Report
June 2020

Estimates of Station Usage 2018/19: Methodology Report

Office of Rail and Road
Our ref: 23285203
Client ref:

steer

Report
June 2020

Estimates of Station Usage 2018/19: Methodology Report

Prepared by:

Steer
28-32 Upper Ground
London SE1 9PD

+44 20 7910 5000
www.steergroup.com

Prepared for:

Office of Rail and Road
25 Cabot Square
London
E14 4QZ
Client ref:
Our ref: 23285203

Contents

Tables

Appendices

Executive Summary

Introduction

This report explains the information contained within the Office of Rail and Road's (ORR) Estimates of Station Usage dataset ('Estimates of Station Usage 2018-19.xlsx') and provides guidance on the methodology followed during the process of creating this dataset for the financial year 2018/19. It also includes a summary of the validation checks undertaken as part of the production process.

The Estimates of Station Usage dataset and associated reporting has been produced by Steer[1] on behalf of the ORR.

The Estimates of Station Usage dataset (referred to in the rest of this report as "Station Usage dataset") consists of estimates of the total numbers of people:

- Travelling from or to the station (entries & exits); and
- Interchanging at the station (interchanges).

Information is given for all the national rail stations in England, Scotland, and Wales based on tickets sales data and are the most recent in a series produced for the ORR since 1997/98. The spreadsheet containing the estimates is in a similar format to those published in previous years.

The statistics on usage are necessarily estimates based on a methodology which utilises data on ticket sales. This is then supplemented with other data and adjusted to more appropriately represent passenger movements across the national rail network. The methodology is reviewed annually and enhancements to the methodology are specified and implemented to address known issues. Often these enhancements utilise new sources of data that were not previously available.

Methodology

The Station Usage dataset is generated from the Origin Destination Matrix (ODM), a comprehensive matrix of rail flows between stations throughout Great Britain (GB). This is also produced by Steer, and based largely on data produced for the MOIRA2.2 rail planning tool which itself is derived from LENNON, the rail industry's ticketing and revenue system. **This does place some limitations on the data of which users should be aware and these are detailed in this report.**

The MOIRA2.2 matrix provides an estimate of journeys on the GB (England, Scotland and Wales) rail network for the duration of a financial year (April 1st – March 31st). It includes all journeys associated with point to point flows and includes overlays ("infills") to reflect travel on Travelcards in the London area, Passenger Transport Executive (PTE) sponsored tickets in the major urban areas outside London and travel on some selected 'Rover/Ranger' products (e.g. Anglia Plus). The methodologies for the production of the PTE and 'Rover/Ranger' infills included in MOIRA2.2 were largely developed as part of Steer's work on the ODM.

[1] In July 2018 Steer Davies Gleave changed its name to Steer - https://www.steergroup.com/insights/we-are-steer-read-story-behind-our-name-change For clarity all references to Steer Davies Gleave in this report have been updated to Steer.

The production of the ODM and the Station Usage dataset involves making a number of further adjustments to the MOIRA2.2 matrix to address known issues across the network that impact on the Station Usage estimates. The adjustments included in the Station Usage dataset are:

- Allocation of demand associated with tickets sold to 'London Terminals' between those terminals;
- Allocation of demand between individual stations within station groups outside central London. For example, where tickets are sold to/from 'Dorking BR' it is necessary to estimate how these journeys are distributed between Dorking West, Dorking and Dorking Deepdene stations;
- Allocation of demand between stations to account for specific known issues, for example adjustments are made to account for situations where passengers buy season tickets from a station other than the one they generally travel from, to allow additional flexibility.

Methodological Development

Consistency with past datasets is important to enable comparisons to be made over time. However, stakeholders have indicated that they are keen to see improvements, even where this reduces consistency with historic data, provided any changes are clearly explained.

In the 2018/19 Station Usage dataset the following methodological improvements were made:

- Updated application of 'Season ticket journey allocation' adjustments (see paragraphs 3.13 - 3.16); and
- Updated allocation of journeys between selected Group Stations following implementation of recommendations from a programme of passenger count surveys at selected stations (see paragraph 3.17 onwards).

Revision to 2018/19 estimates

In March 2020 an issue was identified with how journeys on the 'Multiflex' ticket, used on Transport for Wales (TfW) services, were recorded in LENNON. This issue meant that the number of journeys associated with this ticket and therefore the station usage estimates for 2018/19 were overstated at stations in Wales - and to a lesser extent some stations in England served by TfW services.

The 2018/19 Station Usage estimates have been updated to account for this issue and were published in June 2020. Further details of the adjustment made is provided in chapter 3.

Revisions to 2017/18 estimates

Subsequent to the publication of the 2017/18 station usage estimates it was identified that the impact of increased demand in the West Midlands had not been fully accounted for in the published numbers. As a result the 2017/18 estimates for 179 stations, predominantly in the West Midlands, have been restated. These revised estimates have been included in 2018/19 data tables. An adjustment has also been made to the 2017/18 station usage estimates relating to a double counting of station usage at Ilkeston station in the original publication of the Estimates of Station Usage. Further information on both these updates can be found in chapter 3.

Limitations of the data

In the absence of a fully gated system or comprehensive count data, the use of ticket sales data (LENNON) as the primary source of the Station Usage dataset is the best approach available. In particular its national coverage makes it suitable as a basis for the production of Official Statistics such as those reported by the ORR.

Nonetheless, this data does have weaknesses when utilised for this purpose and, although some of these are catered for in the methodology and the ORR continues to seek improvements to address identified issues, the user should be aware of these acknowledged limitations and bear these in mind when using the data. The key limitations are outlined in chapter 1 with more extensive discussion of some aspects of the limitations of the dataset included in chapter 5.

1 Introduction

Overview

1.1 Steer was appointed by the Office of Rail and Road[2] (ORR) to produce the Estimates of Station Usage dataset for 2018/19, continuing the historic series that dates back to 1997/98. This report accompanies the Estimates of Station Usage dataset for 2018/19 and provides details of the process and outputs used to produce the statistics on behalf of the ORR. In the rest of this report the Estimates of Station Usage dataset is referred to as the "Station Usage dataset."

1.2 The Station Usage dataset is generated from the Origin Destination Matrix (ODM), a comprehensive matrix of rail flows throughout England, Scotland and Wales, also produced by Steer, and based on data produced for the MOIRA2.2 rail planning tool which itself is derived from LENNON, the rail industry's ticketing and revenue system.

1.3 Steer have provided the ORR with an MS Excel file, ('*Estimates of Station Usage 2018-19.xlsx*') containing entries, exits and interchanges made at stations throughout England, Scotland and Wales, for the financial year 1st April 2018 to 31st March 2019. For the entries and exits, figures are split into the three main categories of the available ticket products (Full, Reduced, and Season).

The underlying methodology adopted by Steer in the production of the Station Usage data is consistent with that adopted by Resonate[3] in the production of the Station Usage dataset in the years prior to 2011/12. A number of updates to the methodology have been implemented by Steer over recent years which have been documented in this and previous annual reports. A summary of the methodological updates made by Steer is provided in the accompanying report, '*Station Usage & Origin Destination Matrix 2018/19: Historical Methodological Changes*'.

Use of the Station Usage dataset

1.4 When using the Station Usage data, particularly when comparing with previous years, it is important to be aware of:

- Methodological improvements made to the dataset over time which can impact consistency between years;
- Limitations of the data and specifically factors e.g. some ticket sales not being included, that may mean that demand on particular flows and at stations is underestimated or overestimated; and
- Factors which can affect reporting of entries and exits.

Methodological improvements to the dataset

1.5 Improvements to the dataset made in 2018/19 are set out in chapter 3. A summary of improvements made over recent years are further detailed in the accompanying report, '*Station Usage & Origin Destination Matrix 2018/19: Historical Methodological Changes*'. The ORR continues to work with stakeholders and its own consultants to improve the robustness

[2] The Office of Rail Regulation was renamed the Office of Rail and Road from 1st April 2015.

[3] Resonate were formerly known as 'DeltaRail' and changed their name in August 2016.

of the dataset by implementing methodological changes that demonstrate value and address acknowledged issues.

Limitations of the data

1.6 In the absence of a completely gated system that allows a complete recording of flows through stations or comprehensive and robust count data the use of ticket sales data, LENNON, as the primary source of the Station Usage dataset as described in the following chapter is the best approach available. In particular its national coverage makes it suitable as a basis for the production of Official Statistics such as those reported by the ORR.

1.7 However, this data does have weaknesses when utilised for this purpose and, although some of these are catered for in the methodology, the user should be aware of these acknowledged limitations. The key limitations are outlined below with a more extensive discussion of some aspects of the limitations of the dataset included in chapter 5.

- **Non-Point to point tickets** – An overarching issue is the inherent difficulty and uncertainty associated with estimating the number of journeys associated with many rail products which do not simply represent point to point single or return journeys and furthermore the distribution of those journeys. This is a particular issue for the London Travelcard Area and Passenger Transport Executive (PTE)[4] areas. The Origin Destination Matrix (ODM) does include 'infills' that are estimated in order to account for journeys made on many of these products;
- **Concessionary travel** – Transport for London (TfL) and most PTEs subsidise some form of free travel for certain types of users including those over a certain age, students and those with disabilities. This creates a substantial additional element of demand which is very difficult to include in the ODM as information on the level and distribution of journeys associated with these free travel products is not recorded and will not even have point of sale information. The current approach to this in the ODM is to include this demand where data has been made available by TfL/PTEs which would generally be estimates based on surveys;
- **Non-LENNON Sales** – A significant proportion of sales is either not passed directly through LENNON (sold at non-railway sales points) or is included in LENNON in a format which requires additional processing and assumptions i.e. is not associated with a station to station flow;
- **Group stations** – Many products to major destinations are sold with the origin or destination as a group of stations (e.g. London Terminals, Manchester BR stations).

[4] There are six metropolitan counties in England. These are Greater Manchester, Merseyside, South Yorkshire, Tyne and Wear, West Midlands, West Yorkshire. Formerly, each of these areas had a Passenger Transport Executive (PTE), which was a local government body with public transport responsibilities. They were accountable to Integrated Transport Authorities (ITAs), which were formerly known as Passenger Transport Authorities (PTAs) prior to 2008 and the Local Government Act 2008. Following enactment of the Local Democracy, Economic Development and Construction Act 2009, all Integrated Transport Authorities have now been reformed into Combined Authorities, some with a larger geographic coverage than the ITA they replace. Some Combined Authorities (Greater Manchester, Merseyside, North East, South Yorkshire) continue to have a free-standing transport executive, whilst in others (West Midlands and West Yorkshire) the transport executive has been incorporated within the Combined Authority. In Scotland the Strathclyde Partnership for Transport is the equivalent body covering the region of Strathclyde. For convenience, in this report we continue to refer to these seven areas as PTEs.

Current industry data does not distinguish between the component stations and therefore a split between these stations has to be estimated during the production of the ODM; and

- **Ticketless travel** – Journeys associated with ticketless travel are not included in the datasets but as with journeys made on other products excluded from the datasets, some journeys would be observed in passenger counts. This is likely to be an issue on some flows and in some areas where ticketless travel is significant. As more stations have become gated over time and rail operators focus on revenue protection activities this is likely to be less of an issue than in the past in contributing to a shortfall in journeys. Finally, there is a strong argument that it is inappropriate to include ticketless travel in the Station Usage dataset as its purpose is to record bona-fide journeys on the rail network and inclusion of ticketless travel could distort business cases for new investment where these are reliant on Station Usage data.

1.8 It is important to remember that in aggregate the underlying data, from LENNON, is a rich and comprehensive data source and importantly covers the entirety of Great Britain. The issue is that when using the data source (in particular for Station Usage statistics) the data is being pushed significantly beyond what it was originally designed for which was primarily to report and allocate revenues across train operators.

Factors which can affect reporting of entries and exits

1.9 There are numerous factors that can affect the reporting level of usage at a station from year to year which could reflect local or more widespread trends or factors that have an impact on an individual or group of stations. Such factors should be considered by users of the statistics should consider such factors. Of particular note in the 2018/19 statistics:

- Infrastructure changes can significantly affect recorded journeys. New and improved routes and changes in service patterns and new connections between stations can lead to increases in the number of recorded journeys and distribution across stations. For example:
  - the completion of Thameslink infrastructure through Central London led to many more services through the Thameslink core from May 2018, along with changes in interchanges due to new direct journey opportunities.
  - the electrification of the Chase Line between Walsall and Rugeley was completed in December 2018.
- Engineering and upgrade work can result in temporary line and station closures. In 2018/19 this includes:
  - the blockade at Liverpool Lime Street in Summer 2018 which had a large impact in the city, and also on stations further across Merseyside and the North West; and
  - the electrification works in the North West of England at Blackpool and Bolton continued to have a detrimental impact on passenger journeys at specific stations in 2018/19.
- Industrial action can have an impact on train services and consequently on rail journeys that are undertaken. Northern rail were impacted by Industrial Action in 2018/19.

Adverse Weather and Consequential Impacts

1.10 Cases of extreme adverse weather may cause disruption to normal railway operations and can impact on travel patterns. There were several named storms in 2018/19 which led to

disruption on the railways, including storms Ali, Bronagh and Callum in Autumn 2018, and storms Freya and Gareth in Spring 2019.

Major incidents

1.11 Major incidents affecting services such as those at Southall (1997), Ladbroke Grove (1999), Hatfield (2000) and Grayrigg (2007).

Gating Schemes

1.12 Installation of ticket gates can significantly affect not only the usage figures at that station, but also those at neighbouring stations. The gates help to ensure that customers purchase tickets, but customers may also alter their travel patterns to avoid gated stations. We would expect travel patterns to be most affected in the months following the installation of the gates.

Change in Service Pattern

1.13 Alterations in service frequency or stopping patterns would be expected to alter station usage figures. This is particularly apparent where a group of stations along a line show similar increases or decreases. Again, this can be a long-term trend.

Ticket Issuing Facilities Changes or Product Changes

1.14 Some London stations have both underground and National Rail trains operating. LENNON does not directly capture tickets sold by London Underground, only those sold by Train Operating Companies (TOCs). Changes in ticket facilities provided by TOCs, for example the provision of ticket machines, can therefore increase the ticket sales captured by the system.

1.15 Changes to on-train and station ticket sales can have an impact as well as product changes can have an effect on passengers' purchasing patterns at rail outlets thus affecting Station Usage data. For example, the introduction of Oyster cards and, more recently, Contactless Payment can affect stations inside the Travelcard boundary in the London area.

Engineering Work

1.16 Engineering work can alter customers' travel patterns, either causing passengers to not travel, use an alternative mode or use an alternative rail route. The works can range from overnight possessions through to weekends, week-long or longer periods e.g. the blockade at Liverpool Lime Street in Summer 2018.

Advance tickets

1.17 Advance tickets can be sufficiently cheap to incentivise travellers to purchase a number of tickets but only use one dependent on how their circumstances change, creating an inflated number of trips in the ticket sales data. This can be particularly true for business travel and could overstate actual journeys.

Tourism/Leisure

1.18 Stations near to tourist and leisure attractions may show significant changes in usage as a result of weather, promotions or other factors, which affect tourists' journeys.

New/Special Stations

1.19 Some stations serve a particular activity or business. Some fluctuation in usage of such stations is reasonable. Such activities include:

- Sporting Events e.g. The 2018 Open Golf tournament at Carnoustie in Scotland,
- Special Events e.g. Birmingham International (for the National Exhibition Centre), Exhibition Centre station in Glasgow (for the Scottish Exhibition and Conference Centre);
- Airports, where rail demand is closely linked to airport passenger numbers e.g. Gatwick Airport, Stansted Airport and Southend Airport.

1.20 In addition, where there are new stations, ramp up effects can cause large demand increases over a number of years.

Trend of Growth or Decline

1.21 For stations with a history of growth or decline, it is reasonable to expect this trend to continue. There are many possible reasons for these trends, such as demographic and employment changes (new developments in the vicinity), changes in rail service levels or new stations abstracting demand.

Errors in recording of Sales of Individual Ticket Types

1.22 Miscoding of ticket information entered into LENNON can alter the statistics, although this would not be reflecting an actual change in customers' journeys.

Changes in journeys per ticket

1.23 Sales of tickets are assumed to correspond to a number of journeys. In the case of Single and Return tickets the relationship is very clear but for period tickets, e.g. weekly/monthly/annual seasons, there are a set of journey factors have been used within the LENNON system that have remained fixed for a number of years. Whilst they were likely based on reasonable estimates of ticket use made in the past it could be argued that with lifestyle and working practice changes, e.g. greater flexible working, and ticketing arrangements that they are not as representative for today's market.

2 Methodological Overview

Introduction

2.1 All estimates of entries, exits and interchanges included in the Station Usage dataset are derived from the Origin Destination Matrix (ODM), also produced by Steer for the ORR. The ODM is in turn derived primarily from a matrix of journeys and revenue that is produced by Resonate for inclusion in MOIRA2.2.

2.2 The MOIRA2.2 matrix includes a comprehensive representation of travel on the national rail network. The base data for the MOIRA2.2 demand matrix is LENNON ticket sales, with the addition of "infills" for London Travelcards, some specific tickets to/from airports and multi-modal and zonal products sponsored by Passenger Transport Executives (PTEs). The current MOIRA2.2 matrix now includes some of the methodological enhancements that have been previously developed for inclusion in the ODM, for example a revised methodology for the PTE infills – see the accompanying report, '*Station Usage & Origin Destination Matrix 2018/19: Historical Methodological Changes*' – and the 'Other' infills relating to selected Rover and Ranger products.

Base Data

LENNON and MOIRA2.2

2.3 The underlying matrix of ticket sales and associated journeys and revenue used in MOIRA2.2 is derived from LENNON. It is based on an extract from LENNON, produced by Worldline, of total sales revenue and journeys for the year, broken down by flow (origin and destination National Location Code (NLC)), route code and by product type (CTOT). However, as there are known omissions in this data in respect of Transport for London (TfL) and PTE sponsored tickets, and non-National Rail tickets on some airport services, there needs to be a "matrix infilling" exercise undertaken. This enables the estimation of a more complete origin-destination matrix and include the associated journeys and revenue that either do not appear in the underlying matrix at all or at a sufficiently disaggregate flow level.

2.4 There are three main cases:

- Tickets with non-geographical destinations, e.g. zonal products, Rovers;
- Tickets sold at some non-National Rail[5] outlets, e.g. newsagents; and
- Tickets which do not appear in LENNON at all. This includes some TOC tickets on airport flows and tickets for TOCs which fall outside the Rail Settlement Plan (RSP).

2.5 Certain tickets with destination codes that are not national rail stations are included in the MOIRA2.2 demand matrix, being mapped to the corresponding rail station. These 'Rail Links' usually include a third-party element, such as to a bus zone, or tourist/leisure attraction. The MOIRA2.2 demand matrix includes the journeys and the net revenue associated with such tickets.

[5] Not part of Rail Settlement Plan (RSP)

2.6 Data excluded from the MOIRA2.2 demand matrix is set out in chapter 5.

Ticket Type Definitions

2.7 Within the base demand matrices, journeys and revenue have been sub-divided into the following four ticket types, each of which is further split by First & Standard Class:

- Full: all walk-up undiscounted single or return tickets, whether or not issued with a status discount (child, railcard etc);
- Reduced: all walk-up discounted single or return tickets, whether or not issued with a status discount (child, railcard etc);
- Advance: all advance-purchase tickets; and
- Seasons: all multi-use tickets.

2.8 It should be noted that for the purposes of the Station Usage dataset, Advance products are included in the Reduced ticket category and First and Standard classes are combined.

Infills for London Travelcards, Major Urban Areas (PTE) & Airports

2.9 Infills are included within the MOIRA2.2 demand matrix to add in the missing journeys and revenue identified in para 2.4 in three key areas:

- **Within London Travelcard area.** Whilst the underlying matrix includes an estimate of journeys made on Day Travelcards / Travelcard seasons purchased at National Rail stations, it does not include a significant number of national rail trips made using Travelcards purchased at Tube stations, travel shops and newsagents. From 2015/16, a new methodology has been used to represent 'in-boundary' Travelcards based on Transport for London's (TfL) Oyster Clicks Model (OCM) – see the accompanying report, 'Station Usage & Origin Destination Matrix 2018/19: Historical Methodological Changes', for more detail. Also from 2015/16, the matrix has included journeys associated with Freedom Pass (these were previously added as an infill within the ODM).
- **Within Passenger Transport Executive (PTE) areas.** The underlying matrix excludes virtually all rail trips made on PTE-sponsored tickets, which are usually zonal and often multimodal. From 2015/16 the 'infills' representing these journeys in MOIRA2.2 have been based on a methodology originally developed for the ODM.
- **Trips to/from Airports.** The underlying matrix includes many trips to/from airports, but excludes all Heathrow Express journeys, and some tickets sold for Gatwick Express, Stansted Express and other airport operators.

2.10 There are also other ticket sales which are not included in the MOIRA2.2 demand matrix, but these are generally much less significant. It should also be noted that journeys with no associated ticket sales such as staff travel, and particularly fare evaders, are not included in the MOIRA2.2 demand matrix and therefore are not included in the ODM either.

2.11 The most significant "infills" are for the London Travelcard area (sales made by TfL), and for PTEs, since in both cases a substantial proportion of the rail journeys use multimodal travelcard-type tickets.

2.12 The third infill, for airports, estimates the significant number of rail journeys on both Gatwick and Stansted Express, made on tickets sold outside of the RSP system i.e. not sold by National Rail outlets. Journeys on Heathrow Express are excluded from the MOIRA2.2 demand matrix.

Origin Destination Matrix (ODM)

2.13 The MOIRA2.2 demand matrix is used as the starting point for the production of the ODM and as part of this process a number of adjustments and overlays are included which can be categorised as follows:

- Overlays (in addition to those already included in the MOIRA2.2 matrix relating to the London Travelcard Area and Airports – see paragraph 2.9)
  - PTE infills – although included in the MOIRA2.2 matrix these are developed as part of the work undertaken to produce the ODM and are provided to Resonate for inclusion in the MOIRA2.2 matrix. The methodology development work to produce the revised infills was undertaken between 2011/12 and 2014/15 and the methodology has remained largely unchanged in the current year statistics.
  - Ranger/Rover infills – Methodological development was undertaken to include a representation of passenger flows on a selected number of Rover and Ranger products from 2011/12. Since 2015/16 this infill has also been included in the MOIRA2.2 matrix.
- Adjustments:
  - Allocation of demand associated with tickets sold to 'London Terminals' between those terminals;
  - Allocation of demand between individual stations within station groups outside central London. For example where tickets are sold to/from 'Dorking BR' it is necessary to estimate how these journeys are distributed between Dorking West, Dorking and Dorking Deepdene stations. To support this part of the methodology there is a programme of station counts that are undertaken on an annual basis at selected stations;
  - Unknown destinations: Ticket sales do not always tell us where a passenger is travelling, for example where the Origin or Destination is a London Travelcard. Unknown destinations are converted into an estimate of the actual stations that passengers are travelling to. The full detail of this part of the methodology appears in Appendix A;
  - Individual station adjustments: There are a number of cases where adjustments are made to selected stations to account for specific known issues:
    - Adjustments at a number of stations are made to reflect circumstances where there are significant numbers of season tickets sold at a particular station (where the passenger travels from) for travel to London that allow for travel to/from a different origin station to provide flexibility. This leads to a situation where station usage, as estimated by ticket sales, can be under- or over-estimated and journeys involving those stations needs to be adjusted to reflect actual usage. Since 2014/15, an adjustment has been at selected stations where this issue has been identified. Further details on these adjustments can be found in in the accompanying report, '*Station Usage & Origin Destination Matrix 2018/19: Historical Methodological Changes*'.
  - The 'Digby & Sowton' adjustment – described in the accompanying report, '*Station Usage & Origin Destination Matrix 2018/19: Historical Methodological Changes*', and first included in the 2014/15 dataset – relates to journeys associated with a season ticket product for students which are being made to Exeter Central and Exeter St. David's on tickets with a recorded destination of Digby & Sowton.

2.14 Further details relating to the overlays and adjustments outlined above can be found in chapter 3 of this report, and in the accompanying report, '*Station Usage & Origin Destination Matrix 2018/19: Historical Methodological Changes*'.

Interchanges

2.15 In addition to entries and exits at stations an estimate of the number of people interchanging at each station is included in the dataset. This is obtained by combining the number of journeys made on each flow (from the ODM) with the information on passenger journeys taken from the Central Allocations File (CAF).

2.16 The CAF is an output of the ORCATS system which predicts passenger choices of rail route and train used, and determines the allocation of passenger revenue between TOCs. Since ORCATS is a model, the CAF contains estimates rather than actual journeys. However, it is used throughout the rail industry, so it is an appropriate source of data to use for this purpose. Since CAFs are updated with the timetable, not with financial years, no CAF will match the ticket sales data exactly. The December 2018 CAF is used in the creation of the 2018/19 Station Usage dataset.

2.17 The CAF contains:

- Origin and destination;
- Route alternatives for each origin and destination, including all interchange points;
- Ticket type data; and
- For each flow, the proportion of passengers who choose to travel on each route alternative as calculated by the ORCATS model.

2.18 An overview of the ORCATS allocation process can be found in in Appendix B.

3 Methodological Changes in 2018/19

Introduction

3.1 Consistency with past datasets is important to enable comparisons to be made over time. Nonetheless, stakeholders have indicated that they are keen to see improvements, even where this reduces consistency with historic data, provided any changes are clearly explained. Steer has worked with the ORR to scope and implement methodological enhancements to address identified issues and utilise new data as it is made available whether this is from primary data collection (e.g. passenger counts at stations), or industry systems such as TfL's Oyster Clicks Model (OCM).

3.2 The number of changes made to in the 2018/19 dataset is relatively small in comparison to previous years. The changes made to improve the dataset are explained in the rest of this chapter, together with some quantification of their impact.

Revisions to 2017/18 Data

3.3 During the production of the 2018/19 statistics, a routine update was made to the West Midlands PTE infill to reflect updated ticketing data provided by the PTE. Each year, infills are updated using a combination of PTE sold and LENNON ticketing data.

3.4 After updating both these data for 2018/19, a high growth rate was identified between the 2018/19 and 2017/18 infills for the West Midlands. While growth in the region was expected to be high, the apparent growth rate exceeded expectations and did not closely align with other published sources.

3.5 On reviewing the 2017/18 processing, it was identified that the PTE infill was likely understating demand in the region. The LENNON sold portion of the infill was reflecting demand from 2017/18, but the PTE sold demand data reflected 2016/17 levels.

3.6 The 2017/18 dataset was subsequently rerun, and the estimated impact was an increase in usage of 2.3% in the West Midlands PTE area, or around 2.75m entries and exits. This change only impacted stations within the West Midlands PTE area.

Table 3-1: West Midlands PTE – 2017/18 update

PTE Urban Area	Entries & Exits 1819	Entries & Exits 1718 - Rerun	Entries & Exits 1718 - Published	Increase - Published vs Rerun, abs	Increase - Published vs Rerun, %
West Midlands	127,930,142	118,104,796	115,338,144	2,766,652	2.3%

3.7 At a national level, this represents a change of less than 0.1% of total entries and exits.

3.8 For consistency, throughout this document, growth rates and 2017/18 totals are calculated using the re-run 2017/18 dataset, as this reflects a best estimate of demand in 2017/18.

3.9 It was also identified that the Ilkeston station usage estimates for 2017/18 had been double counted. This wasn't obvious at the time of producing the original 2017/18 estimates as this was the first year Ilkeston station had been operational, however the overstatement became apparent when compared with the 2018/19 estimates. This has been updated in the revised 2017/18 ODM and station usage dataset.

Key PTE Changes

3.10 Each year, PTE infills are prepared by Steer (West Midlands) and Mott MacDonald (Greater Manchester, Merseyside, South Yorkshire, Tyne and Wear, West Yorkshire).

3.11 These infills are subject to annual improvements, which normally represent a simple update, but some years contain a step change in the methodology.

3.12 For 2018/19, Concessionary ticketing data was available for Greater Manchester PTE for the first time. The table below shows the split of sales by PTE product.

Table 3-2: Summary of TfGM Infill change in 2018/19

Ticket Type	2017/18 Journeys	2018/19 Journeys	Abs Change 2017/18-2018/19	% change 2017/18-2018/2019
Traincard	3,387,129	3,479,328	92,199	2.65%
Countycard	1,548,081	1,462,476	-85,605	-5.85%
Wayfarer	126,983	81,438	-45,546	-55.93%
GMPTE Accompanied Child	150,958	125,845	-25,113	-19.96%
DaySaver	53,615	44,454	-9,161	-20.61%
Rail Ranger	12,902	9,078	-3,824	-42.12%
Concessions	0	3,705,977	3,705,977	100.00%
Total	5,279,668	8,908,596	3,628,928	40.74%

3.13 This led to a total increase of 3.6m journeys, or 7.2m entries and exits, across the Greater Manchester area.

Season ticket journey adjustments

3.14 In the production of previous years' statistics, adjustments were made to account for situations where passengers buy season tickets for travel to/from a station other than the one they generally travel from, in order to allow additional flexibility. This issue and the previous adjustments are described in detail in the accompanying report, 'Station Usage & Origin Destination Matrix 2018/19: Historical Methodological Changes'.

3.15 In some areas, multiple stations have identically priced season tickets to London. As a result, London season tickets are generally sold as being from the furthest station, regardless of the actual origin of travel. This means that the ticket sales data shows that there are more people travelling to/from this station than is actually the case.

3.16 LENNON sales data was used to estimate the number of tickets where the issuing office was at a branch line station but the ticket origin showed a station further along a line. In these cases, it was assumed that the journey was actually being made from a point on the branch line and not the recorded origin.

3.17 For the production of the 2018/19 statistics, the analysis underpinning this reallocation was updated with 2018/19 LENNON data. Table 3-3 shows the scale of the adjustments.

Table 3-3: Summary of adjustments in 2018/19

Station TLC	Station Name	Station Group	Adjusted Journeys	Total Journeys
SOV	Southend Victoria	Southend Victoria Branch	-293,456	2,129,590
HOC	Hockley		-360,173	724,288
RLG	Rayleigh		493,169	1,832,722
RFD	Rochford		124,590	615,582
PRL	Prittlewell		23,430	220,626
SIA	Southend Airport		9,644	611,256
WIC	Wickford		2,796	2,261,210
REI	Reigate	Reigate/Redhill	-155,388	1,436,558
RDH	Redhill		155,388	3,787,090
CHW	Chalkwell	Southend Central Branch	-66,071	1,913,550
BEF	Benfleet		31,435	3,664,766
LES	Leigh-On-Sea		31,361	2,355,650
SOC	Southend Central		3,275	3,446,092
GRV	Gravesend	Gravesend	-55,733	3,087,304
EBD	Ebbsfleet International		55,733	2,106,802
EGR	East Grinstead	East Grinstead/Lingfield	-60,033	1,586,800
LFD	Lingfield		60,033	503,444
ELD	Earlswood (Surrey)	Earlswood/Redhill	-63,349	404,288
RDH	Redhill		63,349	3,787,090
CTM	Chatham	Chatham/Rochester	-58,095	2,730,416
RTR	Rochester		58,095	2,056,936
HIB	High Brooms	Tonbridge	-12,104	1,256,908
TON	Tonbridge		-34,225	4,554,198
TBW	Tunbridge Wells		46,330	3,838,532
GRY	Grays	Thameside	-55,038	4,165,432
CFH	Chafford Hundred Lakeside		13,235	2,911,958
OCK	Ockendon		20,496	1,160,354
PFL	Purfleet		21,307	686,726
GLM	Gillingham (Kent)	Medway	-38,556	2,747,444
CTM	Chatham		-58,095	2,730,416
RTR	Rochester		58,095	2,056,936
SOO	Strood (Kent)		38,556	1,194,020

Demand allocation at Group Stations

3.18 In order to validate and improve the allocation of journeys between stations within groups (e.g. Dorking BR), passenger counts are routinely carried out at selected group stations on the network. The most recent counts were carried out in Autumn 2018 and Spring 2019 at the following station groups:

- Brighton Main Line Stations
- Brighton BR
- Canterbury BR
- Colchester BR
- Dorking BR
- Folkestone BR

- Hertford BR
- Portsmouth BR
- Wakefield BR
- Bristol BR
- Exeter BR
- Guildford BR
- Edinburgh BR
- Reading BR

3.19 These counts were used both to validate existing data, and to create or update station group counts splits, with three new sets of station counts splits created. The proportion of journeys split between stations based on routine counts are shown in Table 3-4 below. Note that not all counts splits are applied, hence the table below does not include all the stations noted above.

Table 3-4: Count-based adjustments to 2018/19 statistics

Station Name	Station Group	Proportions	Year Undertaken
Colchester	COLCHESTER BR	85.0%	2013/14
Colchester Town		15.0%	
Bedford Midland	BEDFORD BR	96.0%	2014/15
Bedford St.Johns		4.0%	
Dorchester South	DORCHESTER BR	74.0%	2015/16
Dorchester West		26.0%	
Deepdene	DORKING BR	25.0%	2014/15
Dorking		72.0%	
Dorking West		3.0%	
Farnborough (Main)	FARNBOROUGH BR	82.0%	2013/14
Farnborough North		18.0%	
Maidstone Barracks	MAIDSTONE BR	11.0%	2013/14
Maidstone East		55.0%	
Maidstone West		34.0%	
Newark Castle	NEWARK BR	46.0%	2015/16
Newark North Gate		54.0%	
Portsmouth & Southsea	PORTSMOUTH BR	49.0%	2014/15
Portsmouth Harbour		51.0%	
Wakefield Westgate	WAKEFIELD BR	82.0%	2014/15
Wakefield Kirkgate		18.0%	
Canterbury East	CANTERBURY BR	30.0%	2014/15
Canterbury West		70.0%	
Edenbridge	EDENBRIDGE BR	49.0%	2014/15
Edenbridge Town		51.0%	

Station Name	Station Group	Proportions	Year Undertaken
Falkirk Grahamston	FALKIRK BR	44.0%	2014/15
Falkirk High		56.0%	
Helensburgh Central	HELENSBURGH BR	98.0%	2014/15
Helensburgh Upper		2.0%	
Worcester Foregate Street	WORCESTER BR	72.0%	2015/16
Worcester Shrub Hill		28.0%	
Southend Central	SOUTHEND BR	46.0%	2015/16
Southend Victoria		28.0%	
Southend East		26.0%	
Warrington Bank Quay	WARRINGTON BR	44.0%	2015/16
Warrington Central		56.0%	
Wigan North Western	WIGAN BR	52.0%	2015/16
Wigan Wallgate		48.0%	
Folkestone Central	FOLKESTONE BR	57.1%	2018/19
Folkestone West		42.9%	
Hertford East	HERTFORD BR	51.1%	2018/19
Hertford North		48.9%	
Guildford	GUILDFORD BR	88.0%	2018/19
London Road Guildford		12.0%	

Revision to 2018/19 data for 'Multiflex' issue

3.20 In March 2020 an issue was identified with how journeys on the 'Multiflex' ticket (product code 2BRS), used on Transport for Wales (TfW) services, were recorded in LENNON. This ticket was previously 6 return products, but in P08 of 2018/19 was converted to 12 single products. This change was incorrectly reflected in LENNON, with each of the 12 single products being allocated 12 journeys. This meant for the 12 single products, LENNON recorded 144 journeys. Station usage numbers in Wales (and to a lesser extent some stations in England served by TfW) for 2018/19 were therefore overstated. It is worth noting that this only affected journeys and not revenue and was corrected in LENNON in P01 of 2019/20.ORR asked Steer to adjust the 2018/19 station usage statistics and ODM to account for this error. Analysis was undertaken on LENNON data to calculate how much each affected flow should be adjusted by and hence a revised ODM and station usage dataset were produced and published in June 2020.

3.21 This adjustment resulted in an overall reduction of total station usage of 0.2%, with the region most affected being Wales (8.3% reduction in station usage). This has also explained some of the high growth rates for Welsh stations, for example Rhiwbina previously had a growth rate (vs. 2017/18) of 51% but after the Multiflex adjustment was applied this reduced to 4%.

4 Validation

Introduction

4.1 Checks undertaken on the Station Usage dataset encompass a number of elements, including:

- Investigation of large increases and decreases for individual stations;
- Checks at different geographical levels; and
- Validation against other data sources.

Data Checks

Large increases and decreases

4.2 Table 4-1 shows the 10 stations with the largest proportional increases in total usage for stations with more than 10,000 entries and exits.

Table 4-1: Top 10 Increases in 2018/19

Station Name	18/19 Entries & Exits	17/18 Entries & Exits	Change	Reason
Styal	13,980	5,926	135.9%	Some stations usage increased during disruption in Merseyside in Summer 2018, possibly due to passengers taking alternate routes into Liverpool
London Road (Guildford)	1,021,802	540,978	88.9%	Methodological Change – adjusted by observed station counts
Walton (Merseyside)	450,148	248,896	80.9%	Some stations usage increased during disruption in Merseyside in Summer 2018, possibly due to passengers taking alternate routes into Liverpool
Rice Lane	580,620	324,336	79.0%	
Hall Road	460,036	266,608	72.6%	
Deptford	1,785,086	1,057,530	68.8%	Related to increases in Southeastern and Thameslink usage
Cambridge North	812,972	488,878	66.3%	New station that opened part way through 2017/18
Wanstead Park	886,990	563,186	57.5%	Reflects growth at all stations along Gospel Oak to Barking line following completion of electrification and service improvements

Station Name	18/19 Entries & Exits	17/18 Entries & Exits	Change	Reason
Carnoustie	200,460	127,364	57.4%	Carnoustie hosted the Open Golf tournament in Summer 2018
Heswall	91,736	58,452	56.9%	Some stations usage increased during disruption in Merseyside in Summer 2018, possibly due to passengers taking alternate routes into Liverpool

4.3 Table 5.2 shows the 10 stations with the largest proportional decreases in total usage for stations with more than 10,000 entries and exits.

Table 4-2: Top 10 Decreases in 2018/19

Station Name	18/19 Entries & Exits	17/18 Entries & Exits	Change	Reason
Halewood	53,594	135,832	-60.5%	Impacted by Liverpool Lime Street closure and knock-on impacts in Merseyside
West Allerton	43,258	101,296	-57.3%	
Whiston	192,828	411,222	-53.1%	
Mossley Hill	125,558	262,736	-52.2%	
St Helens Central	657,274	1,301,488	-49.5%	
St Helens Junction	212,050	413,616	-48.7%	
Huyton	595,444	1,150,124	-48.2%	
Thatto Heath	134,750	257,132	-47.6%	
Roby	182,748	338,272	-46.0%	
Garswood	190,210	350,520	-45.7%	

4.4 As in the 2017/18 dataset two flags have been included in the published dataset identifying:

- Stations with more than 10,000 entries and exits a year where entries and exits have increased or decreased by more than 10% (Large station change flag); and

- Stations with less than 10,000 entries and exits a year where entries and exits have increased or decreased by more than 25% (Small station change flag).

4.5 These flags have been used to identify stations where further investigation should be carried out to ensure, where possible, the reported changes reflect reality. The limits set are demanding (10% of 10,000, for example could represent just two extra season ticket holders per year) and investigations have been focussed on the most significant changes but where obvious explanations for less significant changes are available these have been included in the Station Usage dataset. In total 624 stations were captured by one of the two flags.

4.6 Whilst reasons for large changes at some stations are specific to that station, in many instances there are groups of stations where there is a common cause for the changes seen. We have identified a number of reasons that affect multiple stations in the 2018/19 statistics. These are shown in Table 4-3.

Table 4-3: Summary of identified reasons for large changes

Trend	Description
Methodology	These changes relate to methodological updates rather than changes to the underlying demand (see Chapter 3 for detail).
Growth or decline trend	These stations are those that consistently experience strong growth or decline in usage year on year. These can be due to a variety of exogenous and endogenous reasons.
Line specific trend	Consistent growth and decline has been observed on some lines. This can be due to service changes or potentially changes to ticketing (and therefore recording).
Local Factors	Demand at some stations is strongly linked to demand for nearby airports, industrial/employment sites, or new housing developments. Furthermore, if multiple stations are available nearby, passengers may switch to an alternative nearby station.
Weather disruption	These relate to impacts of weather-related disruption, including landslips and floods that caused line closures. These impacts can be negative (reduced usage due to disruption), or positive (usage bounce-back following end of disruption)
Rail operator disruption	These impacts relate to operator issues, for example poor performance or reduced timetable offering.
Industrial action disruption	These impacts relate to industrial action, as seen on Northern in 2018/19 (and growth observed in areas with less industrial action than in previous years).
Timetable change	These relate to improvements to journey opportunities or quality.
Facility change	These relate to improvements to passenger environment improvements (e.g. improved rolling stock or stations).
Fluctuating usage at small stations	Small stations are often flagged due to the high percentage change implied by a small change in absolute usage.

Validation against alternative data sources

Comparison with ORR journey data on the ORR data portal

4.7 The ORR produces passenger journey data by sector and TOC and makes this available on the ORR website via its data portal and as a National Statistics release[6]. This dataset shows an increase in overall journeys from 2017/18 to 2018/19 of 3.0% at the national level for all operators. The Station Usage dataset shows an increase of 3.1% over the same period.

Comparison with passenger count data

4.8 The Department for Transport (DfT) collects passenger count data for major cities throughout Great Britain. The method of collection means that for through stations it is often not possible to calculate boarders and alighters but for terminal stations this is usually possible. Using data published by DfT we have compared growth rates at the major London termini covered by the all-day arrivals and departures count data with those seen in the calculated Station Usage dataset.

4.9 In addition, Steer also carried out validation against Network Rail gateline data for their managed stations, in order to validate the total entries and exits at each of these stations. This data is based on infra-red sensors mounted above station gatelines, and differentiates between passengers accessing platforms and the station concourse, hence is a robust comparison against the Station Usage's LENNON ticket sales methodology.

Table 4-4: Comparison of Station Usage and reported growth rates of all-day passenger counts at London Terminals 2017/18–2018/19

Station	Note	SU Growth (2017/18 – 2018/19)	DfT Growth, Autumn 2017 vs 2018	Comment
London Blackfriars	Measured at Elephant and Castle	12.4%	19.2%	Change is generally in the same direction. Given the disruption after the May 18 timetable introduction, and the overall scale of the change, this seems reasonable
London Euston		3.1%	-0.2%	
London Fenchurch Street		0.6%	-2.3%	
London King's Cross		2.2%	-6.9%	Large difference – potentially related to the impact of Thameslink services, and passengers switching between king's Cross and St Pancras
London Liverpool Street		3.8%	-2.2%	

[6] Passenger Rail Usage, available at: **http://orr.gov.uk/statistics/published-stats/statistical-releases**

Station	Note	SU Growth (2017/18 – 2018/19)	DfT Growth, Autumn 2017 vs 2018	Comment
London Bridge		26.5%	-1.3%	Large difference – likely that the dates for DfT counts did not capture the difference between the fully open station and the station partially closed for Thameslink Engineering work
London Marylebone		-3.3%	-5.9%	
Moorgate	Measured at Old Street	10.3%	5.5%	
London Paddington		4.4%	0.5%	Station usage figures exclude Heathrow Express so comparisons are not like for like
London St Pancras International		3.9%	4.4%	
London Victoria		-0.3%	6.0%	
London Waterloo	Measured at Vauxhall	-0.2%	4.3%	

Source: *Rail passenger numbers and crowding on weekdays in major cities in England and Wales: 2018. Table RAI0201- "City centre peak and all day arrivals and departures by rail on a typical autumn weekday, by city: 2018"*
Available at: **https://www.gov.uk/government/statistics/rail-passenger-numbers-and-crowding-on-weekdays-in-major-cities-in-england-and-wales-2018**

4.10 Where there is significant variation between observed growth from counts data and the station usage growth, there are some general points to be noted, and some points specific to individual stations (which are noted in the table above).

- Firstly, the count data may also include interchanging passengers and therefore does not provide an exact comparison with the station usage entry and exit data;
  Secondly, the period over which data has been collated is different, as the counts are from Autumn 2017 and Autumn 2018, whereas the station usage figures aggregate annual passenger numbers for the years beginning April 2017 and April 2018. In this case, any external event during the counts period, or any external event during the rest of the year which was not present during the counts period, will drive a difference between the two data sources.

5 Station Usage Dataset Limitations

Limitations of the LENNON data

5.1 The LENNON database captures ticket sales for the entire national rail network from many different input machines. It is consequently a very large dataset. With all large data sources there will always be input errors resulting in a certain amount of invalid data. Generally, such errors will be small, and are more likely to occur in the journeys rather than revenue fields.

5.2 Checks are performed on the data when the MOIRA2.2 demand matrix is compiled, but due to the size and complexity of the dataset it is not possible to validate each and every entry.

5.3 We have used similar information extensively in the last ten years or more, and have found the data to be reliable, particularly when examining the data at an aggregated level.

5.4 There are a number of areas where we know that LENNON does not capture the data correctly, or instances where it is not possible to derive passenger journeys from ticket sales data. These areas are expanded upon below.

Known Problems of Data Capture

5.5 The data in LENNON from which the ODM is derived is based on ticket transactions. In order for the data to be included in the ODM it must include an origin station and a destination station. However if this is not the case then the data will automatically be excluded.

5.6 Human error at the point the ticket sale is entered into the input machines will also produce invalid data in LENNON.

Travelcards

5.7 As Travelcards are for multi-modal travel they allow the purchaser to make journeys on the rail system and on other modes. Equally, tickets purchased elsewhere on the local transport system will be valid for rail travel. Therefore LENNON gives only a partial picture of the rail travel in conurbation areas, such as: London, Birmingham, Glasgow, Leeds, Liverpool, Manchester, Newcastle and Sheffield.

5.8 The ODM contains reasonably robust estimates of journeys within London and other conurbation areas where travelcards are widely used. An infill for London Travelcards has been included in the ODM since 2006/07, and an infill for PTE tickets is included from 2008/09. Both these infills have been subject to methodological enhancements in recent years.

Return and Single Journey Tickets

5.9 It is possible that on certain routes the cost of a return ticket could be lower than a single ticket. This leads to the cheaper return ticket being purchased even though the passenger has no intention of making the return journey by rail. This results in two journeys being recorded instead of one.

Multiple Tickets

5.10 It is possible to buy special cheaper tickets between certain stations for example under a promotion by one of the train companies. In these cases a local ticket may be bought to gain access to a main station and a second ticket bought for the rest of the journey. This results in

two journeys being recorded in the ODM and will not accurately represent the journey undertaken.

Rail Staff Passes

5.11 Prior to the privatisation of the rail network, British Rail employees and their families were eligible to various levels of free or reduced rate rail travel. When the various rail companies were converted to private companies, this benefit often continued.

5.12 If you consider the network as a whole, the effect of staff passes is unlikely to be significant. However, it may be significant on certain routes, for example on routes out of Derby due to large concentration of companies in Derby relating to British Rail both pre and post privatisation.

Ticketless Travel

5.13 On every route on the network there will always be passengers who travel without purchasing a ticket. This is referred to as ticketless travel. As LENNON data is derived from ticket transactions it cannot reflect this travel.

Other Rail Systems

5.14 There are a number of rail systems in operation in the country that are not covered by LENNON. For Heathrow Express and Eurostar revenue and journeys data were not available.

Journey Factors

5.15 Ticket transactions are converted into an estimate of the number of journeys made by applying a series of ticket type journey factors. Single and return tickets unambiguously translate into one and two journeys respectively, for season tickets, the factors used represent a rough historic estimate as set out in Table 5-1.

5.16 Ticket periods of other lengths are converted to a number of journeys using a proportion of the monthly journey factor.

5.17 Therefore the journeys data in the ODM represents an assumed number of journeys made based on the ticket type sold and the above journey factors. In particular it should be noted that the journeys data has not been cross-checked against other data sources of the actual number of journeys made on the network.

5.18 These journey factors have been used within the LENNON system for a number of years at their current values. The source of the factors is unclear, and there is some indication that they were based on reasonable estimates of ticket use made in excess of fifteen years ago. It can therefore be argued that these journey factors do not provide an accurate estimate of the number of journeys that result on the rail system at present, or in any ODM.

Table 5-1: Journey Factors used in LENNON

Single Journey Ticket	1
Return Journey Ticket	2
Return Journey 2 Persons	4
3 Day Return/ 6 Single Journeys	6
4 Day Return/ 8 Single Journeys	8
5 Day Return/ 10 Single Journeys	10
6 Day Return	12
5 Day Single	5
1.5 Journeys	1.5
Weekly Ticket	10.3
10 Day Return/ 20 Single Journeys	20
2 Weekly Ticket	22
Seasons-Variable Periods	***
Monthly Ticket	45
Not Used	0
3 Monthly Tickets	135
Not Used	0
6 Monthly Tickets	270
Summary Group Codes	***
Annual Ticket	480
8 Day Ticket	22
22 Day Ticket	44
14 Day Ticket	30
50 Journeys	50
10 Weeks	103

Data Excluded From Estimates of Station Usage

5.19 Some of the LENNON data has been excluded from the MOIRA2.2 Demand Matrix, and subsequently from the ODM.

5.20 All products that are classified into the 'miscellaneous' ticket pot is excluded. These products are:

- Car Parking;
- Railcard Sales;
- Penalty/Excess Fares;
- Seat Reservations;
- Sleeper Supplements.

5.21 Also excluded from the analysis were all the flows that had either an Origin or Destination that did not represent a geographical location (these are mainly "I codes"), e.g.

- Rover and Ranger Tickets (except those included in the new 'Other' Infill in 2011/12 and subsequent years);
- BritRail Tickets;
- Gate passes usually used by staff;
- Passenger Charter Discounts;
- Headquarters Input Items, other than those which can be identified as TfL or PTE.

5.22 Finally for flows that have either Origin or Destination as a Private Settlement Code some are included and some are excluded.

- PTE tickets and TfL sold London Travelcard records from LENNON are removed, and replaced with an estimate of all rail travel using these tickets via 'infill's to the MOIRA2.2 demand matrix.
- PlusBus – all significant flows have been included since 2007/08 and minor flows are excluded.
- Attractions – the rail element of the significant flows have been included since 2007/08, which include:
    - Bluewater Shopping Centre;
    - Alton Towers;
    - Whipsnade Zoo;
    - Chatsworth House.

5.23 All other flows involving Private Settlement are excluded, e.g. Irish Stations.

Appendices

A Methodology: Tickets with Unknown Origins or Destinations

A.1 Ticket sales do not always tell us where a passenger is travelling. Ticket sales can be divided into the seven categories listed in table below. Ticket sales data has been converted into an estimate of the actual stations that passengers are travelling from/to.

A.2 The processing of ticket sales data is undertaken in the creation of the MOIRA2.2 demand matrix, and then subsequently in the creation of the ODM. For each of the flow categories, the table below states where the flow is processed: MOIRA2.2 or ODM.

Table A.1: Categorisation of ticket sales in LENNON

Flow Category	Description	Processing
Category 1	Origin and Destination Stations Known	No processing required
Category 2	Origin or Destination a Group Station (excl. London BR)	ODM
Category 3	Origin or Destination is London Terminals	ODM
Category 4	Origin or Destination a London Travelcard including Zone 1	ODM
Category 5	Origin or Destination a London Travelcard excluding Zone 1	MOIRA2.2 Demand Matrix
Category 6	Origin or Destination a London Travelcard Boundary Zone	MOIRA2.2 Demand Matrix
Category 7	Non-National Rail Stations	MOIRA2.2 Demand Matrix

A.3 In the descriptions below any reference to the methodology used prior to 2011/12 is drawn from documentation produced by Resonate when they were the ORR's consultants producing these statistics. From 2011/12 onwards a number of changes have been made in the methodology in order to better represent the distribution of demand between Group Stations (Category 2) by using passenger count data as described in chapter 3 of this report.

Category 1 – Origin and Destination Stations Known

A.4 Both the origin and destination were known stations so no further processing is required for such flows.

Category 2a – Origin or Destination a Group with all Stations Having a Ticket Office

A.5 In 2005/06 all origins or destinations that were a group station (with the exception of London BR) were changed to the major station within the group. For example, all ticket sales to or from Reading BR were recoded to Reading.

A.6 In 2006/07 the ODM was based on the journeys from ticket sales to the individual stations within a group. We assumed that passengers travelling to the stations in a group would act in the same way as passengers travelling from the stations in that group. It was believed that this was, in general, a valid assumption to make, and no bias would be introduced into the journey figures.

A.7 From 2007/08 onwards this process is still used where all stations in the group have ticket offices, so that the relative flows from the individual stations are credible.

A.8 For example, in 2006/07 the journeys between stations in the 'Manchester BR' group and Crewe and vice-versa are shown by the column "jnys" in the table below. First the proportion of journeys from each of the individual Manchester stations to Crewe is determined, as shown in column "%split."

A.9 Then these proportions are applied to both the 'Manchester BR to Crewe' and 'Crewe to Manchester BR' flows, giving the breakdowns to individual stations shown in column 'BR portion'. These are added to the base values to give "Total Journeys", before the 'Manchester BR to Crewe' and 'Crewe to Manchester BR' flows are deleted, to avoid double counting. The slight discrepancy between the 'Grand Totals' is due to rounding error.

Table A.2: Example of breaking down journeys to/from a BR group of stations

Orig	Dest	Origin Name	Destination Name	Jnys	%Split	BR portion	Total Jnys
2963	1243	DEANSGATE	CREWE	83	0.32%	85	168
2966	1243	MANCH OXF RD	CREWE	5,464	21.03%	5,580	11,044
2968	1243	MANCH PICC	CREWE	19,733	75.95%	20,152	39,885
2970	1243	MANCH VICT	CREWE	700	2.69%	714	1,414
0438	1243	MANCH BR	CREWE	26,533		Remove	
1243	2963	CREWE	DEANSGATE	207		1,478	1,685
1243	2966	CREWE	MANCH OXF RD	2,262		97,287	99,549
1243	2968	CREWE	MANCH PICC	8,017		351,349	359,366
1243	2970	CREWE	MANCH VICT	343		12,464	12,807
1243	0438	CREWE	MANCH BR	462,578		Remove	
		Grand Total:	525,920			525,918	

A.10 The above methodology has been applied to all flows with more than 1,000 journeys in total, based on sales data, leaving the individual group stations (i.e. not including the 'BR Group NLC to destination' flow). For the smaller flows an average split is applied based on the flow with more than 1,000 journeys.

A.11 Since 2011/12 a number of station passenger counts have been undertaken at individual stations within some of the BR station groups in order to support a revision to how the total demand is split between the individual stations. Since 2012/13 progressively more station groups have a count based methodology for apportioning total demand amongst its member stations. In the 2015/16 Station Usage dataset the following Group Stations use passenger counts to calculate the split between individual stations:

- Bedford BR (Bedford Midland, Bedford St. Johns);
- Canterbury BR (Canterbury East, Canterbury West); ;
- Colchester BR (Colchester, Colchester Town);
- Dorchester BR (Dorchester South, Dorchester West)
- Dorking BR (Deepdene, Dorking, Dorking West);
- Edenbridge BR (Edenbridge, Edenbridge Town);
- Falkirk BR (Falkirk Grahamston, Falkirk High);
- Farnborough BR (Farnborough Main, Farnborough North);
- Helensburgh BR (Helensburgh Central, Helensburgh Upper);
- Hertford BR (Hertford East, Hertford North);
- Maidstone BR (Maidstone Barracks, Maidstone East, Maidstone West);
- Newark BR (Newark Castle, Newark North Gate);
- Portsmouth BR (Portsmouth Harbour, Portsmouth & Southsea);
- Southend BR (Southend Central, Southend East, Southend Victoria);
- Wakefield BR (Wakefield Kirkgate, Wakefield Westgate); and
- Worcester BR (Worcester Foregate Street, Worcester Shrub Hill).
- Bicester BR (Bicester Village, Bicester North)
- Birmingham BR (Birmingham New Street, Birmingham Moor Street, Birmingham Snow Hill)

Category 2b – Origin or Destination a Group with some Stations Having no Ticket Office

A.12 For this class of stations the above process breaks down because the proportion of journeys **to** the group stations with no ticket offices will tend to be estimated as zero because the sales **from** those stations are necessarily zero. For these groups bespoke methodology has tended to be used based on the best available data. This year entries and exits for the majority of stations in this group have been obtained by apportioning total station group entries and exits using count data.

A.13 For the remaining stations splits between stations have been fixed at an origin and destination and route code level at the proportions estimated in the 2010/11 dataset.

Category 3 – Origin or Destination is London BR

A.14 This category contained all flows that had London BR as either the origin or destination. In order to assign an appropriate London station on flows where either the origin or destination is London BR (NLC=1072) or a London Travelcard involving Zone 1, we analysed responses

from the 2001 London Area Travel Survey (LATS). For journeys from any given station, we established the percentage of passengers using each London terminus.

A.15 For example, if the flow was from Ashford International to London BR, we used our pre-generated table showing the percentage spilt between the alternative London termini for passengers starting at Ashford International. From this we apportioned the exits between London Bridge, Charing Cross, Victoria and other London termini.

A.16 Stations with small sample sizes were removed from the 2001 LATS data. Where there was insufficient data in the 2001 LATS to generate the split for a particular station, a similar process with the Non London Groups methodology was applied. Firstly for all the flows with more than 1000 journeys leaving London BR and having as a destination the particular station we used split factors as above. However, if the sum of journeys was less than 1000 we assigned to the flow the top origin from the London BR stations.

Category 4 – Origin or Destination a London Travelcard including Zone 1

A.17 All origins and destinations that were London Travelcard Zones that include Zone 1 were converted to 'London BR' under the assumption that they will travel to the same stations as point-to-point passengers and then transfer to another mode. The methodology set out above for Category 3 was then applied.

Category 5 – Origin or Destination a London Travelcard excluding Zone 1

A.18 This category contained all Travelcards that did not include Zone 1, for example Zone R2345 London.

A.19 For flows with origin or destination a London Travelcard (excluding zone 1) we use a set of assumptions based on survey responses from the 2001 LATS. They use the starting station to work out which stations it is possible for the passenger to be travelling to, and also give the proportion of passengers travelling to each of these stations. This is based on the assumption that a passenger holding a Zones 2-6 Travelcard would travel as far as Zone 2.

A.20 This processing is undertaken during the production of the MOIRA2.2 demand matrix.

Category 6 – Origin or Destination a Boundary Zone

A.21 All origins and destinations that were a London Travelcard Boundary Zone were converted to 'London Travelcard including Zone 1' under the assumption that a passenger travelling from or to a Boundary Zone will hold a Travelcard that includes Zone 1. The methodology set out above for Category 3 was then applied.

A.22 This processing is undertaken during the production of the MOIRA2.2 demand matrix.

Category 7 – Non-National Rail Stations

A.23 This final category contains all those flows in the original ticket sales data that do not fall into one of the above categories. Refer to chapter 5 for a detailed description of this data and what has been included and excluded from the ODM.

A.24 This processing is undertaken during the production of the MOIRA2.2 demand matrix.

B Interchanges - Overview of the ORCATS allocation process

B.1 This appendix gives an outline of the Central Allocations File (CAF), which is used in producing the interchange figures, and the ORCATS process which is used to create the CAF.

B.2 Most of the train tickets that are sold are inter-available – the customer has a choice of routes and operators. For example, when a customer buys a ticket to travel from Leicester to Leeds, that customer may travel on various combinations of East Midlands Trains, East Coast, CrossCountry Trains and Northern, and may interchange at Doncaster, Sheffield, Derby or Nottingham. LENNON captures the sale of the ticket, but unless the ticket has stringent route restrictions, the route actually taken by the customer is not recorded.

B.3 The route taken by any particular customer may never be known, but some route options are more attractive than others. The customer is more likely to choose a faster, more frequent service than a slower, less frequent one. This likelihood can be translated into the proportions of customers choosing each route option, on a particular flow. (A 'flow' represents all journeys from a given origin station to a given destination station, irrespective of the route taken.) The revenue received from all customers on that flow should be split between different operators to reflect the proportion of customers which each operator carried.

B.4 ORCATS was developed to model the choice made by the customers, and to allow revenue to be split between operators. It applies passenger choice modelling to the train timetable, to determine the relative attractiveness of different route alternatives. It then weights the results by journey mileage.

B.5 For any given timetable, ORCATS works out the possible routes between each origin and destination and calculates the percentage of the passengers that are expected to choose each route based on the services in that timetable.

B.6 The output from ORCATS is the CAF. This lists the proportion of journeys on each flow (or origin-destination pair) estimated to be made by each route alternative. For journeys involving interchanges, each leg of the journey is listed. By combining this information with the ODM data, which contains journeys for all flows, the number of interchanges occurring at individual stations has been estimated.

steergroup.com

steer