



# Estimates of Station Usage: Quality and Methodology Report

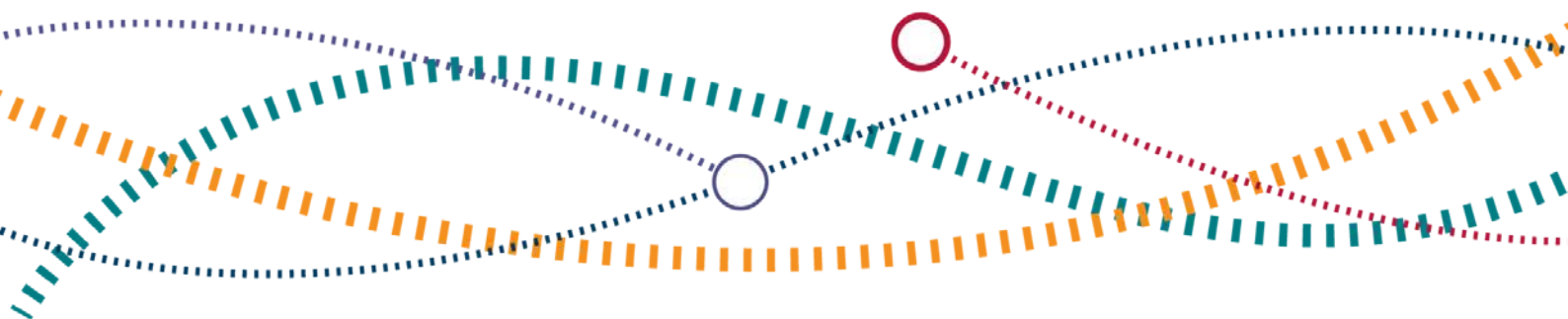
01 December 2020

**Contact information:**

Responsible statistician: Jay Symonds

Public enquiries: [rail.stats@orr.gov.uk](mailto:rail.stats@orr.gov.uk)

Media Enquiries: Tel: 020 7282 2094



# Contents

---

<b>Introduction</b>	<b>2</b>
<b>Data sources, methodology and definitions</b>	<b>3</b>
<b>Historical background</b>	<b>10</b>
<b>National Statistics accreditation</b>	<b>11</b>
<b>Relevance to users</b>	<b>12</b>
<b>Accuracy and reliability</b>	<b>15</b>
<b>Timeliness and punctuality</b>	<b>19</b>
<b>Accessibility and clarity</b>	<b>20</b>
<b>Coherence and comparability</b>	<b>21</b>
<b>Annex 1 - Details of methodology changes over time</b>	<b>23</b>

# Introduction

This is a report on the quality and methodology used to produce the annual Estimates of station usage. It helps users to understand the quality of our statistics, and also ensures ORR is compliant with the three quality principles in [the Code of Practice for Official Statistics](#) - Q1: Suitable data sources, Q2: Sound methods, and Q3: Assured quality. This report also provides information on the methodology and data sources used to produce the statistics.

This report covers the following areas:

- Data sources, methodology and definitions – detail on the various data sources, methodology used to compile the statistics and definitions;
- Historic background – background to the time series (summary of methodological improvements contained in annex 1);
- National Statistics accreditation – the recent assessment of these statistics by the Office for Statistics Regulation;
- Relevance to users – the users of the statistics, and our engagement;
- Accuracy and reliability – the accuracy, data coverage and quality assurance of the statistics;
- Timeliness and punctuality – our timescales for the production and publication of the statistics;
- Accessibility and clarity – the format of our statistics and where they can be found;
- Coherence and comparability – similar statistics published elsewhere and the degree in which the statistics can be compared over time;
- Annex 1 – Details of methodology changes over time (2006-07 to 2019-20).

**ORR commission [Steer](#) to produce the Estimates of station usage dataset covering the latest financial year (1 April to 31 March). More detailed information on methodology and changes each year are available in two reports produced by Steer: [Estimates of Station Usage 2019/20: Methodology Report](#) and [Station Usage & Origin Destination Matrix 2019/20: Historical Methodological Changes](#).**

# Data sources, methodology and definitions

The Estimates of station usage dataset consists of estimates of the total numbers of people:

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

The estimates of entries & exits are further split by ticket type (full price, reduced price and season tickets). There is also a range of station attribute information included too, e.g. geographic data.

Time series of entries/exits and interchanges by station are available back to 1997-98.

## Data sources

Estimates of station usage are primarily based on sales data from LENNON, the rail industry's ticketing and revenue system. This is supplemented with some local ticketing data.

Below is a list of all the data sources used to create the Estimates of station usage:

### Entries/exits:

- LENNON, Transport for London (TfL) data and train operator data (Gatwick Express and Stansted Express) as an input to the MOIRA2.2 base matrix
- Local ticketing data from Passenger Transport Executives (PTEs)
- Manual station counts
- Heathrow Express ticketing data
- Additional LENNON data

### Interchanges:

- Central Allocations File (CAF)

### Station attribute data:

- Geographic data – [Office for National Statistics geoportal](#)
- National Rail Enquires

## Methodology

These statistics on usage are **estimates** based primarily on tickets sales using the methodology described below. This methodology is the best approach possible given Britain does not have a fully gated rail network or comprehensive and robust count data at every station. However, this data does have weaknesses when utilised for this purpose and, although some of these are catered for in the methodology and we continue 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 detailed in the ‘*Accuracy and reliability*’ section of this report.

Estimates of station usage are primarily based on data originating from LENNON, the rail industry’s ticketing and revenue system. LENNON contains the majority of National Rail tickets purchased in Great Britain. However, it excludes some tickets sales e.g. London Travelcards. These ticket sales, together with LENNON data are used to derive a matrix of journeys and revenue which is an input to the MOIRA2.2 rail planning tool. Steer take this MOIRA2.2 base matrix and supplement it with updated local ticketing data for Passenger Transport Executive (PTE) areas. Various adjustments are made to the data to deal with a range of issues (see below) to create a comprehensive matrix of passenger flows throughout Great Britain, the Origin Destination Matrix (ODM). The ODM is used to derive the number of entries and exits at each station in Great Britain.

A new column, “Data source/adjustments” has been added to the Estimates of station usage dataset for 2019-20 to summarise the basis of each station’s estimate, i.e. if the data is supplemented by PTE data or an adjustment has been made to improve the estimate. These adjustments are described below with the abbreviation used in the dataset shown in brackets:

### Infills and adjustments:

- PTE infills (“**PTE**”) – local ticketing data provided by PTEs and prepared by Steer (for West Midlands) and Mott MacDonald (Greater Manchester, Merseyside, South Yorkshire, Tyne and Wear, West Yorkshire, Strathclyde). These infills are subject to annual improvements, which normally represent a simple update, but some years contain a step change in the methodology;
- Ranger/Rover infills (“**Other**”) – Rover and Ranger products, e.g. Ride Cornwall, Cambrian Coaster, Anglia Plus, are tickets with non-geographic destinations so

estimates are developed to represent passenger flows on a selected number of these tickets;

- London adjustment (“**London**”) – Allocation of demand associated with tickets sold to ‘London BR’ between the London terminals. This is for tickets that have an origin and destination on the ticket of ‘London Terminals’, which is then reallocated between the component stations based on known sales between specified origins/destinations from LENNON;
- Allocation of demand between individual stations within station groups (“**Group**”) – Many products are sold with the origin or destination as a group of stations, e.g. London BR, Birmingham BR. Current industry data does not distinguish between the component stations and therefore a split between these stations must be estimated. For example, where tickets are sold to/from ‘Birmingham BR’ it is necessary to estimate how these journeys are distributed between New Street, Snow Hill and Moor Street stations. For London BR this is done using the 2001 London Area Travel Survey (LATS). For other group stations, this is done by apportioning to individual stations based on known ticket sales between specified origins/destinations from LENNON. A consequence of this adjustment is that all stations in a group will have the same annual percentage change, unless the station split assumptions change in a given year;
- Manual counts used to allocate demand between individual stations within 21 station groups outside central London (“**Counts**”) – Steer undertake a programme of manual counts on an annual basis at selected group stations to calculate the splits where the above ‘Group’ adjustment isn’t representative. 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. A consequence of this adjustment is that all stations in a group will have the same annual percentage change, unless an updated count adjustment is applied in a given year. Since 2012-13 progressively more station groups have a count based methodology for apportioning total demand amongst its member stations;
- Season ticket adjustment (“**Season**”) – There are a number of cases where adjustments are made to selected stations to account for specific known issues. For example, 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 using the origin and destination on the ticket can be under- or over-estimated and journeys involving those stations needs to be adjusted to reflect actual usage;

- The 'Digby & Sowton' adjustment ("**D&S**") – 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. Adjustment uses data provided by Great Western Railway;
- Heathrow adjustment ("**HEX**") – not all ticket sales to/from Heathrow stations are included in LENNON, i.e. some Heathrow Express tickets. Therefore an adjustment is made to add in the non-Lennon ticketing data supplied by Heathrow Express and apportion this between the three Heathrow stations using the splits from Lennon's Heathrow Express journeys.

In 2019-20, we also needed to make the following adjustments to data originating from LENNON:

- Amendment of data to reflect refunds of season tickets. Two adjustments were made:
  - Firstly, for refund requests that were received and processed in the final weeks of March 2020 (i.e. at the start of the national lockdown), all the refunded journeys and revenue would appear in the 2019-20 financial year when in 'reality' those refunded journeys should be distributed over the season ticket's remaining validity. Therefore, an adjustment was required to remove the proportion of refunded journeys that shouldn't have applied to 2019-20.
  - The second adjustment related to the refund requests that were processed after March 2020. These refunded journeys would all be allocated to the 2020-21 financial year, whereas it is likely that some of the journeys should be backdated to mid-March 2020 when the national lockdown commenced. This meant that a small proportion of these refunded journeys should be added to the 2019-20 estimates (i.e. the 2019-20 usage figures reduced) and Steer undertook this adjustment by using refunds data from LENNON for the first 5 periods of 2020-21 (i.e. to 22<sup>nd</sup> August 2020) to calculate the adjustment required for each flow.
- Amendment of journeys associated with the TfW Multiflex product to address known issue with overstatement of journeys associated with the ticket in LENNON.

Further details relating to the infills and adjustments outlined above can be found in Chapter 3 of [Steer's methodology report](#).

## Methodology changes

The methodology to produce the Estimates of station usage is reviewed annually and enhancements are implemented to address known issues. Often these enhancements utilise new sources of data that were not previously available and improve the estimates.

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. ORR has worked with Steer 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).

A number of improvements to the methodology have been implemented over recent years (see Annex 1). These changes should be taken into account when considering year-on-year changes in journeys for some stations as it may not reflect an actual change in demand. A new column, "Quality limitations", has been added to the Estimates of station usage dataset to highlight if data for 2019-20 are not comparable to the previous year, and/or any other limitations or quality considerations.

In the 2019-20 dataset the following methodological improvements have been implemented:

- Updated season ticket journey allocation adjustments;
- Updated allocation of journeys between selected Group stations based on Spring 2020 passenger count surveys. These updates were implemented for: Dorchester BR, Edenbridge BR, Warrington BR and Worcester BR. In addition, updated splits for the Manchester BR group stations were provided by the PTE, Transport for Greater Manchester, and implemented;
- Inclusion of off-network tickets for Merseyside PTE and inclusion of concession (senior and disabled) tickets for South Yorkshire PTE;
- Inclusion of the three Heathrow stations and subsequent improvement to the London Paddington estimates and some other local stations.



## Interchanges

An estimate of the number of people interchanging at each station is included in the Estimates of station usage 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).

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 2019 CAF is used in the creation of the 2019-20 Estimates of station usage dataset.

An overview of the ORCATS allocation process can be found in in Appendix B of [Steer's methodology report](#).

## The impact of COVID-19 on the 2019-20 estimates

To understand the likely impact of COVID-19 at a station level we produced an 'alternative' entries and exits dataset for the period March 2019 to February 2020. This dataset was produced on a consistent basis with the main dataset (April 2019 to March 2020). The estimated number of entries and exits for each station for the year March 2019 to February 2020 can be found in [Table 1410](#) (column M).

This alternative dataset has been produced to be as close to the main (financial year) dataset as possible, whilst largely being unaffected by the impact of COVID-19. Looking at the differences between the main dataset and the alternative dataset provides some insight into the impact of COVID-19 at a station level. However, it should be noted that differences between these datasets may be explained by other factors impacting usage in March 2019 and compared with the excluded March 2020, e.g. (non-COVID related) station closures.

More information can be found in Chapter 3 of [Steer's methodology report](#).

## Definitions

- LENNON – 'Latest Earnings Networked Nationally Over Night' is the rail industry's ticketing and revenue system. It contains information on the majority

of national rail tickets purchased in Great Britain. However, it excludes some tickets sales.

- MOIRA2.2 base matrix – produced by Resonate as an input into the MOIRA2.2 rail panning tool, it provides an estimate of journeys on the Great Britain rail network for the duration of a financial year. It includes all journeys associated with point to point flows and includes overlays (“infills”) to reflect travel using tickets not included in LENNON (e.g. London Travelcards and some specific tickets to/from airports and multi-modal and zonal products sponsored by PTEs).
- Origin Destination Matrix (ODM) - a comprehensive matrix of passenger flows throughout Great Britain.
- Passenger Transport Executive (PTE) - 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 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.
- Ticket Types:
  - 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). All advance-purchase tickets are also included in the category;
  - Seasons: all multi-use tickets.

For more detailed information on the methodology see [Steer's Estimates of Station Usage 2019/20: Methodology Report](#).

# Historical background

## **1997-98 to 2002-03:**

Estimates of station usage were calculated from CAPRI (Computer Analysis of Passenger Revenue Information) which was the rail industry's former central ticketing system.

## **2003-04:**

No estimates produced.

## **2004-05 to 2019-20:**

From 2004-05 onwards, LENNON (Latest Earnings Networked Nationally Over Night), which is currently the rail industry's central ticketing and revenue system, has been the basis for calculating these statistics.

Several improvements to the methodology have been implemented over the years. A summary of methodology improvements between 2006-07 to 2019-20 can be found in Annex 1. Also see [Steer's historical methodological changes report](#).

Steer has been contracted by ORR to produce these statistics since 2011-12. Between 2005-06 and 2010-11 DeltaRail (now known as Resonate) were the contractor.

## **2019-20:**

National Statistics accreditation.

# National Statistics accreditation

During April to November 2020 Estimates of station usage were assessed by the [Office for Statistics Regulation](#) (OSR) against the [Code of Practice for Statistics](#) for National Statistics (NS) designation. National Statistics status means that official statistics meet the highest standards of trustworthiness, quality and public value.

OSR identified several ways in which we should develop the statistics in order to achieve National Statistics status. These are described in the [assessment report](#) published on 3 November. All the requirements were addressed for the 2019-20 publication and OSR confirmed in a [letter on 1 December that these statistics have been designed at National Statistics](#).

In their report OSR said: *Estimates of station usage provide a valuable, trusted source of information about the use of the rail network across Great Britain. They are used by a broad range of people, who appreciate the granularity of the data and the ability to use these statistics for further analysis.*

As part of their assessment, OSR sought and received feedback from a wide range of users of our statistics<sup>1</sup>. This feedback informed their requirements and has been considered by us in developing an [improvement plan for estimates of station usage statistics](#). All the actions in the plan have been successfully implemented for the 2019-20 publication. For example, replacing the factsheet with a more informative statistical release; a new quality and methodology report; inclusion of two new columns of information in the data table, the first one highlighting any data quality issues or methodology changes, the second listing the data sources/adjustments used to estimate usage at each station.

---

<sup>1</sup> Feedback was received from the following users: DfT, Welsh Government, RDG, Network Rail, RSSB, Transport Focus, London TravelWatch, Transport for Wales, TfL, Transport For West Midlands, Cambridgeshire County Council, BBC, All the Stations, Community Rail Network, TSUG, academics.

# Relevance to users

The degree to which the statistical product meets the user needs in both coverage and content.

## Uses and Users

Estimates of station usage are the only source of published information on the use of all stations in Great Britain. They are used by a broad range of people, e.g. by government departments, local transport authorities and community groups, and transport watchdogs. The statistics are also of interest to journalists, academic researchers and members of the public.

One of the main strengths of Estimates of station usage is that they provide a data series going back to 1997/98. This means that users can explore trends over time and combine the data with their own local knowledge to understand the impact of infrastructure projects or changes to the usage of the rail network.

User engagement during the assessment by the Office for Statistics Regulation identified the following uses of Estimates of station usage:

- Monitoring station use to understand capacity and identify potential issues
- Informing business cases for station or service developments
- Planning, monitoring and evaluating infrastructure projects
- Used in the production of other official statistics, such as those from the National Rail Passenger Survey
- National and local media outputs that inform public interest
- To facilitate academic research
- Personal interest

## How these statistics can and cannot be used?



- Monitoring the number of annual entries and exits or interchanges at individual stations e.g. to understand demand
- Monitoring how usage at individual stations changes over time (subject to methodology changes) and insights as to why
- Comparing the relative usage of stations within local areas, regions or across the whole of Great Britain
- To gauge the use of different ticket types e.g. season vs reduced



- Monitoring passenger rail usage at a national level, by train operating company or by ticket type (refer to [Passenger rail usage statistics](#))
- Monitoring the number of passenger journeys between and within regions (refer to [Regional rail usage statistics](#))
- Exploring rail journey flows between origin and destination stations
- Volume of entries compared to exits at an individual station (methodology makes these equal)

When using Estimates of station usage data, 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, e.g. infrastructure improvements; temporary line or station closures.

## User satisfaction

A number of changes and improvements were made to the 2019-20 statistics following feedback from OSR and users. These are summarised in *the 'National Statistics accreditation'* section above.

ORR's last [user survey](#) took place from mid-January to mid-April 2020. The aim of the survey was to gather feedback on ORR's new data portal; this includes statistical releases, data tables and other supplementary material. There were 42 responses to the survey. ORR created an [implementation plan](#) following the 2020 user survey.

More detailed information on users of ORR statistics and meeting the needs of users is available on our [user engagement webpage](#).

# Accuracy and reliability

The proximity between an estimate and the unknown true value.

These statistics on usage are **estimates** based primarily on tickets sales using the methodology described above. This methodology is the best approach possible given Britain does not have a fully gated rail network or comprehensive and robust count data at every station. The methodology's national coverage makes it suitable as a basis for the production of these National Statistics accredited statistics.

## Data coverage

The data presented in this release are for all open mainline stations in Great Britain, i.e. those with a timetabled train service.

## Limitations

A number of improvements to the methodology have been implemented in recent years. These changes should be taken into account when considering changes in usage between years, as it may be a result of improved methodology, rather than reflecting an actual change in demand at stations. These improvements, and the reasons for them are documented in [Steer's historical methodological changes report](#).

As these statistics are primarily based on ticket sales, there are a number of limitations that users should be aware of which are summarised below. There is more detail on these in Chapter 5 of [Steer's methodology report](#).

**Stations with known under estimates or no estimates** - Eurostar ticket sales are not the rail industry's ticketing system (LENNON). Therefore our estimated usage at St. Pancras, Ashford International and Ebbsfleet stations will not be a true reflection of the total usage at these stations. There are two other stations which we do not have estimates for as tickets sales are not recorded in LENNON: Corfe Castle and Manchester United Football Ground.

**Concessionary travel** - 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 these estimates 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. Currently concessionary travel data are included for London, Greater Manchester, Merseyside, South Yorkshire, and West Midlands areas. No information is available for Strathclyde, West Yorkshire or Tyne and Wear, therefore there will be some under estimates of usage at some stations in these areas.

**Ticketless travel** - As the estimates of station usage are based on ticket sales, journeys associated with ticketless travel are not included in the data. This is more likely to be an issue on some flows and where ticketless travel is significant. As more stations have been gated over time and train operating companies (TOCs) focus on revenue protection activities, this is likely to be less of an issue than in the past. It can be argued that it is not appropriate to include ticketless travel in the dataset as its purpose is to record genuine journeys on the rail network. The inclusion of ticketless travel could distort business cases for new investment where these are reliant on the estimates of station usage data. It is worth noting that ticketless travel also includes an element of individuals who are legitimately travelling for free, such as the British Transport Police or some rail industry employees.

**Group stations** - Many products to major destinations are sold with the origin or destination as a group of stations (e.g. London BR Terminals, Manchester BR). Current industry data does not distinguish between the component stations and therefore a split between these stations has to be estimated. These estimates are currently apportioned to individual stations based on a combination of known sales between specified origins/destinations, survey and/or count data.

**Season ticket fare zones** - 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, giving the passenger additional flexibility for no increase in fare. This means that the ticket sales data shows that there are more people travelling to/from this station than is the case. Therefore for stations where this is a known issue the data is adjusted to reflect actual usage at each station, for example, stations along the Southend Victoria and Central branches.

There may be similar issues at stations where we don't make adjustments to the data. One recent example is for some end of the line stations in Wales which their usage estimates are likely to be an over estimate, e.g. Barry Island, Radyr, Pengam.

**Split ticketing / multiple tickets** – The complexity of the fares system has led to opportunities where a combination of shorter-distance tickets can be combined to provide a cheaper fare than the advertised price for the end to end journey ('split ticketing'). The split tickets are sold as the individual components, and therefore will give higher usage

(over estimates) at the intermediate stations where the tickets are 'split'. Also, 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.

**Season ticket journey factors** – Ticket sales in LENNON are converted into an estimate of the number of journeys made by applying a series of ticket type journey factors. Therefore station usage estimates are based on an assumed number of journeys made based on the ticket type sold. The journeys factors used for the main season tickets are as follows:

<b>Season ticket validity</b>	<b>Journey factor</b>
Weekly	10.3
Monthly	45
3 monthly	135
6 monthly	270
Annual	480

## **Quality assurance**

ORR has a quality assurance framework for Estimates of station usage. This sets out the steps and methods used to quality assure these statistics throughout the production process by both Steer and ORR. Also included is a risk log detailing any high risk points around data quality during the process and how these are mitigated.

A summary of the quality assurance framework is below:

Steer produce draft station usage estimates following a prescribed set of processing steps. Quality assurance checks are undertaken at each processing step including various comparisons of the draft estimates with other data sources for validation purposes. These comparisons are shared with ORR and subsequently published in [Steer's methodology report](#). Once a draft dataset is ready Steer and ORR meet to discuss the draft estimates and any emerging issues are investigated and resolved before the dataset is formally passed to the lead statistician in ORR.

For the 2019-20 statistics, ORR shared the draft estimates with the stakeholders who had provided supplementary ticketing and usage data, e.g. the PTEs. These recipients were asked to review the estimates, provide reasons for large changes in usage at individual stations or to flag any estimates where they had quality concerns. ORR also shared draft estimates with other key stakeholders, e.g. Transport for London, Transport for Wales, for their review. Following the external review outlined above, estimates are reviewed internally by analysts and experts within ORR. Colleagues are asked to use their own knowledge or investigate reasons for large changes in usage and to flag any estimates where they had quality concerns. Any estimates where the internal or external reviews raised quality concerns were shared with Steer and underwent further investigation.

The final data are then prepared for publication by the lead statistician. The process includes quality assuring the tables and charts produced and providing supporting commentary regarding the key trends. These are subject to peer review by another analyst following a well documented process which is followed for all ORR statistical releases. The final stage of the quality assurance process is a sign off by the Head of Profession for Statistics confirming the statistical release and associated outputs, e.g. data tables, meet quality standards and are fit for publication.

## **Revisions policy**

ORR's statement on [orderly release and revisions policy](#) outlines ORR's revision policy. Details of any revisions are available in the [revisions log](#). Further information on revisions and data series breaks can also be found in the data tables.

### **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 were updated to account for this issue and were published in June 2020. An adjustment was also made for this in the 2019-20 estimates (with the error corrected in the base data from May 2019 onwards).

# Timeliness and punctuality

Timeliness refers to the time gap between publication and the reference period.  
Punctuality refers to the gap between planned and actual publication dates.

ORR aims to publish the Estimates of station usage as soon as possible after the end of the financial year, which is currently around nine months after in December. It is important to take time to implement the detailed methodology, collect additional data (e.g. from PTEs) and carry out a series of quality assurance checks to ensure the final data are as accurate as possible.

ORR will continue to work with our consultants and stakeholders to shorten the time between the reference period and publication.

The [publication schedule](#) available on the data portal outlines the publication dates for National Statistics quarterly and annual statistical releases and other official statistics up to 12 months in advance.

ORR is committed to releasing its statistics in an open and transparent manner that promotes confidence.

# Accessibility and clarity

Accessibility is the ease with which users are able to access the data, also reflecting the format in which the data are available and the availability of supporting information. Clarity refers to the quality and sufficiency of the metadata, illustrations and accompanying advice.

All rail statistics data tables can be accessed free of charge on the [ORR Data Portal](#). Commentary about the statistics and trends are provided in the statistical releases.

The tables and other outputs currently published on the [Estimates of station usage theme page](#) of the Data Portal:

- Passenger entries and exits and interchanges by station (2019-20) – *Table 1410* (ods and csv)
- Time series of passenger entries and exits and interchanges by station (1997-98 to 2019-20) – *Table 1415* (ods)
- Interactive dashboard (Power BI)
- Animated charts (MP4) and infographics (PDF)
- Frequently Asked Questions (FAQs) document (PDF)
- Quality and methodology report (PDF)
- [Steer's methodology report](#) (PDF)
- [Steer's historical methodological changes report](#) (PDF)

# Coherence and comparability

Coherence is the degree to which data that are derived from different sources or methods, but refer to the same topic, are similar. Comparability is the degree to which data can be compared over time and domain.

## Related data

### Passenger rail usage (ORR):

[Quarterly statistics reporting the volume of passenger journeys, kilometres and revenue on the mainline network in Great Britain](#). Statistics are presented by ticket type, sector, and train operating company. Long-running time series on passenger journeys (Table 1220) and passenger kilometres (Table 1230) are updated annually.

### Regional Rail Usage (ORR):

[Annual statistics providing passenger journeys data for each region](#) of Great Britain, including journeys between regions, within regions and between England, Scotland and Wales. These estimates are also produced by Steer based on the ODM.

### Station footfall (Network Rail):

Quarterly data on [station concourse footfall](#) at 18 of Network Rail's managed stations only. The footfall data captures the numbers of people entering and exiting these stations.

### Passenger numbers and crowding (Department for Transport):

[Rail passenger numbers and crowding statistics](#) provides information on the number of passengers travelling by rail into and out of major city centres in England and Wales. The statistics represent passengers on National Rail services on a 'typical' weekday.

For more information on **COVID-19 impacts** see:

- [Transport use during the COVID-19 pandemic](#) (Department for Transport)
- [All Change? Travel tracker](#) (Department for Transport)
- [Coronavirus and the social impacts on Great Britain](#) (Office for National Statistics)
- [Weekly travel during COVID-19 survey](#) (Transport Focus)

- [Public transport journeys by type of transport](#) (Transport for London)

## Comparability

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. See *Methodology* section above and Annex 1 for more information.

## Length of comparable time series

Measures	Start of time series	Any break in time series
Estimates of entries, exits and interchanges	1997-98	2003-04 (no data)  Methodology changes each year affect some stations

# Annex 1 - Details of methodology changes over time

## Summary of methodology improvements 2006-07 to 2019-20

### 2006-07:

Additional estimates for rail travel using TfL sold travelcards and airport links were included.

### 2008-09:

The generation of the ODM was integrated with the demand matrix in MOIRA, a software tool used by the industry to model the impact of timetable changes on the rail market. In addition to having LENNON data, MOIRA also provided more robust estimates of rail travel on TfL sold travelcards and airport links. It also included estimates of rail travel in PTE areas which had previously been excluded from the ODM due to a lack of data.

### 2009-10 / 2010-11:

From January 2010, rail travel using Oyster pay-as-you-go (PAYG) was included in LENNON so these data were included in the ODM from 2009-10 with the first full year of data being 2010-11.

### 2011-12:

- Improved estimates of travel in the West Midlands (Centro) PTE area included.
- Estimates of rail travel made using a small number of Rover and Ranger products included. The tickets included were: St Ives Day Rangers, Valleys Night Rider, and Cambrian Coaster Ranger. Whilst volumes of travel on these products are relatively small, in the specific area of use they can be significant.

### 2012-13:

- An improved PTE infill was included for two more PTEs – West Yorkshire (WYPTE) and Greater Manchester (GMPTE/TfGM).
- Estimates of rail travel using TfL's concessionary product, the 'Freedom Pass', were included for the first time.
- A further five Rover and Ranger products were included: Anglia Plus, Devon Day Ranger, Devon Evening Ranger, Ride Cornwall, and Freedom Travel Pass (West of England product).

### 2013-14:



A number of changes were made to improve the representation of journeys on PTE-sponsored tickets in South Yorkshire, Merseyside and Strathclyde.

**2014-15:**

- An improved infill for the Tyne & Wear PTE area was included.
- An adjustment process was made to account for the change in LENNON treatment of PAYG journeys to make the statistics more consistent with previous years. This adjustment was a one off as in 2015-16 it was included in the MOIRA base matrix.
- An adjustment was required due to changes in journey patterns as a result of the London Bridge works. Data from Transport for London's (TfL's) Oyster Clicks Model (OCM) was used to estimate the number of journeys 'to London Bridge' and the number of journeys 'to London Terminals'.
- Journeys using a season ticket product for students have been redistributed to Exeter Central and Exeter St. David's from Digby & Sowton to better reflect actual journey destinations.

**2015-16:**

- London (In-boundary) Travelcard Methodology - Oyster Clicks Model (OCM) data used to allocate journeys made wholly within the London Travelcard Area to individual London stations rather than based on a survey from 2001.
- London Terminals Demand Allocation - improved due to MOIRA base matrix now disaggregated by individual London Terminal where possible, such as where a ticket is bought to a specific terminal rather than to the generic 'London Terminals.'
- St. Ives Branch Line Counts - Passenger counts were carried out at all five stations on the St. Ives Bay line (St. Erth to St. Ives) in August 2016 and the results of these counts were used to produce a more accurate allocation of entries and exits from sales of ranger or rover tickets across the stations.
- Season Ticket Journey Adjustment (Southend) – An adjustment to the allocation of usage at stations around Southend was made to account for season tickets issued for travel to/from Southend Victoria which were actually being used to travel from alternative stations on the branch, as the price of a season ticket is the same.

**2016-17:**

- London BR Allocation Update - Reallocation of some journeys for Kensington Olympia due to previous over estimates.

- Season Ticket Journey Adjustments (expanded number stations) - In the production of the 2014-15 and 2015-16 statistics, some 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. For the production of the 2016/17 statistics additional LENNON analysis was conducted and discussions with train operators to identify and include additional stations in the adjustment to better reflect their usage.
- Updated Demand Allocation at Group Stations – In order to validate and improve the allocation of journeys between stations within groups (e.g. Worcester BR), passenger counts were carried out at selected group stations on the network. These counts were carried out in Autumn/Winter 2016 and have informed the allocation of demand at the following station groups: Dorchester BR, Newark BR, Southend BR, Warrington BR, Wigan BR and Worcester BR.

#### **2017-18:**

- Season Ticket Journey Adjustments - Similar to previous years, adjustments were made to account for situations where passengers buy season tickets from a station other than the one they generally travel from. The analysis underpinning this reallocation was updated with 2017-18 LENNON data.
- Updated Demand Allocation at Group Stations – Passenger counts were carried out in Autumn 2017 and have informed the allocation of demand at the following station groups: Bicester BR, Birmingham BR, Farnborough BR, Southend BR, Warrington BR, Wigan BR and Worcester BR.

#### **2018-19:**

- Concessionary travel in Greater Manchester - Concessionary ticketing data were available for Greater Manchester PTE for inclusion in the ODM. This led to a total increase of 3.6m journeys, or 7.2m entries and exits, across Greater Manchester.
- Season Ticket Journey Adjustments - Similar to previous years, adjustments were made to account for situations where passengers buy season tickets from a station other than the one they generally travel from. The analysis underpinning this reallocation was updated with 2018-19 LENNON data.

#### **2019-20:**

- Merseyside PTE – Off network sales (commercial retailers, non-commercial retailers and Merseytravel centres) of Saveway and Trio tickets included. Previously only sales at stations and on trains were included.
- South Yorkshire PTE - Concessionary tickets (senior and disabled) included.

- Season Ticket Journey Adjustments - analysis underpinning this reallocation was updated with 2018-19 LENNON data.
- Updated Demand Allocation at Group Stations – Passenger counts were carried out in Spring 2020 and have informed the allocation of demand at the following station groups: Dorchester BR, Edenbridge BR, Warrington BR and Worcester BR. In addition, updated splits for the Manchester BR group stations were provided by TfGM and implemented.
- Estimates for the three Heathrow stations were included for the first time. This addition also improved the estimates for London Paddington station and other local stations.

See [Steer's historical methodological changes report](#) for more detail.



© Office of Rail & Road 2020

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://nationalarchives.gov.uk/doc/open-government-licence/version/3).

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available on the [ORR data portal](#)

Any enquiries regarding this publication should be sent to us at [orr.gov.uk](https://orr.gov.uk)

