Rail Passenger Customer Experience Survey – Method Review

Final recommendation report

13th November 2023

Contents

Page no.

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. The project has been managed by the rail industry comprising members of:

- The Department for Transport Rail,
- The Rail Delivery Group,
- Transport Focus,
- Network Rail, and
- GBRTT.

The method review has evaluated a wide range of potential methodologies to use in the enhanced customer experience programme. There were six phases to the method review:

- Phase 1 - development of criteria by which to evaluate different methodologies.
- Phase 2 - development of a weighting process to aggregate scores from the different criteria.
- Phase 3 - analysis of existing documentation, leading to a listing of all possible methodologies.
- Phase 4 - scoring of all possible methodologies against all the criteria agreed in phase 1 and the production of an aggregate score for each methodology using the weightings derived in phase 2.
- Phases 5 & 6 - sampling and weighting of field trial data, field trial results of the top methodologies, final recommended (optimal) detailed methodology and proposition for field trials.

For each phase, a technical report has been produced which documents the processes that were used in that phase and confirms the major conclusions that were derived and agreed.

At the end of all six phases, these reports were aggregated into a complete technical report which will be peer reviewed by an independent assessor. Further to the aggregated technical report a final recommendation report is required that details the recommendation, evidence and specification for the enhanced approach to measuring rail customer experience which is included in this document (as part of phase 6).

2. Part A. Proposal for methodology of the future customer experience survey

2.1 Executive summary of field trials for the future customer experience survey

From the field trials, intercepts on board emerge as the superior methodology. Based on the 21 criteria on which both the on board and the at station intercept approaches were evaluated during the field trials, on board was the better performing on nine criteria, at station the better performing on three criteria and the two methods performed equally well on nine criteria. As such, we recommend the following approach for the pilot and future continous survey:

- on board intercepts;
- offering email and QR code as response options (natural fallout); and
- conducting six hour shifts; although as mentioned in Part B. we recommend that six hour shifts are trialed first.

Sampling and weighting

In the pilot survey it is planned to append MOIRA data to the RDG timetable data thus enabling selection of train services with probability proportional to number of passengers (and not based on the number of services as was done in the field trials). This will select more busy services, including those at peak times, and fewer lightly used services and should bring the response profile more in line with that derived from the analysis of MOIRA data (see section 3.1.2 in Part B.).

The sample can be drawn as far in advance as possible, however, timetable changes have to be taken into consideration, not just the major timetable changes that happen twice a year but also changes due to planned maintenance and seasonal changes. A sample from say, rail period one could be replicated for each succeeding rail period - this would certainly make the issue of identifying return trips a lot easier (which at the same time would reduce costs) but we could run the risk of selecting a service that doesn't run in future months. To avoid that, the sample could be checked nearer the date against a current timetable, so a repeated sample has some attractions. It would also reduce some of the random variation in the results. The original sample needs updating every time there is a timetable change but the dates for these are known well in advance. Picking a sample that was replicated every rail period until the next major timetable change would have great benefits, with a routine run a couple of weeks before each rail period to check the selected services still

existed (see also further down in Part A. about cost reduction).

Weighting for the pilot survey is proposed to be similar to the weighting of the field trials. Weights would be applied to the following categories:

- Journey volume by TOC.
- Day of the week (weekday ranges and weekend days) by TOC.
- Station size bands by TOC.
- Age and gender at an overall level although this could also be done by TOC (from the footfall counts).

It might also be necessary to weight by ticket type. In the future continuous survey it is recommended to weight data for each rail period since that equals a reporting period. It has to be agreed how often weighting criteria should be updated to take into account passenger seasonality. Reviewing weighting criteria each rail period will take time, but it might not have to be that often. Saying that, passenger profiles will be different in, say, August compared to September.

A key point for the weighting is that MOIRA needs to be updated with more recent load factors to ensure weighting is done to realistic profiles (see section 3.1.2 in Part B.). We understand that this is not likely to happen for some time, but this will not change our recommendation. It is better to sample probability proportional to the number of passengers rather than sample trains at random, as this will select busier trains even if the passenger estimates are out of date. Needless to say it would be better if the passenger estimates were updated.

Questionnaire

We suggest that the questionnaire undergoes a further review to shorten it (see section 3.2.3 in Part B.). We understand that this is being worked on. A shorter questionnaire is likely to improve data quality and also response rate. Feedback from the qualitative element of the field trials was that the questionnaire is too long.

We recommend not to use any paper questionnaires in the pilot survey. They add costs and time and any benefits (which is mainly that a few older people select them) are far outweighed by its drawbacks (see section 3.2.5 in Part B.).

It is our understanding that the fact that some respondents complete the survey before the end of their journey is not considered to be a significant risk. The proportion of those doing so when recruited on board is 32% (see section 3.2.4 in Part B.). Needless to say, that excluding

early completers from the survey would reduce response rate and add to costs for achieving a specific sample size per rail period. The issue mainly occurs amongst those that select QR code as a response option (who are also younger) and it is our recommendation that a survey access delay for QR respondents is trialed at a larger scale in the pilot survey than it was in the field trials to understand how this impacts response rate.

The Journey Picker Tool (JPT) was hugely successful. When used correctly, the journey picker tool significantly improved the accuracy of journey reporting. By accessing the live database of trains to which it is linked, respondents were able to select the exact journey they travelled on, eliminating the potential for subjective interpretations or misinterpretations of route information.

It caused only a negligible proportion to drop out (1.8% of those who used it) and only 12% of those recruited on board did not manage to find their journey in the tool (see section 3.2.6 in Part B.). We believe that the improvements proposed in the next section and in section 3.2.6 in Part B. can further reduce these numbers. As such we recommend the use of the tool in any future continuous survey.

Fieldwork

In the field trials we selected low volume services to the same extent as high volume services for on board sampling (as mentioned previously). For the pilot survey we propose appending estimated passenger numbers from MOIRA for each train service, using an algorithm (or write a programme) to do this when the service does not exist on MOIRA. We also propose incorporating all the scheduled cancellations into the extraction process, so that the services extracted are all those planned to run for each day of the fieldwork period at the point in time of the sample selection. Plus, it would almost entirely eliminate the issues with on board fieldwork as mentioned in section 3.3.2 although there will always be a few short notice train changes or cancellations. For such scenarios we propose a clear set of instructions to fieldworkers specifying what to do e.g. if trains are cancelled or delayed or if they are denied access to travel and/or work on a train.

Strikes would be dealt with in the same way in that shifts on those days could, with the help of the aforementioned programme, be changed to non-strike days and return journeys will again be automatically selected.

There needs to be clear instructions for interviewers on entering and recording test surveys so that they can be easily identified and deleted at data cleaning stage. A possible solution would be to use a specific shift number for tests.

The issue of respondents not being able to scan the QR code when it is sunny and there is

a glare on the screen (see section 3.3.2 in Part B.) was mainly experienced at station. It will be easy to avoid sun when on board and as such we do not see a problem here.

It would be immensely helpful and time saving, and therefore cost saving, if the DfT could either arrange permissions for fieldworkers to travel and work on trains directly or produce an overarching document that allows fieldworkers to do so.

As mentioned previously we strongly suggest trialing six hour shifts and use them rather than three hour shifts. Section 3.3.3 in Part B. lists the benefits of both three hour and six hour shifts and benefits of six hour shifts far outweigh those of three hour shifts. A key advantage of six hour shifts is that the cost per complete is notably lower than for three hour shifts.

It is proposed that footfall counts are continued going forward since there is no data on gender and age available from other sources and some differences between survey sample and footfall count were apparent during the field trials (see section 3.3.5 in Part B.). For six hour shifts we recommend conducting two sets of footfall counts (e.g. three hours apart) on different trains for on board where possible. If the methodology uses an on board approach, we recommend that the foot fall count profiles age and gender in a manner that results in robust data for each TOC.

From the footfall counts 1,000 passengers per TOC would be needed to generate robust profiles. In the field trials, the average number of passengers per shift that contributed to the footfall count was around 70. In a six hour shift, the plan is still to undertake two counts during the shift so about 15 shifts would be needed to generate a 1,000 sample. Some TOCs will have less than 15 shifts in the pilot survey so this would involve counts on every shift; the maximum recommended sample size for a TOC is 400 which would involve around 30 shifts so for such a TOC footfall counts would only be needed at half the shifts to generate a 1,000 sample.

We could shorten the duration of the footfall counts and then do them on more trains which would reduce any clustering.

A potential solution is therefore to do footfall counts on all shifts, one in the first three hours and one in the last three hours. The duration would be 5 minutes on a TOC with a large sample size and 10 minutes on one with a small sample.

Data and analysis

Since both response options (email and QR code) direct respondents to an online survey data becomes available as soon as respondents have completed their survey. It has to be borne in mind though that immediate availability does not mean immediate usability. Several data

processing steps need to be taken to ensure data integrity and validity.

In section 3.4.1 we outline the data processing steps that need to be undertaken to ensure data integrity and validity. These include journey validation, postcode verification, review of responses to open ended questions, journey picker data and weighting.

2.2 Learnings from field trials and earlier studies

There were several learnings from the field trials, some of which were also experienced in earlier studies and as such improvements can be made for any future continuous survey. These learnings are listed below in no particular order.

- The inclusion of the journey picker tool was on the whole successful, identifying 86% of journeys whilst causing a negligible proportion of respondents to drop out of the survey (1.8%). However, 14% failed to use the tool and did not find their journey. This might be low and we think the proportion could be reduced further. One recurring issue was that respondents tried to find their whole, multiple leg, journey rather than the specific train they were intercepted on. Wording around explaining about a single leg journey could be improved or respondents could be allowed to enter multiple leg journeys which could then be filtered down to a single leg journey. Also, the departure times appearing in the drop down are forward looking when we ask respondents about a journey in the past and it is not clear that they can type in a journey time (see section 3.2.6 in Part B.). These improvements should be implemented.

  We might consider weighting data by ticket type in the pilot survey or at least compare ticket type data from the survey to LENNON. We already compared ticket type data from the field trials to LENNON data. For that, we attempted to assign each ticket type from LENNON to one of the categories in the questionnaire. "Other" codes and non-response cause a problem in the comparison. One learning from the field trials should be that there needs to be better alignment of the ticket type categories in the questionnaire so that they match the LENNON codes better (see end of section 3.1.2 in Part B.).

- During the field trials fieldworkers encountered a number of challenges like a return (or outbound) train being cancelled or severely delayed or not being given permission to travel and work on board a train or at a station. In such scenarios fieldworkers tried to call their regional managers or the BVA BDRC field team to enquire about what to do. This resulted in wasted time and cancelled shifts that needed to be rebooked which in turn resulted in additional costs (see section 3.3.2 in Part B.). For any future continuous survey we recommend producing a clear set of instructions of what to do in such situations to optimise a fieldworker's time and cost efficiency. These instructions will, of course, be

agreed with the DfT and stakeholders.

- Fieldworkers reported device signal issues on trains which meant that they could sometimes not sync their tablets until they got to their destination or back home. This resulted in delayed invites to some participants (see section 3.3.2 in Part B.). There is little that can be done about that for the pilot survey. Fieldworkers can only mention the issue to recruits at the time and apologise for the delayed email invitation which is what they also did in the field trials. It has to be borne in mind though that some respondents will be lost because of that.

- Sometimes fieldworkers were not able to complete their counts fully when carriages were crowded or when there are no through doors between carriages which means they cannot move to the next carriage to carry on with the count until the next stop by which time the 10 mins might have elapsed (see section 3.3.2 in Part B.). For the pilot survey we suggest that fieldworkers conduct footfall counts as well as they can when carriages are crowded. If it's impossible to carry on, they stop and record what has been counted in the shorter time period. If there are no through doors between carriages we propose that fieldworkers count one carriage and record the details for the shorter time period. If there is a stop within the 10 minutes, they go to the next carriage and continue the count until the 10 minutes are over. Recruitment was also impacted by crowded trains although to a lesser extent since it's still possible to recruit those around a fieldworker and shuffle along (whilst during a count the fieldworkers just cannot see everyone). During pilot survey busier trains can be inspected to see how recruitment goes and we can see how recruitment figures differ for crowded trains (if at all).

- Respondents who complete the survey before the end of their journey have been an issue in the field trials, with 30% completing early (see section 3.2.4 in Part B.), but also in previous surveys like the multi-method study that BVA BDRC conducted on behalf of Transport Focus in 2021. These respondents take their experience from a similar previous journey. Early completers are mainly those that select QR code as a response option. Those that select email rarely complete early because the invitation email sent out is already delayed by 30 minutes. A similar delay could be implemented for QR respondents. This was one of the interventions that worked but it resulted in quite a reduction in completes per shift. As such, we propose to trial this option on a larger scale and with some wording modifications. If such a survey access delay does not work for QR respondents the DfT and its stakeholders have to decide how important it is for the proportion of early completers to be reduced.

- Another learning from the field trials, and here primarily from the qualitative element,

is that the questionnaire is deemed too long. Which might also impact on response rate and therefore costs (see section 3.2.3 in Part B.). The questionnaire requires a further review and it is our understanding that this will be done and finalized prior to any future continuous survey is initiated.

- Paper questionnaires as a response option do not add anything to the survey other than costs and a delay in producing final data. Only a small proportion of older respondents who do not want to or cannot answer the survey digitally opt for paper (see section 3.2.5 in Part B.). As such we recommend not offering them as a response option in any future continuous survey.

- As per the previous section, it is our recommendation to conduct six hour shifts rather than three hour shifts in the pilot survey, after a trial. Cost calculations shows that costs per complete are notably lower for six hour shifts than for three hour shifts. The field trials have also shown other issues with three hour shifts (more details are provided in section 3.3.3 in Part B.):
    - Quite a bit of time is lost for the on board approach when journeys stop notably before the end of the shift but no further journey can be fitted in; six hour shifts are much more flexible in that respect.
    - Especially for long distance TOCs shifts sometimes must be extended because a return journey does not fit into three hours. This results in additional costs.
    - 3 hour shifts are less attractive for fieldworkers and therefore it's harder to book them out.
    - They are less environmentally friendly because they require more journeys from fieldworkers travelling to and from shifts.

2.3 Elements that require further trialing or investigation plus possible options to what was done in the field trials

We have divided elements that require further trialing or investigation and options to what was done in the field trials into two groups: those that may impact response rate and those that are process related which means they save executive and field team time or are linked to data processing.

At the beginning of Part B. we list elements that we recommend to trial and the following will have an impact on response rate, however, it is not clear to what extent:

- Offering a modular approach which came out as the preferred intervention in the qualitative element but had the largest dropout rate of all interventions; it is expected that this will lower response rate.

- QR code access delay to the survey with those scanning the QR code being directed to a page that counts down before allowing access to the survey; the aim would be to reduce the proportion of those that complete the survey before the end of their journey but it needs to be understood how this will impact on response rate on a larger scale; it is, however, expected to reduce response rate.

- Interviewers not saying at the time of recruitment that the survey needs to be completed after the journey which seems to have put people off initially although the instructions can be kept in the script and email message; it is expected that this will improve response rate but this needs to be balanced against how important this aspect is. This was tested as part of the response rate experiments whose results can be seen in the *Final Report on Response Rate Experiments*.

- Survey Team hi vis jackets to be provided to all fieldworkers as fieldworkers noticed it has an impact on the way they are perceived by rail users (they look more official) (see section 3.3.2 in Part B.); this may improve response rate but would require an initial monetary outlay.

- Allocating return journeys and checking trains (i.e. that they still run and that the time has not changed) takes up a considerable amount of time from the executive and field teams. Therefore, we propose that a programme or routine is developed that does this for the teams and reduces their time. It needs to be investigated how long this would take and what the costs are but the ultimate benefit of such a programme or routine (and subsequent time and cost savings) is likely to outweigh the initial time and budget investment. It will take too long to investigate costs and time for implementation of such a programme and include these in this report. Hence they have been omitted.

- The weighting efficiency per TOC could not be fully established from the field trials for the on board and at station methods since only 12 TOCs were included. The weighing efficiency needs to be looked at during the first rail period in the pilot survey to understand if weighting can be applied as was planned for the field trials (please see *Pilot Weighting Guide* used for pilot survey).

- The clustering effect for six hour shifts on board is still a bit of an unknown although we do not foresee this to be an issue in the forthcoming pilot survey and any future continuous survey. But it should be calculated as soon as data allows.

2.4 Justification of the proposed method based on factors used in the method review and emerging factors in the field trials

The following table provides a summary of the proposed method of face to face intercepts on board (offering email and QR natural fallout as response options) based on each criterion from Part B. and the evidence that is provided in Part B. A comparison of which method performs better on each criterion can be found at the beginning of Part B.

Table 1: Summary of performance of face to face intercepts on board train against different criteria

Category	Criteria	Summary of on board intercepts performance
Sampling and weighting	Coverage of the required universe	On board, by definition, all those intercepted are making a journey on the selected day; age must be confirmed but the proportion of those approached that are aged below 16 is negligible (less than 0.5%). On board performs better than at station.
	Ability to generate a random representative sample	In the field trials, the on board sample was a random sample of services rather than one proportional to the estimated number of passengers, so low volume services may have been over sampled. The response profile for on board recruitment better matches both footfall counts and MOIRA than the at station sample. Early services are underrepresented in the at station sample but not on board. On balance both methods perform equally well. QR code and email even out possible age differences.
	Ability to generate large samples (scalability)	An on board approach spreads the sampling across a very large number of services, even if these are sampled with probability proportional to the estimated number of passengers on the train. Scalability is more straightforward than for at station; double the number of services sampled for a TOC to double the sample. On board performs better as scalability is simple and does not require top up shifts outside the main sample.

Category	Criteria	Summary of on board intercepts performance
	Ability to generate required sample sizes of key subgroups	All responses from a selected on board fieldwork shift are for the selected TOC and controlling sample sizes by TOC is easy by selecting either fewer or more services to meet the required sample size target. TOCs could be subdivided into lines of route and the same process can be used to select specified sample sizes for each line of route. As reaching TOC sample size targets is a crucial element of the pilot survey, an on board approach has significant advantages in enabling this.
	Response rate	On board achieves a higher number of completes per shift than at station. Partly this is due to a higher number of recruits per shift. But some is clearly a better response per recruit. The response rate achieved for both methods was considerably lower than what was expected. For on board this was 17% vs. 30%. Completes per shift are also higher on board for the vast majority of TOCs and for each TOC type. As such on board performs better on this criterion. Those that select QR code as a response option are much more likely to drop out of the survey generally and drop out earlier in the survey than those that select email as a response option.
	Weighting efficiency	In the field trials it was necessary to merge some weighting cells (due to no sample achieved for some cells) and it was not possible to weight by TOC size. Bearing in mind these two conditions, on board achieved a weighting efficiency of 64.2, whereas at station achieved a weighting efficiency of 55.5. The clustering effect is very similar across both methods. Overall, on board performs better.
Questionnaire	Knowing the exact train the person was travelling on	For the on board approach the interviewer enters the TOC of recruitment into the recruitment script. This information comes from the sample plan and the sample plan will have 100% accuracy. Once a respondent is recruited on board, the TOC of recruitment is automatically piped into their script. With on board we are much more likely to know the train the respondent was travelling on since the fieldworker also travelled on that train and details can be checked against the shift plan and the information entered into the recruitment script by the fieldworker. As such on board performs better.

Category	Criteria	Summary of on board intercepts performance
	Accuracy of other information about the journey undertaken	A larger proportion of those recruited on board completed the survey on the same day as their journey compared to at station (77% vs. 71%) which implies greater accuracy. The share of those that complete the survey before the end of their journey is slightly higher amongst those recruited on board but amongst at station recruits more answer the survey before their journey even starts (7% vs. 1%). On board generates greater accuracy.
	Questionnaire length	On board can deal with longer questionnaires and so can the at station approach. It is recommended to keep the questionnaire length to 10 minutes.
	Completion after journey	The share of those that complete the survey before the end of their journey is slightly higher amongst those recruited on board (32% vs.28%) but amongst at station recruits more answer the survey before their journey even starts (7% vs. 1%). The issue with completing early seems to lie rather with the response option than with the recruitment method with those selecting QR code to access the survey much more likely to complete the survey before the end of the journey than those that are emailed a survey link (because the invitation email was delayed). On balance both methods perform equally well.
	Paper surveys	There is no advantage of one method over the other with regard to paper surveys although the inclusion of paper surveys has some benefits but these are outweighed by its drawbacks.
	Journey picker tool	A slightly higher proportion of those recruited on board are able to find their journey on the journey picker tool. As such the on board approach performs a little better.
Fieldwork	Ability to recontact participants	Both methods perform equally well for recontacting respondents but it has to be borne in mind that respondents selecting QR code over email at recruitment are less willing to be recontacted since they are mainly younger. In that respect email performed better than QR code.
	Practicability / Feasibility	The on board approach has significant practical and logistical issues such as finding return journeys and being impacted by the return journey being severely delayed or cancelled. These can be minimised but not completely eradicated and this will feed through into higher costs in some areas than for the at station approach. So at station performs better on this criterion.

Category	Criteria	Summary of on board intercepts performance
	Shift length	6 hour shifts have a number of benefits over three hour shifts. six hour shifts are better suited for on board shifts since they provide a much greater flexibility with regard to making an outbound and inbound journey. This is particularly the case for journeys on long distance TOCs where sometimes the first stop of a journey is not within 90 minutes which then results in overtime for three hour shifts. This would not apply for six hour shifts - there is typically a stop within three hours so that the fieldworker can easily get back to their departure station. Both methods score equally well on this criterion.
	Fieldwork force	Many of the points stated in the 'Practicality / Feasibility' section could be repeated for this criterion since they equally apply to the fieldwork force. Based on that at station performs better.
	Footfall counts	Most footfall counts (80%+) were completed for either approach but the percentage undertaken and the number of passengers profiled is higher at station than on train. On the other hand, at station counts cannot be split by TOC. On balance, the two methodologies perform equally on this criterion.
Data and analysis	Speed of generating topline results	There is no difference in the speed of generating topline results for either of the two methods as long as the survey is completed online. Online is the fastest way to produce topline results
	Cutting data by different time periods	The process of cutting data by different time periods would be the same for on board and at station and would generally work well.
	The ability to merge with other data	Here too, the process would be the same for on board and at station and would generally work well. On board data collection is more likely to have accurate data on the train itself but in all other respects the two methods perform to a similar level on this criterion.
Cost	Cost	If trains run as scheduled and there are no strikes, then on board is more cost effective than at station because it achieves a higher number of completes per shifts. This is the case despite of the fact that on board requires more executive time (which can be reduced). On board is more easily impacted by short notice changes (i.e. on the day or a day or two in advance). This would result in additional costs.

3. Part B. Evidence from the field trials

Based on the methodology evaluation in phase 4, the following methods were selected for the field trials:

Method 1

- face to face intercepts on board,
- email/QR code natural split, and
- some of the fieldwork to incorporate a paper backup.

Method 2

- face to face intercepts at station,
- email/QR code natural split, and
- some of the fieldwork to incorporate a paper backup; although this was ultimately changed to some of the fieldwork to incorporate a QR code postcard or leaflet as backup if passengers were in a hurry.

As such the following four shift types were included in the field trials:

1. Blue shifts – face to face intercepts on board no paper; recruit passengers using email or a QR code natural split.

2. Yellow shifts – face to face intercepts on board with paper back up; recruit passengers using email or a QR code natural split with paper questionnaires as back up.

3. Red shifts – face to face intercepts at station no paper; recruit passengers using email or a QR code natural split.

4. Green shifts – face to face intercepts at stion with QR postcard; recruit passengers using email or a QR code natural split with QR code leaflets as back up if passengers in a hurry.

The fieldwork was carried out across nine weeks, from Monday 17th April - Monday 19th June 2023. The first half of fieldwork was from 17th April - 14th May and this consisted of only the main four core methodologies that were tested. The second half of fieldwork was from 15th May - 19th June, and this included the main methodologies, with some of these shifts incorporating interventions to prevent respondents from completing the survey before the end or their journey (see later in this section).

The fieldwork did not include three Monday bank holidays in May (1st May, 8th May, 29th May); five rail strike dates (Friday 12th May, Saturday 13th May, Wednesday 31st May, Friday

2nd June, Saturday 3rd June), and the morning shifts that followed each strike day. In any future continuous survey however, the DfT and stakeholders may want to include holidays and strike days to understand the experience on such days.

In the field trials 12 TOCs were included which covered a span of typologies (commute, airport, interurban, high speed, long distance and non franchised):

1. Avanti West Coast,
2. CrossCountry,
3. East Midlands Railway,
4. Grand Central,
5. Heathrow Express,
6. London Overground,
7. Merseyrail,
8. Northern,
9. Scotrail,
10. South Western Railway,
11. Thameslink, and
12. Transport for Wales.

Following the field trials, the two methods have been compared across 21 criteria selected by the DfT, its stakeholders and the peer reviewer to understand which of the two performs better on each. Please see table 2 below. A short section with evidence from the field trial data or other sources (with the relevant reference) which provides justification for this conclusion has then been written for each criterion.

As mentioned in Part A. our recommended method is:

- **an on board approach, offering both email and QR code as response options and conducting six hour shifts.**

The onboard approach performs better on nine out of the 21 criteria. The methods perform equally well on nine out of the 21 criteria with at station performing better on only three criteria.

The 21 criteria include the 16 criteria that were developed in phase 1 by which we initially evaluated different methodologies (except 'incidence rate' although 'footfall counts' could possibly be classified as such), plus some additional criteria that emerged during the field

trials.

Table 2: Best performing methods against delivery criteria following field trials

Category	Item	Criteria	Better performing method based on field trial findings	Weight in phase 2	Higher scoring method in phase 4
Sampling and weighting	1	Coverage of the required universe	On board	8.7%	On board
	2	Ability to generate a random representative sample	Equal	8.6%	On board
	3	Ability to generate large samples (scalability)	On board	8.2%	Equal
	4	Ability to generate required sample sizes of key subgroups	On board	8.5%	Equal
	5	Response rate	On board	5.4%	On board
	6	Weighting efficiency	On board	8.2%	At station
Questionnaire	7	Knowing the exact train the person was travelling on	On board	6.1%	On board
	8	Accuracy of other information about the journey undertaken	On board	5.0%	On board
	9	Questionnaire length	Equal	4.6%	Equal
	10	Completion after journey	At station	n/a	n/a
	11	Paper surveys	Equal	n/a	n/a
	12	Journey picker tool	On board	n/a	n/a
Fieldwork	13	Ability to recontact participants	Equal	2.8%	Equal
	14	Practicability / Feasibility	At station	6.6%	On board
	15	Shift length	Equal	n/a	n/a
	16	Fieldwork force	At station	n/a	n/a
	17	Footfall counts	Equal	n/a	n/a

Category	Item	Criteria	Better performing method based on field trial findings	Weight in phase 2	Higher scoring method in phase 4
Data and analysis	18	Speed of generating topline results	Equal	4.8%	Equal
	19	Cutting data by different time periods	Equal	6.3%	Equal
	20	The ability to merge with other data	Equal	6.1%	Equal
Cost	21	Cost	On board	5.9%	Equal

Table 3 below shows a summary of the characteristics for both email and QR code as a response option. Bearing in mind the preference of one option over the other by different age ranges it is our recommendation to keep offering both options going forward. Those that select QR code as a response option are more likely to drop out of the survey (from the time of scanning) and are also more likely to complete the survey before the end of their journey. There is, however, no significant difference on overall journey satisfaction amongst those that complete the survey before the end of their journey compared to those that complete after their journey. Furthermore, those that select QR code are more likely to complete the survey on the same day as their journey which may result in greater accuracy.

Table 3: Characteristics of email and QR code as response options

Characteristics of email as response option	Characteristics of QR code as response option
Used more by older passengers (45+ years)	Used more by younger passengers (16-44 years)
Respondents less likely to complete the survey before the end of their journey (30 minute delay implemented before email is sent)	Respondents more likely to complete the survey before the end of their journey (although a delay could also be implemented as for email) and referring to similar journey made previously
Cannot start survey immediately if 30 minute delay set	Enables immediate start of the survey
Small proportion of those recruited drop out of the survey	Larger proportion of those recruited drop out of the survey
Less likely to complete on the same day of the journey which may make responses less accurate	More likely to complete the survey on day of the journey which may result in greater accuracy
Less speedy at recruitment stage	Speedier at recruitment stage
More willing to be re-contacted and to provide contact details	Less willing to be recontacted and to provide contact details
	Difficulty to scan QR in very bright conditions (although could automate screen brightness)

Following the field trials it is recommended that the subsequent elements are trialed further to understand their impact on response rate fully. This could potentially be done during the pilot or additional fieldwork.

1. Prioritising email over QR code as a response option to reduce the proportion of those that complete the survey before the end of their journey and to reduce drop out.

2. Modular approach similar to intervention 5 (which came out as the preferred intervention in the qualitative element but had the largest dropout rate of all interventions).

3. Asking five key questions (including some key metrics) at the time of recruitment before inviting respondents to the full survey; this will provide a much larger sample on key metrics (which should include recommendation and overall satisfaction) but may result in fewer

full surveys; this will also have cost implications in that it could save budget.

4. QR code access delay to the survey with those scanning the QR code being directed to a page that counts down before allowing access to the survey; the aim would be to reduce the proportion of those that complete the survey before the end of their journey but it needs to be understood how this will impact on response rate on a large scale; this is similar to intervention 3.

5. Conducting six hour shifts rather than three hour shifts to reduce the cost per complete; it needs to be understood how this will impact on clustering although it is believed that this will not be an issue.

6. Interviewer not saying that survey needs to be completed after the journey which seem to have put people off initially although the instructions can be kept in the script and email message; it depends on how important this aspect is.

In our view six hour shifts are the most pressing element to trial due to their likely considerable impact on response rate relative to costs with no real disadvantages.

The analysis in the following sections is based on the total sample of validated completes from the field trials (and any subsegments where applicable), regardless of whether respondents completed the survey before the end of their journey or after. The issue of completing the survey prior to the end of the train journey respondents were recruited for, is analysed in detail in section 3.2.4.

3.1 Sampling and weighting

3.1.1 Coverage of the required universe

The criterion for inclusion in the field trials and subsequent surveys is that a respondent must be making a rail journey on the day selected. At stations, some individuals will not be making a train journey – some will be meeting a passenger or seeing someone off, some may be using the retail outlets and services at the station, and some will just be passing through the station. These individuals need to be screened out. In the field trials 741 out of 23,870 passengers intercepted at station were not making a journey on the selected day, which is 3.10%. Although this is not a large percentage, the recruitment question needs to be asked of all those intercepted to ensure they qualify, so this does add time to the recruitment process for every individual at station. On board, by definition, all those intercepted are making a journey on the selected day.

All those taking part must be aged 16 or over. This does require the same recruitment question for each method of recruitment. Of those approached, 0.43% were aged below 16 (of course, fieldworkers will not be approaching those that obviously fail to meet the age threshold).

In summary, ensuring that all those recruited do meet the required criteria involves an extra recruitment question at station, which will impact on the numbers recruited per shift (although there was no specific feedback from interviewers on this). Table 4 below shows the average number of recruits per shift, with on board sampling generating significantly more recruits than at station (although there will be factors other than the length of the recruitment script affecting this).

Table 4: Average number of recruits per shift for different methodologies in the field trials

	At station no paper	At station with QR postcard	On board no paper	On board with paper back up
Recruits per shift	34.4	37.2	41.8	50.1

Conclusion: On board performs better than at station since at station fieldworkers can intercept people who do not make a journey on the day and these need to be screened out.

3.1.2 Ability to generate a random representative sample

Historically the National Rail Passenger Survey (NRPS) generated samples which were biased towards older respondents, as response rates for these individuals tended to be higher. The gradual transition of NRPS from paper to a methodology which was roughly evenly split between paper and email addressed this to some extent and the field trials, with their bias towards digital methods, were expected to further improve this.

There is no data on the age and gender distribution of those making train journeys and so footfall counts were included as part of the field trials to assess how well those responding to the survey match the profile of those making train journeys. Note that any analysis of footfall counts within this report includes intervention shifts.

Table 5: Age profile by response option in the field trials

Red *and a* minus (-) sign or blue and a plus (+) sign show statistical significance at 95% confidence.

	Field trial respondents (including interventions)					Footfall counts
Age group	QR	Email	QR Leaflet	Paper	Total	
Sample size	4357	2671	12	92	7132	71160
16-18	4% (+)	1% (-)	0%	0%	3%	
19-24	16% (+)	4% (-)	17%	2% (-)	11%	40%
25-34	28% (+)	10% (-)	42%	4% (-)	21%	
35-44	21% (+)	13% (-)	8%	3% (-)	18%	
45-54	16% (-)	19% (+)	25%	9%	17%	
55-59	6% (-)	12% (+)	0%	9%	8%	47%
60-64	4% (-)	13% (+)	0%	17% (+)	8%	
65-69	2% (-)	12% (+)	0%	14% (+)	6%	
70-80	1% (-)	13% (+)	8%	33% (+)	6%	13%
81+	0% (-)	1% (+)	0%	7% (+)	1%	
Prefer not to say	1%	1%	0%	2%	1%	
Don't know/ not sure	0%	0%	0%	0%	0%	

As the table shows, the combined respondent profile across all response options does not generate quite enough 16-34 year olds (35% versus 40% from the footfall counts using observable age). In contrast the combined respondent profile generates too many 35-64 year olds (51% versus 47% from the footfall counts). However, the proportion of those aged 65+ matches across the two sources. It is interesting that paper alone, albeit based upon just 92 respondents, generates a much higher percentage of those aged 65+ (54% against 13% for the overall field trials profile) and the inclusion of this option helps the response profile align better with the footfall counts.

Table 6: Gender profile by response option in the field trials

Red and a minus (-) sign or blue and a plus (+) sign show statistical significance at 95% confidence.

	Field trial respondents (including interventions)					Footfall counts
	QR	Email	QR Leaflet	Paper	Total	
Sample size	4357	2671	12	92	7132	71160
Male	45%	44%	25%	37%	45%	52%
Female	52% (-)	54%	75%	59%	53%	48%
I identify in another way	1% (+)	0% (-)	0%	0%	1%	
Prefer not to say	1%	1%	0%	4% (+)	1%	
Don't know/ not sure	0% (+)	0% (-)	0%	0%	0%	

The percentage of males from respondents to the field trials is lower than from the footfall counts. This analysis confirms previous NRPS analysis that response rates tend to be lower for males than females.

Comparing at station to on board methodologies in Tables 7 and 8 show no major differences in age or gender profiles by method.

Table 7: **Age profile by recruitment method in the field**

Blue and a minus (-) sign or Red and a plus (+) sign show statistical significance at 95% confidence.

	Field trial respondents (including interventions)						
Age group	At station no paper	At station with QR postcard	On board no paper	On board with paper back up	Total	NRPS 2019	
Sample size	2547	475	3314	787	7132		
16-18	3%	3%	3%	2%	3%	16-18	2%
19-24	11%	12%	11%	10%	11%	19-25	8%
25-34	21%	22%	22%	19%	21%	26-34	13%
35-44	18%	17%	18%	17%	18%	35-44	16%
45-54	17%	16%	16%	17%	17%	45-54	21%
55-59	8%	8%	9%	8%	8%	55-59	11%
60-64	8%	8%	7%	10%	8%	60-64	10%
65-69	6%	7%	6%	7%	6%	65-69	8%
70-80	6%	6%	6%	7%	6%	70-80	9%
81+	0%	0%	1%	2%	1%	81+	1%
Prefer not to say	1%	1%	1%	2%	1%		
Don't know/ not sure	0%	0%	0%	0%	0%		

Table 8: **Gender profile by recruitment method in the field**

Blue and a minus (-) sign or Red and a plus (+) sign show statistical significance at 95% confidence.

	Field trial respondents (including interventions)					
	At station no paper	At station with QR postcard	On board no paper	On board with paper back up	Total	NRPS 2019
Sample size	2547	475	3314	787	7132	
Male	43%	46%	47%	43%	45%	44%
Female	54%	53%	51%	54%	53%	56%
I identify in another way	1%	1%	1%	1%	1%	
Prefer not to say	1%	0%	1%	2%	1%	
Don't know/ not sure	0%	0%	0%	0%	0%	

It is clear from this analysis that, although the respondent profiles from the field trials are closer to the footfall counts than traditionally produced from NRPS, there is still a need to use other data to correct response bias. There may be potential in future to use data from mobile telephone companies but in the meantime, footfall counts provide this information. This correction is required for both at station and on board methodologies.

It is worth pointing out that the interventions did not have a different profile compared to the non-intervention sample, with only a percentage point difference for some groups in some interventions.

Time of travel by recruitment method

There are other metrics where we can evaluate the profile of the sample generated, principally by day of week and daypart (see table 9).

Table 9: Profile of sample by daypart for different recruitment methodologies and data sources

Summary	All	At station	On board	Footfall
Off peak	**40%**	**38%**	**42%**	**38%**
Morning peak	**14%**	12%	15%	12%
Evening peak	**18%**	19%	17%	16%
Late	**6%**	9%	4%	6%
Sat	**12%**	12%	19%	13%
Sun	**10%**	9%	11%	14%

The first column here shows the profile of the total main sample (i.e. excluding interventions), the second and third split this by at station and on board. The fourth column shows the data from the footfall counts.

There are fewer responses on the weekends compared to the footfall data. In comparison to MOIRA data, there are more responses at off peak times and on Sundays and fewer in peak hours). In making any comparisons, it should be borne in mind that MOIRA uses load factors to convert train data to passenger numbers which are based upon pre pandemic data. The generally accepted wisdom in the rail industry is that for this reason MOIRA considerably overstates peak passenger numbers.

Some of the peak hour shortfall will likely be driven by lower response rates among commuters and weighting by the MOIRA data may well correct this, but ideally MOIRA needs updating so that the load factors are more representative of current behavior.

We should also be aware that in the field trials sampling, the on board sample was a random sample of services rather than one proportional to the estimated number of

passengers, so low volume services will have been over sampled. For at station sampling, shift times are assigned at random and start at 06:00 and end at 19:00. As a result, train services starting at 06:00 and 07:00 are underrepresented in the sampling plan. For on board sampling there is no such issue and even without sampling these proportional to passenger numbers, the response profile better matches both footfall and MOIRA than the at station sample.

In conclusion, footfall data is likely to be needed to correct differential demographic response rates. In the pilot survey it is planned to append MOIRA data to the RDG timetable data thus enabling selection of train services with probability proportional to number of passengers. This will select more busy services, including those at peak times, and fewer lightly used services and should bring the response profile more in line with that derived from the analysis of MOIRA data. However, MOIRA needs to be updated with more recent load factors to ensure weighting is done to realistic profiles.

Ticket type – LENNON data vs. field trial data

We also compared the ticket type data from the field trials to the LENNON data. We attempted to assign each ticket type to one of the categories in the questionnaire. Some ticket types are overrepresented and some underrepresented. Weighting by daypart might improve this. One learning from the field trials should be that there needs to be better alignment of the ticket type categories in the questionnaire so that they match the LENNON codes better. "Other" codes and non-response cause a problem in the comparison.

Table 10: comparison of ticket type profiles of field trials to sample to LENNON data

Ticket type	LENNON	Field trials
Advance	7%	11%
Annual Season Ticket	2%	1%
Anytime single/ return	31%	34%
Day Travelcard	12%	4%
Flexi Season Ticket - Flexible Season Ticket	0%	1%
Monthly Season Ticket	4%	1%
Off-Peak or Super Off-Peak (single or return)	36%	24%
Other (please specify)	0%	22%
Special Promotion Ticket i.e. Rover Ticket	0%	0%
Weekly Season Ticket	7%	2%

Coding the “other” responses from the survey gives the following ticket types (those of 5%+ shown for “other” – See table 11)

Table 11: Ticket types within the 'other' category

Description	%
Staff travel card/ticket	9.4
Answer not relevant to question (e.g. abc)	8.4
Other (than those coded)	8.4
Trainline	7.9
Yearly pass/ticket	7.9
Oyster card	6.3
Other passes	5.8
Season pass	5.2

Conclusion: Both recruitment methods perform equally well in producing a representative sample. As for response option QR code is considerably more popular amongst younger respondents whereas older respondents are more likely to select email to get to the survey.

3.1.3 Ability to generate large samples (scalability)

In principle both at station and on board methodologies can be upscaled to generate larger samples. At station, however, a point will be reached at the largest stations, where fieldworkers will be present most days during a fieldwork period. This might create customer resistance to being approached too often (a situation like this emerged for NRPS in sampling the Island Line, which was undertaken on board and led to concerns that the same passengers were being approached all the time). In contrast, an on board approach spreads the sampling across a very large number of services, even if these are sampled with probability proportional to the estimated number of passengers on the train.

To meet TOC targets using an at station approach, top up shifts targeting the smaller TOCs will be required at key stations (e.g., for Heathrow Express at London Paddington and Grand Central at London Kings Cross). The larger the sample, the more of these top up shifts that will be needed.

For on board sampling, scalability is more straightforward; double the number of services sampled for a TOC to double the sample.

Table 12 below confirms these points, with completed questionnaires per shift at station considerably lower than on board for the smaller but also some other TOCs.

Table 12: Completed questionnaires per shift by TOC and recruitment methodology

	Completes per shift at station	Completes per shift on board
Avanti West Coast	4.1	7.0
CrossCountry	2.7	8.5
East Midlands Railway	2.6	7.7
Grand Central	1.3	11.5
Heathrow Express	1.6	5.1
London Overground	3.7	4.4
Merseyrail	4.7	4.7
Northern	6.6	9.2
ScotRail	4.6	6.2
South Western Railway	7.2	7.9
Thameslink	4.7	9.1
Transport for Wales	5.3	9.4

Conclusion: On board is preferable as scalability is simple and does not require top up shifts outside the main sample.

3.1.4 Ability to generate required sample sizes of key subgroups

On board sampling involves fieldworkers boarding specified trains. A sample of train services is selected for each TOC covering different dates within the fieldwork period and different times of days, with train services selected at random in the field trials (but with probability proportional to the estimated number of passengers provided by MOIRA in the pilot and any future continuous survey.)

By definition all responses from a selected fieldwork shift are for the selected TOC and controlling sample sizes by TOC is easy by selecting either fewer or more services to meet the required sample size target. TOCs could be subdivided into lines of route and the same process can be used to select specified sample sizes for each line of route.

Sample sizes by station of origin would not be controlled using this method. Any project requirement to reach a specified number of interviews for individual stations (e.g., to meet sample size targets for Network Rail managed stations) would need to involve top up shifts at those stations. Similarly for specific regions depending on the region definition.

By contrast, at station sampling makes it easy to reach specified sample sizes for

individual stations or specific regions but makes it more difficult to reach required TOC sample sizes. Once a station sample has been selected, it is possible to estimate the likely sample size per TOC using the profile of passengers using the station by TOC, which is available from analysing MOIRA data. However, this process would not generate large sample sizes for smaller TOCs, primarily the open access TOCs such as Grand Central and Heathrow Express but also some of the 'franchised' TOCs (like Gatwick Express and Chiltern). There would inevitably be a requirement to incorporate boost shifts at some stations targeting the TOCs that would be underrepresented in the main (random) sample. And the main sample to meet TOC targets would not necessarily meet station sample sizes needed for Network Rail managed stations and so further top ups are likely to be needed to meet those targets.

Conclusion: As reaching TOC sample size targets is a crucial element of the Rail Customer Experience Survey, an on board approach has significant advantages in enabling this.

3.1.5 Response rate

The table below shows the expected response rates, as outlined in the original proposal, and those achieved (defined as the proportion of those recruited who completed the full survey). All response rates except for the paper back up on board achieved lower response rates than expected (and it should be borne in mind that 270 out of 304 respondents given the option of a paper version chose the paper option).

Table 13: Expected and Achieved response rates for different recruitment methodologies

	At station - expected	At station - achieved	On board - expected	On board - achieved
Email	25%	20%	30%	18%
QR code	25%	15%	30%	17%
Paper			25%	31%

Response rates are similar between the at station approach and the on board approach. We had seen earlier that recruitment rates on board were significantly higher than at station (42/50 per shift on board versus 34/37 per shift at station) and so this combined with similar response rates means that the number of completes per shift is a lot higher on board. The overall response rate for the on board option with paper back up performs better than digital only; some of this is a better recruitment rate (50 vs 42) but some is clearly a better response per recruit.

Table 14: Completed questionnaires per shift per sheet for different recruitment methodologies

	At station no paper	At station with QR postcard	On board no paper	On board with paper back up
Completed questionnaires per shift	5.8	5.3	7.2	8.6

The impact of questionnaire length on response rate including dropouts

One factor impacting on response rates is the length of the questionnaire. Assessment of interview length as given by the data needs to be careful and the single best estimate comes from looking at responses by email to the main survey. This omits all the interventions and also those completing by QR code where there may (and should) be a delay between scanning the QR code and starting to complete the questionnaire.

The median for the online survey completion is 15 mins 58 secs. This is too long for an online questionnaire and the following table shows how drop off occurred by the stage of the questionnaire.

Overall, 74% of those recruited, dropped out of the survey at some point. The largest proportion of respondents drop out of the survey after landing on the language selection screen (just under a third at 30%). Another 19% then leave the survey after the introduction screen ('Thank you very much for taking part in this important survey…') and a further 26% drop out of the survey after the journey details section.

The stage at which respondents drop out varies considerably by response option, with those selecting QR code over email more likely to drop out generally (81% versus 35%) and also more likely to drop out early on e.g. at language selection and the survey introduction screen. The reason for this is not known – it could be that a proportion of QR recruits never actually intended to complete the survey. QR recruits are counted as respondents as soon as they scanned the QR code since that took them straight to the survey (see also next paragraph).

Allowing respondents to stay on the first screen (language selection) for five seconds changes the picture quite notably. In such a scenario 66% drop out in total which breaks down into 72% for QR respondents and 33% for email respondents and the journey details section becomes

the section after which the greatest share leaves the survey (40% compared to 26% previously – see above). Table 15 provides the percentage of dropouts at different stages of the questionnaire and Table 16 outlines the overall dropout rate when different interventions were applied.

Table 15: Percentage and number of dropouts by stage of questionnaire

Questionnaire order	Dropped out after…	%	n
1	Lang screen - Please select your preferred language / Dewiswch eich iaith ffafriedig os gwelwch yn dda.	30.0%	5045
2	Intro at start of survey - Thank you very much for agreeing to take part in this important survey about the train journey you completed...	19.1%	3220
3	Overall intro - We would like to ask you some questions about the train journey you were making when you were invited to take part in this survey...	1.1%	188
4	Journey details section	25.6%	4309
5	Assistance section	0.3%	55
6	Disability section	0.7%	113
7	Decision and Planning intro page	0.2%	38
8	Decision and planning section	3.7%	616
9	Ticket purchasing intro page	0.3%	43
10	Ticket purchasing section	4.3%	730
11	Getting to the station and Q52 intro page	0.5%	80
12	Getting to the station section	0.3%	44
13	Journey experience at the station intro page	0.2%	39
14	Journey experience at station section	5.3%	896
15	Journey experience on the train intro section	0.3%	55
16	Journey experience on train - section	2.7%	457
17	Intro page for Journey experience at the arrival station	0.5%	84
18	Journey experience at arrival - section	2.7%	454
19	Delays, Complaints and Compensation intro page	0.1%	13
20	Delays, complaints, compensation section	0.4%	75
21	Overall satisfaction and experience intro page	0.0%	5
22	Overall satisfaction section	1.3%	223
23	About you intro page	0.0%	4
24	Demographics section	0.3%	45
	Total	100.0%	16,831

Table 16: Overall dropout rate per intervention

Intervention 1	Intervention 2	Intervention 3	Intervention 4	Intervention 5	All interventions
78%	81%	81%	73%	86%	80%

This implies that proportionately intervention 4 caused the fewest dropouts, however, it cannot be said that the drop out was only because of the intervention or for some other reason.

Change of response rate during fieldwork

The table below shows how many shifts were conducted within each week of fieldwork – in the 2nd column. The 3rd column shows the number of completes that were recorded for each of the shifts conducted within that week. And the 4th column provides the number of completes achieved per shift within each week. There is no evidence that further on during fieldwork, a greater number of completes were achieved per shift based on fieldworkers becoming more familiar with the project and therefore more efficient. This may, however, be different over a longer fieldwork period.

Table 17: Number of shifts, completed questionnaires, and completed questionnaires per shift in each week of fieldwork.

Week of fieldwork	No. of shifts conducted in that week	No. of completes within that week from the shifts conducted in that week	No. of completes per shift within that week
17/04/2023-23/04/2023	78	561	7
24/04/2023-30/04/2023	100	590	6
01/05/2023-07/05/2023	85	678	8
08/05/2023-14/05/2023	71	493	7
15/05/2023-21/05/2023	85	640	8
22/05/2023-28/05/2023	113	728	6
29/05/2023-04/06/2023	58	357	6
05/06/2023-11/06/2023	146	1007	7
12/06/2023-18/06/2023	141	797	6

The following tables (tables 18 and 19) show response rate and completes per shift by individual TOC as well as TOC type for the two methodologies. On board performs almost always better except for the two regional TOCs of Merseyrail and Northern for which at station generated a higher response rate in the field trials. Variations are driven by various factors such as crowdedness, interviewer experience, etc. Crowdedness is likely to be the most impactful factor and in the pilot survey and we can create links to passenger volume

from MOIRA.

Table 18: Strike rate* and completed questionnaires per shift by TOC for different recruitment methodologies

**Strike rate is the total completed questionnaires from that TOC / the total number of recruits from the TOC shifts (even if passengers from other TOCs were recruited, especially for station shifts).*

	Strike rate* at Station	Strike rate* on board	Completed questionnaires per shift at station	Completed questionnaires per shift on board
Avanti West Coast	8%	16%	4.1	7
CrossCountry	5%	22%	2.7	8.5
East Midlands Railway	6%	25%	2.6	7.7
Grand Central	2%	31%	1.3	11.5
Heathrow Express	3%	15%	1.6	5.1
London Overground	8%	13%	3.7	4.4
Merseyrail	15%	15%	4.7	4.7
Northern	14%	35%	6.6	9.2
ScotRail	16%	22%	4.6	6.2
Southwestern Railway	15%	18%	7.2	7.9
Thameslink	7%	20%	4.7	9.1
Transport for Wales	12%	28%	5.3	9.4

Table 19: Strike rate* and completed questionnaires per shift by TOC type for different recruitment methodologies

**Strike rate is the total completed questionnaires from that TOC / the total number of recruits from the TOC shifts (even if passengers from other TOCs were recruited, especially for station shifts).*

TOC type (as per proposal)	Strike rate* at station	Strike rate* on Board	Completed questionnaires per shift at station	Completed questionnaires per shift on board
Commute	12%	18%	5	6.4
Airport	3%	15%	1.6	5.1
Interurban	13%	25%	5.3	8
High speed	8%	16%	4.1	7
Long distance	5%	22%	2.7	8.5
Non franchised	2%	31%	1.3	11.5

The following table shows recruits and completes on board per shift by TOC split out by peak, off peak and weekends (late excluded). There is no clear pattern in that for some TOCs recruits and completes per shift is highest (green figures) during peak times, for others they are highest during off peak and yet for others they are highest at the weekend. Plus there is a group of TOCs where recruits are highest during peak times and completes are highest during off peak which makes sense since commuter (peak time travelers) are generally less likely to complete the survey compared to off peak travelers (who are mainly leisure). As such there is no clear indication that on board intercepts perform less well during peak times.

Table 20: Recruits and completed questionnaires per shift for each TOC split by peak and off-peak times

	Peak (6.00-8:59am, 16:00-18:59pm)		Off peak (00:01am-5:59am, 9:00am-15:59pm)		Weekends	
	Recruits per shift	Completed questionnaire per shift	Recruits per shift	Completed questionnaire per shift	Recruits per shift	Completed questionnaire per shift
Avanti West Coast	49	7	56	8	56	4
CrossCountry	52	10	50	9	55	9
East Midlands Railway	41	8	51	8	47	8
Grand Central	72	15	58	13	57	10
Heathrow Express	52	6	45	6	49	5
London Overground	38	4	42	5	59	6
Merseyrail	33	4	29	6	31	4
Northern	53	10	46	8	35	6
ScotRail	31	8	26	6	34	7
South Western Railway	48	7	42	8	51	8
Thameslink	58	10	72	11	59	8
Transport for Wales	52	13	41	9	35	7

Recruits per shift can be as high as over 180 for on board and up to around 90 for at station, so a larger number of recruits per shift can be achieved on board (in line with the higher average

number of recruits). There are no identifiable patterns regarding recruitment ranges by TOC.

Completes per shift range from 0 to 28 for on board and from 0 to 26 for at station. A larger proportion of shifts achieve 20+ completes on board than at station (in line with the higher average number of completes). Here too, are no identifiable patterns regarding completes by TOC.

Conclusion: On board sampling generates a higher response than at station (but this is largely due to a higher number of recruits). The intervention that delays email completion has an even better response rate and almost certainly provides better quality data. Those that select QR code as a response option are much more likely to drop out of the survey generally and drop out earlier in the survey than those that select email as a response option.

3.16 Weighting Efficiency

Weighting efficiency in this instance comprises two distinct elements:

- The impact from weighting the data to daypart and station size band from MOIRA and age group and gender from the footfall counts (plus potentially ticket type from LENNON if the initial data suggests there is a response bias for this).

- The impact from using a two stage cluster sampling process in the sample design.

Relating to the first point, weighting data always reduces the effective sample size upon which any confidence intervals should be based. Most market research software allows effective sample sizes to be calculated for any subgroup of the data and the weighting efficiency generated (weighting efficiency is the ratio of the effective sample size to the actual sample size and should ideally be above 70%). It should be borne in mind that with similar sample size targets per TOC, the combined results across all TOCs will be severely weighted due to the very different passenger numbers in each TOC. This should not be of great concern as the key outputs of the survey are results for each TOC. However, the impact of weighting on non TOC results, such as regional or national totals will be substantial.

When weighting the field trial data (as per separate weighting document), a couple of issues were encountered:

1. It was not possible to weight by TOC size based on passenger journeys because three of the TOCs made up over 60% of the total journey volume which would have required extreme upweighting of the respondents that used them; weighting by TOC size was also

unrealistic for the field trials since only 12 TOCs were sampled and the share of volume would have been very different had all TOCs been included. Furthermore, the number of shifts per TOC in the field trials were balanced, whereas in a future survey shift numbers would be more aligned to TOC size.

2. Two of the at station weighting cells did not have any sample (e.g. EMR at station size 1) which means that for the at station approach some of the weighting cells had to be merged i.e. station size 1 and 2, station size 3 and 4, weekend days and off peak and late

When merging these weighting cells and without weighting by TOC size, on board achieved a weighting efficiency of 64.2, whereas at station achieved a weighting efficiency of 55.5.

It should be borne in mind that the formulae used to calculate effective sample sizes and hence weighting efficiencies are based upon the assumption that the sample design is a simple random sample. This, of course, is not the case and for both at station and on board; the sample design is a two stage cluster sample. The clusters are a station on a given day at a given time or a specific train service and for each of these a sample of the passengers using the station/train is selected. The cluster design adds further variance to the effective sample size calculation.

As an example, assume an on board sample comprises two trains and 10 passengers complete questionnaires on each. Train 1 is delayed and all respondents rate punctuality as poor; train 2 runs to its schedule and all the passengers rate the punctuality as very good. There is no variance of data within each cluster and so the effective sample size is two rather than 20. If there is variation within each journey, then the effective sample size will be between two and 20 but it will only get to 20 if there is no variation between the two clusters (i.e., each train journey has the same % rating as good).

We can calculate the variance of the cluster sample design from first principles and have done this for the punctuality question for each TOC participating in the field trials, for at station and on board separately.

Table 21: Variance of responses to the punctuality question for different recruitment methodologies by TOC

Punctuality/clustering effect`	At station	On board
Avanti West Coast	58.74%	48.36%
CrossCountry	53.69%	53.94%
East Midlands Railway	54.56%	61.53%
Grand Central	46.73%	44.17%
Heathrow Express	61.16%	67.52%
London Overground	63.85%	58.27%
Merseyrail	58.94%	66.12%
Northern	65.32%	60.88%
ScotRail	62.55%	68.50%
South Western Railways	67.24%	53.69%
Thameslink	56.11%	60.28%
Transport for Wales	58.96%	58.13%
mean	58.99%	58.45%
median	58.95%	59.28%

The figures shown are the ratio of the cluster sampling confidence interval to that from a simple random sample of the same total sample size. Across all TOCs the clustering effect averages just under 60% for both at station and on board (the lower the figure the bigger the impact of the clustering). For a few TOCs there are differences in the effect between at station and on board but for most the results are similar. The figures are much lower for Grand Central, a TOC with few stations and few trains, where the potential for a problem at a station or on a train is likely to be greater than for a TOC which has lots of services calling at lots of stations. Having said that, the results for Heathrow Express, another TOC with few stations but lots of trains, are much higher.

The prior expectation was that on measures such as punctuality, on board would have a greater clustering effect than at station, as many passengers would be taking the same journey. However, for the at station approach, if there is a problem at a particular station on a selected sample day, the same issue will apply. In addition, not all individuals have the same value system and one may rate a 15 minute delay as poor and another as satisfactory. These effects seem to generate similar effects for at station and on board sampling.

The routine to estimate the clustering effect is time consuming in terms of calculations, as it is processing a lot of information to create the results of the formulae. A spreadsheet has been developed where any question can be selected, and any TOC, and the spreadsheet works out the clustering effect for both at station and on board for that combination of question and TOC. Assessments of other questions can be undertaken using this tool on request and that for overall satisfaction is shown in table 22.

Table 22: Cluster effect of overall satisfaction question by TOC and recruitment methodology

Overall satisfaction/clustering effect	At station	On board
Avanti West Coast	67.96%	56.12%
CrossCountry	55.75%	54.32%
East Midlands Railway	65.73%	61.57%
Grand Central	50.65%	50.48%
Heathrow Express	60.40%	67.18%
London Overground	62.44%	59.96%
Merseyrail	57.71%	69.42%
Northern	67.66%	63.03%
ScotRail	58.12%	69.50%
South Western Railways	67.15%	53.67%
Thameslink	62.46%	62.03%
Transport for Wales	60.16%	61.20%
mean	61.35%	60.71%
median	61.42%	61.39%

The clustering effects are slightly more modest but the patterns are the same – no difference between at station and on board with Grand Central considerably lower than the other TOCs.

Conclusion – The weighting efficiency without weighting by TOC size and by merging some of the weighting cells is notably higher for on board, but the clustering effect is very similar across both methods.

3.2 Questionnaire

3.2.1 Knowing the exact train respondent was travelling on

The journey details determine which TOC the respondent data should be attributed to and which part of the TOC's network route the passenger evaluates. This in turn will help the TOC, among other things, to make improvements to the specific section of its network where necessary. As such it is very important to know the exact train the respondent was travelling on.

When this criterion was developed in phase 1 the question emerged if 100% accuracy is a realistic target since all stakeholders ideally wanted to get 100% accuracy. Using the on board approach 100% accuracy (or a figure very close to it) will be possible.

For the on board approach the interviewer enters the TOC of recruitment into the recruitment script. This information comes from the sample plan and the sample plan will have 100% accuracy (based on where the data is taken from – see 'Sampling and weighting' section). Once a respondent is recruited on board, the TOC of recruitment is automatically piped into their script. There are three reasons as to why the respondent train could not be 100% accurate:

- The interviewer entered the incorrect TOC into the recruitment questionnaire – very unlikely.
- The respondent selected the incorrect time for their journey from the Journey Picker Tool (JPT) (e.g. if the journey is run every 30 minutes or even more frequently) – TOC still correct; very rare.
- The respondent enters another journey from the same TOC – TOC still correct; very unlikely.

From the field trial data it cannot be verified how often the first incidence occurred (respondents selecting the incorrect time or another journey is also rare – see 'Journey Picker Tool' section). But the assumption can be made that it would not be very often, if at all, which means that for the on board method journey accuracy will be 100% or very close to that.

With at station recruitment the accuracy can be lower, since for that approach the TOC of recruitment is not piped into the respondent's script but only the station of recruitment (which can also be entered incorrectly by the fieldworker). So, if multiple TOCs depart from the recruitment station that go to the same destinations for part of their network, the probability of the respondent selecting an incorrect journey would be higher. Even though respondents select their journey (or the journey they think they made) from a relatively small range of options in the JPT, they can still select an incorrect journey or believe their journey is not listed and therefore not find their journey.

Based on the field trials data 684 respondents could not find their journey in the JPT. This equals 10% of all respondents who used the JPT, and older passengers (60+ years) are more likely not to find their journey in the JPT than younger passengers (19-34 years) seemingly struggling with the tool a little more. It has to be borne in mind that with the paper questionnaire the respondent's journey is not validated in the JPT.

Conclusion: With on board we are much more likely to know the train the respondent was travelling on since the fieldworker also travelled on that train and details can be checked against the shift plan and the information entered into the recruitment script by the fieldworker.

3.2.2 Accuracy of other information about the journey undertaken

Within previous rail customer experience surveys like the NRPS only one leg of a journey has been assessed to a large extent so that the response can be attributed to a single TOC. If assessing a whole journey (as opposed to a single leg) this can no longer be attributed to a single TOC if more than one TOC was used and it would make it more difficult to verify the journey with the Journey Picker Tool (JPT) through which respondents can only select journey legs rather than whole journeys (which is one reason why some respondents cannot find their journey because they are looking for the whole journey rather than a leg). Consequently it would make more sense to continue with a journey leg assessment rather than the assessment of a whole journey (even though a leg can be the same as a whole journey if there is no change in train).

For those that complete the survey after their journey is finished, it can be assumed that accuracy of other information about the journey can be very high regardless of the method employed since the majority would have made and completed the journey on the same day. Saying that, amongst those that were recruited on board, 3,076 (77%) completed the survey on the same day as their journey, of which 640 (21%) were by email and 2,436 (79%%) were by QR. This compares to 2,158 (71%) amongst those recruited at station, of which 733 (34%) were by email and 1,425 (66%) were by QR. Based on this, it would appear that on board provides greater accuracy since respondents tend to complete the survey sooner after their journey. Those that selected QR code as a response option are also more likely to answer the survey sooner after their journey.

One of the main challenges during the field trials was that respondents completed the survey before the end of their journey, meaning that some of their responses did not relate to the journey they were recruited for but to the same journey made previously (e.g. the previous day, two days ago, the previous week, etc. or generally to previous experience with the same journey). This issue had also emerged in previous trials, e.g. the multi-method approach

project conducted on behalf of Transport Focus in 2021. By method the proportion of those that completed the survey early shows a little variation with 32% completing early if recruited on board and 28% completing early if recruited at station.

If recruited at station, respondents are able to complete more stages of their journey early i.e. from another journey due to the fact that they are recruited so early in their journey compared to those that are recruited on board. Those recruited on board will have had at least some on train experience whereas those that are recruited at station may not have had any if they completed the survey before getting onto the train. On that basis greater accuracy can be expected from the on board recruitment approach.

Neither approach is more bias than the other based on responses given since the script was the same for both methods and answers were equally randomised. There is a greater likelihood that the paper questionnaire generated bias data for questions where answer options were not randomised.

Curiously, there was a very small proportion of respondents (1%) recruited on board that answered the survey prior to the start of their journey which implies that they did not respond for the leg on which they were recruited for but picked a later train journey (although for the same TOC) from the JPT. This proportion compares to 7% of respondents who were recruited at station and completed the survey before the start of their journey.

Conclusion: On board generates greater accuracy about journey aspects than at station.

3.2.3 Questionnaire length

Both intercept approaches can deal with longer questionnaires equally well. From the qualitative element, feedback was that the questionnaire was very long. It also caused some respondents frustration that the interviewer said at the point of recruitment the questionnaire would be 10 minutes long and then it turned out to be closer to 15 minutes when respondents completed the survey. It appears that response rate could be improved if the questionnaire was shortened.

The median for the online survey completion is 15 mins 58 secs. For those who selected email as a response option it's 17 mins 02 secs, for those who selected QR code 15 mins 05 secs. QR respondents being considerably younger than email respondents is likely to be the main reason for the faster completion within that group.

Breaking the questionnaire down

A modular approach was not tested in the field trials but came out as a suggestion from the

qualitative element. Breaking down the questionnaire could take a number of forms.

- The questionnaire is broken down into two versions; key information and metrics will be asked in both versions, but other 'secondary' data is only asked in one version or the other; this is expected to reduce questionnaire length by around a third.

- Respondents are asked to answer key information and metrics and then asked if they are happy to answer further questions about their journey; this more detailed part will be entirely 'voluntary'; it is expected that the first 'key' part will take around seven minutes to complete with the second voluntary part to take another 6-8 minutes.

    - This modular approach might create a bias towards more polarized experiences i.e. mainly those who had a particularly poor or good experience will complete the voluntary questions whilst those who had an 'average' experience less willing to complete the second additional part although this did not emerge from intervention 5 which was a modular approach; but the approach would need to be trialed on a larger scale to confirm there is no bias.

- The interviewer asks the passenger five or so key questions and then invites them to complete additional questions about their journey online.

    - This approach may not be practical for a rail customer experience survey, but BVA BDRC uses this approach for the annual Motorway Services User Survey conducted on behalf of Transport Focus and motorway service area operators; the interviewer would enter the answer to the five key questions on their device and for recruitment on board will also be able to add the correct TOC.

    - This approach would only work with on board recruitment since a TOC could not be allocated to the respondent unless they then completed further questions online.

    - This approach, too, would need to be trialed to understand its practicality.

Questions that generate similar data

Within the current questionnaire there are some questions that generate very similar responses and for which responses are strongly correlated. Following some examples are provided.

Table 23. Comparison of responses to questions "Q50. Overall how easy or difficult did you find it to purchase your ticket/pay for this journey?" And "Q51. And overall, how satisfied or dissatisfied were you with the ticket buying process/paying for this journey?"

Red and a minus (-) sign or blue and a plus (+) sign show statistical significance at 95% confidence.

Column % n	Very easy	Fairly easy	Neither easy nor difficult	Fairly difficult	Very difficult	NET
Very satisfied	83% (+)	13% (-)	5% (-)	2% (-)	0% (-)	60%
	3731	211	13	2	0	3957
Fairly satisfied	13% (-)	73% (+)	23% (-)	19% (-)	6% (-)	29%
	598	1221	60	21	2	1902
Neither satisfied nor dissatisfied	3% (-)	11% (+)	55% (+)	19% (+)	6%	7%
	114	180	145	21	2	462
Fairly dissatisfied	1% (-)	3%	13% (+)	45% (+)	22% (+)	3%
	31	43	35	49	8	166
Very dissatisfied	0% (-)	1%	4% (+)	15% (+)	67% (+)	1%
	21	14	11	16	24	86
NET	100%	100%	100%	100%	100%	100%
	4495	1669	264	109	36	6573

Both questions produce similar results, and it should be reviewed if both are indeed needed.

Table 24. Comparison of responses to questions "Q37. Overall how easy or difficult did you find it to plan your journey? " and "Q38. And overall, how satisfied or dissatisfied were you with the information received when planning your journey?"

Red and a m**inus (-)** sign or blue and a **plus (+) sign show statistical significance at** 95% confidence. No answers and don't knows excluded.

Column % n	Very easy	Fairly easy	Neither easy nor difficult	Fairly difficult	Very difficult	NET
Very satisfied	80% (+)	17% (-)	2% (-)	2% (-)	0% (-)	52%
	2504	317	4	2	0	2827
Fairly satisfied	17% (-)	72% (+)	41% (-)	15% (-)	9% (-)	37%
	547	1355	96	16	3	2017
Neither satisfied nor dissatisfied	2% (-)	8% (+)	43% (+)	29% (+)	9%	7%
	76	152	101	30	3	362
Fairly dissatisfied	0% (-)	2%	12% (+)	44% (+)	11% (+)	2%
	10	40	29	46	4	129
Very dissatisfied	0% (-)	1% (-)	3% (+)	10% (+)	71% (+)	1%
	6	12	6	10	25	59
NET	100%	100%	100%	100%	100%	100%
	3143	1876	236	104	35	5394

Here to, results are similar, and it is recommended to review if both questions are essential. No other questions produce equally similar results, however, to be absolutely sure it is recommended to run a statistical correlation across all questions with similar answers like those in the examples above.

Conclusion: Both approaches can accommodate longer questionnaires equally although a questionnaire of longer than 10 minutes is not recommended.

3.2.4 Completion after journey

An important objective of the rail survey is to gather feedback from passengers based on their actual journey experience at the time of recruitment. Previous research conducted by Transport Focus (multi-method approach project conducted on behalf of Transport Focus in 2021) has highlighted the following issue: when passengers are recruited face to face, either at the station or on board, some individuals tend to complete the survey before reaching the end of their journey. This poses a problem as the questionnaire encompasses various stages of the journey, including arrival station details and post-journey metrics. Moreover, the survey aims to capture any disruptions experienced. If respondents finish the questionnaire early, they are unable to provide accurate information about their complete journey.

The field trials shed light on differences about early completion based on how passengers are recruited i.e. either on board or at the station. For the purpose of analysis, paper questionnaires have been excluded and only respondents who successfully found their journey in the journey picker tool have been included which results in a sample size of 5,000 respondents.

Among respondents recruited at the station, 28% completed the questionnaire before completing their journey. This compares to 32% who were recruited on board and completed the survey before the end of their journey.

Table 25: Questionnaires completed before end of station by recruitment method

	Total Station
Completion in time	1578
Completion before the end of journey	602
	28%

	Total Onboard
Completion in time	1893
Completion before the end of journey	902
	32%

It is important to point out that a proportion of 7% of respondents recruited at the station completed the survey even before their train departed. This implies that these respondents rated the onboard experience, arrival station, and post-journey metrics without having boarded the train for the journey they were recruited for but took their responses from a similar previous journey.

Early survey completion by response option

The data from the field trials shows a notable difference of early completers by response option, with those selecting QR much more likely to complete the survey before the end of their journey compared to those who selected email to get to the survey.

Table 26: Questionnaires completed before end of station by response option

	Email	QR
Completion in time	1699	1769
Completion before the end of journey	48	1455
	3%	45%

Among respondents who chose to access the survey using a QR code, 45% completed the questionnaire before reaching the end of their journey. This high percentage can be attributed to the immediate access provided by the QR code, enabling respondents to start the survey as soon as they scan the code.

By contrast, respondents who opted to receive survey invitations via email experienced a delay of 30 minutes between the time of agreeing to participate and when the email invitation was sent out. This deliberate time gap proved to be effective in ensuring that individuals completed the survey after their journey (only 3% completed before). The delay allowed respondents to board the train/settle into their journey and have a more accurate and comprehensive understanding of their experience before initiating the survey.

These findings highlight the importance of considering the mechanisms employed for survey completion. While the convenience of QR codes appeals to respondents (and some that would generally not take part in rail surveys), it also increases the likelihood of early completion.

There is, however, the option to add a time delay parameter to the QR code links which would hold the respondent at a landing page. The landing page would display a countdown and then automatically redirect them to the survey. This option would need to be trialed to assess how this will impact on response rate.

Reasons for early survey completion from the qualitative element

In parallel to the field trials, some qualitative research was conducted to try to understand why these passengers were completing the questionnaire before the end of their journey and how they answered questions about elements they had not yet experienced.

Respondents who had completed the survey prior to the end of their journey, generally expressed that they were able to do so based on their knowledge of the train and stations and from previous journeys. Many of them were frequent travelers and believed that they had sufficient experience to answer the survey adequately.

Their reasons for completing the survey before ending the journey included:

- Scanning the QR code and completing the survey immediately to get on with it.
    - "Scanned the QR on the platform and then just started it."
- Having familiarity with the journey and feeling confident that they could answer all the questions.
    - "The previous day I had travelled the same route, so I kind of knew the questions about the station without being physically there."
    - "In some sections I had to use my own knowledge, which is quite good because it is a trip I take 4 times a week. But if I had been there maybe I would have got more information. "
    - "Did it on the train home - based response on previous journeys made of the same trip." Being unaware that the survey was intended to be completed after the journey
    - "I completed it before because I didn't know it would be best for me to do it after, at the end of the journey."
    - "Don't remember being told I should finish it once at destination - received a link via email and started."
- Choosing to utilise downtime on the train to complete the survey instead of postponing it until later, e.g. during delays, as when reaching their destination passengers returned to their 'busy' lives again and are less likely to take part.
    - "Maybe when you get off the train you don't want to do a survey, especially if it's going to take a while, and you forget about it, and don't care about it (anymore). "

The impact of completing early on accuracy about train punctuality

An important aspect that affected respondents completing the questionnaire before the end of their journey is their ability to accurately record whether their train experienced any delays. It is crucial to assess the extent of this issue to understand its impact on data reliability.

Analysis of the field trial results demonstrates that out of all the completed questionnaires, only 60% of respondents recorded the status of their train correctly. This indicates that a considerable proportion of respondents may have experienced challenges in accurately capturing their train's delay status.

Table 27: Respondents recall of train delay status by completion time

	Overall	Completion in time	Completion before the end of the journey
Incorrect train status with regard to being delayed	3114	2287	830
All	5185	3695	1441
	60%	**62%**	**58%**

58% of respondents who answered the questionnaire before the end of their journey incorrectly recorded their train's status, compared to 62% of those who waited until the end. This is based on absolute punctuality (i.e. includes those that recorded it incorrectly because their train was one minute late). If allowing a few minutes leeway i.e. adding the next range option either way, the proportions change to 84% (amongst those completing early) and 88% (amongst those completing after the journey). The pattern is similar by recruitment method (on board or at station).

There is only a small difference between those completing early versus those completing after their journey when it comes to recording the correct delay status of their train, indicating that respondents who completed the questionnaire early were still able to provide relatively accurate information regarding train delays.

Amongst those whose train was delayed (by any time i.e. even one minute) only 35% recorded it as delayed. The reason for that is that within that group a high proportion experienced a delay of less than five minutes of which 79% recorded their train as not being delayed. Taking those with less than a five minute delay out of the equation, then amongst those whose train was delayed 70% record it as delayed. Amongst all respondents 22% record no delay when their train was actually delayed (because of those with less than a five minute delay).

The impact of the interventions on early survey completion

Another objective of the field trials was to prevent respondents from completing the survey before the end of their journey. To address this, in the second half of fieldwork, five interventions were implemented to test if they were able to prevent early completion. These were:

1. Delayed email: length determined by sector used i.e. longer delay for journey on a long distance TOC.

2. Warning message if respondent tries to complete less than 30 minutes after recruitment.

3. Next button disabled for 30 minutes once respondent accessed the intro page.

4. On post journey sections, add a message to ask for the questions to only be answered after the journey.

5. Asking passengers at the beginning of the survey where they are at in their journey and only ask them about their experience up to that point; they are then asked to record their email address and are sent a link later for them to finish the survey off once their journey is over; the email is sent automatically after a certain number of minutes.

Table 28 below shows a summary of the interventions' success. Subsequently each intervention is analysed in turn.

Table 28: Summary of performance of interventions on early survey completion

Intervention 1 - Email delay	Intervention 2 - Warning message when clicking next button amongst those that select QR code	Intervention 3 - Next button disabled amongst those that selected QR code	Intervention 4 - Key message to answer post journey sections only after the journey	Intervention 5 - Allow part survey to be completed later
Was used only on shifts conducted on board train; delay dependent on TOC type	Was used during shifts conducted on board trains and at station	Was used during shifts conducted on board trains and at station	Was used during shifts conducted on board trains and at station	Was used during shifts conducted on board trains and at station
not particularly effective but that is mainly due to the fact that the issue of completing early is less of an issue amongst those that select an email as a response option but amongst those that select QR.	Did not particularly help in increasing the proportion of respondents completing the survey after their journey and had a negative impact on the number of QR completes, reducing the response rate of total validated respondents that completed after the end of their journey by a small number of percentage points.	Helped notably in preventing people from completing the survey before the end of their journey, but it reduced the response rate to such an extent that ultimately it is better to not apply this intervention as the overall response rate of those completing the survey after the end of their journey is higher without the intervention.	Was not every effective in preventing respondents from completing the survey before the end of their journey.	Was successful at reducing the number of respondents completing before the end of their journey but it does reduce the total number of completes.

Intervention 1 – email delay

This intervention was used only during shifts conducted on board trains. This is because at the station there was no way of knowing which TOC the passenger was about to take. In the recruitment questionnaire that piece of information was not asked. For onboard shifts we knew which train the interviewer was boarding (and therefore which TOC was used) and could then feed that information through for this intervention.

In total 21 shifts were conducted for this intervention. The intervention was only used for respondents who chose to receive a link to the survey via emails. Those that chose QR proceeded as normal.

If a respondent chose email, the system checked which TOC that respondent was on and delayed the email invitation based on the sector the TOC operated in:

- Initial invitation - Intervention 1 - Long distance: 120 min after recruitment
- Initial invitation - Intervention 1 - London & SE: 60 min after recruitment
- Initial invitation - Intervention 1 - Regional: 90 min after recruitment

Across the 21 shifts, 354 agreed to take part via email of which 74 completed the survey. That is a response rate of 21%, slightly better than for other onboard shifts without interventions.

Table 29: Response rates for intervention 1 and relevant field trial type of method

Screener	Blue - on board no paper	Yellow - on board with paper back up	Inter 1 - email
# Shifts	340	91	21
Email recruited	4623	1226	354
Email complete	778	208	74
Total response rate	17%	17%	21%
Recruits per shift	13.6	13.5	16.8
Completes per shift	2.3	2.3	3.5

Of the 74 completed questionnaire 48 were done after the end of the journey and one was filled in before the end of the journey. (The remainder use the JPT incorrectly, so it was

impossible to determine whether they completed before or after the end of their journey.) This means 98% of validated completes were done in time after the end of the journey. This is in line with the results from other on board shifts (without interventions):

Table 30: Rate of completed surveys in-time for intervention 1

Screener	Blue - on board no paper	Yellow - on board with paper back up	Inter 1 - email
Total email complete in time	624	179	48
Total email completes early	5	2	1
Total completes unknown (no JPT info)	*149*	*27*	*25*
Total share of completes in time (from known arrival time)	99%	99%	98%
Total verified completes in time response rate	13%	15%	14%

This means that the intervention is not particularly effective but that is mainly due to the fact that the issue of completing early is less of an issue amongst those that select email as a response option but amongst those that select QR.

Intervention 2 – Warning message when clicking next button amongst those that selected QR code

This intervention was used during shifts conducted on board trains and at station.

In total 53 shifts were conducted with this intervention. The intervention was only impacting respondents who chose to access the survey through a QR code. Those that chose email proceeded as normal.

If a respondent selected QR code, the survey script checked that the respondent waited for 30 minutes before starting the survey. There was a message in the introduction page asking respondents to wait before starting the survey to ensure they had experienced all aspects of their train journey. If a QR respondent tried to proceed before the end of the 30 min, a warning message would appear asking respondents to wait a little longer. If the respondents tried to proceed despite of this, they were allowed to access the survey.

Across the 54 shifts, 1,377 respondents agreed to take part via QR of which 208 completed the

survey. That is a response rate of 15%, in line with the other interventions.

Table 31: Response rates for intervention 2 and relevant field trial type of method

Screener	Red - at station no paper	Green - at station with QR postcard	Blue - on board no paper	Yellow - on board with paper back up	Inter 2 - QR next button message	Inter 2 - at station	Inter 2 - on board
# Shifts	357	89	340	91	54	*24*	*30*
QR recruited	7976	1955	11013	3026	1377	*643*	*734*
QR complete	1115	260	1813	483	208	*89*	*119*
Total response rate	14%	13%	16%	16%	15%	*14%*	*16%*
Recruits per shift	22.3	22.0	32.4	33.3	25.5	*26.8*	*24.5*
Completes per shift	3.1	2.9	5.3	5.3	3.8	*3.7*	*3.9*

Of the 208 completes 103 filled in the survey after the end of their journey and 80 completed the survey before the end of the journey. This means 56% of the validated completes were done in time. This is in line with results from other onboard shifts (without interventions).

Table 32: Rate of completed surveys in time for intervention 2

Screener	Red – at station no paper	Blue - on board no paper	Inter 2 - QR next button message	*Inter 2 - at station*	*Inter 2 - on board*
Total complete in time	561	866	103	*45*	*58*
Total completes early	439	704	80	*32*	*48*
Total completes unknown (no JPT info)	*115*	*243*	*25*	*12*	*13*
Total share of completes in time (from known arrival time)	56%	55%	56%	*58%*	*55%*
Total verified completes in time response rate	7%	8%	7%	*7%*	*8%*

It can be concluded that this intervention did not particularly help in increasing the proportion of respondents completing the survey after their journey and had a negative impact on the number of QR code completes, reducing the response rate of total validated respondents that completed after the end of their journey by a small number of percentage points.

Intervention 3 – Next button disabled amongst those that selected QR code

This intervention was used during shifts conducted on board trains and at station.

In total 54 shifts were conducted with this intervention. The intervention was only impacting respondents who chose to access the survey via QR code. Those that chose email proceeded as normal.

If a respondent selected QR code, the survey script checked that the respondent waited for 30 minutes before starting the survey. There was a message on the introduction page asking respondents to wait before starting the survey to ensure they had experienced all aspects of their train journey. Respondents couldn't proceed for 30 min as the 'Next' button was disabled.

Across the 54 shifts, 1,374 respondents agreed to take part via QR code of which 85 completed the survey. That is a response rate of 6%, around half of what was achieved in shifts without interventions.

Table 33: Response rates for intervention 3 and relevant field trial type of method

Screener	Red - at station no paper	Blue - on board no paper	Inter 3 - QR next button disabled	*Inter 3 - at station*	*Inter 3 - on board*
# Shifts	357	340	54	*31*	*23*
QR recruited	7976	11013	1374	*688*	*686*
QR complete	1115	1813	85	*39*	*46*
Total response rate	14%	16%	6%	*6%*	*7%*
Recruits per shift	22.3	32.4	25.4	*22.2*	*29.8*
Completes per shift	3.1	5.3	1.6	*1.3*	*2.0*

Of the 85 completes 74 did the survey after the end of their journey and four filled the survey in before the end of the journey. This means 95% of validated completes were done

in time. This is substantially higher than in other shifts (without interventions).

Table 34: Rate of completed surveys in time for intervention 3

Screener	Red - at station no paper	Blue - on board no paper	Inter 3 - QR next button disabled	*Inter 3 - at station*	*Inter 3 - on board*
Total complete in time	561	866	74	*34*	*40*
Total completes early	439	704	4	*1*	*3*
Total completes unknown (no JPT info)	*115*	*243*	*9*	*4*	*3*
Total share of completes in time (from known arrival time)	56%	55%	95%	*97%*	*93%*
Total verified completes in time response rate	7%	8%	5%	*5%*	*6%*

As such the intervention helped notably in preventing people from completing the survey before the end of their journey, but it reduced the response rate to such an extent that ultimately it is better not to apply this intervention as the overall response rate of those completing the survey after the end of their journey is higher without the intervention.

Intervention 4 – Key message to answer post journey sections only after the journey

This intervention was used during shifts conducted on board trains and at station.

In total 47 shifts were conducted with this intervention. The intervention impacted all respondents. Additional intro screens were added to the script reminding respondents they should only complete the questionnaire section after they had experienced that stage of their journey.

Across the 47 shifts, 1,912 respondents agreed to take part via email or QR code of which 333 completed the survey. That is a response rate of 17%, in line with shifts without interventions.

Table 35: Response rates for intervention 4 and relevant field trial type of method

Screener	Red - at station no paper	Blue - on board no paper	Inter 4 - key message	*Inter 4 - at station*	*Inter 4 - on board*
# Shifts	357	340	47	*26*	*21*
Total recruited	12638	15636	1912	*1106*	*806*
Total complete	1989	2591	333	*178*	*155*
Total response rate	16%	17%	17%	*16%*	*19%*
Recruits per shift	35.4	46.0	40.7	*42.5*	*39.4*
Completes per shift	5.6	7.6	7.1	*6.8*	*7.4*

Of the 333 completes 208 were completed after the end of their journey and 65 were done before the end of the journey. This means 76% of the validated completes were done after the journey. This is in line with results from the other shifts (without interventions).

Table 36: Rate of completed surveys in time for intervention 4

Screener	Red - at station no paper	Blue - on board no paper	Inter 4 - key message	*Inter 4 - at station*	*Inter 4 - on board*
Total complete in time	1286	1490	208	*102*	*106*
Total completes early	467	709	65	*41*	*24*
Total completes unknown (no JPT info)	*236*	*392*	*60*	*35*	*25*
Total share of completes in time (from known arrival time)	73%	68%	76%	*71%*	*82%*
Total verified completes in time response rate	10%	10%	11%	*9%*	*13%*

This means the intervention was not very effective in preventing respondents from completing the survey before the end of their journey.

Intervention 5 – Allow part survey to be completed later

This intervention was used on shifts conducted on board trains and at station.

In total 53 shifts were conducted with this intervention. The intervention was impacting all respondents. An additional question was placed upfront in the questionnaire asking passengers to answer the survey up to where they were in their journey (at the station, on board, or at the arrival). If the respondent said they were part way in their journey, the script allowed them to complete the survey on elements they had experienced already. Then it asked them for their email and sent a link with the rest of the questionnaire 30 min later.

Across the 53 shifts, 2,079 respondents agreed to take part via email or QR code of which 180 completed the survey. That is a response rate of 9%, considerably lower than in shifts with no interventions.

Table 37: Response rates for intervention 5 and relevant field trial type of method

Screener	Red - at station no paper	Blue - on board no paper	Inter 5 - allow part survey	*Inter 5 - at station*	*Inter 5- on board*
# Shifts	357	340	53	*32*	*21*
Total recruited	12638	15636	2079	*994*	*1085*
Total complete	1989	2591	180	*86*	*94*
Total response rate	16%	17%	9%	*9%*	*9%*
Recruits per shift	35.4	46.0	39.2	*31.1*	*51.7*
Completes per shift	5.6	7.6	3.4	*2.7*	*4.75*

Of the 180 completes 149 were completed after the journey end and 10 were completed before the end of the journey. This means 94% of the validated completes were done after the end of the journey. This is a lot higher than in shifts with no interventions.

Table 38: Rate of completed surveys in time for intervention 3

Screener	Red - at station no paper	Blue - on board no paper	Inter 5 - allow part survey	*Inter 5 - at station*	*Inter 5 - on board*
Total complete in time	1286	1490	149	*67*	*82*
Total completes early	467	709	10	*6*	*4*
Total completes unknown (no JPT info)	*236*	*392*	*21*	*13*	*8*
Total share of completes in time (from known arrival time)	73%	68%	94%	*92%*	*95%*
Total verified completes in time response rate	10%	10%	7%	*7%*	*8%*

This mean this intervention was successful at reducing the number of respondents completing before the end of their journey, but it does reduce the total number of completes.

786 passengers started the survey with Intervention 5, of which 268 were at their departure station and 386 were on board (with the rest having completed their journey - 132 respondents). Stopping respondent halfway through the survey to ask for their email meant a considerable increase in dropouts (at the question when asking respondent for their email) – the dropout rate was 38%.

Feedback on the interventions from the qualitative element

From the qualitative depths/groups that were conducted, this was by far the most discussed intervention.

- "Complete in stages would work well."
- "It wouldn't have discouraged me. I would want to finish it."
- "It would be like a double hurdle to go through, rather than number 1, that you can do once you receive it."

Notably, intervention 5 received the highest number of comments and stood out prominently due to its ability to facilitate completion by stage and provide timely reminders to ensure survey completion. In the quantitative data intervention 5 recorded the largest proportion of dropouts.

- "The preferred one would be number 5, I can come back to it but not have to think about it."
- "5 could work well. If there's a way, we could complete half and then come back to it with a link sent then that could work well."
- "I think intervention 5 would be the best. I don't have to think or take action, the link will remind me."
- "You let the passenger complete as much as possible, fully stop them and then follow up by email."
- "You could complete in stages which would work well."
- "The fifth one is the one that stands out. It does a couple of things, it breaks up the survey, you don't get the feeling you sat there for ages answering it."

In the group discussions, respondents engaged in a short brainstorming activity. Respondents were asked to jot down any ideas they had for improving the interventions or suggesting new interventions that would prevent people from completing the survey before reaching the end of their train journey. These ideas were later explored and discussed.

The following suggestions emerged from the discussion:

Share mobile numbers to receive a text message or a WhatsApp reminder, instead of relying solely on email to address the concern of emails potentially going to the spam folder.

- "Maybe sharing my mobile number to get a reminder to complete the survey, instead of the email because it could get spammed or (land) in junk mail."

- "WhatsApp/text message received soon after the journey, to avoid email going to the junk. Also, maybe social media in general."

There is a risk that those returning to complete the survey are those that have something to say/had a bad experience with their journey. But looking at the overall satisfaction that is not the case.

Table 39: Comparison of satisfaction for intervention 5 respondents with all respondents

Q89: Overall, taking everything into consideration, how satisfied or dissatisfied were you with your train journey from… ?		
Column %	All respondents (no interventions)	Respondents from Intervention 5 shifts that started the survey before finishing their journey and that came back to the link 30min later to finish the Survey
n		
NET: satisfied	85.5%	84.9%
	4,980	79

Conclusion: As mentioned in a previous section, by method the proportion of those that completed the survey early shows a little variation with 33% completing early if recruited on board and 29% completing early if recruited at station. The issue with completing early seems to lie rather with the response option than with the recruitment method with those that select QR code to access the survey much more likely to complete the survey before the end of the journey than those that are emailed a survey link.

3.2.5 Paper surveys

The table below shows that the paper option used on board did attract a different demographic, much more skewed towards older users of the rail network. However, only 92 of the 787 respondents who were offered a paper questionnaire as a backup option took this option (to be fair the digital options were prioritised).

The first column – no paper option – gives the profile of all respondents on board who were just given digital options. The next column – paper option – shows the profile of those respondents on board who were given a paper option as a backup. The next two columns split this into those that used the digital options and those that used the paper option.

Table 40: Demographic profile of paper options compared to digital options for on board shifts 47%

Red and a minus (-) sign or blue and a plus (+) sign show statistical significance at 95% confidence.

Age group	No paper option	Paper option	Paper	Digital	Footfall counts
Sample size	3314	787	92	695	71160
16-18	3%	2%	0%	3%	40%
19-24	11%	10%	2% (-)	11% (+)	
25-34	22%	19%	4% (-)	21% (+)	
35-44	18%	17%	3% (-)	19% (+)	47%
45-54	16%	17%	9% (-)	18% (+)	
55-59	9%	8%	9%	8%	
60-64	7%	10%	17% (+)	9% (-)	
65-69	6%	7%	14% (+)	6% (-)	13%
70-80	6%	7%	33% (+)	4% (-)	
81+	1%	2% (+)	7%	1% (-)	
Prefer not to say	1%	2%	2%	1%	
Don't know/ not sure	0%	0%	0%	0%	

There is some evidence that offering a paper option as a backup does result in more older respondents taking part, as the percentages in the paper option column from those aged 60+ is a little higher overall. And when we split this total by the method of completion actually used, those using paper are a much older profile. So, having a paper back up does attract an older demographic but actually pushes the overall % of those aged 65+ a little above what is shown in the footfall counts (16% to 13%, compared to 13% for the no paper option).

As we have seen earlier, the response rate for paper questionnaires is higher than any of the

digital methods, so adding a paper option is likely to increase response rates a little overall. The downside is that paper questionnaires cannot have any routing control, need to be returned by post with subsequent manual data entry and add about one week to the time by which data is available for analysis. The number of responses per shift is substantially higher for the paper back up option – 8.6 compared to 7.2 for the no paper option.

The no paper option attracts more respondents aged 60+ than the digital responses from the paper option (13% v 11%) and although the overall response per shift is a little lower, offering paper does appear to switch some potentially digital responses into that mode.

There is no evidence that a paper questionnaire option results in a larger proportion of disabled respondents as the table below shows. In fact, the proportion of disabled respondents is much higher for QR code and email.

Table 41: The impact of paper on attracting disabled passengers or those that require assistance

Q29: We would like to ask whether you have any disability, which is classified as 'sensitive information'. For this we need your consent. Are you happy for this information to be collected? It's no problem if you prefer not to answer this.					
Column % n	QR	Email	QR Leaflet	Paper	NET
Yes	47%	50%	25%	34%	47%
	2028	1324	3	29	3384
No	53%	50%	75%	66%	53%
	2329	1347	9	57	3742
NET	100%	100%	100%	100%	100%
	4357	2671	12	86	7126

Similarly, there is not a higher share of those travelling with someone who needed assistance or of those travelling without a helper, personal assistant or carer (and for that reason needed assistance) amongst paper respondents compared to QR or email respondents. Again, it is rather the other way around (although in this instance minimally so). As such it can be concluded that a paper questionnaire is not necessary to cover these types of respondents.

Responses from those that completed paper questionnaires are generally considerably more positive than responses from those that completed online. This is linked to the fact that paper respondents are substantially older and older respondents give generally more positive ratings. Saying that, for overall satisfaction the difference is not statistically significant at a 95% confidence level: amongst online respondents total overall satisfaction is 85% (very and fairly satisfied) whereas amongst paper respondents it's 90% (sample size for paper

respondents is n=92). But paper respondents give a much higher rating for trusting the rail industry and for likelihood to recommend travelling by train. Satisfaction for overall punctuality also records a substantial difference between online and paper answers but neither satisfaction with train frequency nor satisfaction with overall value for money show any significant difference.

When we asked paper questionnaire respondents for the reasons they completed the survey on paper, almost half said that completing on paper is their preference. A fifth said they did not have access to the internet and only a small number said they have difficulty completing online surveys.

Conclusion: There is no advantage of one method over the other with regard to paper surveys although the inclusion of paper surveys has some benefits, but these are outweighed by its drawbacks.

3.2.6 Journey picker tool (JPT)

As part of the field trials, we introduced a new survey tool known as the "journey picker" to enhance the accuracy of the journeys respondents were reporting on. This tool is integrated with a live database containing information on all active and past journeys across the British rail network. Its primary objective is to improve journey details provided by respondents and it also allows for additional information to be appended, such as the actual disruption status, number of coaches, and the RID of the train.

The introduction of the journey picker tool aimed to address potential inaccuracies and ambiguities when respondents reported their journey details. Previously, with the NRPS, all journeys had to be manually checked and validated. The JPT allows instant validation.

By providing a comprehensive database of all journeys, respondents had access to reliable information to accurately select and describe their individual journeys.

To evaluate the success of integrating the journey picker into the questionnaire, we analysed the data from the field trials and the following key findings emerged:

- The journey picker generated 2,733 dropouts (35% of all of those who used the tool). This represents 16% of total dropouts and compares to 19% dropping out at the intro screen. 26% dropped out after the journey detail section of which the JPT is part. So, it caused the majority of dropouts in that section but by far not all of them.

- 86% managed to use the tool properly and find their journey.

- 85% when recruited at the station, and
- 88% when recruited on board the train.

This means about 14% failed to use the tool and did not find their journey.

Those who didn't find their journey on the journey picker were asked to leave a comment to help us identify their journey. The common issues respondents had were:

1. Wanting to select the full journey when it involved multi legs.
2. Not finding the exact train time.
3. Not finding the date (when completing the survey a day or so later).

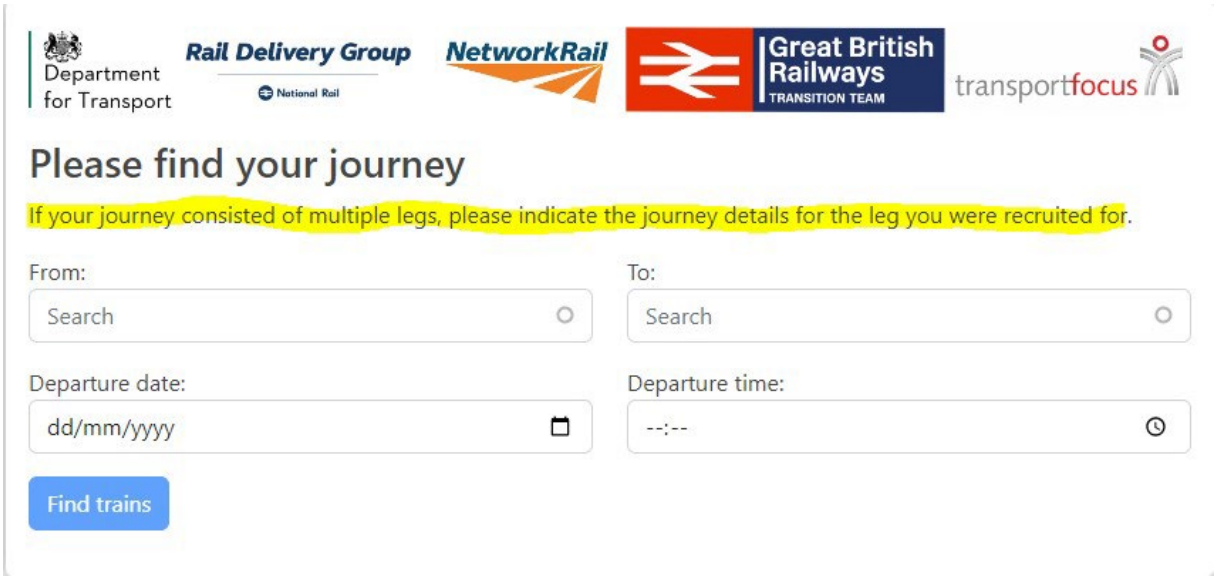
Age ranges amongst those that could not find their journey in the JPT were fairly evenly spread, with 31% 16-34 years old, 40% 35 to 59 years old and 27% 60 years or older. Females are more prevalent amongst those not able to find their journey in the JPT than males at 53% compared to 44%. When used correctly, the journey picker tool significantly improved the accuracy of journey reporting. By accessing the live database, respondents were able to select the exact journey they travelled on, eliminating the potential for subjective interpretations or misinterpretations of route information. This increased precision in journey selection enhances the reliability and quality of the data collected.

The journey picker tool's integration with the live database provided an opportunity to capture additional journey details beyond the standard questionnaire. Respondents could now provide information such as the actual disruption status experienced during the journey, the number of coaches on the train, and the RID of the specific train. This supplementary information enriches the survey data and provides valuable insights into the passenger experience, enabling a more comprehensive analysis of journey related factors.

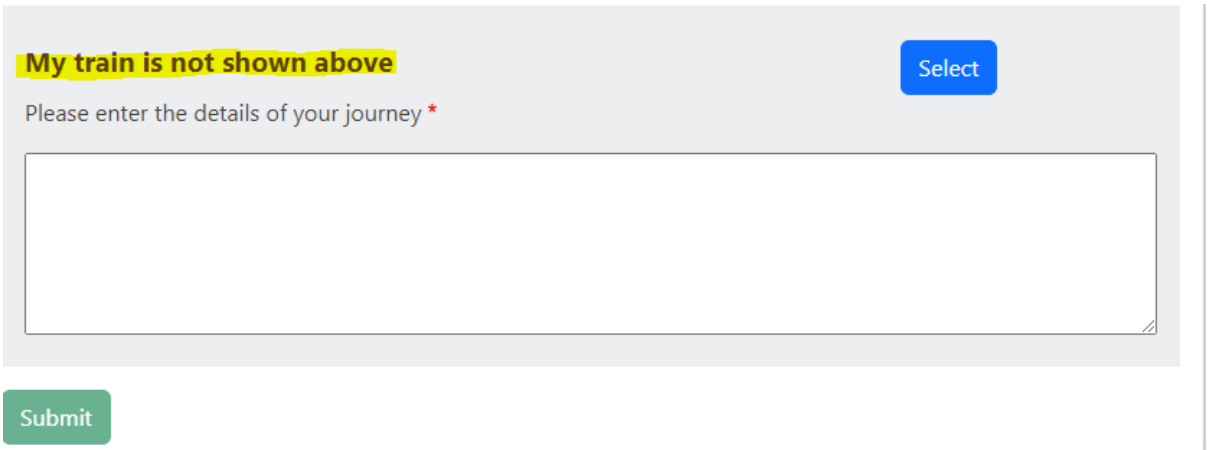
Improvements to the journey picker tool

For the pilot survey we recommend using the tool, but we should improve some elements to help respondents to identify their journey more easily.

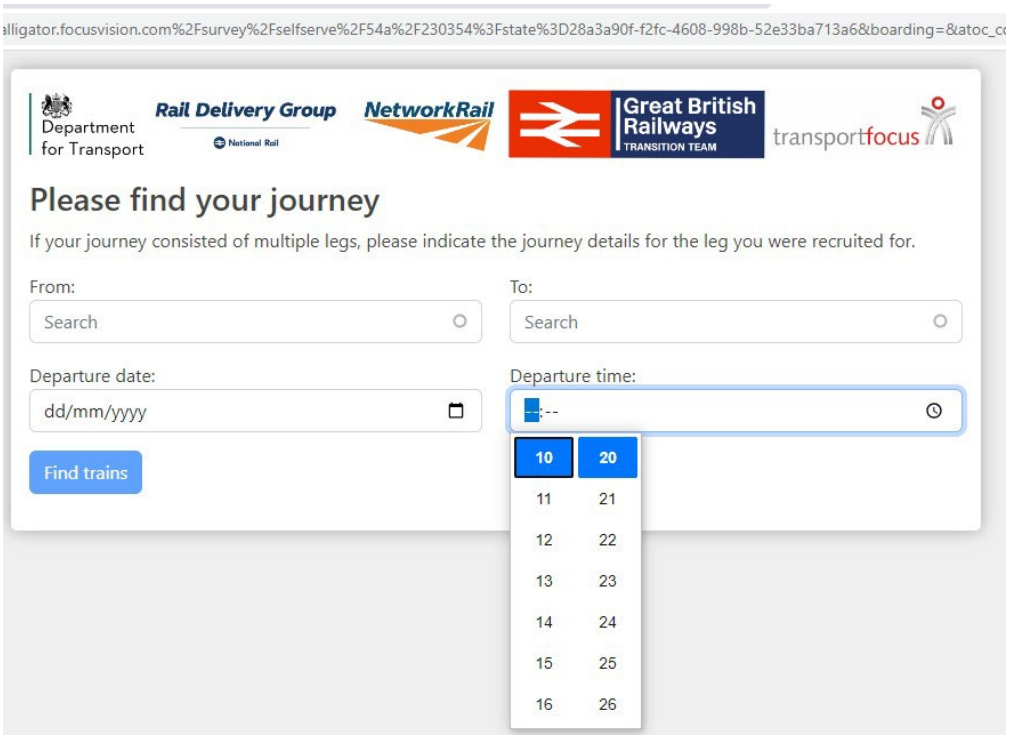
To address the multi leg issue, we had a note on the JPT screen asking respondents to only select the leg they were recruited on.



We recommend adding some text on the JPT if a respondent fails to find a journey to reiterate this and maybe give an example i.e. if you were travelling from XXX to XXX via XXX and you talked to our interviewer at XXX then you need to select the leg from XXX to XXX. Alternatively, we could change the logic of the JPT to allow for multi leg selection and then prompt the respondent to select the leg they were recruited to talk about. Our preference would be the former option since it would require less programming and might be easier to understand for the respondent.



For not finding the exact time/date, we need to update the tool so that the journey picker is not just offering journey times that are in the future but times that are in the past.



In the example above (when using the JPT at 10:20 am), the time selection is showing times that are in the future which may confuse respondents since they are not aware that they can just type in the journey time and don't have to select from the dropdown. It would make more sense to show time in declining order because respondents would have made their journeys already.

Conclusion: Using the journey picker tool was very successful and makes the data considerably more reliable. The on board approach performs slightly better with a marginally larger share of on board respondents able to find their journey. Even though the JPT seems to cause a notable proportion of dropouts, it cannot be said with certain that these respondents would not have dropped out anyway.

3.3 Fieldwork

3.3.1 Ability to recontact participants

Respondents were asked at the end of the questionnaire whether they would be willing to be recontacted about the survey, or other research about rail travel, to improve policy development and deliver a better service. They were asked to provide their name, email and phone number in a textbox if they agreed to this. This question was asked to all respondents regardless of whether an intervention was applied to their survey or not.

Just under 10% of those that left details provided their phone number, the rest an email. Since the details were entered into a textbox, the proportion is not absolutely exact but very close.

Out of all respondents, 28% provided their contact details which seems a decent level. This was not significantly different among those recruited on board versus those recruited at station – 27% and 29% respectively.

Leaving contact details by response option, age and journey purpose

There was, however, a considerable difference amongst those that scanned the QR code versus those that provided an email to be sent the survey link – 22% compared to 38%. Males were more likely to leave their recontact details than females (30% vs. 27%).

There was also notable difference across age ranges, as the table below shows, with the younger less willing to be recontacted than the older (which in turn reflects on the lower level amongst those who scanned a QR code, since they are also younger).

Table 42: Age profile of respondents providing recontact details.

	16-34	35-54	55-64	65+
Yes (Email / Phone number)	23%	27%	35%	37%
No	77%	73%	65%	63%

By journey purpose there was little difference between commuters and leisure travelers (29% vs. 28%), with business travelers somewhat less willing to be recontacted at 25%.

Recontacting respondents for the qualitative element

For the qualitative element 301 respondents were contacted, the majority by email since we only had phone numbers for a small proportion and respondents needed to fit a certain specification. Out of the 301 respondents 28 (9%) responded to the recruitment team, saying

they would be interested in participating in the qualitative exercise. 22 respondents (7% of those contacted) were ultimately recruited for the qualitative element, including two no shows. Of the 301 contacted 15 respondents (5%) actively refused to participate in the qualitative element.

Proportionately contacting respondents by phone was more successful since they cannot only be called but also be contacted by SMS which is an option if they cannot be reached with a call. An email is typically more easily ignored than a call or an SMS.

Even though recontacting respondents on their mobile number is more successful than re-contacting them by email, it is not recommended to ask respondents only for their phone number for recontact since only such a small proportion is willing to provide a mobile number over an email address. A combination of the two generates the best results.

Conclusion: Both methods perform equally well for recontacting respondents, but it has to be borne in mind that respondents selecting QR code over email at recruitment are less willing to be recontacted since they are mainly younger. In that respect email performed better than QR code.

3.3.2 Practicability / Feasibility

Sampling

For the at station methodology in the field trials, we used the ORR Passenger Numbers Report by station to select stations with probability to the estimated number of passengers and then assigned dates and time periods at random to ensure a good spread. For the field trials this process was undertaken separately for each TOC using the RDG Electronic Timetable data on number of services and an average load factor to turn the total number of passengers at a station to the estimated number of passengers for each TOC at the station.

For the pilot survey, where all TOCs will be included, we would not need to sample separately for each TOC but would be able to select a sample of stations overall with probability to the estimated number of passengers. We recommend using MOIRA data rather than ORR data to do this in future as this enables the number of passengers to be broken down by day of the week and time of day, allowing a single sampling process to select stations on a given day of the week at a given time of day proportional to the estimated number of passengers. This is an improvement on appending days and times at random although the limitations of MOIRA need to be continually borne in mind (in particular that although the services included are reasonably up to date, the routines to convert services to passengers are not).

The sampling process can be virtually fully automated allowing updated MOIRA data to be pasted into the routine and a sample of stations/days of week/time of day selected with probability proportional to the number of passengers using that station on that day of the week at that time of day to reach an overall target sample size. The MOIRA data can also be used to estimate the number of passengers per TOC that will result from this sampling plan. This allows any shortfalls against TOC targets to be assessed and boost shifts created. It also allows estimates of sample sizes for each Network Rail managed station to be produced and again boost shifts can be created as necessary.

As MOIRA data is only updated every six months, the sample for the next six months could be created in one pass each time MOIRA data is updated.

The station estate is very stable and new station openings are known about well in advance, as are plans to close stations. For new stations, estimates would need to be made of likely passenger numbers, but we assume these will have been made as part of the case for opening the station. Any stations due to be closed can be deleted from the MOIRA database prior to sample selection.

The field trials confirmed that obtaining permissions for fieldworkers to be at stations has problems both in advance and on the day itself. This appears to be a particular problem at Network Rail managed stations.

For the pilot survey we propose appending estimated passenger numbers from MOIRA for each train service, using an algorithm to do this when the service does not exist on MOIRA. We also propose incorporating all the scheduled cancellations into the extraction process, so that the services extracted are all those planned to run for each day of the fieldwork period at the point in time of the sample selection.

The on board sample for the field trials had problems with services that were selected being cancelled. This is:

1. Partly due to not including planned cancellations.
2. Partly due to the length of time between the sample selection and the fieldwork date.
3. Partly due to short term cancellations.

The first issue will be dealt with by removing cancellations planned at the point of sample selection from the process. The second issue can be minimised by running a check of the selected sample against an updated version of the timetable closer to fieldwork. The last issue is difficult to solve and it was a particular problem for TOCs which cancelled substantial

numbers of services at short notice.

An additional element of the on board sampling process is to identify a return train for the fieldworker to use to get back to their original start point for their shift. A process was developed to address this (see Part C.).

This process was set up in Excel and was run for each TOC using a macro to hunt through the timetable one selected service at a time and identify the appropriate return journey for each. Several issues arose with this process:

- Sometimes there was long wait at the destination or intermediate station before embarking on the return journey, increasing the shift time and decreasing recruitment time.
- Sometimes the first stop was longer than 90 minutes (this was an issue for long distance services) and a different threshold was used to facilitate a return trip, again increasing the shift time above three hours.
- Virtually all return journeys needed to be checked manually to cater for timetable changes and potentially select a different intermediate station to reduce waiting.
- The Excel process is time consuming involving very large files with many iterations (one per selected service).

For the pilot survey and any future continuous survey, we would look to improve the automation and speed of this aspect of the sample selection. In conclusion, although large parts of the on board sampling process can be automated, updates to the planned timetable and short term cancellations make managing the sample a time consuming process.

Disruptions like strike and train cancellations

During the field trials strikes took place which heavily impacted on the services run by the TOCs included. Some services came to an absolute standstill on the actual strike days and services did not always run fully on the day immediately after the strikes or on the day running up to the strikes. The on board method was a little more affected by this than the at station method. At station shifts can still be conducted around and during strike days even though the response rate will be considerably lower than at other times when services run as planned. On board shifts, however, cannot take place when the train services on which a shift is scheduled, do not run.

The following table gives a summary of shifts that were impacted by cancellations, strikes, station closures, etc. including those that could not be allocated. As can be seen the vast majority of shifts were conducted with valid recruitment data and completes or partials.

Shift summary:

1,086 shifts (94%) – All which went ahead and for which there is both recruitment data and partials/completes.

7 shifts (1%) – Shifts that were not completed, either because the interviewer booked was unable to cover the shift or the shift could not be allocated.

20 shifts (2%) – Shifts that were allocated but abandoned due to trains being cancelled, station closures, engineering works or permission refusal.

25 shifts (2%) – Shifts that were allocated but did not take place as shifts were rescheduled and interviewers were not available for the new dates (19 due to strikes and six due to services being amended.)

7 shifts (1%) – Shifts that were rebooked because reported figures from the original interviewer had not been received, poor interviewer performance and in one case the interviewer mixed up dates and conducted their first shift prior to fieldwork start.

Further logistical points – on board shifts:

- When everything goes well it's the method which offers the best recruitment and responses rates which in turn gives higher satisfaction to fieldworkers.
- Staff on board have been welcoming with fieldworkers and very helpful in regard to warning them of disruptions and delays to services.
- Device signal issues on the trains meant that fieldworkers could sometimes not sync their tablets until they got to their destination/back home which meant invites to some participants choosing email were delayed.
- Shifts being cancelled or changed at short notice meant we lost a proportion of fieldworkers (estimated 10%) who no longer wished to work on the project as they lost out on work at short notice.
- Fieldworkers getting stranded on long distance journeys (especially on Grand Central services) when the lines were disrupted and services cancelled.
- Fieldworkers sometimes not being able to complete their counts fully when the carriages were crowded or when there are no through doors between carriages which means they cannot move to the next carriage to carry on with the count until the next stop by

which time the 10 mins might have elapsed (see also 'Footfall counts' section).

- No clear set of rules on what to do when there is a problem with the services; fieldworkers need a clear set of instructions on what to do if the services (outbound or inbound) they are meant to travel on are delayed or cancelled to optimise the use of their time and cost efficiency as all shifts have to be paid that are affected that way.

Further logistical points – at station shifts:

- Shifts are easier to cover; on balance, it is probably the preferred method from fieldworkers' perspective and from an operational point of view.
- Less impact by service disruptions/train cancellations (as mentioned previously.)
- Less productive, especially at smaller stations during quieter times which results in less fieldworker satisfaction and engagement.
- Unsafe to send fieldworkers for late shifts at unmanned stations especially in the winter when it's dark; could be addressed by shift times when station has a minimum number of services alighting.
- Issue with staff not granting fieldworkers permission to work which seems to be mainly at Network Rail stations (as mentioned before.)
- No parking available at stations for fieldworkers which means they had to use public carparks, often with high costs.

Other logistical aspects to consider and possible solutions:

- Clear protocol and instructions for interviewers on entering and recording test surveys so that they can be easily identified and deleted at data cleaning stage e.g. using a specific shift number for tests.
- Very early shift starts (e.g. 6am) are difficult to cover as fieldworkers rarely live close to the station and have to get up sometimes as early at 4.30am in order to get to the station before the start of the shift; there is usually no public transport available at these times which reduce the pool of potential fieldworkers even further to those who own a car; it's not really attractive to leave home at 4.30am and be back at 10.30am for a three hours shift pay (see also 'Shift length'); six hour shifts will make such shifts more attractive for fieldworkers.

Fieldworkers have reported that participants are unable to scan the QR code on the tablet when it's sunny and there's glare on the screen. Passengers don't always have the

time to find a place with shade to try again and more often than not will decline to take part after that when the fieldworker offers them to share their email instead (although this issue primarily occurred at the station); this could potentially be solved with an automatic function that makes the screen as light as possible when a QR code is generated (similar to what airlines do in their apps with ticket QR codes); although a downside would be that this shortens battery life which could become an issue in a six hour shift.

- Survey Team hi vis jackets should be provided to all fieldworkers as fieldworkers noticed it has impact on the way they are perceived by rail users (they look more official) which would have a positive impact on recruitment rate.

- Thank you leaflets detailing the purpose and importance of the survey as well as details on how to find out more about the research should be handed out to all participants choosing the online option (QR or email); this should be trialed, however, since there will be cost implications and to understand if this impacts on response rate.

Both methods can cover all metrics in the questionnaire equally well (see key metric results before). What has to be borne in mind is the accuracy and the share of respondents completing the survey before the end of their journey as stated in previous sections.

A summary of issues, comments and complaints from fieldworkers, respondents and TOC staff can be found in appendix A.

Conclusion: The on board approach has significant practical and logistical issues which can be minimised but not completely eradicated and this will feed through into higher costs in some areas than for the at station approach. As such the at station approach is rated more highly on this criterion.

3.3.3 Shift length

Three hours used to be the standard length of a fieldwork shift in the NRPS. At the time it was effective in striking a balance between cost efficient fieldwork and avoidance of clustering around certain routes and stations on particular days. However, six hour shifts have a number of benefits over three hour shifts as shown in the table below.

Table 43: Comparison of benefits of three hour fieldworker shifts with six hour fieldworker shifts

Benefits of three hour shifts	Benefits of six hour shifts
Reduces the risk of clustering but not to the extent that six hour shifts are not a very good option	Six hour shifts are 43% more expensive than three hour shifts but assuming double the number of completes per shift they are more cost effective – it is recommended that this is trialed
Could be used for at station since the fieldworker can just work straight through the three hours	Fewer shifts (in actual numbers) reduce the time of project management, not only to allocate them but also for the fieldworker to honour them (see point below) – which also makes six hour shifts more cost effective
	More sustainable (less travel and therefore carbon emissions)
	More attractive for interviewers – guarantees them a day's pay
	Easier to secure the services of more experienced interviewers
	Shift length used by other TOCs for CX surveys is mostly six hours, which would make the Rail Customer Experience Survey more directly comparable
	Makes it more likely to get all shifts allocated in a survey with a high number of shifts like the Rail Customer Experience Survey

Six hour shifts are the norm for the majority of projects these days and interviewer fatigue is not expected since interviewers are used to work for that amount of time with just a short break. It is recommended though, as mentioned in the table, to trial six hour shifts if possible to understand fully their response rate across the two methods and how much the greater productivity would impact on costs.

Based on this comparison it seems obvious to use six hour shifts for the pilot survey and it is recommended to do that.

With regard to three hour shifts one could argue that this is better suited for at station intercepts since with that approach the fieldworker can just work straight through the three hours. With on board intercepts there are not always outbound and return journeys that fit into a three hour shift i.e. the journey out and back can be longer or shorter. 25% of on board shifts were longer than 3 hours 15 minutes while 24% were shorter than 2 hours 45 minutes. Plus, the fieldworker spends some time changing trains and having to wait for the journey back. Saying that, the response rate per shift is still higher for on board than it is for at station.

Six hour shifts are better suited for on board shifts (compared to three hour shifts) since they provide a much greater flexibility with regard to making an outbound and inbound journey. This is particularly the case for journeys on long distance TOCs where sometimes the first stop of a journey is not within 90 minutes which then results in overtime for three hour shifts (to get the fieldworker back to their departure station). This would not apply for six hour shifts – there is typically a stop within three hours so that the fieldworker can easily get back to their departure station.

Without trial, it is difficult to say if six hour shifts at station could be less productive than on board shifts (based on the increase of completes per shift compared to three hour shifts), especially when a station is not very busy. Passenger churn on trains seems more likely even on quieter routes which would provide fresh targets for the interviewer to approach and therefore generate greater productivity.

Conclusion: Benefits of six hour shifts far outweigh benefits of three hour shifts. six hour shifts seem more suited for on board intercepts than three hour shifts. Without trial it is hard to say if six hour shifts would be more productive on board than at station compared to three hour shifts, particularly when it is less busy. We have therefore scored both methods as equal.

3.3.4 Fieldwork force

Many of the points stated in the 'Practicality / Feasibility' section could be repeated here since they apply to the fieldwork force.

Overall, a continuous survey makes it easier to manage the fieldwork force, particularly on a larger scale, since the fieldwork team will be in regular contact with fieldworkers. Plus, fieldworkers will know that a continuous stream of fieldwork is forthcoming and based on experience will then very often contact the fieldwork team rather than the other way around.

This, in turn, means that fieldwork shifts can be booked further in advance than for a point in time survey, depending on when the sample is drawn (see also 'Sampling and weighting' sections).

A continuous survey also warrants regular fieldworker training, briefings and conferences which are project specific and can be combined into one event several times a year e.g. every quarter. At such an event, best practices can be exchanged and the agency, if possible together with the client, can provide updates around the project, e.g. any changes made to processes, new content, etc. or inform fieldworkers how the data has been used. From experience such events are much appreciated by the field force and keeps them engaged with and enthusiastic about the project.

On a more practical note, signed off project materials and the final recruitment and survey links should be received by fieldworkers at least three weeks prior to a fieldworker's shift to ensure that fieldworkers have sufficient time to familiarise themselves with the materials and grow in confidence with the recruitment script and survey, which can only have a positive impact on productivity. This is particularly important at the beginning and early stages of the project but also when changes like improvements are made to processes or content.

For larger and ongoing surveys we monitor fieldworker performance by key metrics analysis with underperforming fieldworkers being retrained. If no improvement was seen after they were retrained, they would be taken out of the field force pool.

A summary of issues, comments and complaints from fieldworkers, respondents and TOC staff can be found in appendix A.

Conclusion: If taken into account what is mentioned under 'Feasibility / Practicality', then at station would perform better on this criterion as well.

3.3.5 Footfall counts

From earlier sections it seems clear that the sample generated from the field trials exercise has differential response from various groups and also does not, in some cases, give each time of day its right chance to be selected (e.g., using three hour shifts means services starting at 06:00 and 07:00 have less chance of selection). On train does not have this problem but for both methods it seems likely that some external weighting is likely to be necessary to ensure the returned sample has the correct profile.

Footfall counts were undertaken at a random time during the shift, the interviewers script

prompted the fieldworker to start doing this for a 10 minute period using a pop up routine to assist with the counting process. Passengers were assigned to one of six boxes with a gender split by three age groups. Fieldworker feedback was that the counts were more difficult to undertake on board and data in the table below seems to confirm this. Note that analysis of footfall counts includes intervention shifts.

Table 44: Comparisons of footfall counts performed during on board shifts and at station shifts

	Done	Total Number of shifts	%	Average count
At station	439	500	87.80%	76.91
On board	422	509	82.91%	67.05

The first column shows how many counts were done, the second the total number of shifts and the third the percentage of shifts where a count was undertaken. 83% of scheduled counts on board were undertaken compared to 87% at station. The average number counted in the 10 minute period was 67 on board and 77 at station. The on board count may well have been limited by passenger numbers although similar logic would apply to a count at a small station. Overall, the evidence is that footfall counts did not work quite as well on board as at station. On the other hand footfall counts would have been more accurate on the train since it is clearer who to count on the train (i.e. everyone within the 10 minutes) whereas this is less clear cut at station.

As highlighted earlier, footfall counts suggest that the response profile has slightly too few 16-34 year olds, slightly too many 35-59 year olds and about the right percentage of those aged 60+. The response profile has too few males.

Table 45: Age profile of field trial respondents

Red and a minus (-) sign or blue and a plus (+) sign show statistical significance at 95% confidence.

	Field trial respondents (including interventions)					Footfall counts
Age group	QR	Email	QR Leaflet	Paper	Total	
Sample size	4357	2671	12	92	7132	71160
16-18	4% (+)	1% (-)	0%	0%	3%	
19-24	16% (+)	4% (-)	17%	2% (-)	11%	40%
25-34	28% (+)	10% (-)	42%	4% (-)	21%	
35-44	21% (+)	13% (-)	8%	3% (-)	18%	
45-54	16% (-)	19% (+)	25%	9%	17%	
55-59	6% (-)	12% (+)	0%	9%	8%	47%
60-64	4% (-)	13% (+)	0%	17% (+)	8%	
65-69	2% (-)	12% (+)	0%	14% (+)	6%	
70-80	1% (-)	13% (+)	8%	33% (+)	6%	13%
81+	0% (-)	1% (+)	0%	7% (+)	1%	
Prefer not to say	1%	1%	0%	2%	1%	
Don't know/ not sure	0%	0%	0%	0%	0%	

Table 46: Gender profile of field trial respondents

Red and a minus (-) sign or blue and a plus (+) sign show statistical significance at 95% confidence.

	Field trial respondents (including interventions)					Footfall counts
	QR	Email	QR Leaflet	Paper	Total	
Sample size	4357	2671	12	92	7132	71160
Male	45%	44%	25%	37%	45%	52%
Female	52%	54%	75%	59%	53%	48%
I identify in another way	1%	0%	0%	0%	1%	
Prefer not to say	1%	1%	0%	4%	1%	
Don't know/ not sure	0%	0%	0%	0%	0%	

If footfall counts are undertaken, they need to be at a frequency and intensity that generates sufficiently large sample sizes for each TOC. At station counts cannot be partitioned into TOCs, whereas on board can, so this is another argument for on board. Mobile phone data may be closer to providing workable data at station level but it is expensive and an ongoing cost.

Doing a count within the shift at a random time works well and should be a feature of each

shift. Most counts were done (80%+) and it might be worth considering giving the fieldworkers incentives to do their count when prompted on each shift.

Conclusion: Footfall counts help correct age and gender bias in the sample. At station counts cannot be split by TOC but the percentage undertaken and the number of passengers profiled is higher than on train. On balance, the two methodologies perform equally on this criterion.

3.4 Data and analysis

3.4.1 Speed of generating topline results

An essential objective for future rail customer experience surveys is to ensure the swift delivery of data. The rail industry seeks continuous, in-the-moment feedback to act to and react on customer experiences more promptly. In examining the field trials, we found no significant differences between shifts conducted at station and on board. This similarity is attributed to the prevalent modes of completion, which is online by either scanning a QR code or providing an email address to which a survey link is sent. As a result, completed surveys were readily available in their raw format as soon as respondents had completed them.

However, it's important to note that immediate availability does not equate to immediate usability for topline results. Several data processing steps need to be undertaken to ensure data integrity and validity.

1. Journey validation: Manual validation is needed for surveys that failed to select a journey using the journey picker tool.

2. Postcode verification: Postcodes recorded in the surveys need to be checked to ensure respondents did not enter the full postcode, which could compromise respondent privacy.

3. Open-end review: Thorough review of open-ended responses is necessary to eliminate profanities and identify surveys that lack truthfulness as well as deleting personal identifiable information.

4. Journey picker data: Additional information from the journey picker tool is appended during the data processing stage which takes some time.

5. Weighting: Once the data is consolidated, weighting is applied to ensure accurate representation of the target population.

Differences in the data processing timeline were observed based on the method of survey access chosen by respondents. With email and QR code completion, respondents participate online, resulting in little to no delay.

- QR respondents tend to start answering the survey immediately at the time of recruitment, as they gain instant access to the survey.
- On the other hand, email respondents need to wait for the invitation, typically for around 30 minutes.

Despite this slight difference, both methods generally enable prompt data availability.

The impact of paper surveys on the speed of generating toplines

In one of the field trials, paper surveys were introduced as an option for respondents who preferred not to participate online. These respondents received a paper questionnaire to fill out and a first class return envelope. On average, paper questionnaires were returned after approximately six days. Consequently, at the end of the fieldwork period, we had to wait a few additional days before the cut off. This resulted in delays in processing the data and making it available for industry stakeholders. If paper was to be kept for the pilot survey, we recommend allowing seven days (instead of 9) after the end of fieldwork to accommodate the return of first class paper questionnaires.

After receiving the completed paper questionnaires, further processing steps are undertaken, which can add a few extra days to the overall timeline. Once the data is successfully incorporated into the system, the aforementioned data processing steps are performed to ensure the data's quality and accuracy.

In conclusion, while the survey data is quickly available in its raw form, timely generation of topline results depends on whether paper questionnaires are part of the survey as this requires additional thorough data processing and validation to ensure the reliability of the insights.

Conclusion: Both on board and at station perform equally well, however, paper questionnaires would delay the generation of topline data considerably.

3.4.2 Cutting data by different time periods

The recommendation for a future continuous survey is to select a sample for each four week rail period. Each such sample will be weighted to the universe profiles derived from MOIRA, LENNON and footfall counts. Any analysis for any four week period or any combination of four

week periods should thus be using a correctly weighted sample of responses.

There may, however, be a need to analyse data within rail periods or for specific time periods that span rail periods (for example, to assess the impact of a particular event on a particular day such as a rise in fares or the introduction of a new timetable). There is no guarantee that the data selected from such time periods will be representative in the same way as will be the case for each four week rail period.

One approach would be to reweight the data for the selected time period to the universe profiles. We would not recommend this, as it adds another complication into the data analysis and would inhibit users from undertaking their own analysis from any datasets provided to them (e.g. through an online portal). There is also the danger of having two sets of estimates, one from the original weighting process and one from reweighting. We therefore recommend only using the data with its original weights derived from the four week rail period in which that data was collected. The only exception to this approach would be the analysis of Network Rail managed stations which, in any event, will be based upon a different dataset incorporating boost shifts.

There need to be controls on the minimum sample size for which any ad hoc analysis can be undertaken, by prohibiting the production of analyses based upon sample sizes below an agreed threshold. This can easily be controlled on an online data portal (and has been the case for NRPS).

Conclusion: The process would be the same for both methods therefore they perform equally on this criterion.

3.4.3 Ability to merge with other data

There are likely to be requests to merge the customer experience data with other data sources to enhance the analysis options. These other sources might include:

- Information on the train – configuration, type of rolling stock etc. – which might come from the RDG Electronic Timetable and/or the Journey Picker Tool (JPT.)
- Train performance information – actual departure and arrival times.
- Estimated number of passengers on the train – from MOIRA.
- Total revenue or average revenue per passenger generated by the train – from LENNON.
- Information on the origin and destination station from Network Rail e.g. facilities available.

To merge with these sources requires a linking device of some sort. Some of the data sources have a train code (the RDG timetable has a unique eight digit code), the JPT has a head code etc. All data sources should have the original and departure stations of the train and the scheduled departure time and the date and so it should be possible to create a combination variable that contains these fields in a standard format which might be XXXYYYHHMMDDMM where:

- XXX is the TLC of the departure station of the train.
- YYY is the TLC of the arrival station of the train.
- HHMM is the scheduled departure of the train.
- DDMM is the day and month the train ran.

LENNON uses this convention (except the date) to create a unique train code and all four components should be available in all data sources, albeit the TLC may be replaced by a TIPLOC code or an alphabetical station name. Creating lookup files to turn all the alternatives into TLC's should be fairly straightforward (and indeed we have done this previously to link the RDG electronic timetable data to MOIRA codes and ORR codes).

On board data collection is more likely to have accurate data on the train itself but in all other respects the two methods perform to a similar level on this criterion.

Conclusion: The two methods perform equally well on this criterion.

4. Part C. Implementation details for pilot survey methodology

The following sections describe the processes that were followed during the field trials and highlight if any changes are recommended for the pilot and future surveys. Some future changes have already been mentioned in Part A. and Part B. and are not necessarily repeated in Part C.

4.1 Sampling

The pilot survey will utilise on board sampling, conducting six hour shifts. Based upon the analysis of the field trials, each such shift should generate an average of 15 complete interviews. The pilot survey will cover all TOCs, including both operators with DfT rail contracts and those without.

The sample size per TOC has yet to be agreed but there is likely to be a minimum sample size of 100 for the smaller TOCs going up to around 500 for the more complex and larger TOCs. The process to select a sample for a TOC will be the same whatever the sample size and will operate as described in this section.

Stage 1 – List all services that run during the time period fieldwork

The services are downloaded from the latest version of the RDG Electronic Timetable. This can be downloaded from Data Download | data.atoc.org using the timetable feed. To access this data, the user needs to register at National Rail Data Portal and then through the LINKS option on the data download menu to the RSP feeds. Registration is free to individuals or companies with a legitimate need to use the data.

Downloading the timetable feed generates a large zip file containing eight text data files. The filename reflects the date of the download (files downloaded on 22/12/22 have a file name ttisf585 whilst those downloaded on 28/12/22 have a filename ttisf592). The structure of the eight files is well described in this file:

RSPS5046 timetable information data feed interface specification.pdf (raildeliverygroup.com)

The eight text files cover the following content:

Table 47: Eight text file names and content from RDG electronic timetable

Filename	Content
ttisf585.ZTR	Z Trains file
ttisf585.REJ	TTIS Rejects
ttisf585.SET	Common Interface File Set
ttisf585.FLF	Fixed Links
ttisf585.MCA	**Basic Timetable Detail**
ttisf585.MSN	**Master Station Name File**
ttisf585.ALF	Additional Fixed Links
ttisf585.TSI	TOC Specific Interchange Times

The two emboldened files are those used – basic timetable data and station names. The file structure for each is shown in appendix C. The link between the two is a TIPLOC code, which identifies the station name in the master station file and is used to identify locations in the Basic Timetable file. From these files two derived files are created (only those services that run during the relevant rail period are selected).:

- A file with data for each train service and each day of the rail period including departure time, origin station, destination station, arrival time at destination station and operator.

- A file with a record for each station where each service calls including station name, days run and operator.

To create the first file above, the following record types are needed from the .MCA file:

Table 48: record types needed from .MCA file

Code	Record Contents	Count
BS	Basic Schedule	At least one record per train service
BX	Basic Schedule Extra Details	Exactly one record per train service
LO	Origin Location	Exactly one record per train service
LT	Terminating Location	Exactly one record per train service

- The BX record type contains the TOC code for the train service.

- The LO record type contains details of the origin station including station code and departure time.

- The LT record type contains details of the destination station including station code and arrival time.

- The BS record type contains one of more specifications of the train service including the start and end date of the service and on which days of the week it runs.

In the BS record, column 80 – the STP indicator – must be taken into account as follows:

- P (permanent schedule),
- (STP overlay to replace P schedule on specific dates),
- C (STP cancellation of the P schedule on specific dates),
- N (New STP schedule that runs on specific dates).

So, if column 80 is P, the service runs between the stated dates on the prescribed days of the week. If column 80 is N, the same applies.

If column 80 is C, the service does not run between the specified dates. If column 80 is O, a different service runs between the specified dates.

All records of type BS therefore need to be taken into account to identify whether the service runs or not, for each day of the survey period. If there is more than one BS record for a train service, these do not necessarily appear together in the data file. Any analysis procedure therefore needs to pick up these records from wherever they are in the data file to create an overall list of the days on which the services are scheduled to run.

To create the second file, the LI record types are required. Stations where the service stops are identified by having both an arrival and departure time in columns 11-15 and 16-20.

Stage 2 – Append the estimated passenger numbers for each service

Data extracts from the MOIRA system contain estimates of the number of passengers boarding each service at each station where it calls for a typical weekday, Saturday and Sunday. From this data, the total number of estimated passengers using the train can be calculated (by adding up the number of passengers boarding at each station).

MOIRA uses a train identifier made up from the departure time, origin TLC and destination TLC. This train code can be also generated from the electronic rail timetable data to enable a match to be made.

Each rail service on the file extracted from the RDG Electronic Timetable can have estimated passenger numbers appended using this matching process. Where a service does not have a MOIRA estimate of the number of passengers (typically around 10% of services), this will be generated using an algorithm based on the MOIRA algorithm. In this way, every train service running on each day during the selected rail period will have an estimate of passenger numbers using that train service.

Stage 3 – Select a sample of train services with probability proportional to the number of passengers on the train

For each TOC, the complete list of services will be extracted from the overall file and sorted into date order and time within date. A sample of services will then be selected with probability proportional to the estimated number of passengers. Part of the sampling process will be to profile the selected services by daypart and day and compare to the profile of all services for the TOC. If the profile is significantly different, the sampling process will be repeated until a sample is generated whose profile by daypart and day does match the overall profile. This sample will then constitute the list of train services to be sampled for that TOC. The process will then be repeated for each TOC.

Stage 4 – Select return trips

Once the sample has been selected, an appropriate return trip will be identified. With a shift length of six hours, all journey times of just under three hours should enable the fieldworker to travel to the final destination of the selected train and then come back using the same TOC.

Where journey times are much shorter, the routine will identify the first train back and the fieldworker repeats the process of outward and return journey a pre-specified number of times. For example, for a 50 minute journey, we would specify three return trips to take place during a six hour shift.

Stage 5 – Check the selected services still run

At a pre-specified time before fieldwork starts (2-3 weeks would seem appropriate), the list of selected services is compared to the current timetable. Any selected train services that do not now run due to timetable changes would be flagged and a substitute service suggested.

Stage 6 – Select samples for future rail periods

Major timetable changes are made in May and December each year, always on a Saturday

at the end of a rail period. We recommend that for any future continuous survey the sample selected for the first rail period surveyed is used again for each succeeding rail period until the next major timetable change (subject to the checks at stage 5). This process ensures some stability in the sample undertaken in each rail period, reducing random variation between rail periods and also reducing costs (as the process of checking at stage 5 takes much less time than the process of selecting and then checking a completely new sample).

As an example, the sample of train services selected for the rail period commencing 10th December 2023 would be used for each of the five rail periods until that commencing on 28th May 2024, when an entirely new sample would be selected, which in turn would be used until December 2024. This process runs the risk that some passengers, particularly commuters, may be approached each rail period, if they always use the same train service. One compromise for weekday rail services could be to select the same train service but on a different day of the week.

4.2 Scripting the survey

There are two scripts for this survey – the recruitment script and the survey script.

The recruitment script

The recruitment script guides the fieldworker through intercepting and recruiting respondents. Prior to the shift the fieldworker has to enter a few details into the recruitment script, e.g. the shift number, the TOC name for on board shifts, etc. During the intercept and recruitment process the fieldworker has to enter a few more details into the recruitment script but the recruitment script is mainly used to create a QR code (if that is what the respondent selects as their response option) or to trigger an email invite with the survey link.

The recruitment script is created by our partner Sign Up Anywhere (SUA). SUA have worked with us on a number of recruitment questionnaires, not only for rail but also for other surveys e.g. amongst bus and road users.

Based on the information entered into the recruitment questionnaire by the fieldworker, SUA creates a link. The information from the recruitment questionnaire can be collected with or without an internet connection and the link is provided to the fieldworker as a QR code (which can then be scanned by the respondent) or to the respondent in an email (which then gives them survey access). SUA uses Mailchimp to send out the email invitation and any subsequent reminders and Zapier to send any information to Mailchimp.

The link created from the recruitment questionnaire (either as a QR code or in the invitation email) contains a string of information which is piped into the survey script like the shift

number, respondent ID, TOC, TOC type, shift type, etc. plus a time stamp. The email response delay is handled by SUA i.e. 30 minutes after recruitment which was standard in the field trials. The intervention delays were also handled by SUA.

To send out reminders SUA connects to the Decipher API (the survey script is written in decipher) and looks for specific information to add a respondent either to the reminder list or the partial reminder list.

The survey script

The questionnaire was developed in a separate workstream and provided by the DfT as a Word document. The questionnaire then needs to be converted into an online script. That is first done by the executive team who produce a script instruction document in Word for the digital team. That document includes all questions and routing instructions. Subsequently the survey is scripted by the digital team in Decipher. Once that is completed the script is thoroughly checked by the digital and executive teams, as well as the client.

The survey is offered in English and in Welsh; in the field trials respondents were asked on the first screen which language they preferred but it is intended for the pilot and future surveys not to have a survey question about language preference but to offer a Welsh option on the introduction screen as a toggle that will allow respondents to switch languages e.g. language option at the top right of the screen.

Journey Picker Tool (JPT)

The Journey Picker Tool (JPT) was built into the survey script so that respondents could validate the train journey they assessed. The JPT has been developed by Journeycall. A redirect is built into the survey script and respondents do actually not notice that they leave the survey. The redirect URL is provided by Journeycall.

The JPT accepts various parameters via the query string:

Table 49: Journey Picker tool query string parameters

Query String Parameter	Description
destination	This is the URL the JPT should redirect to after the user has selected a journey. This parameter is required. The digital team at BVA BDRC provides that.
lang	The language the train picker should be rendered in. Values are en (English) or cy (Welsh). If omitted, this will default to en.
boarding	An optional boarding station's CRS code. If provided, the boarding station will be fixed, and the passenger cannot change it.
atoc_code	If you wish to limit search results to a particular TOC, then this parameter should be used to specify their ATOC Code. If omitted, results for all TOCs will be included.

Once the user has selected their journey (or specified that their journey can't be found), they will be redirected back to the URL specified in the destination parameter and include information about the journey in the query string.

If the user has selected a valid journey, a string of information about their journey will be available at the destination URL. If the journey could not be found then only some limited information will be available at the destination URL.

If the user selects a valid journey, the information about that journey can be retrieved by making an API call from the digital team's backend system. To do this, the id parameter sent back by the JPT needs to be used. Journeycall provides a manual that specifies the whole process.

For the pilot survey it is also intended to build a database with station facilities into the script, so that respondents are only asked about facilities available at their departure station. This is currently work in progress. The database is held by RDG.

4.3 Paper questionnaires

As mentioned in section 4.2 the questionnaire is developed in a separate workstream and provided by the DfT as a Word document. The questionnaire then needs to be converted into a printable version. For the field trials we initially had a paper version that included all questions, then, for the 2nd half of the fieldwork, we reduced the paper questionnaire to a shorter version. The printable version is produced by the field team, thoroughly checked by the field and executive teams and then signed off by the DfT. Due to its nature there is limited routing within the paper version. The signed off version is sent to the printers who print the required number of copies which are then sent back to the field team. Depending on what is agreed with the client, paper questionnaires are then sent with the shift packs to the interviewers. In the field trials paper questionnaires were offered to respondents only in selected shifts, so only a limited number of interviewers needed to take paper questionnaires with them on their shifts. In the pilot survey, the DfT considers the option of having a limited number of paper questionnaires available for each shift, e.g. 10 or 20, for those respondents who would like to participate in the survey but either cannot or don't want to complete the survey online.

It needs to be agreed with the client whether a Welsh option should be available for the paper version. In the field trials the paper version was only available in English.

4.4 Data cleaning

Once fieldwork is completed (and all paper questionnaires have been entered) four data files are produced.

1. A data file that contains all the recruitment data.
2. A data file that contains the online survey data.
3. A data file that contains the footfall counts.
4. A data file that contains all the responses to open ended questions.

Each data file is then checked to ensure the data is correct. Where necessary the data is cleaned.

Recruitment data

The recruitment data is produced in Excel with a row per recruit and each piece of data across. Checks are run to ensure interviewers have entered the correct interviewer number,

shift number, station (for at station shifts), train operator (for on board shifts) and the dates of the recruitment. These items are checked against the shift plan in a pivot table.

If there are any discrepancies it may be necessary to get back to the field team to ask them to confirm data with the interviewers. In a worst case scenario, some recruitment data might need to be removed e.g. if an interviewer went to the wrong station.

Survey data

The system already checks that there are no duplicates in the survey data (see also previous section). But a manual check will still be done to ensure this is not the case. A check will also be run to confirm that each online survey respondent has recruitment data.

As mentioned in Part B. not all respondents can find their journey in the Journey Picker Tool (JPT). These respondents will have entered their journey details manually and it needs to be checked that those journeys actually exist. Following that, additional JPT data (15 variables) needs to be appended to the survey data. This includes the final status of the journey (with regard to punctuality), number of coaches loading capacity, etc.

Once these cleaning steps have been conducted the data can be weighted as described in the next section.

Footfall data file

It needs to be ensured that the shift number is valid so that the data can be linked correctly (i.e. to the right station and TOC). It also needs to be checked that there is not more than one count per shift.

Spot checks need to be carried out that the data makes sense with regards to counts i.e. are the highest and lowest counts possible or realistic. For the footfall counts, as well as the recruitment data, there will be test data. In the field trials test data needed to be identified by time – going forward we propose to have a test shift number for easier identification (as mentioned in Part B.).

Open ends

Responses to open ended questions did not require a check for the field trials. However, for the pilot and any future surveys it needs to be checked that responses do not contain any personally identifying information. If postcodes will be collected, full postcodes need to be shortened. Depending on requirements, it might be necessary to check open end responses for swear words and poor grammar. A decision also needs to be made whether to deliver open ends in a standa lone file or whether responses need to be merged back into the

survey data linked to the respective respondent.

4.5 Weighting

Data from the pilot survey needs to be weighted to ensure it reflects the overall profile of rail journeys made on the network.

Weights would be applied to the following categories for each TOC:

- Dayparts (different dayparts for weekdays and Saturday and Sunday overall),
- Station size bands,
- Passenger journey volume,
- Age and gender.

Stage 1 – Weighting by dayparts for each TOC

MOIRA is used to provide targets for each TOC by weekday/weekend and time of day. Early analysis of the field trials data shows that the key satisfaction measures do vary by day of week and time of day. We have therefore constructed the following dayparts so that weighting by these dayparts does counter any response rate differences.

- Weekday - morning peak (trains starting between 06:00 and 08:59).
- Weekday - evening peak (trains starting between 16:00 and 18:59).
- Weekday – late (trains starting from 19:00 onwards).
- Weekday – other (between 09:00 and 15:59; no shifts to be conducted prior to 06:00).
- Saturday.
- Sunday.

We recommend using the profiles that emerge from MOIRA analysis covering these day parts and station size bands and to weight the data from the pilot survey.

It should be noted that the factors used in MOIRA to estimate the numbers boarding and alighting a train service at each station are based upon patterns of travel that existed before the COVID pandemic. These will not reflect current travel patterns and comparison with other sources and suggests that MOIRA does overstate peak hour travel. There is an urgent need

to update MOIRA data to reflect travelling behaviour but until this is done, we recommend using the existing data.

Stage 2 – Weighting by station size band for each TOC

Satisfaction with many station attributes vary by size of station. The data from the pilot survey will therefore be weighted by station size band for each TOC. As MOIRA contains estimates of the number of passengers boarding at each station for each train service the TOC runs, this data can be aggregated by station and then into station size bands. These are four (roughly) quartiles, so that when stations are ranked into decreasing passenger numbers, the first quartile contains the stations that take the first 25% of passengers, the next quartile the stations that generate from 25% up to 50% and so on. The quartiles will not be exact but they do define groups of stations with the estimate proportion of passengers that those stations represent.

Stage 3 – Weighting by TOC passenger numbers

ORR publishes estimates of passenger journeys for each TOC on a quarterly basis. However, the data is somewhat in arrears; at the time of writing (11th August; the latest data is for the January- March 2023 quarter and data for April-June 2023 is due to be published on 8th October). This means the data is considerably out of date.

Up to date estimates of the passenger journeys for each TOC can be provided by LENNON. However, LENNON does not include ticket sales from some sources (particularly PTE's) and has no data for Heathrow Express. However, by comparing ORR data for the latest quarter with LENNON data for the same period, we can calculate a factor to take account of ticket sales not covered by the national ticketing system. The factors generated can then be applied to LENNON data for the rail period being analysed.

Stage 4 – Weighting by age and gender

In the pilot survey, two counts will take place in each six hour shift, profiling passengers on a service by age and gender. All the counts for a TOC will be aggregated to generate an overall age and gender profile for the TOC. This aggregation could use the estimated passenger numbers on the service to weight the data, otherwise a lightly used service might be selected too frequently.

Stage 5 – Aggregating all the weights

The four weights need to be combined into a rim weighting regime that generates weights which match all the input profiles. This should then ensure the sample profile for each TOC matches our best estimates of universe data by passenger numbers, station size band, daypart

and day of week and age and gender.

5. Appendices

A. Summary of issues, comments and complaints received during fieldwork

Feedback from the fieldwork team

- Issues with interviewers being refused permission to work at Network Rail stations (particularly at [station] and [station]).
- [TOC name] - Trains being cancelled last minute. Lines being closed for extended periods of time.
- [TOC name] - Trains being cancelled last minute. Interviewers struggled to return to departure station as return trains were cancelled. We had this happen on a few instances on the on board shifts from [station] to [station]. One particular time the interviewer had to pay for an expensive [TOC name] ticket (£70+) in order to get back to London.
- A male interviewer did not end up sending us the exact details of the train he was on. He was subject to homophobic abuse from a group of young people. The incident occurred on 2nd May on an early morning [TOC name] train from [station] to [station].
- Interviewers also mentioned that there was often no parking at the stations so they were having to find parking elsewhere (primarily private car parks) that charged very high prices.

Emails received during the research from respondents (to our direct email)

Issues with connection:

"I agreed to take the survey and then my train came. I got cut off the survey by the internet connection ([TOC name]) from [station] to [station], but am happy to still do it if needed."

Length of survey:

"I spent about 15 minutes to get 30% of the survey done. Whoever designed this survey is a complete idiot. I will not be completing it."

"I did this survey twice and it takes time. Not prepared to do it again. It is not a well thought out survey."

"The survey takes way too much time!!!"

Inputting a train service that cannot be found:

"I tried starting this survey a number of times… your website does not recognise the time of the train I enter… & won't let me proceed!! So hey ho…. I tried … but your website doesn't want to play!"

"Just to say your survey is not easy to complete. The trains that you have don't fit with the ones I went on because it was a two part journey complicated by a train being cancelled so I actually had to get three trains. Anyway, life is too short. After 10 minutes I gave up. Your survey is very badly designed so I doubt what you discover has any meaning."

"This is terrible. I bet your completion rate is tiny from a self selecting non representative sample."

"Get a grip: asking the exact time of a train!"

"I cannot complete the form. The link does not allow me to set the departure time."

"I have tried to complete your survey, however, your system will not allow me past my journey details telling me "journey not recognized. Then comes up with a different journey option????"

"Couldn't do it as it's a joke. I went to Llandudno not the junction but wouldn't let me do it."

Layout of survey:

"I agreed to answer questions about my train journey and the first question has nothing to do with this. As I can't pass on to the next question I'm afraid I can't complete the survey."

"It's impossible for me to answer your questions accurately given the way you've set out the survey. Sorry."

Troubles during the survey:

"I could not do this survey because [station] was not recognised as the starting point of my journey. It was not included in the drop down list either."

"Tried to complete but wouldn't get beyond first question!!!!" "I would have done this, but it does not work on my iPad."

"Funnily enough I can't access past the first page - I click on survey in English and then click the next button but nothing happens."

Issues with remembering:

"So sorry I abandoned it. The trouble is that I'm losing my short-term memory, and so I can't really remember enough detail."

Feedback about working at stations and on trains

During the research, we experienced problems with authorisation at some stations. Some of these were in relation to acquiring authorisation prior to the fieldwork, while the remainder were issues experienced during fieldwork. In summary:

- As part of the trials, some TOCs required named passes or fieldworker names of who would be attending their stations/travelling on their trains. While it is understood why some TOCs require this level of detail, this will prove tricker when carrying out a longer, continuous survey where there are likely to be more fieldworkers and names and fieldwork dates that could change over time.

- Linking to the first point above, where dates were moved during fieldwork and no longer aligned with the original letter of authority provided by the TOC, this proved problematic in some cases and on some occasions fieldworkers were refused permission to work.

- The information regarding the field trials was shared with the respective TOCs using contacts provided by the DfT. This proved useful for acquiring the authorisation documentation, however the information was not always funneled down to those working at the stations/on board to ensure they were fully in the loop on what was happening. We had examples of station managers unaware of the exact name of the fieldwork or who was conducting the fieldwork, leading to us sending a further email to explain this. It subsequently led to delays in the beginning of shifts or complete refusals, such as experienced at a London Overground shift.

- Fieldwork at Network Rail stations required some additional permissions and details. For this, names of fieldworkers and dates of fieldwork were required to be sent to all station managers of the stations we would be conducting our shifts at. Despite this detail, we experienced a number of issues at stations where the manager working was unaware of the fieldwork or who was conducting it, such as at [station] and [station].

There were also inconsistencies in the practices regarding authorisation, with some refusing shifts due to no RAMs documentation while other stations were happy to proceed. Health and safety briefings also differed by station and were not always detailed in advance. Together, these created delays in starting a shift or refusals to work.

B. Completes by Network Rail region achieved in field trials and projected against a sample of n=5,000

The following tables shows the completes that were achieved in the field trials across the Network Rail stations:

Table 50: Completed questionnaires for each methodology by Network Rail region

NR Region	At station online	On board
Eastern	731	1334
North West & Central	612	734
Scotland	200	267
Southern	474	495
Wales & Western	445	527

The table below then shows the expected number of completes for each region per rail period for a total sample of n=5,000:

Table 51: expected number of completed questionnaires for each methodology in a sample of n=5000

NR Region	At station online	On board
Eastern	1485	1987
North West & Central	1243	1093
Scotland	406	398
Southern	963	737
Wales & Western	904	785

C. Structure of RDG Electronic Timetable files

Basic Timetable Detail File (.MCA file)

Table 52: Schedule record (type BS)

Field	Field description	Length	Position	Notes
1	Record Identity	2	1-2	With the constant value 'BS'.
2	Transaction Type	1	1-3	N' =New. 'O' = Dente. 'R' = Revise.
3	Train UID	6	4-9	Unique Train Identifier
4	Date Runs From	6	10-15	yymrndd
5	Date Runs To	6	16-21	yyrnmdd
6	Days Run	7	22-28	
7	Bank Holiday Running	1	29-29	
8	Train status	1	30-30	
9	Train Category	2	31-32	
10	Train Identity	4	33-36	
11	Headcode	4	37-40	
12	Course Indicator	1	41-41	
13	Profit Centre Code/ Train Service Code	8	42-49	
14	Business Sector	1	50-50	Now used to contain portion suffix for RSID
15	Power Type	3	51-53	
16	Timing Load	4	54-57	
17	Speed	3	58-60	
18	Operating Chars	6	61-66	
19	Train class	1	67-67	
20	Sleepers	1	68-68	
21	ReservatOns	1	69-69	
22	Connect Indicator	1	70-70	
23	Catering Code	4	71-74	
24	Service Branding	4	75-78	
25	Spare	1	79-79	
26	STP indicator	1	80-80	C' STP cancellation of permanent schedule. 'N' = New STP schedule 'O' = STP overlay O permanent schedule. ' p' = Permanent Read n association with the Transaction Type in Field 2

Table 53: Basic Schedule Extra Details (type BX)

Field	Field description	Length	Position	Notes
1	Record Identity	2	1-2	With the constant value 'BX'
2	Traction Class	4	3-6	Not used - always blank.
3	UIC code	5	7-11	Only populated for trains travelling to / from Europe via the Channel Tunnel. Otherwise Blank
4	ATOC Code	2	12-13	
5	Applicable Timetable Code	1	14-14	
6	Retail Service ID	8	15-22	
7	Source	1	23-23	Not used — always blank.
8	Spare	57	24-80	

Table 54: Origin Location (type LO)

Field	Field description	Length	Position	Notes
1	Record Identity	2	1-2	With the constant value 'LO'
2	Location	8	3-10	TIPLOC + Suffix.
3	Scheduled Departure Time	5	11-15	
4	Public Departure Time	4	16-19	
5	Platform	3	20-22	
6	Line	3	23-25	
7	Engineering Allowance	2	26-27	
8	Pathing Allowance	2	28-29	
9	Activity	12	30-41	
10	Performance Allowance	2	42-43	
11	Spare	37	44-80	

Table 55: Intermediate Location (type LI)

Field	Field description	Length	Position	Notes
1	Record Identity	2	1-2	With the constant value 'LI'
2	Location	8	3-10	TIPLOC + Suffix.
3	Scheduled Arrival Time	5	11-15	
4	Scheduled Departure Time	5	16-20	
5	Scheduled Pass	5	21-25	
6	Public Arrival	4	26-29	
7	Public Departure	4	30-33	
8	Platform	3	34-36	
9	Line	3	37-39	
10	Path	3	40-42	
11	Activity	12	43-54	
12	Engineering Allowance	2	55-56	
13	Pathing Allowance	2	57-58	
14	Performance Allowance	2	59-60	
15	Spare	20	61-80	

Table 56: Terminating Location (Type LT)

Field	Field description	Length	Position	Notes
1	Record Identity	2	1-2	With the constant value 'LT''.
2	Location	8	3-10	TIPLOC +Suffix.
3	Scheduled Arrival Time	5	11-15	
4	Public Arrival Time	4	16-19	
5	Platform	3	20-22	
6	Path	3	23-25	
7	Activity	12	26-37	
8	Spare	43	38-80	

Table 57: Master Station Name File (.MSN file)

Field	Field description	Length	Position	Notes
1	Record Type	1	1-1	Constant value 'A'.
2	Reserved	4	2-5	
3	Station Name	26	6-31	
4	Reserved	4	32-35	
5	CATE Interchange status	1	36-36	Always populated with '1', '2', '3', or '9'
6	TIPLOC code	7	37-43	
7	Minor CRS code	3	44-46	
8	Reserved	3	47-49	
9	CRS code	3	50-52	
10	Ordnance Survey Grid Ref East	5	53-57	Values are in 0.1km units. Format is 'tnnnn' where nnnn is the distance in 0.1km units.
11	Bank/Estimate	1	58-58	Value is blank or 'E' if grid reference is an estimate
12	Ordnance Survey Grid Ref North	5	59-63	Values are in 0.1units. Format is 'tnnnn' where nnnn is the distance in 0.1km units.
13	Minimum Change Time	2	64-65	A one or two digit number, in minutes, in the range 0-99. This is regardless of whether or not field 5: Cate Interchange status' shows the station as an interchange
14	Reserved	1	66-66	
15	Footnote/Closed/Staff/Not-advertised code	1	67-67	Redundant and not supported in PMS. Will always be blank.
16	Reserved	11	68-78	
17	Sub-sector code	3	79-81	Redundant and not supported in PMS. Will always be blank.

D. Recruits, completes and response rate by TOC

The following table shows recruits, completes and response rate by TOC. Recruits and completes are provided at both an overall level and per shift. Even though Merseyrail records the lowest number of recruits per shift, its response rate is one of the highest. A high number of recruits does not necessarily mean a high number of completes per shift. Some of the commuter TOCs (like Northern, Merseyrail, even though technically regional TOCs, and SWR) record rather high response rates. By contrast, the long distance TOCs (Avanti and CrossCountry) record rather low response rates.

Table 58: recruits, completed questionnaires and response rates by Train Operating Company

	Total shifts	Total recruits	Total completes	Recruits per shift	Completes per shift	Total response rate
Avanti West Coast	60	2977	342	49.6	5.7	11%
CrossCountry	58	2528	279	43.6	4.8	11%
East Midlands Railway	54	2044	239	37.9	4.4	12%
Grand Central	55	2675	336	48.6	6.1	13%
Heathrow Express	58	2339	195	40.3	3.4	8%
London Overground	59	2298	243	38.9	4.1	11%
Merseyrail	60	1842	282	30.7	4.7	15%
Northern	56	2073	463	37.0	8.3	22%
ScotRail	62	1690	307	27.3	5.0	18%
South Western Railway	60	2664	433	44.4	7.2	16%
Thameslink	54	2922	361	54.1	6.7	12%
Transport for Wales	58	2230	438	38.4	7.6	20%