Department for Transport

Ticket purchasing behaviour and preferences among rail passengers - technical report

February 2024

Department for Transport
Great Minster House
33 Horseferry Road
London SW1P 4DR

This publication is also available on our website at
www.gov.uk/government/organisations/department-for-transport

Any enquiries regarding this publication should be sent to us at
www.gov.uk/government/organisations/department-for-transport

Contents

1. Background and Objectives

Background

The Department for Transport (DfT) commissioned independent research agency Savanta to undertake exploratory quantitative research to explore current ticket purchasing behaviour in England and identify barriers or potential barriers to ticket purchasing that may exist or be exacerbated by a shift towards digital ticketing, with particular focus on vulnerable passenger groups. The research also tested likely uptake of possible new initiatives that may mitigate the impact of potential reforms.

Objectives

The objectives for this quantitative research project were to:

1. Explore how rail passengers make ticket buying choices, including the factors that impact their decisions and the weight they attach to these factors. This will build understanding of how rail passengers currently purchase tickets and how this may impact their future behaviour.

2. Explore how vulnerable groups make ticket buying decisions and identify any barriers or potential barriers to rail travel that may exist or be exacerbated by a shift towards digital ticketing. These vulnerable groups include but are not limited to: people that need to pay by cash, people that do not have a bank account (unbanked), people that have no access to or lack the confidence to use the internet or a smartphone, and disabled people.

3. Understand the proportion of rail passengers most likely to be impacted by a shift towards digital ticketing.

4. Identify the common characteristics of those rail passengers that may face barriers to digital ticketing in the future.

2. Methodological Approach

The methodology for this research was designed to achieve as representative a sample of rail passengers in England as possible, while also ensuring robust coverage of all regions and DfT contracted Train Operating Companies (TOCs).

An on-train self-completion approach was adopted to ensure robust and representative coverage of all passenger types within the population of current rail passengers in England. To further optimise coverage of this population, additional completion options were made available, including reply-paid envelopes, online (via QR code), telephone, and (exceptionally) fieldworker assistance to complete the questionnaire. The use of reply-paid envelopes for postal returns of the paper-based survey and QR codes for online completion enabled rail passengers with shorter journeys or insufficient time onboard the train to complete the survey post journey.

Overall approach

The research methodology needed to be as inclusive as possible. An on-train self-completion methodology was chosen because:

- It is a well-established approach used within rail research to provide robust and representative samples of rail passengers.
- By offering paper and pen, digital and the potential for CATI/fieldworker assisted responses, the methodology was inclusive and not biased towards or against certain types of rail passengers (e.g., those without access to the internet).
- It allowed for a sampling plan to be created that had broad geographic and Train Operating Company coverage.
- It allowed for a degree of random stratified sampling to ensure coverage of a random sample of train services covering different times of day, stopping patterns and user types.
- It ensured that passengers invited to complete the survey were boarding at a wide range of stations, to ensure all possible purchasing options were covered (in terms of availability if not actual usage).
- Given the lack of high quality, up-to-date passenger profiling data, the on-train approach allowed for on-board counts to be used to identify, and through weighting rectify, any non-response bias within the data (see Weighting section for more details).

Sampling plan

The goals of the sampling approach were to:

- obtain a representative sample of rail passengers in England and ensure broad coverage of DfT-operated stations
- collect a large enough sample to allow for robust reporting at a TOC and regional level
- collect a sample across different days of the week and times of day, with a mix of morning, evening, and weekend shifts
- collect a sample of sufficient size for robust analysis of key sub-groups including age, disability, payment methods used, bank account access, and internet/smartphone use

Based on the assumption that a six-hour on-train fieldworker shift would yield 50-70 completed questionnaires, it was agreed that a total of 160 shifts would be needed to achieve a minimum of 8,000 completes.

To enable representative regional coverage, while also obtaining a robust sample at the TOC level, the first 140 shifts were allocated to regions representatively and the remaining 20 shifts were used to boost in TOCs where the number of shifts would otherwise be too low to generate a robust sample.

Steps taken to create the plan

Step 1: Estimating passenger journeys originating from each station

Data from the LENNON[1] (Latest Earnings Networked Nationally Over Night) ticketing and revenue system was used to inform the sampling process. LENNON holds information on the vast majority of rail tickets purchased in Great Britain and allocates journeys from those ticket sales to TOCs using the mathematical model ORCATS[2] (Operational Research Computerised Allocation of Tickets to Services).

A summary of passenger journeys allocated to stations operated by each DfT-contracted TOC over a series of baseline periods between June and October was used to estimate passenger journeys originating from each station in a typical week.

[1] LENNON (Latest Earnings Networked Nationally Over Night) ticketing and revenue system holds information on most train tickets purchased in Great Britain and allocates journeys from those ticket sales to TOCs using the mathematical model ORCATS (Operational Research Computerised Allocation of Tickets to Services).

[2] ORCATS (Operational Research Computerised Allocation of Tickets to Services) utilises similar logic to journey planning systems and identified passenger 'opportunities to travel' from an origin station to a destination station using timetable information.

Step 2: Station selection to achieve representative regional coverage

Within each region, stations were stratified by average number of passenger journeys and the following process undertaken to select stations, around which the 140 shifts would be built:

- For each region, a sampling factor (n) was determined by dividing the total number of journeys in that region by the number of shifts required in that region (for example, if 5 shifts were required in a region and there were 500 journeys made within that region then n=100)
- A random number between 1 and n (inclusive) was generated to identify the first station, which would be the starting point for the remaining selection.
- The remaining stations were then selected by taking the station that accounts for every nth 'journey' (based on the 500 journey example, if the randomly selected start point was 25 and the sampling factor was 100, the stations that covered journeys 25, 125, 225, 325 and 425 would be sampled).

Using this approach meant that it was possible to select stations multiple times. For example, if Station A had 175 of the 500 journeys made in the region originating there, then journeys 25 and 125 would fall under this station. In this instance, it would be included twice as the start station within the sampling plan.

Step 3: First 140 shifts allocated to routes

Lists of all weekly timetabled services running through each of the start stations identified in Step 2 were generated. From each list, a scheduled train service was randomly selected, and a sampling plan developed around this. Sampling plans covered a number of trains, within the following criteria:

- They covered journeys within a six-hour period between 7am and 7pm[3]
- They represented a mix of weekday and weekend journeys
- Journeys were chosen on the specific services at random that last, where possible, between 30mins and 1hr

For example, if Oxford station was chosen as a start station from the random stratified region sample, a list of all timetabled rail services that originate from Oxford (including services that stop to collect passengers as an interim stop on their journey) was generated. From this list, one service was randomly selected (e.g., 10:15 Oxford to London Paddington) and a sampling plan was created around this. Based on the rules specified above, this would result in a number of trains being selected that operate on the Oxford to London Paddington route, including the 10:15 service out of Oxford.

Step 4: TOC coverage of first 140 shifts reviewed

The 140 shifts that were allocated at random were reviewed to ensure they provided a good coverage of areas and TOCs. It was noted that the random allocation of stations and

[3] These timings were used as an indicator. Based on the randomly selected service some shifts started before 7am and some finished after 7pm. Also based on train schedules and journey length some shifts were shorter than six hours in total and some were longer.

services had overrepresented Thameslink services (based on the number of services offered by this TOC) whilst underrepresenting others (specifically London and South East TOCs: Southeastern, Southern and Gatwick Express). The shift allocation was, therefore, examined to identify areas where reallocations could be made to provide better coverage.

In total, six shifts were changed as a result of this analysis:

- A shift covering Haywards Heath was already split between Thameslink and Southeastern services. Services covered were amended to only cover Southeastern services (at similar times to those initially selected).
- Five other Thameslink shifts were switched to cover two Southern based shifts, one Gatwick express shift and two Southeastern shifts. In all cases the same 'start' station was utilised as selected as part of the random selection process but the closest departing service on the alternative TOC was selected to replace the Thameslink services randomly selected (such that shifts still covered the same route, originated from the same station and were conducted at the same time of day/day of the week).

One other shift was amended at this stage as it included Transport for Wales services that were not part of the requirement for this survey. These services were replaced with similar West Midlands Rail services running along the same Telford to Shrewsbury route on the same day of week/times of day.

Step 5: Route-based selection of boost shifts to achieve robust TOC coverage

Further analysis of the 140 allocated shifts was undertaken to identify TOCs for which the number of completes was likely to be below a minimum threshold of 100[4]. The 20 boost shifts were then allocated using the following process.

Based on an estimated 50-70 completed interviews per shift[5], it was determined that each TOC should be covered by at least three shifts[6]. Boost shifts were allocated to TOCs with fewer than three shifts resulting from the regional allocation undertaken in Steps 2 and 3.

LENNON data covering a series of baseline periods from Summer 2022[7] was used to estimate the average number of passenger journeys undertaken in a typical week for each route within each DfT-contacted TOC. For each TOC where boosts were required, the route coverage provided within the first 140 shift allocation was examined. Routes that were not already covered were stratified based on the number of journeys made from that station (using the LENNON data described above), and a random start route selected.

---

4 It was agreed between DfT and Savanta that n=100 should be the minimum base size for reporting.

5 This estimate was based on Savanta's established experience from conducting on-train research across a range of clients - results of which are not held in the public domain but does tie in with actual completion rates seen from this survey.

6 It should be noted that the initial sampling was based on region. Therefore, it was possible for shifts to cover more than one TOC where they cover the same route etc. Where this occurred, estimated numbers were based on the number of trains covered by a TOC within these partial shifts.

7 Due to the time constraints required to conduct this element of the sampling this was the most recent data believed to be readily available in a useable format. Whilst actual journey numbers per route may have changed by the time the research was conducted this was not felt to have impacted the actual routes covered by TOCs or the relative size of these routes (with size only being used to stratify sampling and ensure a mix of busier and less busy routes).

This route was the first to be covered as part of the boost. Then, assuming more than one boost shift was required, every nth route was selected to provide a random sample of routes covered within that TOC.

For example, if two boost shifts were needed on a TOC that already had one route covered as part of the first 140 shift allocation and the TOC covered 41 routes, the boost shifts would be selected as follows:

- The route already covered within the original 140 shift allocation was removed to leave 40 routes
- These 40 routes were stratified based on the number of passengers journeys (most to fewest), using the LENNON data described above
- A random start route was chosen from 1 to 20 and then n+20 applied to choose the second route to be covered

Overall, it was decided that the total 20 TOC shifts should be split to include 10 morning and 10 evening shifts. These shifts were then allocated across 15 weekday and 5 weekend shifts. Once the 20 routes had been identified, fieldwork shifts were generated to cover this profile. For example, if it was decided to conduct a morning shift on the Oxford to Paddington route, a station on this route was chosen with a departure time around 7am and a six-hour shift constructed from this start point (with journeys lasting c.30 minutes in length and six trains being covered in each shift).

The resultant TOCs and regions covered once these additional shifts were devised are detailed in the tables below.

Table 1 shows the number of fieldwork shifts by TOC, and the average number of completes achieved per shift within each TOC.

Table 1. Number of shifts and completes per TOC

TOC	Completes	Number of shifts	Average completes per shift
Total	8132	161[8]	51
Avanti West Coast	256	4	64
C2c	153	4	38
Chiltern Railways	211	4	53
CrossCountry	252	4	63
East Midlands Railway	460	5	92
Gatwick Express	220	5	44
Great Northern	153	4	38
Great Western Railway	1085	15	72
Greater Anglia	378	8	47
LNER	249	4	62
London North Western	246	4	62
Northern	705	18	39
South Western Railway	1094	20	55
Southeastern	646	18	36

[8] Due to lower than anticipated weekend numbers on Northern routes an additional shift covering York to Leeds was added to make 161 shifts (this shift was selected using the same method outlined in Step 5).

Southern	589	14	42
Thameslink	700	17	41
TransPennine Express	200	5	40
West Midlands Railway	535	9	59

Table 2 shows the number of completes by (a) region where rail journey started and (b) region where respondent lives. This information was provided by the respondent during the survey.

Table 2. Number of completes per region

TOC	Region where rail journey started	Region where respondent lives
Total	8132	8132
Northern Ireland	-	15
Scotland	82	97
North West	438	473
North East	146	170
Yorkshire and Humberside	598	501
Wales	26	47
West Midlands	684	573
East Midlands	331	525
South West	695	889
South East	1742	1850
East of England	814	589
London	2512	1582
International	6	-
Other	-	228
Prefer not to answer	-	252
No answer/Prefer not to answer	58	341

A2. At which station did you start your journey? I2. Which region do you live in?

3. Fieldwork

Fieldwork took place over a period of five weeks, from 20th February – 26th March 2023 inclusive. During this period 161 fieldwork shifts were undertaken, each lasting approximately 6 hours. During each shift, fieldworkers were responsible for the distribution of questionnaires to passengers on specific train services, determined during the sampling process (as outlined in the previous chapter).

All rail passengers on sampled train services were asked if they were willing to participate in the research. To maximise response and to be as inclusive as possible, respondents were offered several ways in which they could complete the survey:

- A paper-and-pen self-completion questionnaire that could be completed and handed back to the fieldworker or returned in a pre-paid envelope
- Online via a link provided as a QR code
- Telephone completion was also offered, though there were no requests for a follow-up telephone interview during the fieldwork period.

In exceptional cases, the fieldworker could also assist the respondent in completing the questionnaire.

There was a total of 8,132 completed questionnaires, of which 6,798 were completed via paper-and-pen self-completion and 1,334 were completed online via the QR code provided.

Table 3 shows the number of completes by TOC:

Table 3. Completes by Train Operating Company and completion method

	Total	Paper	Online
Unweighted Total	8,132	6,798	1,334
Total	8,132	6,736	1,396
Avanti West Coast	268	243	25
C2c	179	133	45
Chiltern Railways	209	163	46
Cross Country	184	176	8
East Midlands Railway	387	335	51
Gatwick Express	246	207	39

Great Northern	182	172	10
Great Western Railway	1,012	864	149
Greater Anglia	478	337	141
LNER	254	221	33
London North Western	236	230	6
Northern	629	490	139
TransPennine Express	177	141	36
South Western Railway	1,156	927	229
Southeastern	673	514	159
Southern	644	531	113
Thameslink	811	704	107
West Midlands Railway	406	347	59

Table 4 shows the number of completes by Region where the respondent's rail journey started:

Table 4. Completes by Region where rail journey started (A2) and completion method

	Total	Paper	Online
Total	8132	6798	1334
Scotland	82	75	7
North West	438	334	104
North East	146	111	35
Yorkshire And The Humber	598	519	79
Wales	26	23	3
West Midlands	684	585	99
East Midlands	331	304	27
South West	695	647	48
South East	1742	1417	325
East	814	667	147
London	2512	2055	457
International	6	6	0
No answer	58	55	3

A2. At which station did you start your journey?

Table 5 shows the number of completes by Region where the respondent lives:

Table 5. Completes by Region where respondent lives (I2) and completion method

	Total	Paper	Online
Total	8132	6798	1334
Northern Ireland	15	12	3
Scotland	97	88	9
North West	473	375	98
North East	170	137	33
Yorkshire and Humberside	501	431	70
Wales	47	43	4
West Midlands	573	482	91

East Midlands	525	469	56
South West	889	785	104
South East	1850	1470	380
East of England	589	470	119
London	1582	1296	286
Other	228	194	34
Prefer not to answer	252	205	47
No answer	341	341	0

I2. Which region do you live in?

Research was conducted on rail services operated by DfT-contracted Train Operating Companies across 161 shifts. Train Operating Companies were contacted prior to fieldwork to obtain permission for interviewing on trains, and to request passes for travel and/or letters of authority to allow this work to take place.

Each of the 161 shifts lasted around six hours and included coverage of a number of trains on the same route: typically, four to six train services were covered in each shift, although some shifts had as few as two trains and some as many as eight depending on the length and frequency of the services.

The table below shows an example of trains travelled on in a single shift.

Table 6. Example fieldwork shift

Day	TOC	Station Board	Time	Station Alight	Time
Saturday	South Western Railway	Clapham Junction	11:57	Basingstoke	12:36
		Basingstoke	12:54	Clapham Junction	13:57
		Clapham Junction	14:11	Basingstoke	14:48
		Basingstoke	14:57	Clapham Junction	15:36
		Clapham Junction	15:57	Basingstoke	16:36
		Basingstoke	16:54	Clapham Junction	17:57

Shifts were designed to allow fieldworkers the opportunity to cover a number of trains within a six-hour shift and to end the shift at the station they started. For this reason, trains covered within shifts did not always cover the entirety of a route (i.e., they did not all start and finish at the initial origin and final destination stations of particular routes, but often made up a sub-part of a route).

Throughout each shift, researchers walked the length of the train (or changed carriages at station stops) and approached all rail passengers they encountered on the train during each shift. Each passenger was asked the reason for their journey (commuting, business or leisure travel) and this information, plus their observable age and gender, was recorded on “Count Sheets”. This information was used for a non-response bias adjustment (see Weighting section).

Rail passengers were then asked if they were willing to participate in the research and, if so, were given the option of filling out a self-completion questionnaire (to be handed back to the researcher) or taking a QR code to access the survey online. If they chose to take a

QR code, they were provided with a specific train code to enter so that their responses could be tied to a specific fieldwork shift.

Where rail passengers were on very short journeys but did not want to complete the survey online, a pre-paid envelope was provided so they could return the questionnaire by post.

In exceptional cases, there was also the facility for researchers to assist rail passengers to complete the survey on the train.

It was also possible to complete the survey via a telephone interview. However, no passengers requested this option.

4. Weighting

Scaling weights were calculated by comparing overall proportions of the samples achieved per region with the proportions of operating journeys allocated to each region in LENNON data covering June to October 2022 (drawing on the same data used to inform sampling). This adjustment was used to ensure that any differentials in response rates, routes covering multiple regions and the TOC-led boost shifts were adjusted for to ensure accurate representation by region.

A non-response adjustment was also applied to account for differences in the overall profile of rail passengers observed during fieldwork and the profile achieved in the sample. Fieldworkers used count sheets to record data about respondents who took questionnaires or QR codes, and those who refused to participate. Categories recorded were: journey purpose (commuter, business, leisure), observed age bracket (under 35, 35-44, 45-64, 65+)[9], and observed gender (male, female).

See count sheet data below:

Table 7. Count sheet data

Category (observed)	Total	Percentage
16-34	10803	38%
35-44	8030	28%
45-64	6646	24%
Over 65	2709	10%
Total	28188	
Male	15868	52%
Female	14526	48%
Total	30394	
Commuter	8816	34%
Business	3925	15%
Leisure	13264	51%

[9] Age categories on the count sheets were slightly different to those in the questionnaire (36-45, 46-65, 66+), which aligned to railcard groupings. This has been noted but as the count sheets were observed, it is unlikely that this discrepancy would have a significant impact on the weighting outcomes.

Total	26005

As count sheet data was collected via fieldworker observation, there was a variance in completeness by category, with gender being generally easier to classify by observation than age and reason for travel. Fieldworkers were however instructed to provide an estimate of age where possible.

On review, commuter and business journey purpose counts, and counts for the middle two age categories were each combined into single categories[10], giving the following categories used in the final adjustment:

- Age: under 35, 35-64, 65+
- Gender: male, female
- Journey purpose: commuter/business, leisure

The details from the count sheets within each region were compared to the returned profiles within the data. Weights were then applied to account for any differences (i.e., the profile of the survey respondents was adjusted to match the recorded profile obtained via count sheets on trains themselves).

Table 8. Weights applied by region

Region	Commuter/ Business	Leisure	Male	Female	16-35	36-65	66+
East	0.43	0.57	0.51	0.49	0.42	0.49	0.09
East Midlands	0.51	0.49	0.53	0.47	0.42	0.50	0.08
London	0.58	0.42	0.52	0.48	0.36	0.54	0.10
North East	0.27	0.73	0.44	0.56	0.35	0.54	0.11
North West	0.56	0.44	0.51	0.49	0.37	0.55	0.08
South East	0.45	0.55	0.54	0.46	0.39	0.52	0.09
South West	0.30	0.69	0.49	0.51	0.32	0.52	0.16
West Midlands	0.61	0.39	0.59	0.41	0.51	0.41	0.08
Yorkshire and Humber	0.38	0.62	0.51	0.49	0.42	0.50	0.08

There is the potential for statistical bias to be introduced through human error when applying this count method, and these counts do not give us a perfect indication of the population profile of rail passengers. However, there is no other currently available data that would give as accurate a profile of passengers for each TOC, split by age, gender, and journey purpose.

The final dataset was weighted to reflect these passenger profiles within region, and a combination of the two adjustments (scaling for region journey proportion, and non-

[10] The age categorisation included in the count exercise was based on fieldworker observations and it was reported by interviewers that distinctions between these two categories were most difficult to make. It was, therefore, felt that combining them would reduce the risk of misallocation. In terms of journey purpose, Commuter and Business categories were combined to provide a distinction between leisure and work related travel and to reduce the size of weights applied at individual region level.

response bias adjustment) was achieved using a Random Iterative Method (RIM) weighting algorithm, detailed below.

Random Iterative Method weighting

The final dataset was weighted to reflect the count sheet passenger profiles within regions. A combination of the two adjustments (scaling for TOC journey proportion, and non-response bias adjustment) was achieved using a Random Iterative Method (RIM) weighting algorithm.

RIM weighting is a frequently used quantitative market research technique. It is used when sample data needs to be matched to a known profile amongst a number of characteristics, where there is no known relationship between these characteristics. The technique utilises an algorithm that allows for each characteristic to be weighted to the desired profile at the same time, whilst distorting each variable as little as possible. The RIM weighting algorithm proceeds through a number of iterations in order to match the set target values for all included variables.

Rim weighting works by what is known as an iterative target weighting process. Weights are iteratively adjusted for each case until the sample distribution matches the desired population for the variables that the data are being weighted on. For example, if we want to achieve a 40% female and 60% male weighted sample based on our count-sheet profiles, then weights for each observation are adjusted such that the weighted counts from our observations are 40% female and 60% male. Then, the algorithm adjusts the weights so that the weighed counts of our observations are in the right proportion for our age distribution. This will likely mean that the gender proportions are knocked out of balance with our desired (target) proportions, so the algorithm adjusts the weights again, iteratively. This process continues until all proportions of combinations of the characteristics that are being weighting to match our target "population" proportions.

Summary of size of weighting factors applied

A general rule of thumb in survey analysis is to keep weighting factors between 0.5 and 2 (unless there is strong justification for using more extreme weights), so that no individual response is treated as too important or reduced to the point of not contributing. The majority (>95%) of the individual weighting factors applied to this data were within this range.

Overall, individual respondents within the sample received weighting factors of between 0.48 and 3.46. Whilst this does create some high levels of upweighted data this impacted very few respondents, with only 83 out of the 8,132 respondents receiving a weight factor of 2 or higher.

Impact of weighting (effective sample size)

Weighting has an overall impact on the effective sample size at a total level and within individual sub-groups.

Table 9. Effective sample sizes after weighting

Category	Unweighted sample size	Effective sample size, after weighting
All respondents	8132	7382
Commuter	4370	3971
Business	2482	2245
Leisure	4105	3738
Under 26	1915	1757
26-45	3089	2797
46-65	2353	2167
66+	742	701
Male	3736	3434
Female	4185	3827
Avanti West Coast	256	222
c2c	153	146
Chiltern Railways	211	194
CrossCountry	252	232
East Midlands Railway	460	423
Gatwick Express	220	209
Great Northern	153	141
Great Western Railway	1085	992
Greater Anglia	378	358
LNER	249	215
London North Western	246	225
Northern Rail	705	604
South Western Railways	1094	1048
Southeastern	646	621
Southern Railways	589	565
Thameslink	700	651
TransPennine Express	200	164
West Midlands Trains	535	470

5. Confidence intervals

The sampling approach means that the result is not a simple random sample, which could only be achieved with a sample frame of every individual who intended to travel by rail during the fieldwork period.

The limited availability of data to produce the sample and the requirement to boost the sample for certain TOCs means that it also cannot be considered to be a perfectly constructed stratified/cluster sample, which would enable reliable calculation of confidence intervals.

To provide a rough indication of how the confidence limits for results vary by sample size and proportion, the table below shows what intervals would apply for a random sample. Due to the sample design, the intervals for this sample would be consistently a little larger than those shown here (although the exact intervals for this sample method cannot be calculated).

Confidence intervals are provided at a 95% confidence level and based on 10%/90%, 30%/70% and 50% of respondents giving a specific response (as indicated in the table below).

Table 10. Confidence interval

Indicative data cell	Sample size	Confidence interval (to one decimal place)		
		10%/90%	30%/70%	50%
All respondents	8132	+/- 0.7	+/- 1.0	+/- 1.1
50% of sample	4000	+/- 0.9	+/- 1.4	+/- 1.6
25% of sample	2000	+/- 1.3	+/- 2.0	+/- 2.2
Larger region	1500	+/- 1.5	+/- 2.3	+/- 2.5
Smallest region	150	+/- 4.8	+/- 7.3	+/- 8.0

Where differences between proportions are reported to be statistically significant in the report, this is also an indication based on an assumption of randomness in the sample. For this reason, care should be taken in interpreting statistically significant differences since the assumption of randomness is not met.

6. Research limitations

All survey research is subject to some form of non-response bias, whether this be driven by socio-demographic factors/preferences (e.g., older people being less likely to complete on-line surveys/sign up to be on on-line panels) or attitudinal issues (e.g., some people being more predisposed to participate than others).

The methodology chosen attempted to minimise non-response by providing a range of methods for completing the survey. In this way, most rail passengers on the train services within the sample plan had the opportunity to participate and could do so in a way that was most suitable for them.

The passenger profile count exercise highlighted the fact that some groups were underrepresented in the final dataset. Details for this are provided in the weighting section, however, in general terms there were lower response rates for those making business journeys and those aged 36-65. This is in line with passenger count profiling undertaken for the other recent research, conducted on behalf of DfT: "Rail strikes: Understanding the impact on passengers". Whilst this is accounted for by weighting the achieved profiles to match those from the profile counts, it is possible that there could be some differences between those who chose not to participate in the research.

The methodology was devised to deliver a representative sample of passenger journeys. Whilst this is a strength of the research (in that it matches most other data sources available within the rail industry (e.g., LENNON data, Commuter/Business/Leisure profiles used by TOCs to weight NRPS/Wavelength data etc.) it does mean that care is advised in interpreting the findings.

Frequent and infrequent travellers will be included in the research in line with their usage of rail services. For example, 10% of survey respondents equates to 10% of passengers travelling during the fieldwork period. As frequent rail users make more journeys and infrequent rail users make fewer journeys, the survey will not be representative of rail users at a population level where they over or under index in term of frequency of use.

Four percent (4%) of survey respondents were aged 71-79 years old. This means that, on any one train service, it would be expected that 4% of passengers were aged 71-79 years old. However, if these passengers travel less frequently than younger passengers then this does not equate to 4% of all rail users being aged 71-79. In fact, it is likely that the number of 71-79 year old passengers who made ‘at least one rail journey’ over a year would be higher than 4% as they are making less frequent journeys. Therefore, it is true to say 4%

of journeys were made by those aged 71-79 but not that, out of all individuals who made any rail journeys over the year 4% were aged 71-79. It is important to keep this in mind when interpreting the results.