



Rail Passenger Customer Experience Survey - Pilot report

25 April 2024



CONTENTS

1. Introduction	2
2. Objectives for the pilot report	3
3. Evaluating the feasibility and effectiveness of the survey methodology	4
3.1 Statistical analysis of meta data	5
3.2 Qualitative analysis of fieldworkers' experience.....	9
4. Details of each objective	11
4.1 Evaluation of the feasibility and effectiveness of the survey methodology	11
4.1.1 Sampling process	11
4.1.2 Permission process.....	13
4.1.3 Interview shifts.....	14
4.1.4 Recruitment process	15
4.1.5 Data collection methods.....	16
4.1.6 Data flow	17
4.1.7 Obtaining the required sample sizes.....	18
4.1.8 The weighting design and process	22
4.2 Assessment of the questionnaire, the journey picker tool, the footfall counts and other survey components	24
4.2.1 The questionnaire.....	24
4.2.2 The Journey Picker Tool (JPT)	24
4.2.3 Footfall counts.....	25
4.2.4 The interventions.....	26
4.3 Identification of drivers and barriers to the survey's performance	26
4.4 Evaluation of the degree to which the survey design will achieve the policy and research objectives	28
4.5 Recommendations for future rail customer experience surveys.....	29
5. Quality standards & accreditation	31
6. Other details	32

1. Introduction

In summer 2022, the rail industry commissioned a method review with the objective to identify the optimal methodological approach to deliver an ongoing customer experience survey amongst rail passengers in Britain.

Field trials were conducted from April to June 2023 using the top methodologies that emerged from the method review. A method recommendation report was produced with detailed findings from the trials and a recommended approach for the pilot, as well as the future continuous survey.

Following the method review field trials, further tests in the field and pilot were conducted as follows:

1. Response rate experiments in September and October 2023 to measure further approaches of how the response rate could be improved. These included:
 - a. Removing instruction to complete at end of journey.
 - b. A shorter 5-6 minute survey.
 - c. User experience (UX) modifications.
 - d. Completion incentive.
2. Pilot plus a soft launch to the pilot to run the survey as intended in the future.
 - a. The smaller soft launch was conducted during rail period 9 in 2023 and comprised 39 shifts with 828 completed questionnaires.
 - b. The pilot was run during rail period 10 2023/24 but was slightly extended beyond the rail period to make up for no fieldwork taking place during a few days over Christmas and the New Year and for some disruptions; in the pilot we conducted 368 shifts with 7,753 valid completed questionnaires across all train operating companies (TOCs) in Britain.
3. Short survey trials in February 2024 to test further questionnaire modifications (a shorter modulated questionnaire of 8-9 minutes); this comprised of 15 shifts that were selected from the pilot survey based upon their higher footfall. The fieldwork compared how the modified shorter questionnaire compared to the longer 'standard' questionnaire by comparing the equivalent shifts from the pilot i.e. does the modified questionnaire have any positive impact on response rate?

For the response rate experiments and the short survey trials, separate standalone reports have been published.

The following document presents the key procedural findings of the pilot under bullet 2. A separate weighting guide and technical report relating to the pilot survey have been published for the pilot.

2. Objectives for the pilot report

The objectives of the pilot report are the following:

1. Evaluate the feasibility and effectiveness/performance of the selected survey methodology.
2. Assess the questionnaire, the interventions, the Journey Picker Tool, the Footfall Count and other survey components.
3. Identify challenges/opportunities for improving the survey and its performance.
4. Evaluate the degree to which the survey design will achieve the policy and research objectives.
5. Identify ways to improve the value for money.
6. Provide recommendations for future rail customer experience surveys.

3. Evaluating the feasibility and effectiveness of the survey methodology

For the pilot we used the methodology that emerged as superior from the method review field trials. The methodology that was chosen based upon the field trials was intercepts on board trains, rather than at station. Shifts were scheduled for six hours at a time rather than three hours, and a priority was given to electronic data collection except in exceptional circumstances.

As such, passengers were approached by fieldworkers on board trains and asked whether they would like to participate in the survey. As a response option they were offered to scan a QR code that would take them right to the survey online; or to provide their email address to which a survey link was sent pretty much immediately. If passengers were willing to participate in the survey but did not want to complete it online, they were offered a paper questionnaire with a self addressed envelope that could be posted back to BVA BDRC's field team. 10 paper questionnaires were provided for each shift (numbers were limited due to their bulk) and were used for on board shifts only. Once these questionnaires were handed out, those requesting paper were considered as refusals and unable to take part.

At station shifts were included in the methodology as a resolution for handling recruitment during delays or disruptions. A detailed protocol was shared with fieldworkers to follow in the case of delays or disruptions, with at station to be used while waiting for the next appropriate train, or for the full length of the shift if the disruption is ongoing. The detail of these measures followed by fieldworkers is conveyed in a separate technical report. Due to disruptions six shifts were completely conducted at station rather than on board while 63 shifts were partly conducted at station and partly conducted on board, also because of disruptions.

Prior to launching the pilot, a short two week soft launch was run across all TOCs using a smaller number of shifts (42 shifts). This was run to test the survey methodology was running correctly in advance of the large scale pilot survey. In most cases, analysis of the soft launch is not considered in this report.

Fieldwork for the pilot was initially planned to run during rail period 10 (RP10) i.e. from 10 December 2023 to 6 January 2024. The start was slightly delayed and subsequently fieldwork commenced on 12 December. It was extended to 17 January, because of disruption, due to no fieldwork taking place on days over the Christmas and New year period, and fieldworker cancellations due to personal reasons which meant replacement shifts after the end of the rail period were required. The cut off for paper questionnaires and online responses was 22 January. No fieldwork took place on 24-26 December and on 31 December and 1 January. In total 368 shifts were conducted, and we achieved 7,753 validated complete survey questionnaires.

Overall, including the soft launch 12,802 partially completed survey questionnaires were received. Partially completed responses refer to any surveys where the respondent did not complete the survey to the end and were therefore not incorporated into the final total of validated responses. Soft launch partially completed surveys were included here as departure time was used to determine the period of completion and whether the respondent was in the soft launch or pilot. Therefore as not all partially completed responses had a departure time because the passenger had not progressed to this point, separation of partial completes into the soft launch or Pilot was not possible and so a combined total of partially completed responses of the soft launch plus the pilot is provided.

3.1 Statistical analysis of meta data

Focusing on the pilot, the overall response rate was 28%, with 27,593 passengers recruited and 7,753 completing the survey. This resulted in 21.1 completed questionnaires per six hour shift, compared to 7.5 per three hour shifts in the method review field trials. The pilot shifts were twice as long, but achieve almost three times as many completed questionnaires, which is a significant increase.

Table 1: Number of passengers recruited, who completed surveys and the response rate by each respective response option and split by on board and at station recruitment.

	FINAL UPDATE - PILOT		
Number of shifts	-	-	368
Location of recruitment	On board	At station	Total
Total recruited	26546	1047	27593
QR	19542	794	20336
Email	5481	237	5718
Paper	1379	0	1379
Those who refused online and fieldworker ran out of paper (refusals)	144	16	160
Total completed questionnaires	7538	215	7753
QR	5767	160	5927
Email	1235	55	1290
Paper	536	0	536
Total response rate	28%	21%	28%
QR response rate	30%	20%	29%
Email Response rate	23%	23%	23%
Paper response rate	39%	-	39%
Completed questionnaires per shift	-	-	21.1

The response rate and completed questionnaires per shift by TOC are shown in the following table. Broadly, completed questionnaires are higher for long distance TOCs, with response rate particularly low for Heathrow Express, London Overground and the Elizabeth line. This is consistent with the method review field trials. Anecdotal feedback from fieldworkers is that there are a lot of tourists on Heathrow Express, many of whom do not speak English very well, and also a lot of business travellers who are less interested in completing the survey (even though they may agree to participate).

Table 2: Breakdown of key metrics by TOC – total number of shifts, passengers recruited, completed questionnaires, recruited passengers per shift, the response rate, and number of completed questionnaires per shift.

	Number of shifts	Total recruited	Completed questionnaires	Recruits per shift	Response rate	Completed questionnaires per shift
Avanti West Coast	12	936	328	78	35%	27
c2c	14	667	210	48	31%	15
Chiltern Railways	10	853	320	85	38%	32
CrossCountry	12	690	188	58	27%	16
Elizabeth line	22	1537	324	70	21%	15
EMR - East Midlands Railway	14	858	216	61	25%	15
Gatwick Express	4	357	99	89	28%	25
Grand Central	3	156	39	52	25%	13
Great Northern	11	876	302	80	34%	27
Greater Anglia	20	1405	413	70	29%	21
GWR - Great Western Railway	19	2385	642	126	27%	34
Heathrow Express	4	324	52	81	16%	13
Hull Trains	3	182	87	61	48%	29
LNER – London North Eastern Railway	11	688	287	63	42%	26
London Overground	26	1606	322	62	20%	12
Lumo	3	221	77	74	35%	26
Merseyrail	9	464	162	52	35%	18
Northern	20	1088	355	54	33%	18
ScotRail	6	331	93	55	28%	16
South Western Railway	25	1924	532	77	28%	21
Southeastern	28	1558	400	56	26%	14
Southern	21	1408	359	67	25%	17

	Number of shifts	Total recruited	Completed questionnaires	Recruits per shift	Response rate	Completed questionnaires per shift
Thameslink	26	2264	612	87	27%	24
TransPennine Express	13	845	325	65	38%	25
Transport for Wales Rail	5	398	135	80	34%	27
West Midlands Railway	13	1619	657	125	41%	51
London Northwestern Railway	14	1012	217	72	21%	16

3.2 Qualitative analysis of fieldworkers' experience

Overall, the fieldworkers had a positive experience with the project, with most passengers being engaged and willing to participate in the survey. They found that the high visibility jackets helped to separate them from passengers and made it easier to approach people. They also found that explaining the purpose of the survey in a timely and clear manner was important, as many passengers were happy to engage when they understood the benefits to their journey in the future. They also found that having something for passengers to occupy their time made them more willing to participate.

However, there were some challenges that they faced.

One of the most significant challenges was delays and cancellations, which affected the recruitment process. Passengers were often frustrated and anxious during these situations, making it more challenging to approach them for the survey and made it difficult to continue with their shifts.

The fieldworkers found that the Elizabeth line was the most challenging Train Operating Company to work on, as it was similar to trying to interview on board a tube train during rush hour. Additionally, some trains did not have toilets on board, which meant that the fieldworkers had to use the facilities at stations, which could be time consuming.

Footfall counts were conducted to gather information on the age and gender of the passengers on the train to be used for weighting (see section 5.2.3 for further detail). The fieldworkers felt that the 10 minute footfall counts were not sufficient to gather the required information accurately and only provided a rough estimate. They felt that the footfall counts needed to be longer to gather the required information accurately. Logistical challenges sometimes made footfall and recruitment difficult, such as two trains conjoined together, meaning the fieldworker was unable to walk along the whole train to conduct the footfall.

Despite being provided with permission letters or passes in advance of their shifts, there were mixed experiences from fieldworkers with regards to train staff allowing the on board research to go ahead on the day of fieldwork. In most scenarios, the fieldwork could go ahead as normal. Some train managers were aware and supportive of the research and helped fieldworkers to navigate the trains and stations. While other examples shared included a limited understanding of the research, causing subsequent delays to fieldwork starting.

In terms of the methods used for recruiting passengers, some passengers were not comfortable giving out their personal information via the email recruitment method, but after explaining the research and reasons for collecting emails, some seemed more willing to take part.

Despite these challenges, the fieldworkers found the project to be a very enjoyable one to work on.

4. Details of each objective

The following sections outline the details of each objective.

4.1 Evaluation of the feasibility and effectiveness of the survey methodology

In the following section we will look at the feasibility and effectiveness of the survey methodology across the whole network. This will cover the sampling process, permissions process, the interview process, the recruitment process, data collection methods, data flow, obtaining the required sample sizes and the weighting design and process.

4.1.1 Sampling process

To develop the sampling frame an estimate is made of the number of passengers for each service. To do this a data extraction was undertaken involving selecting all train services. The data extraction involves selecting all train services that run during the RP10 period from the Rail Delivery Group (RDG) Electronic timetable, including data for each service on the origin and destination stations, TOC, departure and arrival times and date. MOIRA data is then used to estimate passenger numbers per service. This is an industry model used for analysing demand and revenue changes from rail timetable changes. It aims to predict the number of people who will travel on each train service. MOIRA data on the estimated number of passengers is appended to each train service, either directly if the train service exists in MOIRA or from a model created by BVA to estimate passenger numbers if the train service does not exist in MOIRA. About 80% of train services exist in MOIRA. Data from MOIRA is based upon pre pandemic modelling and ideally should be refreshed with more up to date data, but it is viewed as the best source available for estimating passenger numbers on each scheduled train service.

A sample of train services is then selected for each TOC with a probability proportional to the estimated number of passengers on each service. Sufficient train services are selected to generate a minimum sample size for each TOC with an allowance for a surplus. The target sample sizes can be found in table 3, and these targets were used with the aim of achieving a minimum TOC level analysis and an overall analysis. Services that start after 16:00 are replaced with the earliest service on the same TOC and route after 16:00 (this may be the same train), to enable a full six hour fieldworker shift to be undertaken.

The process then identifies a return journey to the origin station, a second outward journey and a second return journey to the origin station if the elapse time of the first return trip is less than three hours.

Table 3: Sample size target for each TOC for the pilot prior to fieldwork and the sample size that was achieved.

	Sample size Target	Completed questionnaires	Difference
Avanti West Coast	200	328	128
c2c	200	210	10
Chiltern Railways	200	320	120
CrossCountry	200	188	-12
Elizabeth line	300	324	24
EMR - East Midlands Railway	200	216	16
Gatwick Express	50	99	49
Grand Central	50	39	-11
Great Northern	200	302	102
Greater Anglia	300	413	113
GWR - Great Western Railway	300	642	342
Heathrow Express	50	52	2
Hull Trains	50	87	37
LNER – London North Eastern Railway	200	287	87
London Overground	300	322	22
Lumo	50	77	27
Merseyrail	100	162	62
Northern	300	355	55
ScotRail	100	93	-7
South Western Railway	400	532	132
Southeastern	400	400	0
Southern	350	359	9
Thameslink	400	612	212
TransPennine Express	200	325	125
Transport for Wales Rail	100	135	35
West Midlands Railway	200	657	457
London Northwestern Railway	200	217	17
Total	5600	7753	2153

The targeted sample sizes were achieved for all TOCs bar three where they were slightly missed. These were CrossCountry, Grand Central and ScotRail.

The reasons for slightly missing the targets for these three TOCs are as follows:

1. CrossCountry and ScotRail: Services for some shifts were affected by disruption, or no longer ran; when realising this, there was not sufficient time to book additional shifts.
2. Grand Central: Three shifts for Grand Central were in the sample plan with an expected number of completed questionnaires per shift of 15; when realising the

target would not be met, there was not sufficient time to book an additional shift.

These were mainly shortcomings in the process of monitoring meeting field objectives. This is something that could be adjusted for through preliminary stages of a future survey.

In the vast majority of cases, the achieved sample size is significantly over the target, highlighting, that fewer shifts will be required in a future survey for a similar sample target.

4.1.2 Permission process

To acquire permissions for the pilot survey to conduct surveys on board trains and at stations, the Department for Transport (DfT) initially contacted the respective TOCs to inform them of the survey. BVA BDRC then contacted the TOCs individually with any relevant information required. Permissions were required for both on board and at station.

The process of acquiring permissions was not consistent. Seventeen of the TOCs were able to provide permissions quickly in the form of either a letter or a pass. The remaining TOCs proved more complex to acquire permissions from. This was due to each TOC having different requirements. Some required names of individuals or exact journey dates or needed to ensure there were no clashes with other fieldwork before permitting the authorisation.

Permissions were also required for station shifts during delays. This meant understanding which station was owned by which TOC in order to obtain the correct permissions for the shift in case of a delay. It led to many permission letters being given to each fieldworker for their shifts.

For any Network Rail stations, there was a new process introduced for the pilot survey. This allowed for one permission letter for all stations, which was an improvement on the previous requirement for a letter for each station.

Despite achieving all of the permissions in time for fieldwork, there were still issues of fieldworkers being denied access or remaining on trains even when permissions were in place. This is because information was not always relayed to train managers and staff. This meant some fieldworkers faced being removed from trains mid journey, creating unnecessary disruption and impacting the data collection process.

Our recommendation going forward would be to seek a simplified process. This would be seeking one permission from the DfT or industry body, such as RDG, which would apply to all TOCs. This would remove the time spent contacting each TOC individually and would make the permissions required consistent across the TOCs.

4.1.3 Interview shifts

Overall, 44% of the shifts lasted between 5.5 and 6.5 hours. For the remaining, most shifts lasted between 4.5 and 5.5 hours.

At a TOC level, for nine TOCs the mean time spent recruiting was between four and five hours – all the others were averaging over five hours. Generally long distance TOCs are those recording shorter recruiting time on shifts. This is due to their limited number stations and train numbers, meaning fewer passengers are available and making it difficult for the interviewer to spend the full six hours recruiting.

The following tables show the average hours spent recruiting per shifts, for each TOC.

Table 4: length of a shift time in five groupings and the number of shifts within each of these groupings

	Count of shifts
Less than 3 hours	20
3 hours	17
4 hours	78
5 hours	210
6 hours or more	43

Table 5: mean length of time spent recruiting by each TOC (incl. time spent on passenger counts)

	Mean
Avanti West Coast	4:58
c2c	4:20
Chiltern Railway	5:43
CrossCountry	6:01
Elizabeth line	5:07
EMR - East Midlands Railway	4:46
Gatwick Express	5:37
Grand Central	4:42
Great Northern	4:44
Greater Anglia	5:24
GWR - Great Western Railway	5:38
Heathrow Express	5:55
Hull Trains	5:56
LNER – London North Eastern Railway	4:30
London Overground	5:03
Lumo	5:43
Merseyrail	4:40
Northern	5:17
ScotRail	4:11
South Western Railway	5:13
Southeastern	5:08
Southern	5:01
Thameslink	5:22
TransPennine Express	5:18
Transport for Wales Rail	5:14
West Midlands Trains	4:57

4.1.4 Recruitment process

The following section provides qualitative evidence from the fieldworkers on their experience of the recruitment process.

As part of the recruitment process, fieldworkers were asked to engage with passengers on board specific trains they were assigned to. When there were disruptions or cancellations, passengers would resort to a station shift, aligning with the delays and cancellations procedure provided as part of their manual.

The fieldworkers found that engaging with passengers was generally easy, with most people willing to participate in the survey. The fieldworkers also offered multiple options to complete the survey, which gave passengers a choice and did not exclude anyone willing to take part. Some were hesitant at first, particularly with regards to sharing personal information such as email addresses, and the QR did not always work in moments of poor signal. Also, fieldworkers reported that not all days were as easy to recruit on, with Friday or Saturday nights proving trickier to handle when people had been drinking.

In terms of collecting data, fieldworkers reported that it was generally easier to recruit on board trains rather than at stations. They had a captive audience on trains, which made it easier to engage with passengers. Having multiple options to complete the survey was helpful, and working on board the train was generally not a problem unless the train was very busy. In contrast, stations were noisy and crowded, which made it challenging to approach passengers. They also experienced higher rates of refusals at stations.

Delays and cancellations were a common challenge for the fieldworkers during data collection. Below we will outline the key challenges that should be addressed in any future survey fieldworker manual:

- When trains were delayed or cancelled, it often resulted in passengers being frustrated, anxious, or in a hurry, which made it more challenging to approach them for the survey.
- In some cases, passengers were less willing to engage in the survey when there were train delays or cancellations.
- Additionally, dealing with delays and cancellations could be stressful and confusing for the fieldworkers, even when procedures were outlined in the recruitment manual. Some fieldworkers encountered engineering faults or delays that affected all trains in a particular region, which made it difficult to get advice or continue with their shifts.
- Moreover, some fieldworkers found it difficult to determine at which point to start interviewing on the platform when trains were delayed. Multiple delays could result in a loss of recruitment time, and it was challenging to recruit passengers who were already frustrated and in a hurry.

Overall, delays and cancellations presented a range of challenges for the fieldworkers, including dealing with frustrated and anxious passengers, unclear procedures, and a loss of recruitment time. They suggested having clearer instructions on what to do in all given situations.

4.1.5 Data collection methods

The following section provides qualitative evidence from the fieldworkers on the data collection methods used to recruit respondents, and how this aligns with our findings.

The fieldworkers found that the QR code was the most efficient method for data collection, as it was quick and easy to use, and passengers could directly access the survey questionnaire on their smartphones by scanning the QR. This aligns with figures reported earlier at 3.1, where QR achieved the highest number of recruits out of all the methods.

However, some passengers preferred paper questionnaires, especially older passengers who were not comfortable with online options. This was the least used method, though it was limited to 10 paper copies per shift.

Email collection was used by around a fifth of respondents. Fieldworker feedback showed that this was time consuming, and the process was often prone to errors, such as misspelling or duplicating email addresses. This meant that the fieldworkers had to spend time correcting these. Additionally, some passengers were hesitant to provide their email addresses due to privacy concerns, which further slowed down the process.

Overall, the QR code was the preferred method, but having paper questionnaires as an option was important to ensure that no passengers were excluded. The fieldworkers suggested having more paper questionnaires available for better response rates and to capture a wider demographic of passengers.

4.1.6 Data flow

As part of the pilot, the survey data was shared via an application programming interface (API) with GBRTT for their data portal, using the survey platform (Forsta) Decipher where the survey is managed. This API transfer included partial and qualified completed questionnaires, the sharing of verbatims that had been cleaned of any personal identifiable information, and the weight once this was created. The process of delivering the API connection required GBRTT to have access to the survey, which is managed by Decipher, and once the permissions were granted by BVA BDRC this allowed the data transfer in a seamless ongoing process.

However, the data flow to GBRTT experienced several issues, which proved challenging to resolve.

- Verbatim data did not always show in the data downloaded by GBRTT when refreshing the data on their side.
- The parameters for the API were not consistent. The transfer initially used a utf-8 format, which later changed to a window-1252 format which created additional work for GBRTT to debug and backfill the data to make the portal work.

Therefore, the API data flow route was ultimately not successful, and ensuring a robust data flow infrastructure is necessary in any future survey.

4.1.7 Obtaining the required sample sizes

At an overall level, the requirement per a rail period was to obtain 5,600 completed questionnaires. As conveyed in section 3.1, the targeted sample size was achieved overall. At a TOC level the target samples were achieved for the majority and were significantly over for some. In only three TOCs was the target sample not met – CrossCountry, Grand Central and ScotRail. This highlighted that obtaining the required sample size at a TOC level would be achievable with fewer shifts overall but targeting a greater number of shifts towards the three TOCs where completed questionnaires were lower would be needed in order to meet the targets.

Below in table 6, the data conveys the number of completed questionnaires by key metrics. The data shows there is sufficient sample size that meets confidence levels to allow for analysis at a national level across journey purpose, type of TOC, sector, length of journey and age. Focusing on protected characteristics, ethnicity achieves a usable sample size for White and Mixed/Multiple ethnic groups, though this is below 30 for those categorised as Other ethnic group. There is a sufficient sample to conduct analysis on some disability characteristics, but some sample sizes were too low, e.g. dexterity, suggesting the method may not be able to capture all disabilities with sufficient sample.

Table 6: Number of completed questionnaires by journey purpose, type of TOC, sector that a TOC falls into, average length of the survey, age, ethnicity and disability.

	Number of completed questionnaires
Journey purpose	
Commuting	2308
Business	610
Leisure	4467
Type of TOC	
High Speed	615
Commute	5483
Airport	151
Interurban	2603
Long Distance	513
Non-franchised	203
Sector the TOC travels	
London and SE	5461
Regional	745
Long Distance	1547
Mean length of journey (Of those who answered JPT)	56 minutes

15 mins or less	774
More than 15 mins - 30 mins	1557
More than 30 mins - 45 mins	1491
More than 45 mins - 60 mins	975
More than 1 hour - 1.5 hours	1189
More than 1.5 hours - 2 hours	454
More than 2.5 hours - 3 hours	369
3 hours+	158
Age	
16-24	1445
25-34	1928
35-44	1386
45-54	1216
55-64	1024
65+	715
Ethnicity	
White	2376
Mixed/Multiple ethnic groups	383
Other ethnic group	26
Disability	
Yes - with disability	747
Vision	39
Hearing	82
Mobility	150
Dexterity	31
Learning or understanding or concentrating	47
Memory	22
Mental Health	290
Stamina or breathing or fatigue	81
Socially or behaviourally	169
Other	132

Focusing on Network Rail stations, 28% of these stations were captured as departure stations within the fieldwork. Below are the number of completed questionnaires for each station. For any forthcoming rail customer experience survey, boosts will be required to cater for any shortfalls for Network Rail stations. Particular focus could be on Glasgow Central, Guildford, London Charing Cross, and London Cannon Street.

Table 7: Number of completed questionnaires by each of the Network Rail stations.

	Number of completed questionnaires
Network rail stations	
LONDON KINGS CROSS	250
BIRMINGHAM NEW STREET	218
LONDON WATERLOO	186
LONDON LIVERPOOL STREET	182
LONDON EUSTON	168
LONDON VICTORIA	140
ST PANCRAS INTERNATIONAL	143
LEEDS	123
LIVERPOOL LIME STREET	119
EDINBURGH	110
LONDON BRIDGE	105
MANCHESTER PICCADILLY	94
LONDON PADDINGTON	71
BRISTOL TEMPLE MEADS	65
CLAPHAM JUNCTION	63
READING	50
GLASGOW CENTRAL	33
GUILDFORD	33
LONDON CHARING CROSS	29
LONDON CANNON STREET	16

Exploring sample size by day of the week and day part, Thursday is the most represented by number of completed questionnaires, while Monday has the least. The number of shifts per day of the week is likely to be a contributing factor with fewer shifts recorded on Mondays compared to other days of the week. For instance, 77 shifts fell on Thursday versus 33 shifts conducted on a Monday in RP10. Though other effects could be contributing towards this, such as a reflection of current travel patterns, but without an up to date source on travel patterns we cannot test this hypothesis.

Table 8: Number of completed questionnaires at an overall level by day of the week and then split by time of day.

Day of the week	Overall	Peak AM	Peak PM	Off-peak	Late
Monday	432	34	171	201	26
Tuesday	1176	128	313	629	106
Wednesday	1401	233	404	647	117
Thursday	1641	276	342	922	101
Friday	1203	137	243	718	105
Saturday	1283	n/a	n/a	n/a	n/a
Sunday	617	n/a	n/a	n/a	n/a

Peak AM refers to trains departing in the morning peak hours between 06:00 hrs to 08:59 hrs, for weekday only.

Peak PM refers to trains departing in the evening peak hours between 16:00 hrs to 18:59 hrs, for weekday only.

Off-peak refers to trains travelling in the hours between 09:00 hrs to 15:59 hrs or times recorded prior to Peak AM (0:01 hrs to 05:59 hrs)

Late refers to trains travelling from 19:00 hrs onwards until 23:59 hrs.

Saturday and Sunday are treated as a whole day respectively.

4.1.8 The weighting design and process

More detail on weighting will be in the ‘Rail Passenger Customer Experience Survey – Pilot: Technical report’ and the ‘Rail Passenger Customer Experience Survey – Pilot: Weighting guide’.

Analysis of the method review field trials data showed that the key satisfaction measures vary by day of the week, time of the day, by size of station and by demographics. Weights have therefore been applied for each TOC using the following:

1. Dayparts

The sampling process generates totals of estimated number of passengers for each TOC by section of the week that the train is travelling, also known as dayparts, so this can be used for weighting the data. These sections of the week (also referenced above at Table 8.) are split by Saturday, Sunday, and the weekdays are split into peak AM (06:00 hrs – 08:59 hrs), peak PM (16:00 hrs to 18:59), late (19:00 hrs onwards until 23:59 hrs) and off peak hours (09:00 hrs to 15:59 hrs or times recorded between 0:01 hrs to 05:59 hrs). MOIRA data is used to generate the dayparts figures.

2. Station size bands

Furthermore, the data was weighted by station size. Station size can be split into four station size band profiles using the MOIRA data. This is based on the estimated number of passengers boarding at each train service the TOC runs, which is then aggregated together by station and placed into station bands Passenger journey volume.

This is the number of passenger journeys being undertaken for each TOC which are known to vary considerably by Rail Period. LENNON¹/ORR data is used to generate passenger volumes. LENNON data is used as it is available for Rail Periods and can account for the variability by rail period. LENNON data provides up to date estimates of passenger numbers for TOCs included in the LENNON database. However, some TOCs are not included – London Overground, Merseyrail, Heathrow Express. In addition, for TOCs where substantial numbers of tickets are sold outside the channels that LENNON incorporates, that proportion of tickets will be missing. For these TOCs, a factor is applied to estimate the number of journeys made by

¹ Latest Earnings Nationally Networked Overnight is an application used by the rail industry. It provides data, such as ticket sales and franchise earnings, which helps them better understand how the rail network operates.

tickets sold outside the LENNON system. The factor was calculated by comparing ORR data for the last published period with LENNON data for the same period.

3. Age and gender

The data is also weighted by passenger gender and age. Gender and age proportions derive from the footfall counts.

In order to carry out the weighting some new variables needed to be created within the pilot data set (e.g. dayparts) and a decision had to be made what to do with respondents who, for example, under gender answered with 'Prefer not to say/Identify in another way'. Furthermore, station and TOC had to be confirmed manually for respondents that failed to find their journey in the journey picker tool (JPT) and entered their details manually (and not always correctly).

RIM weighting was used for all weights.

The initial weighting efficiency was 20% which is very low (a good weighting efficiency is typically 60% or higher). A significant contributor to the low weighting efficiency is bringing together data for all the TOCs. Some TOCs were oversampled to generate the requested minimum sample size. As a result, the remaining TOCs are undersampled and need large weights to ensure the TOC is represented at its proper level in the national results. This has the impact on the overall weighting efficiency.

To better understand the factors behind the overall weighting efficiency it was decided to analyse the weights in two stages:

1. From weighting each TOC individually.
2. From amalgamating all TOCs into the national dataset so that it is representative overall.

This helped us to better understand the weighing efficiency of each TOC and where extreme upweights and downweights had to be applied (as mentioned above). Given the need for both TOC level performance and overall national results it is important to understand how to address the extreme upweights and down weights. This was not something that could be corrected at the weighting stage, and we would recommend taking this into account at the sampling stage for any future survey applying this methodology.

To improve the weighting efficiency of individual TOCs we looked at where, for example, we could combine some weighting criteria together e.g. Saturday and Sunday or combining some age groups. We went through several rounds to improve the weighting efficiency of TOCs. In the end we decided not to weight one TOC which was Grand Central due to its low overall base size.

Ultimately, we ended up with an overall weighting efficiency of 53.6%. The majority of TOCs also had a weighting efficiency of over 50% (ranging from 22.8% to 84.5%). After

some discussion, it was agreed that a few respondents who had answered don't know on age or prefer not to say for gender were left unweighted on those variables.

The weighting took longer than expected, but now that the sampling process has been set up and applied, possible problems and solutions have been identified with the weighting strategy. Therefore the process of weighting for any future survey will be much quicker. The sample sizes by TOC, however, need to be discussed further to avoid any extreme TOC weights in the future.

4.2 Assessment of the questionnaire, the journey picker tool, the footfall counts and other survey components

4.2.1 The questionnaire

The questionnaire consisted of 183 questions across a core questionnaire, with three modularised elements. The modules were included to help reduce the length of the survey, dropouts and respondent fatigue. The modules were randomised on a least full basis. This means strategically directing respondents to answer modules that are under represented within the survey to ensure each of the modules are evenly represented. This resulted in a 33/33/33 split with Module 1 seen by 2,409 respondents, Module 2 seen by 2,405 respondents, and Module 3 seen by 2403 respondents.

On average (mean), respondents took 10 minutes 31 seconds to complete the survey. Including the soft launch, there were 12,802 partially completed surveys, while 7,753 completed questionnaires during the formal pilot period.

4.2.2 The Journey Picker Tool (JPT)

The Journey Picker Tool (JPT) that was produced for the earlier trials was used within the pilot survey. For the pilot, we introduced modifications to improve the efficiency and effectiveness of the tool. This included clearer instructions so that respondents were aware of what they needed to do on the tool. The instructions detailed what a one journey leg would be by providing examples and explaining what to do if they had changed trains. In addition, if they respondent could not return their journey details using the Journey Picker tool, they were offered a box to provide additional information. Formerly, the details on what to include here were limited to 'please enter the details of your journey'. For the pilot, we included a detailed description of how to rectify their failing on the JPT to see if the tool could pick their journey up, e.g. check the time of departure, and asked them to provide specific details relating to their journey if they could still not find their train journey. The aim of these changes was to minimise the number of respondents not able to return their journey details using the JPT by including multiple legs of their journey, or mis-entering incorrect departure information.

Over RP10, 7% (538) of respondents were unable to find their journey in the JPT. This is an improvement on the method review field trials where 14% were unable to locate their journey using the tool. The majority of the JPT failings fell on the London and South East TOCs, and a third of these were for leisure trips. Often journeys in London and South East included travel on more than one train company to reach a destination and therefore combining multiple legs increases in likelihood.

Table 9: Demographics of Respondents unable to find their journey in the JPT

	Failing JPT							
Demographics	Female	Male	16-24	25-34	35-44	45-54	55-64	65+
	49%	50%	17%	25%	17%	19%	14%	8%

Table 10: Journey purpose and type of route breakdown of those unable to find their journey using the JPT

	Failing JPT					
Journey type	Commuting	Business	Leisure	London and SE	Regional	Long Distance
	28%	9%	63%	73%	17%	10%

There are still areas for improvement for the journey picker tool to make sure it is able to support future rail customer experience surveys. It is important the tool departure information is kept up to date with the latest station information. For example, East Linton was a new TransPennine Express station opened during fieldwork and this was not picked up initially by the tool. Similarly, the tool was able to produce departure stations that were not stations in the industry data, such as Eurostar stations, this then impacted later stages such as weighting processes. It's important the data feed of stations matches those of the industry data to make the data output and cleaning as efficient as possible.

4.2.3 Footfall counts

Footfall counts were conducted at two points during the shift. A footfall count would last 10 minutes each and required fieldworkers to count how many adults aged 16 or more they could visibly see on board the whole train, splitting this by age and gender. This included any that were walking, standing or sitting. The first was conducted roughly 30 minutes into the shift, and the second was conducted three hours and 30 minutes into the shift.

From the footfall count, we found adults were 47% females and 53% males. There were 42% adults aged 16-34 years old, 45% aged 35-64 years old, and 12% aged 65+ years old. Overall, there were 54323 adults observed, and 152 adults on average observed per shift (based on 358 shifts with footfall counts).

4.2.4 The interventions

In the survey, a prize draw incentive of being able to win 1 of 5 x £200 Love2shop vouchers were used to help improve uptake. This was a finding established and brought forward from the response rate experiments. In total, across the fieldwork including the soft launch, 6865 respondents took part in the prize draw. The incentives were drawn once the 15th of every month was over, and the entry for respondents was determined based upon when respondents filled in the survey. We had 1594 respondents in the first draw, 4686 respondents in the second draw, and 585 respondents in the final draw. Of those participating, more females took part, and those aged 25-34 were most likely to take part out of all age categories.

Table 11: Demographics of those participating in the incentives

Demographics	% portfolio of those participating in the incentives							
	Female	Male	16-24	25-34	35-44	45-54	55-64	65+
	53%	46%	18%	26%	19%	16%	13%	8%

The drawing of the names was an easy process that required a randomisation rule in excel to identify each of the winners. Each time a winner was picked out, they were removed from the list and then next name was drawn. This was done until all five winners for that month were selected. These winners were then contacted via email using the railsurvey@bva-bdrc.com email address. The process of sending the first email was smooth. However, following up on responses took time. Often respondents reported the email going to their junk folder, and each draw had at least one respondent who did not respond and therefore a new name had to be selected and emailed. The process of arranging the incentive could be more streamlined, as Love2shop only allowed a maximum of £100 voucher per transaction, meaning each of the five winners had to be sent two vouchers each, and 10 transactions had to be made each draw by BVA BDRC.

4.3 Identification of drivers and barriers to the survey's performance

Following the method review field trials conducted earlier in the programme to establish the methodology to be used in the pilot, the following interventions were implemented to improve the number of completed questionnaires per shift and ultimately, value for money for the Rail Customer Experience Survey:

1. Removing instruction to complete the survey at the end of the journey (instruction both by the fieldworker as well as in any communication and the actual survey).

2. Shortening of the survey.
3. UX modification in communication and the actual survey.
4. Offering an incentive for survey completion.

More detail about and the impact of these interventions are documented in the ‘Rail Passenger Customer Experience Survey – Final Report of Response Rate Experiments’. To summarise, the combination of these interventions has considerably improved response rate per shift to 28% from 17% recorded for on board shifts in the method review field trials. Although there are still other factors that further influence response rate per shift, we feel that with these interventions, completed questionnaires per shift have been optimised and there is little else that can be done to further improve participation. A yet shorter questionnaire may increase the number of completed questionnaires per shift, however, that cannot be completely confirmed based on the short questionnaire used in the response rate experiments (5-6 minutes) and on the short survey trials (eight minute questionnaire).

A higher footfall count (and thus a busier train) typically and unsurprisingly has a positive impact on completed questionnaires per shift, however, we have also seen during the pilot and previous trials that sometimes overcrowded trains can have the opposite affect i.e. the number of completed questionnaires per shift declines on some trains e.g. during peak times on the Elizabeth line and the London Overground (see also completed questionnaires per TOC earlier on). As such, we would not recommend shifting the focus of sampling to a yet greater extent on busier trains than is already the case.

During the pilot a larger sample size than the one targeted was achieved for all but three TOCs: CrossCountry, Grand Central and ScotRail. The reason for this underachievement (which for all three TOCs was only marginal) was primarily due to insufficient time to reschedule shifts following disruption or no longer running services. Sufficient time should be provided to check the timetable, and time should be made available in the fieldwork to enable a change to be made to the shifts where they no longer are running. Completed questionnaires achieved for each TOC should also be monitored on a regular basis to understand if there is a requirement for further shifts to improve the completed questionnaires for a given TOC.

Regarding disruptions, while there could be some greater clarity around some circumstances for fieldworkers during a disruption as noted in section 4.1.4, we do not think this would be likely to optimise the response rates significantly. Disruption is very often out of anyone's control and usually difficult to predict. In a continuous rail customer experience survey it will be challenging (if not impossible) to build in any time for top up shifts during a given rail period. It might be worth oversampling for some TOCs right from the outset to ensure a minimum sample size is achieved.

The total number of shifts initially planned were achieved although some were conducted beyond the time of RP10. The main reasons for shift changes and

replacements were strike action, timetable changes and cancellations. This broke down as follows:

- 25 shifts were changed due to strike action.
- 171 shifts were changed due to timetable changes or trains no longer running.
- 15 shifts were changed due to permission delays.
- 40 shifts were changed due to interviewers' personal reasons.

The permission delays could be prevented with an overall permission (see further down). To keep shift changes due to personal reasons to a minimum we already apply a payment penalty if shifts are cancelled at short notice (although some emergencies have to be accepted).

The fieldworker itinerary (see further down) could also reduce the shifts changed due to timetable issues or train cancellations.

4.4 Evaluation of the degree to which the survey design will achieve the policy and research objectives

The policy and research objectives are to provide a customer perspective, build a meaningful understanding of what influences customer experience and identify any opportunities for improvement. The pilot, along with the method review trials, are there to inform future surveys on a survey design that will enable the industry to monitor and improve rail customer experience. This is detailed further in the annex.

We feel that the policy and research objectives are to a large extent achieved by the survey design. We have achieved a minimum sample, the operational aspects worked well with some minor improvements suggested, and response rates were fairly positive and improved with the inclusion of interventions. In our view the biggest caveat is that the provision of survey responses (by way of pre codes), to some degree limits what respondents can say. However, there are some open ended questions which allows respondents to give more detail.

As such the industry may want to consider the following for further insight from a passenger perspective:

- Regular or intermittent qualitative research with rail passengers about specific topics that emerge from the quantitative survey with a focus on how issues can be improved (if applicable.)
- Providing the options for passengers to submit feedback by way of a selfie video (vox pop) about a specific topic or topics; we use this option successfully in the RDG Infotracker.

- Social media monitoring: a lot of conversations with and about TOCs take place online i.e. on social media platforms, especially X (previously Twitter). Social media analysis and the tracking of sentiment from conversations online can be invaluable and provide very actionable feedback. We used to include such an element in our Rail Reputation Index (to which the DfT and RDG subscribed.)
- It would also be useful to analyse the survey data alongside some external data e.g. investment in rolling stock or technology or line improvements to understand their impact on passenger satisfaction.

4.5 Recommendations for future rail customer experience surveys

While we feel that the process of how the Rail Customer Experience Survey is currently set up is fairly smooth, there is always room for improvement.

In our view the process can be made more watertight with further investment, particularly in technology and automation. The key areas of improvement and thus our recommendations for a future rail customer experience survey are:

- Further investment into and streamlining of the Journey Picker Tool (JPT); a small proportion of respondents still struggle to find their details on the JPT which implies that usage instructions could be further enhanced (they had already been further enhanced for the pilot). Also, it seems that the JPT could be updated more quickly when new stations are opened.
- Automated itinerary for fieldworkers based on the JPT. Fieldworkers would get updates prior to their shifts with the latest itinerary so they would know if journeys have changed e.g. in case of disruption. This would further maximise working time during shifts. The itinerary would be created in such a way that would keep time between journeys to a minimum.
- Improvements to fieldworker manual. Improvements could be made to the fieldworker manual around what to do when disruption occurs while conducting their shift. For example, the difficulties engaging with anxious passengers and that passengers are less willing to complete during a period of disruption. It could outline when the fieldworker should consider interviewing on platforms, and cover examples of what to do when a train cancellation or delay hits multiple trains in the same region, and when they should end their shift.
- The footfall counts were used throughout the survey process but fieldworkers did not feel 10 minutes was sufficient time to review the whole train for the

numbers on board. It was often met with logistical challenges. Reviewing the length of these will be important for a future survey.

- The permission process can also be simplified as mentioned previously. To have one permission from the DfT or an industry body like RDG would be ideal and eliminate the time needed to deal with individual TOCs (which would also reduce costs). This option was considered for the pilot but rejected by the train operators.
- The pilot highlighted issues relating to the data flow via the API. A future survey will need to ensure a robust infrastructure is in place for regular data flows.
- The weighting efficiency could be addressed earlier through the sampling design rather than at the weighting stage. This would help minimise the excessive upweighting and down weighting that was required.

5. Quality standards & accreditation

We are ISO 20252:2019 and ISO 27001:2013 certified, the recognised international quality standards for market research and information security.

- All work will be carried out in conformity to these standards, the MRS Code of Conduct, GDPR, the UK's Data Protection Act, and all other relevant industry codes, legal and ethical requirements.
- Adherence to the standard is independently audited once per year.
- Where subcontractors are used by BVA BDRC, they are assessed to ensure any outsourced parts of the research are conducted in adherence with these same standards.

6. Other details

Images within this proposal have been sourced via Getty Images through subscription / Flaticon / logos from Wikipedia and Wikimedia.