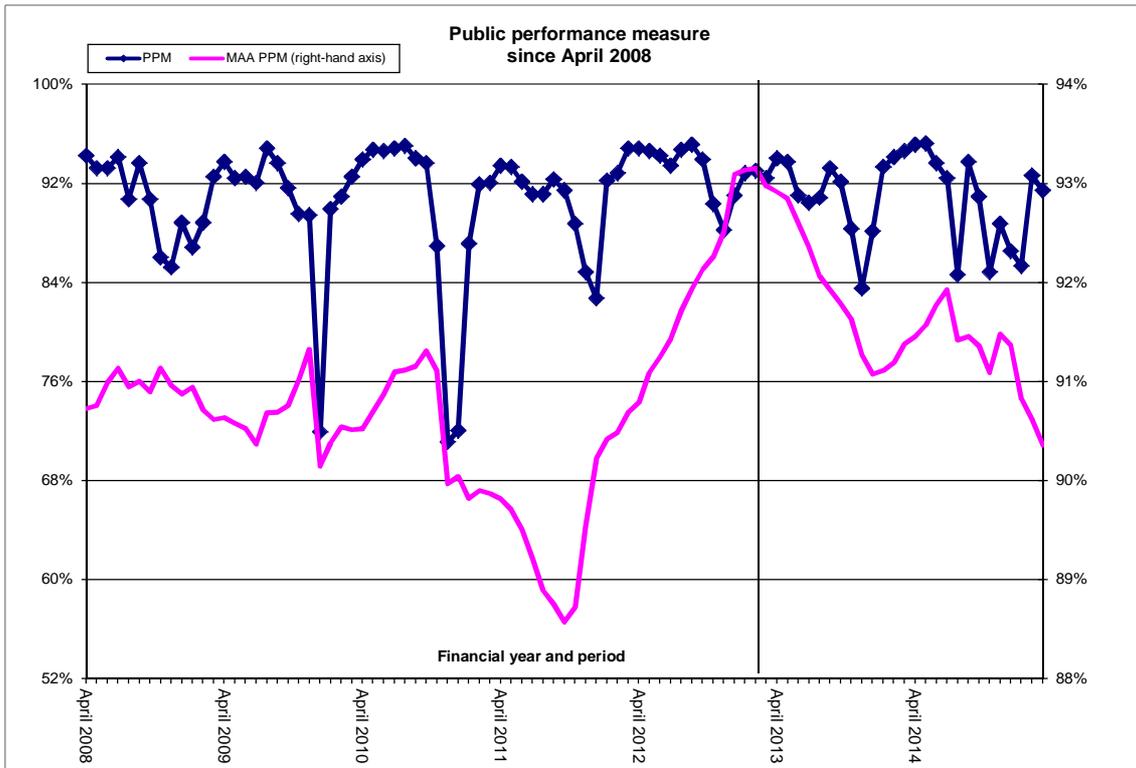
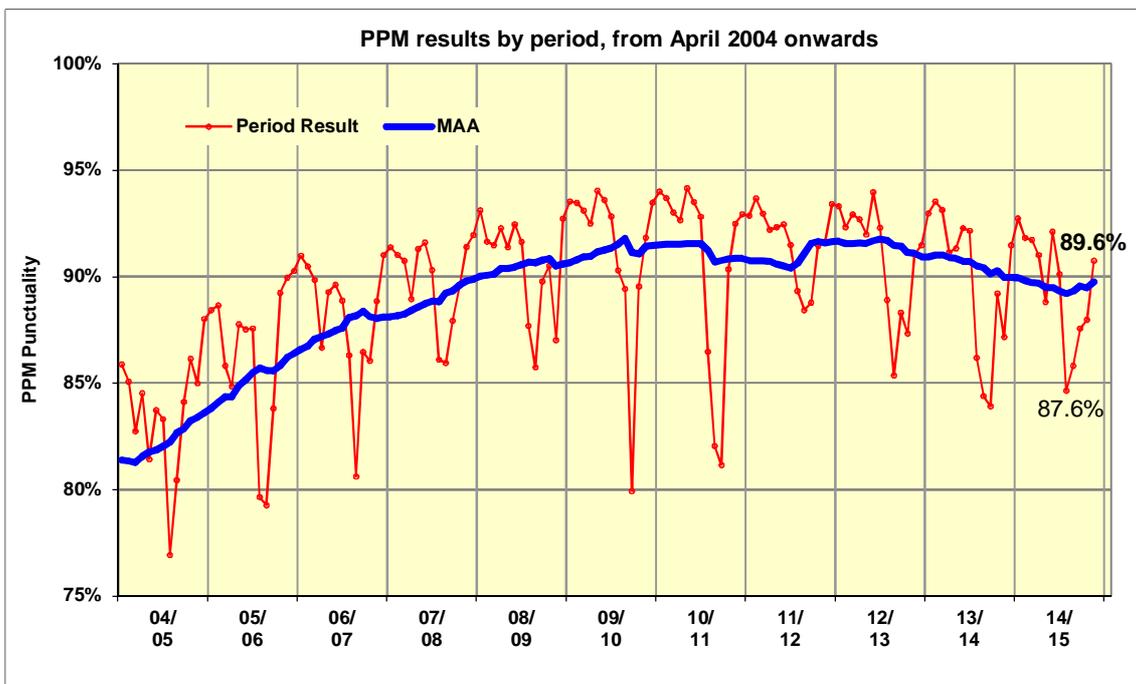


Chart 2b (detail, from April 2008 onwards)



The pattern here reflects, more or less, the situation for the railway network in Great Britain as a whole. In October 2004, overall GB-wide PPM (MAA) was running at 82.2 percent ². This value was slightly below the measure for Scotland at that stage.

Chart 3 Performance in Great Britain as a whole since April 2004



The years shown here are March years; the data are shown up to period 12 (~February) of the 2014/15 year.

2.2 FSR's performance

FSR's performance story over the last ten years can be broken into four chapters:

The steady improvements

The basic story of the first five years of the franchise, as far as performance is concerned, is that of a pretty consistent improvement. This, it must be said, was not restricted to FirstGroup – these improvements were pretty well across the board for the railway industry in Great Britain as a whole, as Chart 3 above illustrates. A rising tide has lifted all the boats, although the improvements in Scotland have been slightly higher than the average.

The Winter War

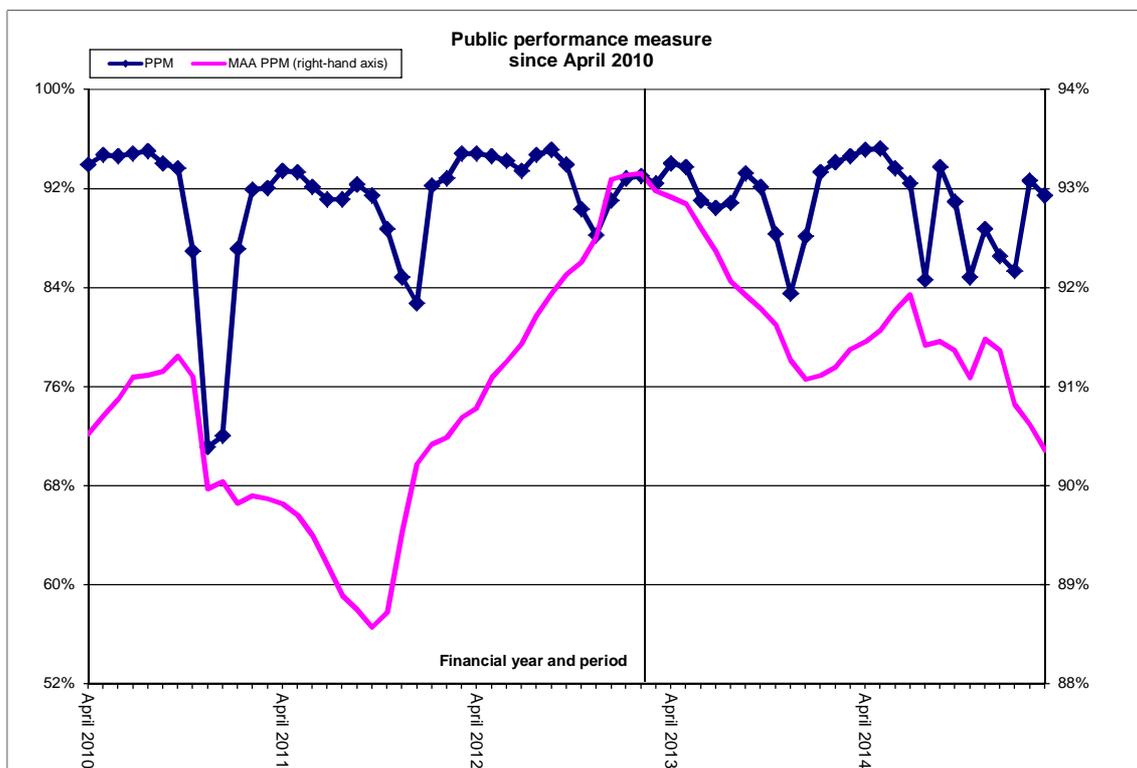
The run of 'mild' winters which had prevailed for several years – more or less, as long as the writer had been in Scotland – came to an end at the end of 2009. Early snowfall (from late-November) turned into a solid period of bad weather and massive disruption; it was later estimated that FSR had incurred an extra 180,000 minutes of delay over what would normally have been expected for that time of year. The following winter was more of the same.

This is illustrated in Charts 2a and 2b above. The navy lines on these charts show the PPM results period by period, which makes evident the degree of normal seasonality in the data. The poor performance over those two winters (2009/10 and 2010/11) is evident both in the period results and in the effect on the moving-annual-average (the mauve line on both charts). On top of that, performance during the next year (calendar-2011) continued well off the pace; that is, most periods had poorer performance than the equivalent period the year before.

As good as it's got

The situation came right with the 2011/12 winter, which was far milder than the year before, and then performance continued to improve through to the end of calendar-2012. This can be seen in the trends for Great Britain as a whole. At the start of calendar-2013 the moving annual average for PPM reached a high of 93.1 percent. This is illustrated in Chart 4 below.

Chart 4 Performance since April 2010



The last two years

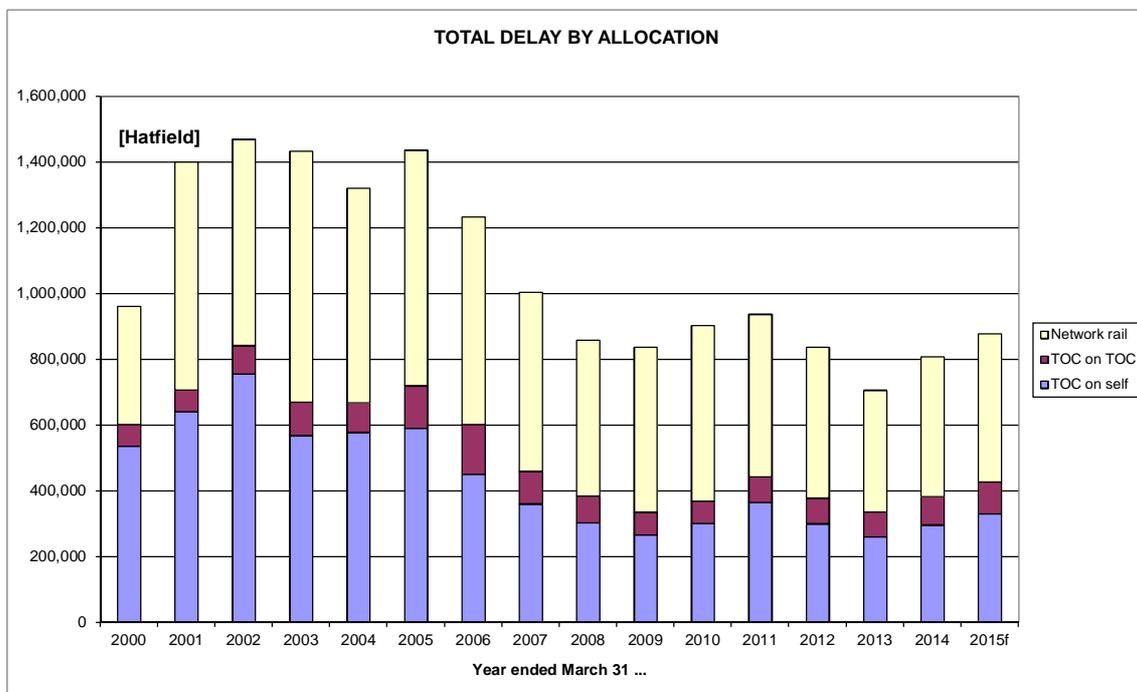
Performance then declined somewhat during calendar-2013 and was only slowly recovering by the time of the Commonwealth Games. At this point PPM had recovered to 92 percent. Unfortunately, the effect of the Commonwealth Games period on performance was significant: in two weeks the demands of running a timetable to cope with so much additional traffic led to some 50,000 additional delay-minutes being incurred over the Games fortnight; according to the ORR this was equivalent to taking a full 0.6 percent off the annual PPM. In the event the decision was very simple – carry the passengers rather than try to run the trains to an effectively-unworkable timetable.

Performance over the final winter (2014/15) had some significant challenges, mainly from a series of weather-events. One of these, early in January 2015, and just as people were getting back to work, effectively closed down the network in Scotland for over a day. So, on handover, the MAA for PPM was tracking to about 90.4 percent *, well down on the specified target of 92 percent (89.6 percent nationally, 91.5 percent for the specific regional franchises).

2.2 Performance in terms of measured delay

Performance can also be reported in terms of delay-minutes, which correspond fairly consistently to the converse of PPM (that is, the count of trains failing to make the PPM standard).

Chart 5a Total delay-minutes incurred since the March 2000 year, by responsible party

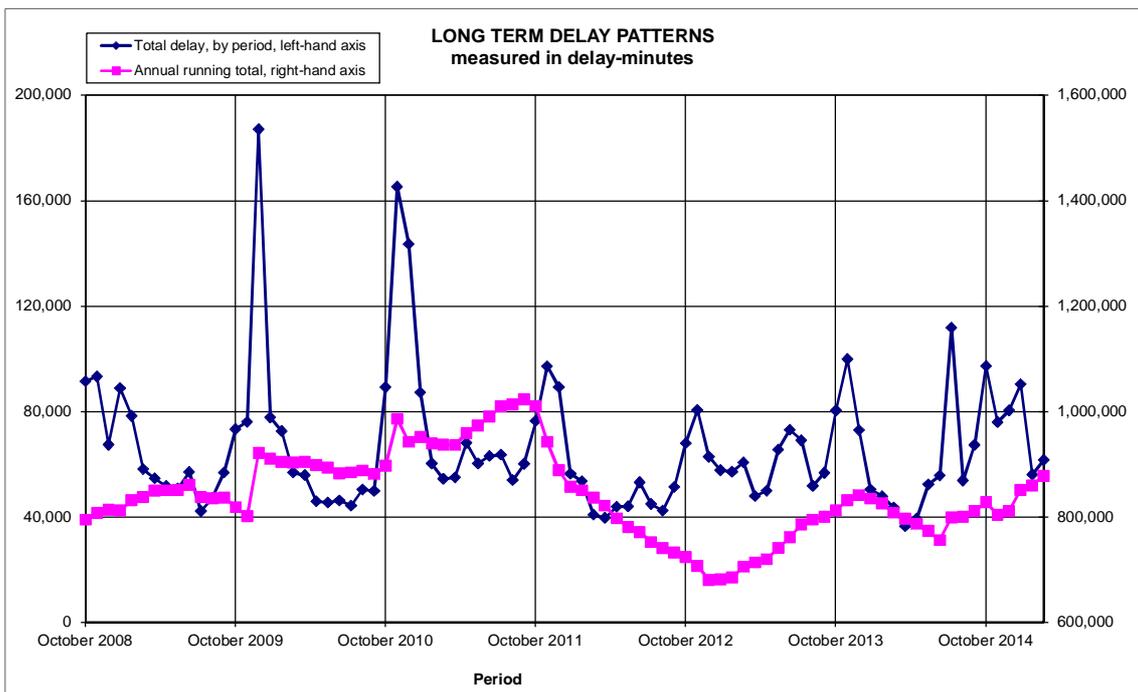


The falloff in PPM over the previous year apparent in the March 2014 year is also apparent in the increase in total delay. Also, delay for the March 2015 year is higher than on the previous two years, which is what would be expected with the decrease in PPM.

To show more recent detail: the trends in total accrued delay since October 2008 are shown in Chart 5b below. This illustrates the effects on the rolling twelve-month total of delay, especially as it tracked from 800,000 to 1m delay-minutes between October 2009 and October 2011. The improvement during calendar-2012 and the worsening since, is shown as well. Again, note that spikes in levels of delay correspond to the spikes in poor performance, as shown in the charts above.

* More precisely, this was the interim calculated value at the point at which this paper was finalised.

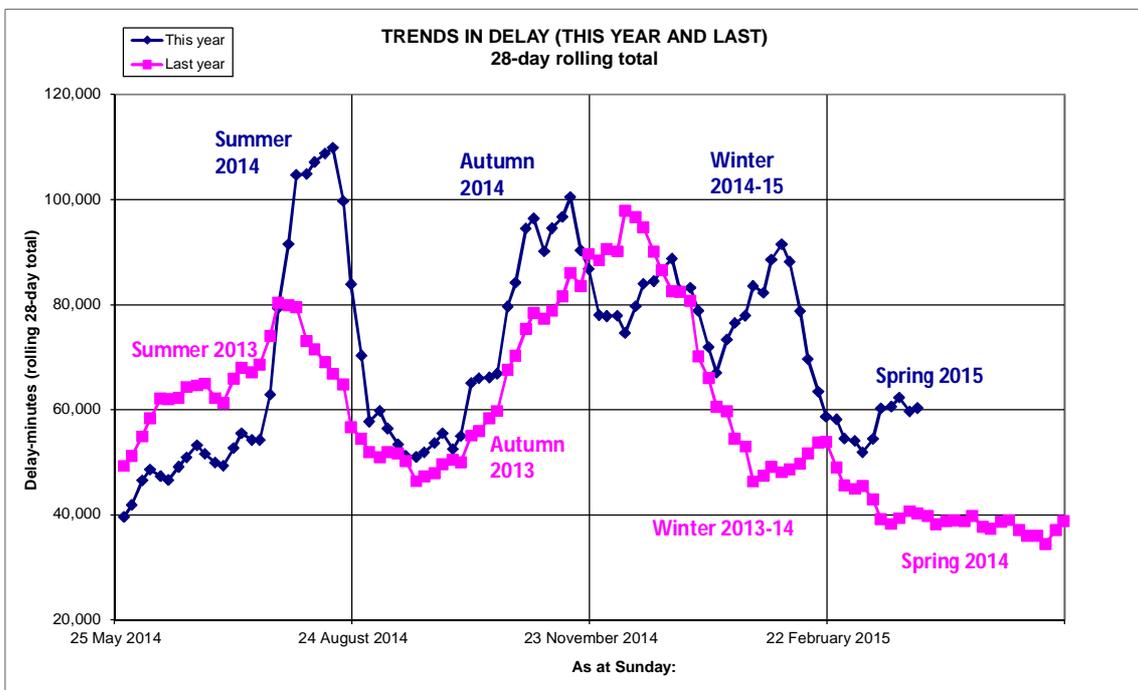
Chart 5b Total delay, by period, since October 2008



Note the effect on the delay totals of the two very bad winters.

The following chart shows the rolling 28-day total for the last two years.

Chart 6 The detail since summer-2013



The sharp spike in Summer-2014 represents the effects of the Commonwealth Games; the almost-as-bad spike in Winter 2014-15 represents the effects of a difficult January, as noted, although performance improved thereafter. Normally the rolling total of delay is falling rapidly at that stage. The spikes in autumn and winter are driven by known seasonal factors.

3. Abellio's tenure

By the time it was necessary to develop a new franchise for Scotland, thinking about performance had moved on significantly from where it had stood in 2003/04 ³. At the time, the performance arrangements applying to the Scottish franchise were more or less 'boilerplate'. Ten years on, ScotRail was the only franchise left using these old Schedule 7 arrangements, as the DfT had developed a variety of other options in the meantime.

Why, then, the change? Essentially, the whole performance picture had shifted, and not just in Scotland. In the first half of the 2000s there was plenty of room for improvement in performance, so it was appropriate to have a performance regime which was designed to pay for that improvement, or to charge the franchise for its absence. Ten years on, the total volume of FSR delay had fallen by about half (and that in the context of an increase in the number of trains run, of about ten percent). In other words, all the 'easy' fixes had been applied; performance could, in theory, be improved by putting more slack into the timetables, but that would clash with the separately-stated aims of the Scottish ministers to reduce timetabled and actual travel time. Another, related, option would be significant investment in infrastructure additional to that under way already, to reduce overall travel times and also provide scope for better timetable compliance; but the cost of this would be very steep. Hence the shift in approach is to something much closer to an 'enforcement' regime, and one which avoids the financial risk, in either direction, of the previous approach. The new franchise also provides the opportunity to align the performance standards set for the franchise, with those set for the network infrastructure.

For the new franchise, the Scottish Ministers stated that their overall policy objectives were as follows:

- Ensure value for money.
- Secure industry alignment to increase effectiveness and reduce costs.
- Exploit utility and capacity of the network.
- Improve journey times and connectivity.
- Manage change effectively.
- Improve passenger satisfaction.
- Improve environmental performance.
- Improve accessibility to services and stations.

The new performance standards can be grouped as follows: ⁴

- Schedule 7.1 – train operating performance
- Schedule 7.2 – KPIs and journey time metrics, including off-peak patronage and passenger satisfaction. There is also an intercity journey time metric included in Schedule 8.1.
- Schedule 7.3 – the SQUIRE (soft-performance) regime

The working of those schedules is explained as follows.

Schedule 7.1 – Train operating performance

3.1 Meeting a PPM standard

Instead of the previous performance standard, the new franchise will be required to meet a PPM standard – and thus, making it accountable for the performance of the network and other operators as well. This standard will be set at **92.5** percent (moving-annual-average) for the franchise as a whole, although this will be applied in such a way to give the new franchise time to get performance to that level; especially given that the performance as the old franchise finished up is well below 92.5 percent. This standard has been set to align with the various targets set by the Office of Rail Regulation for Network Rail ⁵. The significant conceptual change is the move to PPM as the relevant KPI metric, in

order to align better with various policy objectives. Thus, the aim is a stronger focus on alignment of targets with Network Rail to encourage a joined-up industry approach, supporting any future alliance and to avoid the costs of delay attribution. A unified target also has the side-benefit of reducing the costs associated with delay attribution.

Also, subgroups within the franchise will also be reported. The new reporting groups will break out the Edinburgh & Glasgow trains from the wider Express group as a whole⁵. There will be a single 'Rural' group for Highlands and South-western trains; Suburban West (Strathclyde) services will be reported on a peak- and off-peak basis, as at present, instead of being broken out by 'diesel' and 'electric' categories; and the Edinburgh regional trains will be reported within the East service group, as at present.

Also, right-time will be measured and reported, as part of a 'basket' of supporting measures.

3.2 Meeting a cancellations standard

Although the effects of cancellations are built into PPM, cancellations are not normally that major a contributor to a PPM out-turn. Hence, a separate measurement is also required. The cancellations benchmark target will be set at 1.6 percent of all services run; with a part-cancellation counting as a half-cancellation for the purposes of this metric⁶. Note also that this target applies to all cancellations, not just those for which the TOC is responsible.

3.3 Meeting a capacity standard

The previous franchise employed a capacity benchmark which monitored individual services against the specific Train or Capacity Plan. The plan itself requires that a number of trains are provided with more capacity than a base load; this is about 300 trains of the 2,300 run on a 'normal' weekday. When a train was run which did not have the required additional capacity, a charge was raised, proportional to the degree of capacity not supplied. So, if a two-carriage train was run when a four-carriage train was specified, the charge raised was equivalent to half the charge of a full cancellation. If a two-carriage train was run where a three-carriage train was required, the charge raised was equivalent to a third of the charge of a full cancellation.

The recognition that a satisfactory level of capacity must be provided is acknowledged in the new franchise. The change is that the benchmarks now require that a base level of capacity is provided for the 'relevant trains'; if this is not done, then a breach could be raised. The change in practice is thus from the current approach of raising a specific per-train charge when the specified level of capacity is not provided.

The required capacity benchmarks are as follows⁷, expressed at maximum under-provision for a list of the trains specified as needing strengthening:

| | |
|-----------------|--------------------------|
| As a whole | 1.0% |
| Express (E&G) | 1.0% |
| Express (other) | 0.5% |
| Suburban West | 0.8% |
| Suburban East | 1.6% |
| Rural | [no specific target set] |

3.4 Commentary

The change to the PPM and cancellation standards is more significant than at first glance. The former regime incentivised FSR for its *own* performance – the problems which affected FSR which were the responsibility of the network or other operators, were not held to be FSR's responsibility, and in theory they were meant to be managed through the Schedule 8 process anyway. But a shift to PPM means that the new franchise has to manage its own performance much more closely with both the network and the other operators, because its retention of the franchise will then be dependent to at least some extent on how the rest of the system performs. If targets are not met, the structure of 'enforcement' would begin with the preparation of detailed improvement plans and could result in financial penalties or even termination. Again, this is not unlike the procedures available within the old franchise. We are

* This is something we had wanted to do for a while, because of the high 'visibility' of the E&G service.

seeking continuous improvement from the franchise; with performance targets aligned with those of Network Rail, we will see incremental increases in the benchmarks over the next five years. The headline rate demanded is also well above previous levels.

Schedule 7.2 & 8.1 – KPIs and journey time metrics ⁸

3.5 New performance incentives

More general KPIs are also included, which are a significant development in this respect, as they comprise service level targets as distinct from performance targets:

- An off-peak patronage target. This development is of note because it means that the franchise will be incentivised for its *outcomes* (more off-peak travel) as well as its own *outputs* (a better train service).
- A passenger satisfaction measurement. This is based on data provided by the current Passenger Focus research process, but which will be undertaken three-monthly instead of twice a year.
- A journey time measurement. The franchise will be required to maintain average journey times at their current level, if not see them improved. A further measure, included in Schedule 8.1 of the franchise, will also require this specifically of nearly 80 of the intercity and express services ⁹.

Schedule 7.3 – SQUIRE

Refer here for further information: ¹⁰

3.6 The sleepers franchise

For the sleepers franchise, different benchmarks will be in place, partly because the sleepers' operation is such that the trains tend to arrive either extremely early or very late. In this case, then, a right-term metric has been put in place. The sleepers franchise will also have its own regime for SQUIRE ¹¹. The main target for the first three years of the franchise is 75 percent right time performance, and 80 percent thereafter ¹². There are also targets for passenger satisfaction and growth in off-peak patronage.

3.7 A delay-repay mechanism

Also, a delay-repay mechanism has been included within both franchises. This approach is also used in a number of the English franchises; it specifies that when journeys are delayed for more than thirty minutes, half the one-way fare can be claimed back. For journeys delayed for more than an hour, 100 percent of the one-way ticket price can be claimed back; and for delays of more than two hours, 100 percent of the return ticket price can be claimed. For season tickets, any compensation will be proportional to the daily cost of a ticket. More details can be found here: ¹³. This approach can be seen as a significant extension of the much more basic arrangements within the old Passengers' Charter system.

4. Why improved timekeeping matters

Finally, why the reference to McDonald's in the title of this paper? And why does performance matter at all?

That has to do with the nature of service industries generally. The point is – and this is surely a major part of McDonald's success – is that their product and service standards are *consistent*. Customers can rely on that product and service to maintain and sometimes exceed those standards. By analogy: punctuality & reliability, in a railway or indeed any other transport setting, is exactly the same. Trains which can be relied on to keep to a timetable – neither late or even early – are for that reason the more attractive as a transport option, and especially when as a transport option trains face much more market competition, including from the private motor vehicle, than is often appreciated.

In terms of 'where to from here', the responsibility of the industry will be to make the new arrangements work. To that end, Abellio ScotRail and Network Rail are entering into a 'Deep Alliance' ¹⁴, to encourage a pattern of working together which the institutional arrangements of the mid-2000s did not easily allow.

CONCLUSION

This paper is being completed just as the old franchise is being wound up and the new franchise is coming into play. Essentially, the advent of a new Scottish franchise has provided the tendering authority (Transport Scotland) with the opportunity to rethink a great deal about the franchise, including how the 'hard performance' of the franchise is incentivised. The environmental conditions present in 2003 and 2004, when the performance incentives for what became First ScotRail were developed, are not now present. Accordingly, a very different operating environment has required a new incentive regime. The change in the performance incentive environment illustrates the approach which has now been adopted.

Acknowledgements

The assistance of colleagues in the Scottish Government in the preparation of this paper is acknowledged, and was gratefully received, but the usual caveat about the final responsibility for any errors and/or omissions remaining with the author still applies!

Endnotes

¹ Refer here for more information on the Hatfield rail disaster:

http://en.wikipedia.org/wiki/Hatfield_rail_crash

² This chart is prepared by Network Rail's Performance Team, and is used with their kind permission.

³ Refer, e.g., the *Invitation To Tender* for the Scottish franchise:

http://www.transportscotland.gov.uk/system/files/uploaded_content/documents/tsc_basic_pages/Rail/ScotRail%20franchise/ScotRail-franchise-itt-4855286-45.pdf. The performance requirements are discussed from p64 onwards.

⁴ <http://www.transportscotland.gov.uk/system/files/documents/tsc-basic-pages/Redacted%20Franchise%20Agreement%20-%20CU%20version.pdf>; pp389 onwards.

⁵ Ibid., p3.

⁶ Ibid., p397. The service groups used in this context are: Express (E&G), Express (Other), Suburban West, Suburban East and Rural. A distinction for peak services is not made.

⁷ Ibid., p399.

⁸ Ibid., p405ff.

⁹ Ibid., p445ff.

¹⁰ Ibid., p417ff.

¹¹ Refer: http://www.transportscotland.gov.uk/system/files/documents/tsc-basic-pages/Franchise%20Agreement%20EXECUTION%20COPY%2023%20May%202014_Redacted.pdf; pp276ff.

¹² Ibid., p279.

¹³ Refer to the Abellio franchise documentation (my endnote #5, above), at p67.

¹⁴ Refer here for more information: <http://www.networkrail.co.uk/working-with-us/alliances/>. The practice is now becoming more common across the railway network in Great Britain as a whole.
