



# Stations count

Accurate station usage figures are essential if we are to make sound investment decisions for the future of the railway. Yet the current system is a statistical mire with widespread under-reporting, reports Jonathan Roberts

**T**he Office of Rail Regulation (ORR) published its latest annual station usage estimates, for 2010-11, on 29 March 2012. But just how reliable are these figures? This article focuses on the reliability of ORR's estimates, with particular focus on city region stations.

The station data has been published since 2002-03 (there was no survey in 2003-04). Data quality has improved taking one year with the next, but it remains variable. ORR stresses that its estimates are derived from ticket sales held in the LENNON accounting system with subsequent transformation of this raw data, rather than from precise counts at the door.

## Variety in city regions

Inter-city and inter-region travel is mostly on point-to-point tickets, which LENNON is good at tracing. Also, guards on rural trains are also expected to get good data with their ticket checking and sales (however, travel surges might test all conductors).

It is the city regions where advances in travel marketing such as travelcards, zonal products and the sheer volume of rail travel generates the greatest risk of error. It is here that ORR is most in catch-up mode.

Adjustments are made for zonal travel in city regions, to create a notional assessment of individual station use. Details are set out in the ORR publications website, in the yearly reports on methodology prepared by DeltaRail, ORR's consultants. It is useful to summarise the main elements before comparing the results.

## Adjusting the original data

An Origin and Destination Matrix is taken from the LENNON sales database. Since 2008-09, this has been integrated with the rail industry's planning model, MOIRA. That estimates journeys and revenues from zonal products sold by city region organisations, eg Passenger Transport Executives/Integrated Transport Authorities (PTEs/ITAs). Previously this was a major information gap. Ticket transactions are converted into journeys by multipliers, eg a return ticket is two journeys, while others are on a ratio linked to monthly travel.

DeltaRail cautions that 'the journeys data has not been cross-checked against other data sources of the actual number of journeys made on the network... 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 15 years ago'. The consultants query whether the factors

are still reliable. However, it is not their responsibility to change things.

Various types of travel are not estimated, including rail staff passes, London freedom pass usage, ticketless and underpaid travel, off-network sales of tickets to/from airport stations, and Heathrow Express and Eurostar. Also excluded are many national rail trips on Travelcards purchased at tube stations, shops and newsagents.

Until the latest ORR publication year, 2010-11, use of London's Oyster Pay-as-you-go (PAYG) was excluded from national rail statistics. From January 2008, PAYG began to be introduced on national rail in the London area, but it was only fully available across the London Zones from January 2010.

In the final quarter of 2009-10, PAYG was counted in ORR station usage statistics. But before then, Oyster had been a very successful product which stimulated much urban rail travel. Yet this wasn't recorded in official data and, because of passengers converting from point-to-point to PAYG, intermediate years' ORR figures implied *reductions* in station use. The London station data were depressingly unrealistic.

## Matrix infilling

In a rail world increasingly defined by marketing initiatives and travel zones, if you estimate usage by starting with ticket sales you will have greater dependence on 'infill' matrices to be overlaid on the original sales data.

Even if Greater London's travel volume could now be managed via Oyster, PAYG and other

Table 1: Station usage in city regions

ORR station usage	Million passenger entries and exits in city region/other parts of region or devolved administration			Local travel assignment	
	2006-07	2010-11	% change from 06-07	2010-11 entries and exits	
CITY REGION	Total	Total	% change from 06-07	Local journeys	% change from 06-07
Other devolved/Govt. region					
<b>BIRMINGHAM</b>	53	93	77	52	81
Other West Midlands	9	11	32	5	34
<b>BRISTOL</b>	19	28	44	16	31
Other South West	30	39	28	21	31
<b>CARDIFF</b>	27	34	25	27	24
Other Wales	8	11	29	5	30
<b>EDINBURGH</b>	29	35	21	18	20
<b>GLASGOW</b>	95	117	23	91	23
Other Scotland	10	13	25	5	28
<b>LEEDS &amp; SHEFFIELD</b>	66	94	42	78	44
Other Yorks & Humber	8	8	8	7	8
<b>LIVERPOOL &amp; MANCHESTER</b>	100	179	79	159	82
Other North West	15	19	24	15	23
<b>NEWCASTLE</b>	10	12	22	8	22
Other North East	6	7	16	6	16
<b>NOTTINGHAM</b>	18	20	12	12	15
Other East Midlands	11	12	10	7	7
<b>LONDON (whole region)</b>	936	1,070	14	741	15
South East (whole region)	304	342	12	226	11
East of England (whole region)	163	174	7	63	8
<b>Total:</b>					
<b>CITY REGIONS</b> outside London	416	611	47	462	48
Other developed/regions o/s L/SE/East	98	120	23	69	22

Notes: The catchment overlap between some City Regions has led to combination of Leeds & Sheffield, and Liverpool & Manchester. Elsewhere, stations are allocated by geography (eg Peak District, River Severn) and by train frequency and destination. Birmingham City Region includes Hinckley (East Midlands). Leeds & Sheffield City Regions include Dronfield, Chesterfield, Worksop, Retford and Hope Valley line to Edale (East Mids). Liverpool & Manchester City Regions include Bidston line to Wrexham (Wales) and to Buxton, Chinley, Glossop (East Mids)

electronic information on an automated basis (as discussed below), that option isn't yet there for other city regions, with their slower spread of electronic ticketing. Some infill improvements are now made during the production of the MOIRA demand matrix, addressing some of the shortcomings noted above.

**City regions excluding London**

The city region data presented in Table 1 is not just about PTE/ITA station usage. It looks to wider catchments up to 20 / 25 miles distant (and further, to Ayr, for Strathclyde). The headline change in ridership at all stations in the city regions is set out, compared to those elsewhere in devolved territories or Government regions.

The table shows the variability in demand changes between different city regions, from 2006-07 to 2010-11. The earlier date is comparable with London data in Tables 1 and 3.

There have been service improvements on some inter-city and inter-regional routes in this period. So as a sensitivity test, I have attempted an indicator of change in local travel demand, allocating percentages of local or main line travel at each station depending on train service types and their daily service volume. While the absolute numbers are not precise, the percentage growth comparing all-user travel and local travel estimates is broadly consistent in the city regions, with variations only where local service volume is weak.

Within this period, ORR inflated its estimates of PTE urban stations and some others, by an average of around 75%, between 2007-08 and 2008-09. This was a major statistical correction. Growth from start to finish is therefore a notional figure, as the 2006-07 baseline was inaccurate to start with, and then improved two years later. The national increase in entries and exits was around 5% in 2008/09, compared with the previous year. This change represents the inclusion of 60.4million rail journeys on PTE tickets that had not been included in previous years.

The current state of city region data excluding London remains as 'work in progress'. The case for non-London projects such as the Northern Hub is made stronger by the improved urban

Table 2: Station usage in West Anglia - annual entries and exits

Stations in area	No. of stations	ORR 2006-07	% change 0607 to 1011	ORR 2010-11	No. of stations surveyed by WARG	ORR 2010-11 at stations surveyed by WARG
London zone 2	7	2,845,244	+81%	5,136,064	3	2,224,190
London zone 3	9	10,913,565	+12%	12,259,075	6	6,764,678
London zones 456	13	10,422,279	-1%	10,336,281	10	8,986,259
Hertford Line + Roydon	9	6,048,716	-0%	6,046,782	3	2,897,184
Harlow-Cambridge	13	13,626,143	-3%	13,152,412	3	3,606,430
H-Camb excl Stansted Apt	12	8,255,819	+10%	9,078,202		
Waterbeach-Kings Lynn	6		+26%			

ORR Stratford and Cambridge based respectively on 6% and 28% of whole station figures, as this is WA approx % of total train service. WARG 2011-12 counts reduced by 6.2% to reflect lower passenger volumes in London & South East area in 2010/11. WARG autumn counts reduced by 5% to remove effect of seasonality



West Anglia in the metropolis: the 15.01 service from Cheshunt to Liverpool Street arrives at London Fields on 2 August 2011, formed of Class 317/6 No 317664. Brian Morrison

data set. The danger with the present situation is that business case analyses are founded on what might still be argued are shifting sands.

PTEs or their equivalents with statistics indicating low rates of growth from 2006-07 to 2010-11 might reasonably question what is going on. Are structural factors such as economic and demographic limitations affecting demand, or are the station usage estimates understating performance?

The coalition government's interest in accelerating use of electronic ticketing may assist statistical progress, but possibly not for a couple of years or more.

### London region

ORR figures for London have not been trusted, as highlighted above. In theory London's statistical world should now be better, as 2010-11 is the first year with full PAYG data. To take a considered view and contrast before and after, it is desirable to take a year before PAYG began and review against 2010-11. 2006-07 has been selected,

as it is pre-PAYG at most national rail stations in London.

The following direct comparisons are feasible:

- West Anglia stations in 2010-11, both within and outside Greater London, where the West Anglia Routes Group had commissioned station counts with support from the train operating company (TOC) in autumn 2010, and again in autumn 2011, to create a detailed database for use in timetable and line project business cases.
- London Overground station usage data and travel trends, informed by low levels of ticketless and fraudulent travel (now under 3% on estimates from the operator LOROL), can be compared with ORR data for London Overground stations. The Overground data can also be compared at East London line 'stand-alone' stations with the previous London Underground data for 2006-07.

### West Anglia Routes

The West Anglia Routes Group (WARG) commissioned detailed station counts from my company, JRC, in autumn 2010 and again in autumn 2011. The London Borough of Enfield also supported additional counts to inform the business case for the Lea Valley three-tracking project.

The counts were stimulated by the experience since 1996 of the Cambridge Heath & London Fields Rail Users Group, which holds twice-yearly station passenger counts; these always show dramatic variance from the ORR estimates. Supporting counts were also undertaken in autumn 2011 by the Chingford Line Users Association.

The WARG counts were taken at a wide spread of sample stations throughout the West Anglia suburban and main lines to Enfield,

Chingford, Hertford, Stansted and Cambridge. Detailed information is available from the WARG secretariat at [www.westangliaroutes.org.uk](http://www.westangliaroutes.org.uk)

The data comprise mostly direct observations with boarding and alighting counts for each train. At a few busy stations it is instead time-series entry/exit data taken every 5-15 minutes. Counts were taken for most or part of the day, with sufficient time range to allow grossing up to daily and annual volumes. Transport for London moderated the counts. The autumn 2010 data was included in the Department for Transport's data room for the Greater Anglia interim franchise bidders. They are used by TfL and the TOC for planning and business case development, as well as by the Routes Group and its local authority members.

A summary of WARG's direct observed entry/exit data for autumn 2010 and 2011, annualised and compared to ORR's for 2010/11, is shown in Table 2.

What we see in Table 2 is that, despite the ORR 2010-11 data finally recognising travel that shifted from point-to-point to PAYG, the latest ORR figures are *still* a large under-estimate on a like-for-like basis, within the London Travelcard zones and at inner Home Counties stations.

### London Overground

TfL has reported major passenger growth on London Overground: 'London Overground passenger volumes are now two and a half times the level when TfL took over management of the concession. The opening of the extended East London line has contributed a large part of the growth but the existing Overground network also experienced an increase in demand of 80%'.

Passenger growth is strong, and continuing. 116million journeys on an annualised basis was the figure reported by London Rail Chief

WARG 2010/11, or 2011/12 adjusted to 2010/11	Variance between ORR and WARG
3,053,706	37%
11,151,655	65%
13,541,801	51%
3,477,778	20%
3,941,714	9%



London Overground, the urban success story: unit No 378136 at Shoreditch, heading towards New Cross on 23 March 2012. Tony Miles

Operating Officer Howard Smith recently to *Modern Railways*' Fourth Friday Club (p40, May issue), compared to 39million on the old Silverlink Metro and the former East London line. TfL is now bidding to expand the four-car Class 378 trains to five-cars during 2014-19. London Overground's predecessor, Silverlink, was running less frequent three-car trains in 2006-07.

Based on 'stand-alone' ORR station entry/exit data, Table 3 shows Overground station usage for 2006-07 and 2010-11.

The variations in the percentage figures in the final column of Table 3 make the ORR figures difficult to believe statistically. Only the figure for the Gospel Oak-Barking line (Upper Holloway-Woodgrange Park) looks anything like TfL's assessment of an 80% increase in demand on the LO routes.

Furthermore there is a problem with the absolute numbers. Divide column two of Table 3 by two to count total journeys rather than entries and exits (interchange within national rail is excluded). This gives a total of over 25million originating journeys in 2010/11. Yet Howard Smith said total Overground flows are now 116million passengers on an annualised basis. Is the 3½ times gap to be covered by the excluded stations, plus the East London line? This isn't possible.

Let us look at the current data for Dalston Kingsland to Hackney Wick. ORR's assessment of all Dalston Kingsland-Hackney Wick entries and exits was 6.69million in 2006-07 and 6.93million in 2010-11 (4% up). However current weekly London Overground data (x 50 for annual) shows the same stations having 14.5million entries and

exits (this figure is after reducing their use by 10% to give a 2010/11 level, to allow for exceptional Overground growth in the last year). So there is an underestimation by ORR of 110%, despite all the recent efforts to improve data standards.

At the ex-London Underground East London line stations there is another comparison available, between ORR's modified ticket sales data for 2010-11 and LUL's 'click-in click-out' data for 2006-07. The accurate London Underground data shows 100% greater use of these stations, four years before the recent reopening. The ORR data does not use the full 'click-in click-out' information, though it includes MOIRA estimates (so is MOIRA reliable?). In neither case was the line opened throughout the year, but experience shows that passenger flows return quickly to the railway (it took only 10 weeks after the three-

Table 3: London Overground usage according to ORR - annual entries and exits

Section of line (ex-Silverlink Metro)	ORR 2006-07	ORR 2010-11	% ORR change to 2010-11
Watford High Street-Headstone Lane	2,489,884	3,464,024	39%
Kenton-Harlesden	2,033,721	5,158,784	154%
Kew Gardens-Acton Central	3,512,582	3,838,868	9%
Imperial Wharf-Willesden Junction	3,383,155	9,173,238	171%
Kensal Green-South Hampstead	2,361,673	6,192,012	162%
Kensal Rise-Caledonian Road & Barnsbury	11,703,780	11,884,531	2%
Canonbury-Hackney Wick	7,484,113	7,703,430	3%
Upper Holloway-Woodgrange Park	1,852,622	3,508,806	89%

Excludes LUL station entry/exit data Kenton-Queens Park and Kew Gardens-Gunnersbury

Section of line (ex-LU East London Line)	LUL 2006-07	ORR 2010-11	Variance compared to ORR
<b>Stand-alone stns:</b>			
Shadwell, Wapping, Rotherhithe, Surrey Quays	7,098,000	3,543,500	LUL: +100%
Same stations, London Overground 2011-12	10,813,223	3,543,500	LO: +205%

Recent Overground weekday use x 300 for annual, excludes interchange at Surrey Quays. Excludes 'other operator' interchanges such as Richmond, Harrow & Wealdstone, Watford Junction, Highbury & Islington, Clapham Junction, Stratford and Barking, but includes Southern/Overground joint stations and Bushey.

year East London line closure during 1995-98 to return to previous levels).

The East London line was open for more than 11 months during 2010-11 (April 2010 from Dalston). It has a more frequent service than before, including through services to Southern destinations from May 2010. It was opened to Highbury & Islington in February 2011. The local catchments have intensified in housing and jobs. So is the Overground a failure, or is the ORR's way of estimating station usage still wrong? The comparison with recent use of these stations shown in the last line of Table 3 shows it is the ORR data which still has problems.

**What's still wrong in London?**

The scale of local travel volume in London means that any errors in infill processes are magnified. Local journeys in Greater London represent at least 1½ times all other local journeys and exits in all the city regions outside London put together.

A large problem is that the underlying trip distribution matrix for the London region, to allocate much zonal travel on a point-to-point basis, is the 2001 LATS database (London Area Transportation Survey). This is outmoded. It is pre-Oyster and PAYG which have radicalised inter-modal travel between buses, light rail, heavy rail and tube. 2001 is also prior to much repopulation of inner London and the further growth which is forecast to the 2030s in the numbers of homes and jobs. Unfortunately Government decided to save the cost of re-doing LATS in 2011.

There has also been the successful creation under TfL London Rail of the popular and trusted London Overground brand during the last four years, which has unlocked suppressed orbital travel demand on public transport whose upper limit is still not quantified. Possibly it is the more successful because it is a creature of zonal pricing rather than point-to-point pricing.

**Group stations**

Finally, it is useful to see the extremes of station data manipulation, and believability, as demonstrated by the combined ticket sales for 'Group stations'. These stations represent an attempt by ORR to divide multiple-destination ticket sales into point-to-point categories. As an extreme example of a 'group station', National Location Code 1072 is 'London BR': this includes

all London Zone 1 stations except Old Street and Shoreditch High Street.

As a simpler example, Table 4 shows entries and exits at the three stations in Dorking, with interchange noted separately. This is an homogenous Home Counties commuter town with both radial and orbital railways, but statistical consistency is not the order of the day. Look at how the number of passengers entering and leaving Deepdene, for example, is supposed to have been 200 times as many one year as compared to the year prior, as the statisticians have struggled to divide up the figure in the final column fairly between the three stations.

Readers with time could look up other Group stations data on the ORR website for different years, and rationalise what the numbers suggest.

**Assessment**

In a different retail world, supermarkets know at once what you like and how much you have bought, once you have gone through their EFTPOS till systems. This information drives their 'just in time' re-stocking, and their future marketing.

Rail could be like this, including influencing future timetable proposals and better bus-rail integration, but we aren't yet there. To be effective, it will require much more 'click-in click-out' real-time knowledge, particularly in city regions.

By 2010-11, under 15% of local entries and exits outside London/South East/East of England were from beyond the main city regions, so when the city regions move to ITSO smartcard ticketing, the stimulus for statistical change could be strong.

Meanwhile, it isn't clear why maintaining the current matrix infill in London will be much more productive or believable in future years than its predecessors, if factors such as ignoring freedom pass travel, ticketless travel and a 2001 travel matrix are still omnipresent.

It could be argued that London Overground is a phenomenon, with great marketing, a vastly improved product and a city population which is motivated to use rail. But isn't that what many stakeholders want elsewhere across Britain's city regions?

West Anglia isn't a unique product, it is a commuter railway with strong with-flow and

contra-flow local travel seeking to justify vital investment in a local service which will liberate economic development opportunities along the Lea Valley worth over £10billion in gross value added by 2031.

London is also further advanced in electronic ticketing and monitoring than other parts of Britain. Perhaps London's urban rail data should be liberalised soon onto a new 'click-in click-out' baseline, where London Underground information can be more comfortably published alongside its national rail equivalent? LENNON could then be an infill for London, rather than the other way round. There are supporting opportunities, with some train loadmeter data, and automated movement detection in and out of Docklands Light Railway platforms and on Croydon trams.


**Is your number up?**

For city regions the ORR's station usage data is not to be trusted for strategic policies, project planning nor business case analysis. Yet, for all schemes which seek to be successful in securing funding for Control Period 5, the Department for Transport is using this data.

There are the Northern Hub, expansion schemes in Scotland, the West Midlands, and London's Lea Valley and the Cardiff Valleys, to name just a few.

National Passenger Survey (NPS) data is also used increasingly by DfT and Passenger Focus as a tool to measure franchise quality and delivery standards. Guess what station usage data starts the process of weighting their importance for NPS assessment?

The growth of city regions is put forward as the explanation of why urban local railways continue to exist and thrive post-Beeching. But the latest city region rail data appears to undersell the urban stations – and the railway's current and future potential.

Within his recommendations in the Rail Value for Money Study, Sir Roy McNulty has sought 'greater transparency of the industry's finances and cost performance'. He has suggested that the ORR should have 'a new focus on whole-system outputs and with the necessary resources, skills and standing', and 'improved oversight and management of cross-industry information systems'. Station usage looks a strong candidate for attention. 

**Table 4: Station usage in Dorking**

Dorking station group, entry & exit, separate interchange volume	Group lacking ticket offices			All Dorking		
	Dorking (North) entry & exit	Dorking (North) interchange	Dorking (Deepdene) entry & exit		Dorking (Deepdene) interchange	Dorking West entry & exit
ORR year						
2002-03	1,168,693		8,799		154	1,177,646
2004-05	1,238,425	40,916	5,456	156,392	29	1,243,910
2005-06	1,279,034	49,901	1,706	164,966	40	1,280,780
2006-07	1,362,275	56,423	2,474	172,860	79	1,364,828
2007-08 unamended	1,078,443	68,631	495,310	52,049	52	1,573,805
2007-08 amended	1,078,494		495,260		52	1,573,806
2008-09	1,141,369	62,772	443,106	69,465	104	1,584,579
2009-10	1,071,978	63,088	408,757	76,972	1,810	1,482,545
2010-11	1,186,717	75,499	382,194	91,050	22	1,568,933