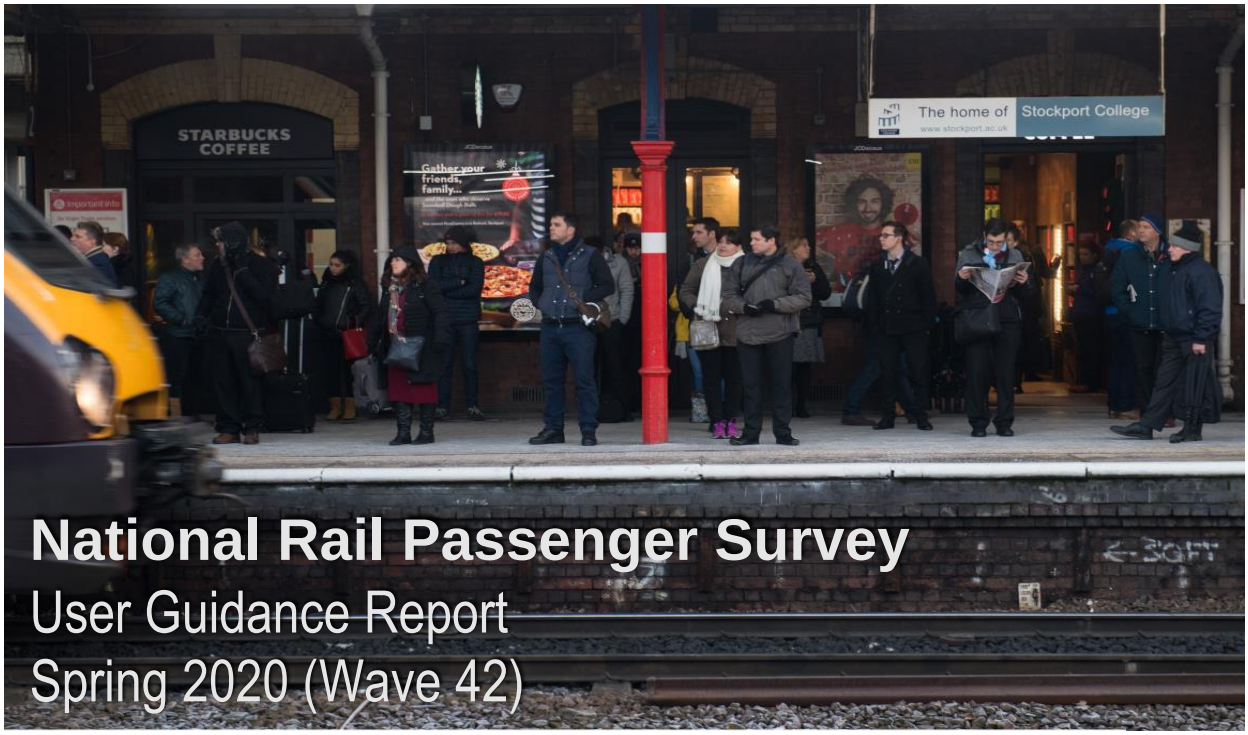


National Rail Passenger Survey

User Guidance Report

Spring 2020 (Wave 42)

transportfocus

Contacts:

David Greeno
Transport Focus
Fleetbank House
2-6 Salisbury Square
London, EC4Y 8JX
Tel: 0300 123 0837
Email: david.greeno@transportfocus.org.uk

David Chilvers
David Chilvers Associates
Devine House
1299-1301 London Road
Leigh-on Sea
Essex SS9 2AD

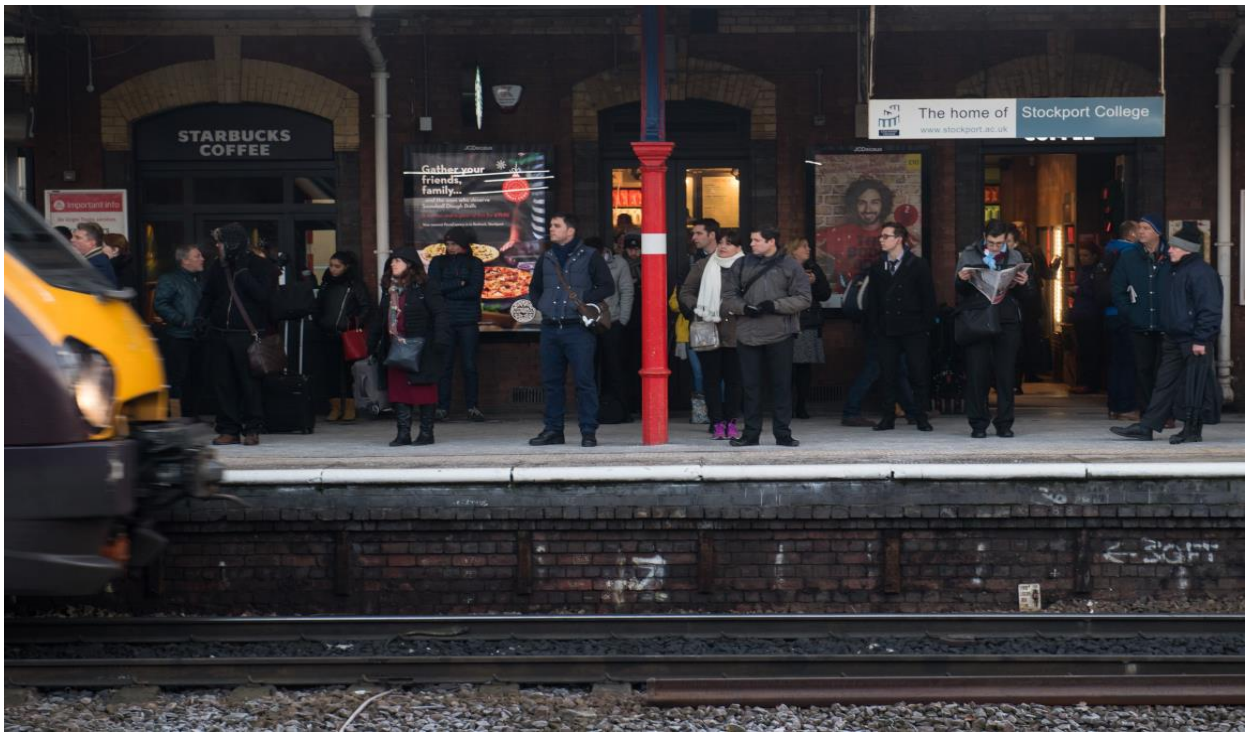
Tel: 07973 671990
Email: david.chilvers@dcaweb.com

David Adley
Watermelon Research
2nd Floor, Francis House
Francis Street
London SW1P 1DH

Tel: 07807 648518
Email: dadley@watermelonresearch.com

Contents

Page No.



WATERMELON

1. Background

Transport Focus (known as Passenger Focus until April 2015), and previously OPRAF and the Strategic Rail Authority set up the National Rail Passenger Survey (NRPS) in 1999. The aim of the NRPS was to provide customer views on rail company performance on a consistent basis, so that comparisons could be made between the various companies over time. Data from the NRPS has been built into the franchising contracts with train companies, making the results an important commercial dimension of running a Train Operating Company (TOC). Given this, the sample design, fieldwork standards and accuracy of assigning journeys to specific TOCs are of the greatest importance. In addition, large enough sample sizes are required for each TOC to ensure that performance changes can be seen in the marketplace.

It is important to note that NRPS is what is called an immediate event-based passenger satisfaction programme; it measures passengers views of individual events that have taken place on the day they were given a questionnaire (or invitation to participate online). As such, the results allow Transport Focus and other users to generate various journey satisfaction measures. The results do not purport to provide an overall view of a TOC, rather an overall view of journeys undertaken on that TOC. The same distinction applies to journeys relating to an subset of NRPS; for example the data for London Euston provides a view of journeys undertaken from that station rather than an overall view of the station itself.

The benefits of immediate event-based passenger satisfaction measures include:

- Focus on a specific journey, rather than some generic view about the train company or station used in that journey that might be gained from a survey of passengers
- Good recollection of the various aspects of the specific journey, again rather than some aggregate view of a collection of journeys undertaken by passengers (which other research has shown will be dominated to an extent by the views of bad experiences)
- The ability to select a sample of journeys and weight up to a recognised universe of train journeys, where the details required for weighting are provided by the industry itself and are open to effective quality assurance. All the data used in NRPS for weighting has been provided by ORR or the TOCs themselves, either directly or via their information systems (e.g. LENNON and the RailPlanner database). In contrast, surveys of passengers can be difficult to weight up to ensure the views of commuters, business and leisure travellers are properly represented

The first NRPS was run in Autumn 1999 and it has been run twice a year since then. The first seven waves were undertaken by The Oxford Research Agency, until the contract was offered at competitive tender in Autumn 2002. In December 2002, Continental Research (now merged to become BDRC Continental) was appointed to run the survey. This survey remained with BDRC Continental as supplier until 2017 when the research agency contract was awarded to Facts International (now branded Watermelon Research). A transitional process followed this change in contract and Watermelon Research have since been responsible for delivery a further five waves of the NRPS.

During this time the weighting and sampling approaches have both undergone significant review to ensure the study remains representative of the UK rail network. This document provides information around the sampling and weighting mechanisms used to control the data collection and production and the types of analysis that can be undertaken. It also gives guidance on the likely levels of accuracy which attach to different subgroups and highlights areas where the data may be subject to wide margins of error.

Much more detail on NRPS processes is included in the NRPS Technical Report; this also details all the quality assurance checks that are made each wave to ensure users can have confidence when using NRPS data.

2. Caveats regarding this wave

Fieldwork on the Spring 2020 NRPS wave was suddenly curtailed in early March due to the onset of Coronavirus. This meant that target sample sizes were not achieved and there was no real opportunity (as is normally the case) to use the last 2-3 weeks of fieldwork to target specific sample cells that were most adrift of target.

The lower sample size naturally leads to wider margins of error. The lack of targeting has led to lower weighting efficiencies for most TOCs reducing effective sample sizes and further widening margins of error. A small number of sample cells had to be merged due to low sample sizes. Apart from widening margins of error, the loss of fieldwork does not appear to have led to any unusual results for the key metrics at national level or for most TOCs. The widening of margins of error has been taken into account in identifying any significant shifts in reported metrics.

However, Merseyrail clearly stands out from other TOCs, with six metrics that show spikes in Spring 2020 (which are not continuations of previous trends) and the TOC featuring on four of the six criteria used to judge whether there are any specific issues with the Spring 2020 NRPS wave. There may be good reasons for the changed perceptions if the TOC has undertaken specific related actions but if this is not the case, we would suggest considerable caution is used when interpreting the results for Merseyrail.

A number of TOCs have low rankings on more than one of the criteria used to judge whether there are any specific issues with the Spring 2020 NRPS Wave. These TOCs are Great Northern, East Midlands Railway, London North Eastern Railway, c2c, Chiltern Railways, Transport for Wales and TransPennine Express. With these TOCs, we would suggest greater than usual care is taken when interpreting the results of the Spring 2020 wave.

For the other TOCs, we see no great evidence that the curtailment of fieldwork has had any major impact on the results and we suggest that the metrics for these TOCs meet the normal NRPS standards, with the caveat that reduced sample sizes and weighting efficiencies will widen the margins of error. These widened margins of error will be in any event used to calculate whether there have been significant changes and we recommend that the NRPS User Guide is updated to reflect these wider margins for this wave. Further details are available from a full Quality Assurance Report, which is available on the Transport Focus website or by emailing David Greeno.

3. Types of information available

The NRPS is designed to generate random samples of journeys for each Train Operating Company (TOC). Used at the TOC level, the normal rules for calculating sampling error for a weighted sample apply and examples are shown in Section 5 below. The NRPS sample is selected for TOC building blocks. All franchised TOCs use route-based building blocks as part of their sample design and as a component part of the weighting targets (the only exception being Northern who amended their weighting process in Wave 39 to be based on the journey start station rather than route-based building blocks).

The NRPS can be used to explore data at a station or route level, which may cover more than one TOC. However, as with any sample, the smaller the subgroup being analysed, the more inherent variability there is in the data and the level of precision decreases. NRPS can also generate data at regional level and this is used extensively in the Stakeholder Report. For some regions, this involves amalgamating data from several TOCs. At present the NRPS data reports on the following sub-categories of data:

- Franchised: Franchised only (excluding Hull Trains, Heathrow Express and Grand Central)
- Sectors: London South East; Long Distance; Regional
- Virtual TOCs (areas that used to be TOCs or part of TOCs): GWR London Thames Valley; GWR Long Distance; GWR West; London Northwestern Railway; Southern including Gatwick Express; TfL Rail East; TfL Rail West; West Midlands Railway
- Network Rail: Network Rail Stations; Non-Network Rail Stations
- PTE: Merseytravel; Nexus; South East Wales; South Yorkshire; Strathclyde; TfGM; West Midlands; West Yorkshire

NRPS is available as a single dataset covering the last ten waves with an SPSS format – a full five year period. For example, for the Spring 2020 Wave (Wave 42) this covers Autumn 2015 to Spring 2020. The majority of ad-hoc requests for analysis received use this data as their basis and are managed through an Extra Analysis Log between the research agency and Transport Focus. It is also possible to acquire NRPS data at respondent level within this dataset. The dataset is very large. Users need a good understanding of analysing large datasets with weighting to be able to use this facility. Further to this a verbatim database is issued containing key variables and verbatim comments at a respondent level. i.e. one respondent per row of an excel document.
All published documents are housed on a secure server and transferred via a SFTP (Secure File Transfer Protocol). Access rights to the secure server are managed by Watermelon Research and Transport Focus. From Spring 2019 onwards the NRPS results are also published via the Transport Focus Data Hub – a central reportal for data dedicated to the key studies delivered by Transport Focus. NRPS data is available on the system back to wave 1 (Autumn 1999).

4. Sample design

NRPS uses a two stage cluster sample design. Firstly, stratifying each Train Operating Company by Building Block and secondly, identifying the stations within the identified Building Blocks.

The main sampling unit is a train station or on-train route. A process is undertaken to select stations for sample based on passenger volumes within a Building Block. Stations are then randomly selected across the different station size profiles to produce a selection of stations across the different building blocks.

Once the station list has been generated for sampling, a further randomised process is applied to determine the day of week and time of day that the fieldworks shifts should be conducted at the selected stations. This process ensures fieldwork shifts cover a broad spectrum of journey and passenger types.

Different sample sizes are set for each TOC that reflect the complexity of routes and the number of passengers the company carries. To arrive at a national dataset that represents all passengers' satisfaction with their journeys, each TOC is weighted to reflect the number of journeys that it contributes to the national rail network.

The NRPS Technical Report covers more detail on how the sampling process operates and the quality assurance checks that have been made prior to release of the data.

Table 1 below shows data on:

- The number of journeys undertaken in the past twelve months for each TOC (derived from the LENNON database)
- The sample size for the TOC
- The ratio of these two numbers, showing the incidence of sampling from each TOC

Some of the smaller TOCs (e.g. Hull Trains and Grand Central) are sampled quite intensively to generate an adequate sample size; some of the larger TOCs (e.g. Thameslink and Southern) are sampled less intensively due to their very high number of passenger journeys.

Table 1

TOC	Number of journeys (000s per annum)	Target Sample Size	Ratio
Avanti West Coast	38285	1000	38.3
c2c	46743	1000	46.7
Chiltern Railways	25376	1000	25.4
CrossCountry	38829	1200	32.4
East Midlands Railway	26433	1000	26.4
Gatwick Express	11680	500	23.4
Grand Central	1389	500	2.8
Great Northern	43247	500	86.5
Great Western Railway	113196	1500	75.5
Greater Anglia	86049	1300	66.2
Heathrow Express	6388	500	12.8
Hull Trains	2074	500	4.1
London North Eastern Railway	21800	1000	21.8
London Overground	187800	1600	117
Merseyrail	43828	700	62.6
Northern	103627	1400	70.4
ScotRail	95875	1300	73.8
South Western Railway (inc Island Line)	216670	2000	108
Southeastern	163597	1500	109
Southern	164028	1300	126
TfL Rail	55713	1000	56
Thameslink	170652	1000	171
TransPennine Express	29520	1000	29.5
Transport for Wales	36125	1000	36.1
West Midlands Trains	76176	1000	76.2

(Note that while this table includes non-franchised TOCs which take part in NRPS, only franchised TOCs contribute to national, regional or sector aggregates for normal reporting.)

5. Weighting

The NRPS data is weighted by TOC to reflect the number of passenger journeys undertaken by each individual TOC. Station sizes and building blocks are identified as the sub-groups within the TOC and the passenger volume weights are proportionately applied across these sub-groups. A second proportional weight is also applied to ensure the data accurately reflects the overall proportions for each TOC for journey purpose and day of week. Within the dataset for the NRPS this weighting is applied to what is referenced as the 'main' dataset (that is reflective of the target numbers in table 1.1).

To allocate the building blocks the questionnaires responses are analysed across the whole route assessing the start station, end station, train origin and train destination. These four factors determine which building block should be applied. The only exception to this approach is for the Northern TOC where, in Autumn 2018, they made the decision to use the start station alone to identify the building block classification and not the whole route (this was partly because of difficulties in producing accurate weighting information based on whole routes). Unlike other building blocks where each station could occur in more than one Building Block, Northern has made each station exclusive to a Building Block. For example, Manchester Piccadilly will be always be in Northern – Central BB and Blackpool South will be Northern – West BB.

Station size is allocated to every station within each building block and is based on passenger volumes within that building block. Station size is categorised as Very large, Large, Medium and Small and the passenger start station is the dependent variable.

The data for each TOC is then weighted by weekday/weekend and journey purpose (Commuter/Business/Leisure), and grossed up to the estimated number of passenger journeys for that TOC building block. This means that the weighted data for a number of TOCs or building blocks can be simply aggregated (e.g. to generate data for a virtual TOC or a regional grouping).

Where boost samples are requested, fieldwork shifts may be set to specifically target boost stations. In this case the data is not part of the weighting calculation and results are looked at separately to the 'main' survey outputs.

In addition to the 'main' weighting there are two separate weighting profiles applied to the NRPS data: PTE weighting; and Network Rail weighting. Details of the likely accuracy of data in these reports is shown in Sections 6.4/6.5 below.

Further detail on the weighting process is given in the NRPS Technical Report.

6. Statistical Accuracy and Evaluation

6.1 Effective sample size (effective base)

When calculating any statistical movements using the NRPS data, the effective sample size is used within the calculation. If the data has been weighted, the effective base is designed to ensure the likelihood of the statistical test producing significant results factors the adjustments made by the weighting. If the weighting sees each respondent have the same weight, the unweighted and effective base will be the same; where weights differ by respondent, as in NRPS, the effective sample size will be lower than the unweighted sample. In the case of the NRPS there will be range of downweights and upweights applied to each individual respondent to ensure the data is reflective of the set weighting requirements (station size, building block, customer journey purpose, and day of week). Each of these will impact the effective base shown in Table 2 below:

Table 2

TOC	Unweighted total	Effective Weighted Sample	weighting efficiency
Avanti West Coast	641	441	68.80%
C2c	619	438	70.76%
Chiltern Railways	818	470	57.46%
CrossCountry	932	714	76.61%
East Midlands Railway	637	370	58.08%
Gatwick Express	416	205	49.28%
Grand Central	395	331	83.80%
Great Northern	314	251	79.94%
Great Western Railway	1032	520	50.39%
Greater Anglia	1016	494	48.62%
Heathrow Express	533	411	77.11%
Hull Trains	427	289	67.68%
London North Eastern Railway	753	331	43.96%
London Overground	1244	898	72.19%
Merseyrail	525	183	34.86%
Northern	1004	611	60.86%
Scotrail	1025	528	51.51%
South Western Railway	1471	1049	71.31%
Southeastern	1345	711	52.86%
Southern	899	774	86.10%
TfL Rail	825	533	64.61%
Thameslink	852	661	77.58%
Transpennine Express	668	404	60.48%
Transport for Wales	781	246	31.50%
West Midlands Trains	675	492	72.89%

6.2 Confidence Intervals

When calculating confidence intervals for NRPS outputs, the effective sample sizes for the relevant subgroups should be used. Confidence intervals also depend upon the statistic being estimated. For percentages close to 0% or 100%, the confidence interval will be narrower than for percentages close to 50%, which is the worst case situation. Below we give estimated confidence intervals for each TOC for statistics estimated at 10/90%, 25/75% and 50%.

The confidence intervals for estimates at 10/90%, 25/75% and 50% for each TOC are shown in Table 3 below:

Table 3

TOC	10%/90%	25%/75%	50%
Avanti West Coast	2.80%	4.00%	4.70%
C2c	2.80%	4.10%	4.70%
Chiltern Railways	2.70%	3.90%	4.50%
CrossCountry	2.20%	3.20%	3.70%
East Midlands Railway	3.10%	4.40%	5.10%
Gatwick Express	4.10%	5.90%	6.80%
Grand Central	3.20%	4.70%	5.40%
Great Northern	3.70%	5.40%	6.20%
Great Western Railway	2.60%	3.70%	4.30%
Greater Anglia	2.60%	3.80%	4.40%
Heathrow Express	2.90%	4.20%	4.80%
Hull Trains	3.50%	5.00%	5.80%
London North Eastern Railway	3.20%	4.70%	5.40%
London Overground	2.00%	2.80%	3.30%
Merseyrail	4.30%	6.30%	7.20%
Northern	2.40%	3.40%	4.00%
Scotrail	2.60%	3.70%	4.30%
South Western Railway	1.80%	2.60%	3.00%
Southeastern	2.20%	3.20%	3.70%
Southern	2.10%	3.10%	3.50%
TfL Rail	2.50%	3.70%	4.20%
Thameslink	2.30%	3.30%	3.80%
Transpennine Express	2.90%	4.20%	4.90%
Transport for Wales	3.70%	5.40%	6.20%
West Midlands Trains	2.70%	3.80%	4.40%
National total	0.50%	0.80%	0.90%

So, if the confidence interval of an estimated statistic of 75% is required, this will be +-0.8% at national level, if the statistic is measured for the complete sample, and range from +-2.8% for London Overground to +-6.3% for Merseyrail.

All the significance tests in the various published reports use the effective sample size for the subgroup concerned. To compute significant upward and downward changes between one NRPS wave and another, the two % scores you are comparing and the two relevant effective base sizes are required to complete the calculation.

At TOC and TOC building block level, most analyses are robust enough to stand up to scrutiny. For TOC building blocks, the sample sizes are smaller than for complete TOCs but the range of weights will also normally be smaller, as a large contributor to the variance of weights is the need to gross each building block up to its passenger numbers. So, for TOC building blocks, the effective sample size will usually be closer to the unweighted sample than for complete TOCs.

6.3 Station level estimates

At station level, the combination of smaller sample sizes and greater variation in weights if more than one TOC is involved mean that analysis at station level is substantially less robust. Ideally, station or route analysis should be based on sample sizes of at least 100, and certainly at least 50. To reach this sample size for some stations or routes, it may be necessary to combine waves.

In addition to the weighting effect at a station level, there is a clustering effect. As most TOC building blocks are sampled based upon stations, the number of different days contributing to data for a particular station may be quite low. The number of returns for a typical three hour interviewing shift is 17 and thus a station with an unweighted sample of size of 50 may be based upon passenger perceptions on just three separate days. If there was a problem at the station on any one of these days, the impact of this will be more pronounced than at a station where sampling took place on a larger number of days.

In addition, the sampling plan for each wave of NRPS tends to be based upon the sampling plan for the previous wave. This has always been the case and is a good way to minimise random changes from one wave to another. However, this also means that when data for several waves is aggregated, there may be concentration of the sampling, at station level, on particular days or particular times of day. This phenomenon is not present at TOC level, as the sample at that level is weighted by weekday/weekend and journey purpose, but at individual station level this type of control will be absent.

Combining the weighting effects and the impact of clustering implies that the data at station level, for anything other than the larger stations, should be taken as indicative of performance rather than being generated from a proper random sample. For all the Category A stations and many of the Category B stations, the sample size and the number of distinct interviewing shifts contributing to the data should render the statistics reasonably robust; but for smaller stations, caution should be exercised in use of the data.

6.4 PTE Reports

For PTE reports, a different weighting regime is used, but the same principles apply to calculating confidence intervals based upon effective sample sizes.

PTE weighting is applied to records that qualify as a PTE category. PTE definitions include those listed in Table 4 below. Please note that whilst London, South East Wales and Strathclyde are treated like PTEs, they are not specifically defined as PTEs, but are included within this weighting. The data identified as PTE has the volume weighting removed and a revised proportionate weighting applied across journey purpose and day of week. Any data reported within PTE reports will use this weighting. Where PTE data is used in the main data set, the standard 'main' weights are applied.

PTE Reports use weights derived from several (usually ten) waves of data, to ensure smoothing of the weighting regime from one wave to another. Each wave, the day of week and time of day profile for each PTE area is compared to the existing weights and a judgement made as to whether the weights should be updated. Further detail on this is shown in the NRPS Technical Report, including information on whether the PTE weights have been updated.

The 95% confidence limits for each PTE area for the Spring 2020 wave are shown below:

Table 4

PTE area	10%/90%	25%/75%	50%
London	1.2%	1.7%	1.9%
Merseytravel	3.7%	5.3%	6.2%
Nexus	5.4%	7.8%	9.0%
South East Wales	5.6%	8.1%	9.3%
South Yorkshire (Travel South Yorkshire)	12.0%	17.3%	20.0%
Strathclyde	3.9%	5.6%	6.4%
TfGM	5.4%	7.8%	9.1%
West Midlands	2.8%	4.1%	4.7%
West Yorkshire (Metro)	6.1%	8.8%	10.2%

6.5 Network Rail Reports

A report is produced that provides data for each Network Rail owned station and a different weighting regime is used to generate this report.

For the Network Rail weighting a separate weighting profile is applied to the Network Rail stations across the main and boost datasets, which is calculated by Populus. The data is weighted by each Network Rail station to reflect the number of passenger journeys by Network Rail TOCs, by day of week and time of day.

The 95% confidence limits for each Network Rail station are shown below

Table 5

NR station	10%/90%	25%/75%	50%
Birmingham New Street	3.0%	4.4%	5.0%
Bristol Temple Meads	4.2%	6.0%	6.9%
Edinburgh Waverley	2.9%	4.3%	4.9%
Glasgow Central	4.7%	6.8%	7.8%
Guildford	3.9%	5.6%	6.5%
Leeds	3.7%	5.3%	6.1%
Liverpool Lime Street	4.4%	6.3%	7.3%
London Bridge	4.3%	6.1%	7.1%
London Cannon Street	4.9%	7.1%	8.2%
London Charing Cross	4.0%	5.8%	6.7%
London Clapham Junction	3.4%	4.9%	5.7%
London Euston	4.3%	6.3%	7.2%
London Kings Cross	3.9%	5.6%	6.4%
London Liverpool Street	2.7%	3.9%	4.4%
London Paddington	3.9%	5.6%	6.4%
London St Pancras International	4.6%	6.7%	7.7%
London Victoria	3.6%	5.2%	6.0%
London Waterloo	3.2%	4.6%	5.3%
Manchester Piccadilly	3.7%	5.3%	6.1%
Reading	4.6%	6.6%	7.7%

There is also a Station Report that covers other stations. All the comments made in Section 6.3 above apply in terms of the impact of sample size and clustering.