

WNTS Pilot 1: Analysis Report

Authors: Eva Aizpurua, Nick Waugh, Peter Cornick & Shane Howe
Prepared for: **Transport for Wales and The Welsh Government**



We believe that social research has the power to make life better.

By really understanding the complexity of people's lives and what they think about the issues that affect them, we give the public a powerful and influential role in shaping decisions and services that can make a difference to everyone. And as an independent, not for profit organisation we're able to put all our time and energy into delivering social research that works for society.

National Centre
for Social Research
35 Northampton Square
London EC1V 0AX
020 7250 1866

www.natcen.ac.uk

A Company Limited by
Guarantee

Registered in England
No.4392418

A Charity registered in England
and Wales (1091768) and
Scotland (SC038454)

This project was carried out
in compliance with ISO20252

designbysoapbox.com

Contents

1. Background	8
1.1 WNTS Pilot 1	8
2. Initial Assumptions – Volume of Journey Data	10
2.1 Is the volume of trips captured in Pilot 1 sufficient to track progress against the targets identified by the Welsh Government regarding active travel and public transport use?	10
Implications	12
3. How often do respondents record inaccurate journeys?	14
3.1 Inaccuracies related to journeys, journey legs, or locations	14
3.2 Inaccuracies related to transport modes	16
4. How often are respondents unable to locate places using the map?	18
5. Distribution of Journeys Across Days of the Week	21
5.1 Is the distribution of days of the week achieved during the web component of the survey balanced to accurately represent journeys across different days?	21
Implications	22
6. How comprehensive is the list of journey purposes?	22
7. Diary Design - Interleaved vs. Roster	24
7.1 Does the interleaved design increase the risk of underreporting journeys in the context of a travel diary?	24
8. Diary Placement – Early vs. Late	26
8.1 How does the placement of the diary within the survey influence the risk of underreporting journeys?	26
9. Diary Design and Placement	29
9.1 Does the placement of the diary moderate the effect of its design?	29
10. Within-household Selection	30
10.1 Assessing compliance with the selection process	31
Sample profile	31
Implications	33

Compliance	34
Indicators of Data Quality from the Questionnaire	37
<hr/>	
10.2 Consent to follow-up studies	37
10.3 Questionnaire length	38
Drop-offs	39
11. Recommendations	40
<hr/>	
11.1 Volume of Data and Approach to Follow-Up Questions	40
11.2 Diary Placement	40
11.3 Within-Household Selection	41
11.4 Summary Screens	42
11.5 Map Usability	43
11.6 Summary of Recommendations	43
12. Analytical Approach	45
<hr/>	

List of Tables

Table 1. Pilot 1 characteristics.	9
Table 2. Expected and achieved volumes of data.	10
Table 3. Number of recorded journeys.	11
Table 4. Spontaneous comments reported by respondents in response to an open-ended question focusing on map performance.	15
Table 5. Summary screen selections by day.	17
Table 6. Responses to map questions.	19
Table 7. Submissions and trips by day of the week.	21
Table 8. Purposes selected across journeys.	22
Table 9. Number of recorded journeys by design.	25
Table 10. Unadjusted response rate by diary placement.	26
Table 11. Number of recorded journeys by placement.	27
Table 12. Journeys by version and placement.	29
Table 13. Achieved response by selection method.	31
Table 14. Sex composition in the achieved samples and the population.	31
Table 15. Age composition in the achieved samples and the population.	32
Table 16. Race/ethnicity composition in the achieved samples and the population.	32
Table 17. Disability status in the achieved samples and the population.	33
Table 18. Welsh fluency in the achieved samples and the population.	33
Table 19. Self-reported household size for selecting participants.	34
Table 20. Computed respondent number from total household size.	34
Table 21. Recruitment outcomes for selected household members who are different from the individual who started the survey.	35
Table 22. Type of adjustment to selection response.	35
Table 23. Respondent selection questions prompting disengagement.	36
Table 24. Response outcomes.	36
Table 25. Consent for further research, overall and by placement condition.	37
Table 26. Completion time (minutes: seconds) by experimental condition.	38
Table 27. Valid responses to the first question in each block.	39
Table 28. Advantages and disadvantages for selection type.	41
Table 29. Summary of Recommendations from Pilot 1.	43
Table 30. Items used in the report by metric.	46

List of figures

Figure 1. Summary screen to review locations.	14
Figure 2. Summary screen to review mode(s) of transport.	16
Figure 3. Map interface.....	18
Figure 4. Follow-up question if respondents indicate that they could not find a location using the map.	18
Figure 5. First question of the Rizzo-Brick-Park within household selection method.....	30

1. Background

The Wales National Travel Survey (WNTS) is a web-first survey employing a combination of web, telephone, and face-to-face interviewing designed to measure travel attitudes and behaviours among residents of Wales.

The instrument consists of two components: (1) a questionnaire and (2) a 2-day travel diary, both of which are programmed using Blaise 5, a software suite developed by Statistics Netherlands for creating, managing, and deploying survey data collection instruments. Two pilots were set to take place in the second and third quarters of 2024 to inform decisions for the project launch in 2025.

1.1 WNTS Pilot 1

The first pilot, which was exclusively web-based, aimed to collect responses from 1,000 individuals. It followed a series of steps designed to identify and model optimal research designs and data collection strategies to satisfy the project's requirements. During these stages, a series of assumptions were made based on previous internal and external research. Some of these assumptions were robust, while others, supported by fewer or less directly relevant evidence, required appropriate testing. These included:

- 1) The volume of usable data captured by a 2-day travel diary programmed within the survey environment.
- 2) The distribution of journeys captured throughout days of the week.
- 3) The performance of two different approaches to follow-up questions for recording journey details.
- 4) The effect of diary placement within the overall survey.
- 5) The impact of within-household selection methods on data volumes and the composition of the achieved samples.

The first two areas (i.e., volumes of journey data and representation of days of the week) were naturally tested in the field, while the remaining aspects were evaluated by embedding three fully crossed randomised experiments in Pilot 1. Further information about the rationale for these experiments can be found in the "Diary Review Report" (Cornick, Aizpurua, & Howe, 2023).

Experiment 1 compared the performance of two diary designs: interleaved and rostered. In the interleaved version, follow-up questions were placed immediately after each filter question, capturing all details about a journey before moving on to the next. In the roster version of the diary, all filter questions were asked upfront, with details of each journey captured only after recording all visited locations in a given day. We hypothesised that the interleaved approach would lead to motivated misreporting, reducing the number of reported journeys.

Experiment 2 explored the impact of placing the travel diary earlier versus later within the survey. This experiment aimed to understand whether fatigue effects would influence responses in a meaningful way. Our expectation was that presenting the diary after the questionnaire would increase fatigue effects,

reducing the number of reported journeys and negatively affecting respondent experience.

Experiment 3 was designed to compare the performance of two methods for the selection of respondents within households. We expected that the non-random, 2-adult approach would result in a higher number of achieved responses, as it allows more than one person per household to participate in the survey. Following a random selection approach, we also anticipated the Rizzo-Brick-Park method would yield a sample more similar to the population than the non-probabilistic up-to-2-adults method.

Table 1. Pilot 1 characteristics.

Fieldwork dates	Friday 3 rd May – Sunday 2 nd June
Mode of data collection	Web only
Languages	English and Welsh
Communication strategy	Invitation (Friday 3 rd May) Reminder 1 (Monday 13 th May) Reminder 2 (Wednesday 22 nd May)
Sampling frame	Postal Address File
Issued addresses	7,500

2. Initial Assumptions – Volume of Journey Data

2.1 Is the volume of trips captured in Pilot 1 sufficient to track progress against the targets identified by the Welsh Government regarding active travel and public transport use?

A key requirement of the Wales National Travel Survey (WNTS) is the ability to detect a change in estimates of modal journey share from 33% to 35% between two survey years as statistically significant. As outlined in our report “Wales National Travel Survey: Additional Sampling Note” (Cornick, Aizpurua, Howe, & Keyes, 2023), achieving this requires an effective sample of 8,800 trips. The overall sample size required to achieve this number is reliant on several assumptions, including the number of productive cases per household, the average number of trips recorded per day, and the design effect of weighting the diary data.

A key aim of this pilot, therefore, was to test initial assumptions about the average number of trips recorded per day. This will be used to refine the diary approach and assumptions for the mainstage survey. Our initial assumptions were: (1) no dropouts between the first and second days of the diary, and (2) the volume of daily trips would be comparable to the English NTS (1.89 journeys on day 1 and 1.84 journeys on day 2).

With these assumptions, we projected collecting 2,000 travel days in the pilot. This projection was supported by the data, with a total of 2,016 days recorded. We expected to gather 3,730 journeys during these 2,000 diary days. However, the number of achieved journeys was 2,743 (-26.5%), below initial assumptions.

Table 2. Expected and achieved volumes of data.

	Expected	Achieved
Responses	1,000	1,008
Total days of diary data recorded	2,000	2,016
Approx. total journeys recorded	3,730	2,743
Average number of trips recorded on Day 1	1.9	1.7
Average number of trips recorded on Day 2	1.8	1.0

Because the number of journeys was based on the count of substantive responses to the main transport mode question (see Table 30 for further details), it is possible that some journeys have been underreported (e.g., respondents who recorded origin and destination but skipped the main mode question), and others likely overreported (e.g., uncodable origin or destinations, or ineligible journeys such as flights or walks under 5 minutes). Later sections in this report provide further evidence that the number of usable journeys will be lower. This is attributed to wrongly entered journeys and uncodable locations (see “How often do respondents record inaccurate journeys?”).

On average, we expected 1.9 journeys on day 1 (yesterday) and 1.8 journeys on day 2 (the day before yesterday). The actual figure for day 1 was lower at 1.7 journeys per person ($SD = 1.6$) and even lower on day 2, where 1.0 journey was recorded on average ($SD = 1.4$). The difference in the mean number of reported journeys between days 1 and 2 was significant, $t(1007) = 11.59, p < .001$.

This discrepancy is likely attributed to underreporting on day 2, potentially influenced by the cumulative burden of completing two full days of travel diary and a questionnaire. This is supported by the increased rate of no trips on day 2 (51.6%, an increase of 77.9% from day 1), which suggests that some respondents who had completed trips did not want to enter them into day 2 of the diary – perhaps because it would have been burdensome to do so. Additionally, the proportion of respondents who reported 3 or more journeys dropped from 39.8% on day 1 to 10.9% on day 2 (see Table 3), further evidencing resistance to reporting trips.

Table 3. Number of recorded journeys.

# of journeys	Day 1	Day 2	Overall
0	292 (29.0%)	520 (51.6%)	186 (18.5%)
1	204 (20.2%)	145 (14.4%)	131 (13.0%)
2	277 (27.5%)	233 (23.1%)	258 (25.6%)
3	111 (11.0%)	49 (4.9%)	127 (12.6%)
4	68 (6.7%)	33 (3.3%)	131 (13.0%)
5	25 (2.5%)	15 (1.5%)	56 (5.6%)
6	15 (1.5%)	10 (1.0%)	45 (4.5%)
7	12 (1.2%)	1 (0.1%)	26 (2.6%)
8	3 (0.3%)	2 (0.2%)	17 (1.7%)
9	1 (0.1%)		14 (1.4%)
10			6 (0.6%)
11			2 (0.2%)
12			5 (0.5%)
13			2 (0.2%)
14			1 (0.1%)
15			1 (0.1%)
	Day 1	Day 2	Overall
	M (SD)	M (SD)	M (SD)
Trip rate	1.7 (1.6)	1.0 (1.4)	2.7 (2.4)

M: mean; SD: standard deviation

The proposed design of WNTS is substantially different to that of the English NTS. Whereas the English NTS uses a face-to-face design and a seven-day travel diary which is collected in real time (i.e., over a period of one week), the WNTS proposes a web-first, mixed-mode design and collects travel data during the survey using a recall approach (i.e., collecting travel behaviour over the previous two days).

Therefore, it is valuable to benchmark the trip rates achieved in the WNTS pilot against other travel surveys. Perhaps the closest comparison is with Scotland. The Scottish Household Survey includes a travel diary module that collects travel behaviour over the previous day using a 24-hour recall. Similar to the WNTS, the diary element is conducted as part of a wider survey and uses a recall-based approach. In the three years prior to 2020 (when the pandemic necessitated a change in survey mode) the (weighted) trip rate recorded in the Scottish Household Survey averaged 1.9. However, this fell to 1.52 in the most recent survey (2022).¹ As such, the trip rates recorded on day 1 of the WNTS diary are in-line with those achieved in the most recent Scottish Household Survey (this is especially true when looking at particular versions of the diary, as outlined later in this report). However, the substantial drop-off in recorded trips for day 2 suggests underreporting, most likely caused by the burden of entering trip information into the diary.

Implications

If this volume of data were to be collected in the mainstage, it is highly likely that it would still be possible to achieve the overall level of precision (based on an effective sample size of 8,800 trips) within the available budget. However, meeting the requirements around facilitating region-level analysis would require an increase in the budget.

Further, while it is possible to compensate for the drop-off in trip rates between day 1 and day 2 with weighting, this would both increase the design effect (significantly increasing the number of diaries required to reach precision targets) and bring into question the robustness and reliability of the day 2 data. This leaves two broad approaches:

- 1) Shortening the diary recall period, from 2 days to 1, whilst keeping a similar level of detail in the follow-up questions.** The advantage of this approach is that initial analysis of the pilot data indicates it would be possible to retain a similar level of detail captured around trips. The disadvantage is that it would significantly increase the required sample size, and therefore costs, required to achieve the desired precision estimates.
- 2) Maintaining the recall period (2 days) whilst substantially reducing the number of follow-up questions in the travel diary.** The advantage of this approach is that we would anticipate it would boost the number of trips reported (although this would need to be tested), meaning that it would likely be possible to meet the required precision targets within

¹ <https://www.transport.gov.scot/publication/transport-and-travel-in-scotland-2022/personal-travel/>

budget. The disadvantage is that we would be collecting less detail on each trip.

Other options to consider include reducing the questionnaire length (although it is unclear how much of an impact this would have on the burden of the diary task), and - if the recall period is reduced to 1 day - there might be an option to collect a second travel day, either immediately after the survey by incentivising it separately, or in the future. There is some evidence later in this report that reducing the length of the questionnaire content could be beneficial in reducing underreporting, particularly for the roster version, where placing the diary before the questionnaire resulted in a higher number of journeys reported, which is consistent with the fatigue hypothesis (for further details, see “Diary Placement – Early vs. Late” and “Diary Design and Placement”).

3. How often do respondents record inaccurate journeys?

Summary screens are commonly included in complex data collection instruments, such as recall-based diaries. Their purpose is to summarise the information entered by respondents, allowing them to confirm its accuracy or amend relevant details. While these screens can help reduce measurement error, they also add to the respondents' burden, as they are required to review the summary screens to confirm or correct their previously provided answers.

For a travel diary programmed within survey software, including summary screens is more challenging than if other software were used (e.g., application software, diary-specific solutions). This is because, for this task, survey software capabilities are limited, as there is a greater focus on promoting a linear flow through the questionnaire, to minimise the risks of respondents looping through the survey and re-answering questions they have already answered. For this reason, Pilot 1 was designed not to ask respondents to correct any information but to only quantify the number of instances respondents reported issues within their trips, including incorrectly entering journeys, locations, and/or transport modes.

3.1 Inaccuracies related to journeys, journey legs, or locations

The first summary screen included in the travel diary displayed the location information for each of the journeys reported by respondents on a given day (see Figure 1). Overall, most responses to the summary screens indicated that the information entered was correct. On day 1, 92.4% of the responses (1,495 instances) confirmed that the location information was correct. On day 2, this figure rose to 95.4% (933). The lower error rate on day 2 could indicate a learning curve where respondents make fewer errors after completing a full day of travel. However, it could also reflect that day 2 captured fewer journeys and, therefore, fewer opportunities to make mistakes, as well as increased disengagement with the screens as they approached the end of the diary.

Figure 1. Summary screen to review locations.

Please review the information below about your journeys from yesterday. You will not be asked to correct any entries, only to specify any errors. If more than one issue applies to a single journey, you can select all relevant options.

Journey 1

From: HOME
To: LLANFYLLIN LEISURE CENTRE, LLANFYLLIN, SY22 5BJ
Please select all that apply.

<input type="checkbox"/> This information is correct
<input type="checkbox"/> The start and/or end location for this journey is incorrect
<input type="checkbox"/> A journey is missing before or after this journey
<input type="checkbox"/> This journey did not take place

Journey 2

From: LLANFYLLIN LEISURE CENTRE, LLANFYLLIN, SY22 5BJ
To: BRYNTEG, CARDIFF ROAD, BARGOD, CF81 8NY
Please select all that apply.

<input type="checkbox"/> This information is correct
<input type="checkbox"/> The start and/or end location for this journey is incorrect
<input type="checkbox"/> A journey is missing before or after this journey
<input type="checkbox"/> This journey did not take place

The most frequently reported errors were issues with the locations entered. On day 1, 65 responses indicated errors with the start and/or end of the journeys. This represents 4% of the total responses and slightly over half of the responses indicating location errors (52.8%). On day 2, this figure dropped to 19 (1.9% of total responses and 42.2% of location errors).

The second most frequent category of location errors referred to journeys that did not take place and were incorrectly recorded. These included 43 instances on day 1 (2.7% of responses and 35.0% of errors) and 21 on day 2 (2.1% of responses and 46.7% of errors).

Respondents reported the need to modify a journey or part of a journey less frequently on both day 1 (15 cases, 0.9% of responses and 12.2% of errors) and 2 (5 cases, 0.5% of responses and 11.1% of errors).

These results point to some errors in the location information that the diary relies on. This information is used in the follow-up questions to provide context about the journey in question and is also used for some estimations, such as distance travelled. For this reason, the number of estimated journeys based on reported mode described in earlier sections overestimates the number of journeys in which location information is not available or accurately reported.

It is uncertain, however, to what extent respondents would have corrected their entries had the summary screens been editable. Pilot 1 was intended to quantify the extent of these problems and did not require editing of the journeys, which could improve accuracy but result in further drop-offs.

Respondents were also asked about their experience of entering location information using the map.

Table 4 describes some of the issues reported by respondents in the open-ended questions regarding map performance.

Table 4. Spontaneous comments reported by respondents in response to an open-ended question focusing on map performance.

Map not loading or displaying properly on certain devices	<p><i>"The maps wouldn't load so had to select the option 'can't find location on the map' in order to proceed".</i></p> <p><i>"It did freeze for a little while but overall was easy to use".</i></p> <p><i>"Maps working alright but zooming in and out extremely tedious".</i></p>
Issues with the scrolling functionality of the retrieved locations	<p><i>"No difficulties except struggled to scroll through to correct flat number".</i></p>
Respondents not able or to find or select the exact location	<p><i>"Yes, I found it difficult. The need to find an address when I was travelling to a beach did not really work. I am not sure if I entered it correctly".</i></p> <p><i>"Tedious in the extreme, also despite selecting 19 [REDACTED] each time sometimes the survey recorded 10 [REDACTED]".</i></p>

Respondents not willing to provide the exact location. Some respondents reported feeling uncomfortable providing specific addresses

"Yes, I only wanted to put the location, for example the town or village name. I shouldn't have to specify the exact location e.g. shop or house I was in".
"It would have been nice to record 'returns' without additional map entries. I would also like to give postcode areas over exact addresses for privacy".
"Not always easy to just select a town or village without choosing a specific address which I considered too personal, possibly giving other people's address".

Difficulty understanding the task, particularly the requirement to enter journeys back as separate journeys

"Don't think I realised in time I needed to record journey back home for journey 1 sorry".
"Entering the detail of the journey was confusing as I started and finished my walk from home, but the destination was to the local lake for a walk. The question made it appear that I'd finished my day at the lake rather than back at home".

Issues locating train stations or bus stops, especially when the journey involves using multiple stages

"No difficulties but maybe seeing the bus stops might make things more accurate."
"1) Entering the name of a railway station did not find it on the map. 2) The first question did not specify if I should include multiple steps of the same journey (e.g. a train journey with multiple changes) and I had to go back to restart as it asked me how I got from a train to a train and wouldn't let me select train".
"It's horrible to choose location, no possibility to specify that you had to change buses several times to reach destination".

3.2 Inaccuracies related to transport modes

The second summary screen was intended to confirm the modes of transport reported for the journeys (see Figure 2).

Figure 2. Summary screen to review mode(s) of transport.

Please review the information below about the modes of transport you used yesterday. You will not be asked to correct any entries, only to specify any errors. If more than one issue applies to a single journey, you can select all relevant options.

Journey 1

From: HOME
To: LLANFYLLIN LEISURE CENTRE, LLANFYLLIN, SY22 5BJ
Main mode of transport: **Car or other motor vehicle**
Other mode(s): **Walking or wheeling,**
Please select all that apply.

This information is correct

The main mode of transport for this journey is incorrect

Any secondary mode of transport for this journey is incorrect

Journey 2

From: LLANFYLLIN LEISURE CENTRE, LLANFYLLIN, SY22 5BJ
To: BRYNTEG, CARDIFF ROAD, BARGOD, CF81 8NY
Main mode of transport: **Car or other motor vehicle**
Other mode(s): **Cycle, Walking or wheeling,**
Please select all that apply.

This information is correct

The main mode of transport for this journey is incorrect

Any secondary mode of transport for this journey is incorrect

Most responses confirmed that the information provided was correct. On days 1 and 2, this option was selected in slightly over 98% of the cases (98.2% on day 1 and 98.8% on day 2). The number of occasions in which issues were reported regarding the main or the secondary mode(s) of transport was, therefore, low (main mode: 15 cases on day 1 and 7 on day 2; secondary modes: 13 cases on day 1 and 5 on day 2).

Table 5. Summary screen selections by day.

	Day 1	Day 2
<u>Location check</u>		
Everything is correct	1495	933
I need to modify a location	65	19
I need to modify a journey/ part of a journey	15	5
This journey did not take place	43	21
<u>Mode check</u>		
This information is correct	1571	963
The main mode of transport for this journey is incorrect	15	7
Any other mode of transport for this journey is incorrect	13	5

Implications

The analysis of the summary screen responses indicates that some journeys, or the information related to them, such as start or end location and, to a lesser extent, the transport modes, were incorrectly entered and would benefit from correction. This correction could be implemented in the form of an editable summary screen that allows respondents to amend the information wrongly entered into the diary or allowing them to indicate that a journey was incorrectly entered at the beginning of the journey loop to skip journey-specific questions.

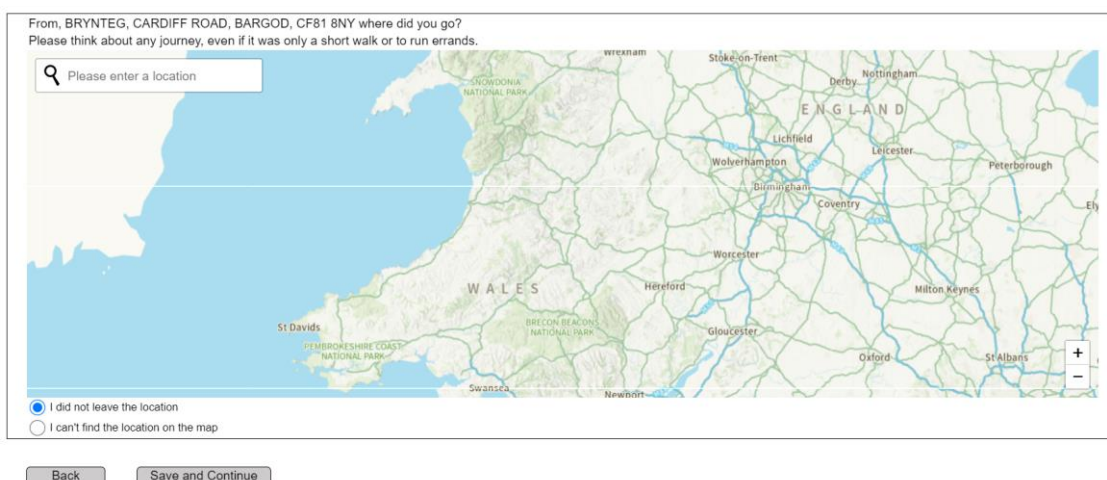
As shown later in the report, the analysis of the “other” responses to the journey purpose question also suggests that respondents who entered journeys that did not take place were unable to answer the follow-up questions, indicating in the free-text questions that the journey should not be included. These results highlight potential issues if mechanisms to allow corrections are not incorporated, including increased random error in the estimations and potential frustration from respondents navigating the diary.

On the other hand, it is clear that the majority of journeys entered were ‘correct’, as reported by respondents, and did not need amending. Adding summary screens can increase respondent burden and, potentially, lead to dropout. It is important to strike a balance between the detail and accuracy of data entered and respondent burden. As such, any introduction of measures to correct erroneously entered journey information should be fully tested.

4. How often are respondents unable to locate places using the map?

In contrast to other surveys, like the English NTS, which collect general location information such as areas of a city or town, the WNTS aims to gather precise location data about the places visited by respondents during the reference period. To achieve this, the survey has an Application Programming Interface (API) integration (using the *Ordnance Survey* Places and Names databases) where respondents are expected to use a map to locate places. Only if they are unable to locate places using the map are they asked to indicate so and describe the location using a free-text open field (see Figure 3 and Figure 4).

Figure 3. Map interface.



From, BRYNTEG, CARDIFF ROAD, BARGOD, CF81 8NY where did you go?
Please think about any journey, even if it was only a short walk or to run errands.

Please enter a location

I did not leave the location
 I can't find the location on the map

Back Save and Continue

Figure 4. Follow-up question if respondents indicate that they could not find a location using the map.

You indicated that you were unable to find your location on the map. Please describe the place you were looking for. For example, "Newport Market" or "Porthmadog Train Station".

Back Save and Continue

While this option is necessary to account for situations in which addresses or places are not registered in the *Ordnance Survey* database, it also increases the risk of receiving unusable answers. For example, if locations are used to calculate distances, and respondents record unspecific places which cannot be pinpointed (e.g., a friend's house), the usability of the information collected from those journeys will be limited. In other instances, respondents might describe a place that can be pinpointed, with some degree of error (e.g., a town or a street with no address attached), reducing the accuracy of the estimates.

When asked to find their location on the map, there were 278 instances whereby respondents selected the "I could not find my location" option. This represents 7.9% of all map entries, a non-trivial proportion of responses which require post-processing work.

Table 6. Responses to map questions.

Map responses	N	%
API match	2,748	78.2
Could not find location	278	7.9
Map loop end	487	13.9
Total	3,513	100

Of the 272 responses to the open-ended fields, four main categories of descriptions emerged:

- 1) Town, city, or region names** was the most common category (>100 times). It included references to towns/cities (e.g., Cardiff, Exeter), regions/areas (e.g., North Rd in Aberystwyth, Thornhill in Cardiff), and in some cases countries (e.g., England, Poland). Some referred to ineligible places outside of the country (e.g., Brussels, Florida, Gibraltar, Majorca, Poland, Washington DC). While eligible places could potentially be coded to a specified point (e.g., an agreed mid-point in that location), they would produce less precise estimates than specific locations and would require additional coding efforts of the open-ended responses. Ineligible journeys would be excluded, reducing the number of usable information.
- 2) Places specific enough that could be pinpointed** was the second, most frequent category (80 instances). These included references to train stations (e.g., Wrexham train station), airports (e.g., Heathrow, Bristol airport), specific stores or coffee shops (e.g., Lidl Supermarket Gorseinon), hotels (e.g., Falcondale Hotel Lampeter), sport centres (e.g., Canolfan Carwyn Sports Centre in Drefach), landmarks or parks (e.g., Castell Coch, Prestatyn Retail Park), and service providers (e.g., Beechouse Surgery Denbigh, CK Post Office Johnstown). These places could be matched to addresses during data processing, although it would extend processing times due to the manual nature of the task.
- 3) Vague place names** was the third most frequent category (30 occasions). It included references to services within towns that were too vague to be pinpointed to specific locations (e.g., B&Q, farm, local park, dentist, Morrisons) or where multiple locations could be a match (e.g., leisure club Llanfynydd, school Tonteg).
- 4) Home** was a common response (28 instances), and one which was ambiguous at times. The reason is that, in some cases, it was unclear whether the respondent was referring to their own home, or someone else's home, as they included some additional information (e.g., "Home, Maes Y Coed, Pontypridd", "Home, near Berriew", "home – Buckley").
- 5) Postcodes and addresses** were relatively uncommon for respondents to record in open fields (<20 cases), likely suggesting that specific addresses are unknown for most locations

Because the free-text open fields had very few restrictions regarding what information respondents could enter, in some instances, multiple places were recorded for a single journey (e.g., "Siop Anrhegaron, Tregaron & Spar Tregaron", "Swansea centre and Dobbies LLansamlet to meet friends"). These

respondents may have treated multiple journeys as one. While this was not a frequent occurrence (< 5), the presence of these answers highlights the need to develop protocols to establish how they might be coded or used when the survey launches.

On a few occasions, journey purposes were recorded instead of locations (e.g., hospital appointment, short walk), or respondents provided uncodable answers (e.g., 2, x). Both make any further information or questions which rely on location data unusable.

In a few other instances respondents did not follow the routing as intended, as they selected "I couldn't find my location" in the map to later indicate that they did not travel (e.g., "nowhere stayed at home", "did not go out").

These latter issues were rare occurrences in the data. If they were more prevalent, however, the amount of available, useful information could be substantially less than anticipated given the number of reported journeys. Such a reduction in information would harm the accuracy of the data and mean that the sample size available could vary depending on the type of information. For example, estimates based on specific location data would be potentially based on a lower analytical sample (due to the lack of useful information) than estimates based on other types of information (e.g., travel frequency items from the questionnaire).

5. Distribution of Journeys Across Days of the Week

5.1 Is the distribution of days of the week achieved during the web component of the survey balanced to accurately represent journeys across different days?

A further key requirement of the project is to produce data that account for seasonality and balance the distribution of journeys across days of the week. Whilst random allocation of the start day would be optimal, the self-completion element makes it impossible to ensure compliance and could negatively impact response rates. For this reason, Pilot 1 was intended to provide information about the expected distribution of web responses if a weekly mailing approach with staggered days for the reminders were to be implemented.

We examined the number of submissions by day of the week and then estimated the number of journeys per day, under the assumption that day 1 would be the day prior to submission and day 2 would have taken place two days before the submission. This might not hold true if respondents completed the survey in multiple sessions.

As shown in Table 7, the proportion of submissions was highest on Tuesday and Wednesday. For reference, the invitation letter was sent out on a Friday and reminders 1 and 2 were sent out on Monday and Wednesday, respectively. The days with the fewest submissions were Mondays, Saturdays, and Sundays. Most journeys were reported on Mondays and Tuesdays, whilst Fridays and Saturdays had the fewest reported journeys.

It is important to note that there were two Bank Holiday Mondays within the Pilot 1 period, which may have affected submission levels on Mondays and consequently the coverage of Saturday and Sunday diary days. Additionally, the first Bank Holiday was just after the first survey mail-out. If the first letter arrived on the Tuesday or Wednesday, this could have contributed to the large initial response rate observed on those days.

Future work when the survey is live will be conducted to explore mailing strategies and interviewer interventions in the face-to-face component to further balance journey representation across days of the week.

Table 7. Submissions and trips by day of the week.

Day	Submissions	Journeys
Monday	70 (6.9%)	646 (23.6%)
Tuesday	279 (27.7%)	541 (19.7%)
Wednesday	233 (23.1%)	364 (13.3%)
Thursday	118 (11.7%)	353 (12.9%)
Friday	148 (14.7%)	255 (9.3%)
Saturday	85 (8.4%)	203 (7.4%)
Sunday	75 (7.4%)	381 (13.9%)

Implications

From the pilot, it is clear that issuing the mailings in one batch per week skews responses towards the middle of the week and away from the weekend. While it is possible to correct this with weighting, it will be important to explore alternative ways to smooth the distribution of responses across the week. For the web-element of the survey, this implies experimenting with issuing the sample in more than one batch per week once the survey launches.

It is currently unclear what impact the use of the face-to-face follow-up option for non-responding addresses will have on distribution by day of the week. The second pilot will shed further light on this element, albeit only on a small scale.

6. How comprehensive is the list of journey purposes?

When respondents record journeys, they are asked to indicate their purpose, with this diary question having the longest list of categories (13 in total). The categories are designed to be comprehensive, reducing the number of “other” write-in answers, while remaining compact enough to ensure that no category receives very few responses. The list was created during the diary design phase, with input from travel modellers, and then tested qualitatively through cognitive and usability interviews. Pilot 1 was, therefore, the first large-scale testing of this item.

As shown in Table 8, the most frequently selected categories were “return home”, “shopping, even if there was no intention to buy”, “go to work”, and “social or entertainment”.

Table 8. Purposes selected across journeys.

Purpose	Count
Exercise or play sports	281
Go to work	379
Holiday or day trip	97
Medical consultation or treatment	70
Personal appointment (e.g., bank, hairdresser, launderette)	30
Personal journey during work	6
Shopping, even if there was no intention to buy	429
Education (e.g., go to school, college)	53
Social or entertainment (e.g., meet friends or family, voluntary work)	327
Informal caring responsibilities (e.g., care for an elderly relative)	30
Accompany someone (e.g., drop someone off, pick someone up)	129
Return home	903
Other	187

The option “personal journey during work” was selected only six times across all journeys. This selection option may not be a viable stand-alone category and responses could be absorbed by similar categories (i.e., personal appointment) or the “other” selection option. The categories “personal appointment” and “informal caring responsibilities” were also among the least frequently selected, each selected 30 times.

The “other” option was selected 187 times and any respondent who selected “other” was asked to describe the purpose of their journey in an open text-field. For 13 (7.0%) of these responses, respondents indicated that the data had been entered by error and the journey did not take place. As one respondent noted, “[Journey] didn't happen; survey wouldn't erase it”.

Slightly over one-third of journey purposes entered in the open text-field ($n = 69$, 36.9%), however, belonged to existing categories and could, therefore, be recoded into them. Such responses belonged to the “social or entertainment” category (e.g., “family visit”, “leisure”, or “voluntary work”), the shopping category (e.g., “to visit shops and leisure centre”, or “I walked to the shop near here for milk”), the “returning home” category (e.g., “to get home”, or “bring food home”), and the “accompany someone” category (e.g., “took my daughter to school”, or “took my daughter to dance class”). These responses suggest that respondents were not carefully reading all response options, likely given the relatively long list (13 categories) and the need to scroll vertically to see all categories on small devices.

While a substantial number of responses entered by respondents in the “other” open text-field belonged to existing categories and could be recoded into them, others were journey purposes not covered by the list ($n = 102$, 54.5%). New categories with over 10 responses included:

- **Collecting/ dropping/ walking/ caring for animals:** “drop-off dog”, “walk dog”, “pick-up parents’ dog”, “vet trip for two dogs”, “to feed my horses”, “took cat to vets”, “to collect the cat”, “feeding ponies for a neighbour”, “check sheep on farm”.
- **Eating/ Drinking/ Collecting food or drinks:** “have dinner”, “eat lunch”, “get breakfast”, “buy coffee”, “food takeaway”, “eat a meal out”, “for coffee and refreshments”.
- **Church-related activities:** “funeral”, “going to the Sunday service”, “mass”, “attend church”, “to attend a chapel service”.

Other relatively common responses (between 5 and 10 instances) included journeys within workplaces (“travel within work to another place”, “driving between work locations”), collecting or dropping parcels (“retrieve a parcel delivered wrongly”, “drop off an Evri parcel for Vinted”, “post a letter”, “visit in-store Post Office”), and dropping or collecting vehicles from the garage.

7. Diary Design - Interleaved vs. Roster

7.1 Does the interleaved design increase the risk of underreporting journeys in the context of a travel diary?

A key decision in designing the diary was the approach to follow-up questions (that is, questions which collect detailed information for each journey). In the pilot, we experimented with two main approaches:

- 1) Interleaved:** Where filter questions (i.e., locations visited) are immediately followed by detailed questions about each journey. This approach gathers all details about a journey before moving on to recording the next journey. The English National Travel Survey is an example of this design.
- 2) Roster:** In this design, all filter questions are asked upfront, and details about each journey are not gathered until all locations have been recorded. This approach is used in other travel surveys such as the Dutch Mobility Panel.

We compared both approaches to test whether the interleaved version would show signs of “motivated misreporting” (i.e., underreporting due to the respondents’ desire to reduce the burden of the survey). For this, we compared the number of journeys recorded in each approach. The rationale was that respondents would learn about the follow-up questions after the first filter question in the interleaved format and underreport trips to reduce their burden. Differences between the interleaved and roster approaches were expected to be greater on day 1, since by day 2, those completing the roster version would also have learned that detailed questions follow responses to the filter questions.

Aligning with our expectations from the literature, on day 1, the average number of journeys was significantly higher in the roster design compared to the interleaved (Roster $M = 1.9$; Interleaved $M = 1.5$; $t(928.6) = -3.65$, $p < .001$; $d = -0.23$). By day 2, the number of journeys reported was nearly identical (Roster $M = 1.0$; Interleaved $M = 1.1$; $t(963.4) = 0.90$, $p = .369$). Overall, there were no significant differences between the roster and interleaved diary in mean journey count across the two days (Roster $M = 2.9$; Interleaved $M = 2.6$; $t(951.0) = -1.92$, $p = .055$)².

The proportion of respondents reporting no journeys was higher in the roster version compared to the interleaved on both day 1 (31.9% vs. 26.2%) and day 2 (58.1% vs. 45.3%). While a greater proportion of respondents reported no journeys in the roster version, on day 1, the proportion of individuals recording 3 or more journeys was substantially larger among those assigned to the roster design (30.9% vs. 16.0% in the interleaved). However, consistent with the results on average number of journeys, this difference diminished by day 2 (11.5% in the roster version vs. 10.5% in the interleaved).

² The results were consistent when using non-parametric tests (Wilcoxon rank sum test with continuity correction), suggesting differences on individual days but not overall (aggregated count for days 1 and 2).

Table 9. Number of recorded journeys by design.

Journeys	<u>Day 1</u>		<u>Day 2</u>		<u>Overall</u>	
	Roster	Interleaved	Roster	Interleaved	Roster	Interleaved
0	157 (31.9%)	135 (26.2%)	286 (58.1%)	234 (45.3%)	104 (21.1%)	82 (15.9%)
1	55 (11.2%)	149 (28.9%)	43 (8.7%)	102 (19.8%)	49 (10.0%)	82 (15.9%)
2	128 (26.0%)	149 (28.9%)	107 (21.7%)	126 (24.4%)	108 (22.0%)	150 (29.1%)
3	77 (15.7%)	34 (6.6%)	23 (4.7%)	26 (5.0%)	73 (14.8%)	54 (10.5%)
4	33 (6.7%)	35 (6.8%)	14 (2.8%)	19 (3.7%)	56 (11.4%)	75 (14.5%)
5	20 (4.1%)	5 (1.0%)	9 (1.8%)	6 (1.2%)	37 (7.5%)	19 (3.7%)
6	11 (2.2%)	4 (0.8%)	7 (1.4%)	3 (0.6%)	20 (4.1%)	25 (4.8%)
7	9 (1.8%)	3 (0.6%)	1 (0.2%)	0 (0.0%)	16 (3.3%)	10 (1.9%)
8	1 (0.2%)	2 (0.4%)	2 (0.4%)	0 (0.0%)	9 (1.8%)	8 (1.6%)
9	1 (0.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	8 (1.6%)	6 (1.2%)
10					3 (0.6%)	3 (0.6%)
11					2 (0.4%)	0 (0.0%)
12					3 (0.6%)	2 (0.4%)
13					2 (0.4%)	0 (0.0%)
14					1 (0.2%)	0 (0.0%)
15					1 (0.2%)	0 (0.0%)
	<u>Day 1</u>		<u>Day 2</u>		<u>Overall</u>	
	Roster	Interleaved	Roster	Interleaved	Roster	Interleaved
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Trip Rate	1.9 (1.8)	1.5 (1.4)	1.0 (1.5)	1.1 (1.3)	2.9 (2.6)	2.6 (2.2)

M: mean; SD: standard deviation

8. Diary Placement – Early vs. Late

8.1 How does the placement of the diary within the survey influence the risk of underreporting journeys?

The placement of the diary was a key consideration during the design and modelling stages of the WNTS. By placing the diary later in the survey, respondents would have had more time to acclimatise to the survey before encountering the more demanding diary element, potentially reducing the risk of early disengagement. However, respondent fatigue was expected to be most intense toward the end of the survey, likely reducing the volume and quality of data collected by late diaries.

Response rate figures (see Table 10) showed that significantly more respondents completed the survey when they were assigned to a version with a late diary placement, compared to versions with an early diary placement ($X^2(1) = 10.90, p < .001$). Respondents were more likely to disengage from the survey when they encountered the diary earlier and, as a result, late placement of the diary contributed to a greater number of submissions.

Table 10. Unadjusted response rate by diary placement.

	Diary Placement	
	Early	Late
Unadjusted response rate	454 (8.1%)	554 (9.8%)

However, as anticipated, placement of the diary also affected journey volumes among those who completed the survey. Placing the diary earlier within the survey resulted in a significantly higher number of reported journeys in both days 1 (Early placement $M = 1.9$; Late placement $M = 1.5$; $t(905.4) = 3.28, p = .001, d = 0.21$) and 2 (Early placement $M = 1.2$; Late placement $M = 0.9$; $t(859.5) = 2.32, p = .020, d = 0.15$). This resulted in the overall count being half a journey higher when the diary was placed earlier than when it was placed later (Early placement $M = 3.0$; Late placement $M = 2.5$; $t(868.0) = 3.51, p < .001, d = 0.23$)³.

The proportion of respondents with no journeys was slightly lower when the diary was placed earlier for both day 1 (26.4% vs. 31.0%) and day 2 (50.7% vs. 52.3%). Similarly, a higher proportion of respondents reported 3 or more journeys when the diary preceded the questionnaire on both day 1 (27.2% vs. 20.3%) and day 2 (13.4% vs. 8.9%).

Both sets of results are consistent with the hypothesis that respondents who were allocated to a late diary would reach the diary feeling greater fatigue, increasing how much they underreported trips. When the diary was placed earlier in the survey the journey count was greater but fewer people completed the survey.

³ The results were generally consistent when using non-parametric tests (Wilcoxon rank sum test with continuity correction), suggesting differences in day 1 and at the aggregate level.

Table 11. Number of recorded journeys by placement.

Journeys	<u>Day 1</u>		<u>Day 2</u>		<u>Overall</u>	
	Early	Late	Early	Late	Early	Late
0	120 (26.4%)	172 (31.0%)	230 (50.7%)	290 (52.3%)	72 (15.9%)	114 (20.6%)
1	81 (17.8%)	123 (22.2%)	59 (13.0%)	86 (15.5%)	58 (12.8%)	73 (13.2%)
2	130 (28.6%)	147 (26.5%)	104 (22.9%)	129 (23.3%)	109 (24.0%)	149 (26.9%)
3	52 (11.5%)	59 (10.6%)	23 (5.1%)	26 (4.7%)	64 (14.1%)	63 (11.4%)
4	38 (8.4%)	30 (5.4%)	17 (3.7%)	16 (2.9%)	51 (11.2%)	80 (14.4%)
5	14 (3.1%)	11 (2.0%)	11 (2.4%)	4 (0.7%)	35 (7.7%)	21 (3.8%)
6	8 (1.8%)	7 (1.3%)	8 (1.8%)	2 (0.4%)	19 (4.2%)	26 (4.7%)
7	9 (2.0%)	3 (0.5%)	0 (0.0%)	1 (0.2%)	16 (3.5%)	10 (1.8%)
8	1 (0.2%)	2 (0.4%)	2 (0.4%)	0 (0.0%)	9 (2.0%)	8 (1.4%)
9	1 (0.2%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	9 (2.0%)	5 (0.9%)
10					3 (0.7%)	3 (0.5%)
11					1 (0.2%)	1 (0.2%)
12					4 (0.9%)	1 (0.2%)
13					2 (0.4%)	0 (0.0%)
14					1 (0.2%)	0 (0.0%)
15					1 (0.2%)	0 (0.0%)
	<u>Day 1</u>		<u>Day 2</u>		<u>Overall</u>	
	Early M (SD)	Late M (SD)	Early M (SD)	Late M (SD)	Early M (SD)	Late M (SD)
Trip rate	1.9 (1.7)	1.5 (1.5)	1.2 (1.5)	0.9 (1.2)	3.0 (2.6)	2.5 (2.2)

M: mean; SD: standard deviation

While the later placement of the diary resulted in a lower volume of journeys (see Table 11), which we attributed to suboptimal response behaviour driven by fatigue, the perceived difficulty was comparable between the two conditions. Respondents in both conditions rated the survey as between “not at all” and “slightly” difficult (Early placement $M = 1.39$; Late placement $M = 1.45$; $t(999.7) = -1.27$, $p = .205$).

Similarly, respondents’ expectations of how long it would take to complete the WNTS were comparable between the two placement groups. Both groups indicated that the survey took “about as long as they expected” (Early placement $M = 2.95$; Late placement $M = 3.01$; $t(950.6) = -0.92$, $p = .358$). Furthermore, perceived time taken to complete the WNTS (reported in minutes) was also comparable between the groups (Early placement $M = 30.18$; Late placement $M = 27.79$; $t(775.6) = 1.12$, $p = .261$). These findings suggest that those who remained in the survey found the task no more time-consuming or burdensome than respondents who encountered the diary later on.

Overall, the evidence here suggests that respondents who encountered an early diary were more likely to disengage and exit the survey, when compared to those who encountered a late diary. However, if they did complete the survey, respondents produced higher quality data on average (as measured by the level of underreporting) and subjectively found the survey to be just as taxing.

9. Diary Design and Placement

9.1 Does the placement of the diary moderate the effect of its design?

While the design and placement of the diary were analysed separately in previous sections, in practice, they might interact with each other, with placement potentially moderating the impact of the diary design. To test this hypothesis, we estimated two negative binomial regression models predicting the number of reported journeys (as proxied by transport mode responses) by version and placement first to analyse main effects, and then included an interaction term between the two to explore potential moderation effects.⁴

The results from the main effects model suggest that, holding the diary version constant, having a late diary placement is associated with a decrease in the total number of journeys compared to an earlier placement (IRR⁵ = 0.824; $p < .001$). Although the model also suggests an increase in the number of recorded journeys for the roster version compared to the interleaved, this effect was not significant ($p = 0.073$), which is consistent with the bivariate results.

The results from the second regression model indicate that the interaction term is significant (IRR = 0.676; $p < .001$). This suggests that the effect of the diary design depends on the timing of the diary placement. While the roster version generally increases the number of recorded journeys, this increase is less pronounced when the diary is placed late.

As shown in Table 12, the interleaved design produced the same average number of journeys irrespective of placement. However, the roster version produced a significantly higher number of journeys when presented early (Early placement $M = 3.5$; Late placement $M = 2.4$; $t(403.4) = 4.70$, $p < .001$; $d = 0.44$).

Table 12. Journeys by version and placement.

Version	Placement	# of journeys M (SD)
Roster	Early	3.5 (2.9)
Interleaved	Early	2.6 (2.3)
Roster	Late	2.4 (2.3)
Interleaved	Late	2.6 (2.0)

M: mean; SD: standard deviation

⁴ Negative binomial regression was used instead of Poisson to account for overdispersion ($z = 9.08$; $p < .001$).

⁵ Incident Rate Ratio.

10. Within-household Selection

For the sampling design to qualify as fully probability-based, random selection must occur at every stage. In interviewer-administered surveys using address-based sampling, the selection of individuals within a household is completed by the interviewer. In self-administered contexts, such as the web component of the WNTS, this step is completed by household members and has been shown to result in high levels of non-compliance. For this reason, it is common for surveys to resort to non-random approaches at the last stage of selection, such as inviting up to two eligible household members to participate. This approach is currently used in surveys such as the British Social Attitudes Survey.

Given the trade-offs between implementation feasibility and self-selection biases, we tested two approaches: the “up to two adults” format versus the Rizzo-Brick-Park random selection method (see for the first question of this within-household selection method).

Figure 5. First question of the Rizzo-Brick-Park within household selection method.

To ensure the project represents the population and everyone has an equal chance to take part, we need to know who lives in your home and would also be eligible to take part.

Including yourself, how many people aged 16 or over regularly live in your home? Please include yourself in the number below.

I am the only person aged 16 or over

Two

Three

Four

More than four (please specify)

Back

Save and Continue

Our expectation was that the two-adult approach would result in a higher number of achieved responses, as it allows more than one person per household to participate in the survey. In principle, up to 7,500 individuals were targeted by the two-adult approach, and 3,750 individuals were targeted by the Rizzo-Brick-Park random selection method. The two-adult approach yielded 562 responses (7.5%), while the random selection method yielded 446 responses (11.9%). Although the difference between these two methods of selection was significant ($\chi^2(1) = 58.80, p < .001$), the response rate reported for the two-adult approach is likely underestimated. This is because the targeting estimate of 7,500 individuals assumed each household had at least 2 eligible individuals, which is unlikely to be true, when considering that 31.9% of all households comprised a single person in the Welsh 2021 Census.

In conclusion, the data suggests that the two-adult method produces a higher number of responses than the random selection method, consistent with our expectations (562 vs. 446). However, the volume of responses is not as high as expected based on Census data, suggesting that in some households with multiple eligible members, only one person might have taken part.

Table 13. Achieved response by selection method.

Selection method	Issued addresses	Estimated eligible respondents	Responses achieved
Rizzo-Brick-Park	3,750	3,750	446 (11.9%)
Two-adult method	7,500	7,500 (0% single person)	562 (7.5%)

10.1 Assessing compliance with the selection process

We can assess the effectiveness of selection process in two ways:

- 1) Comparing the sample achieved between experimental groups. If the selection process has been adhered to it should more accurately represent the target population, and, potentially, result in a lower response rate due to the additional requirement on some households.
- 2) Investigate responses received at each question in the selection process, and the accompanying paradata, to identify any signs that the process has been ignored or amended.

Sample profile

A key hypothesis was that the Rizzo-Brick-Park random selection approach, if properly followed by households, would result in a more representative sample profile. To test this hypothesis, we compared the productive sample profiles for each selection approach using the following categories: Sex, age, ethnic group, disability status, and ability to speak Welsh. For context, these profiles were compared with population estimates for the whole of Wales (aged 16+) taken from the 2021 Census.⁶

Nationally, just over half of the population in Wales aged 16+ are female (51.5%). Both selection approaches over-represented women (58% in the “any two adults” approach and 59% in the random selection approach). This suggests that the random selection may not have been fully complied with.

Table 14. Sex composition in the achieved samples and the population.

Sex (RspSex)	Population figures*	WNTS - Any two adults	WNTS - Random selection
Male	48.5%	42.0%	41.0%
Female	51.5%	58.0%	59.0%

*Source: 2021 Census, TS009 - Sex by single year of age, ONS Crown Copyright Reserved [from Nomis on 14 June 2024] (filtered for 16+)

⁶ Because Pilot 1 used an equal allocation approach to selection of addresses by region (i.e., ensuring an equal number of addresses were selected in each region regardless of the size, rather than a proportional approach) to facilitate our understanding of response rates by region, the true target population will be slightly different to the overall population of Wales. However, the national figures remain useful for context and as a benchmark.

There is also little difference in the age profiles of respondents by selection approach. As outlined in

Table 15, both approaches underrepresent younger people (nationally, 28% of the 16+ population are aged 18-34, compared to 23% in the “any two adults” sample and 20% in the random selection sample) and over-represent older people. This trend is common in all social surveys. However, if the random selection method was administered correctly, we would expect it to produce an age profile more in-line with the population targets.

Table 15. Age composition in the achieved samples and the population.

Age (Age and AgeBnd)	Population figures*	WNTS - Any two adults	WNTS - Random selection
16-24	12.9%	7.7%	6.5%
25-34	15.0%	14.9%	13.9%
35-44	14.1%	16.0%	10.1%
45-54	15.7%	13.5%	13.7%
55-64	16.5%	18.9%	24.9%
65-74	14.0%	17.1%	22.4%
75-84	8.6%	9.8%	7.2%
85+	3.2%	2.1%	1.3%
16-34	27.8%	22.6%	20.4%
35-64	46.3%	48.4%	48.7%
65+	25.9%	29.0%	30.9%

*Source: 2021 Census, TS007 - Age by single year, ONS Crown Copyright Reserved [from Nomis on 14 June 2024] (filtered for 16+).

A large majority of the population aged 16 and over in Wales identify as belonging to the White ethnic group category (95%). As expected, this is reflected in the samples for both selection methods. While the proportion of respondents identifying as white is marginally higher in the random selection group (97% compared to 95%), this difference is not statistically significant.

Table 16. Race/ethnicity composition in the achieved samples and the population.

Ethnic Group (RspEth)	Population figures*	WNTS - Any two adults	WNTS - Random selection
White	94.7%	95.3%	97.0%
Mixed	1.1%	0.7%	0.9%
Asian	2.7%	2.6%	1.8%
Black	0.8%	1.1%	0.0%
Other	0.8%	0.4%	0.2%

*Source: Census 2021: Ethnic group by age and sex (filtered for 16+).

It is vitally important for any survey, especially one collecting information on travel attitudes and behaviours, to accurately represent those with impairments. In Wales, 25% of the population aged 15+ are classified as having a disability (using the Census two-category definition as those who say they have any physical or mental health condition or illness lasting or expected to last for 12 months or more which reduces their ability to carry out day-to-day activities either a lot

or a little).⁷ This figure compares with 21% in the “any-two adult” approach sample and 24% in the random selection sample. While it would appear that the random selection approach better represents those with limiting impairments, the difference between the two groups is not statistically significant.

Table 17. Disability status in the achieved samples and the population.

Disability status (two category) (ImpDay =1 or 2 = disabled)	Population figures*	WNTS - Any two adults	WNTS - Random selection
Disabled	24.6%	20.6%	24.0%
Non-disabled	75.4%	79.4%	76.0%

*Source: Census 2021: Disability in England and Wales, 2021 (filtered for 15+ - disability not available for single year age).

It is also important than any survey of the Welsh population treats Welsh and English languages equally and properly represents Welsh speakers. Nationally, 15% of the population aged 16 and over speak Welsh. Both selection approaches used in the WNTS pilot slightly overrepresent Welsh speakers (18% in the “any two adults” sample and 19% in the random selection sample).

Table 18. Welsh fluency in the achieved samples and the population.

Welsh speaking (WishUse = 2)	Population figures*	WNTS - Any two adults	WNTS - Random selection
Can speak Welsh	15.3%	17.7%	19.0%
Cannot speak Welsh	84.7%	82.3%	81.0%

* Source: 2021 Census, TS076 - Welsh language skills (speaking) by single year of age, ONS Crown Copyright Reserved [from Nomis on 14 June 2024] (filtered for 16+)

Implications

The findings indicate that, when implemented in its current form, the Rizzo-Brick-Park random selection method yields a smaller productive sample size than the “any-two adult” approach. However, based on the observed sociodemographic characteristics outlined above, it does not produce a significantly more representative sample. This is likely the result of either non-compliance with the selection methodology, or the selection methodology causing some groups of respondents to drop out of the survey.

⁷ Please note that this figure is based on the population aged 15+. This is because disability by single year of age has not been published for the 2021 Census at the time of writing.

Compliance

During the selection procedure, the chance of being selected is inversely proportional to the total number of eligible respondents. As such, all single-person households are automatically selected, but those in three-person households would have a 33.3% chance of being the randomly selected person and bypassing the additional selection questions, for instance. Compliance among individuals is therefore more likely in smaller households, as the responding individual has a greater chance of being selected.

Table 19. Self-reported household size for selecting participants.

Household size	n	%
1	147	26.6%
2	310	56.1%
3	60	10.8%
4	28	5.1%
5	5	0.9%
6	3	0.5%
Total	553	100%

There were 553 individuals who engaged at least partially with the selection process. A total of 235, or more than two-in-five respondents (42.5%), who initially accessed the survey survey, were not the randomly selected person requested by the procedure. In other words, these were the people who were required to pass the survey to another household member following the selection process.

Table 20. Computed respondent number from total household size.

Person selected	n	%
1	318	57.5%
2	191	34.5%
3	35	6.3%
4	6	1.1%
5	2	0.4%
6	1	0.2%
Total	553	100%

Among these, 223 participants went through the entire selection process. Of these, 81% reported that they were able to recruit the other selected participant to take part. The remaining 19% were primarily situations where the selected person did not want to take part in the survey (11%), with 8% reporting being unable to handover the device or survey.

Table 21. Recruitment outcomes for selected household members who are different from the individual who started the survey.

Reported action	N	%
I have passed the device	181	81.2%
I am unable to pass the device	17	7.6%
The selected person does not want to take part	25	11.2%
Total	223	100%

It is unknown how many participants reported compliance without passing along the device. However, survey paradata can be used to investigate instances where response options have been amended. There are 52 respondents who interacted with the questions multiple times (often during the same session). For instance, two serials changed their responses a total of 11 times between the three options available ('successfully passed the device', 'selected person refused' or 'selected person unavailable'). There are a total of 29 participants who changed their response option from a response of unavailability or disinterest. This could suggest a level of disengagement with the random protocol.

Table 22. Type of adjustment to selection response.

Type of adjustment	N	%
Other unproductive	2	3.8%
Unproductive to productive	29	55.8%
Productive to unproductive	1	1.9%
No change	20	38.5%
Total	52	100%

It would appear, therefore, that a relatively small minority of those who were required by the selection process to pass the survey onto another household member amended their responses. However, this in itself is unlikely to explain why the sample profile is not more representative of the population than the 'up to two adults' approach. That would imply, therefore, that a proportion of respondents who said that they had passed the survey or device onto another household member had not done so.

A further hypothesis regarding the impact of the selection process is that it would, in itself, increase respondent burden and lead to drop-out. However, this does not appear to have been the case. There are 16 instances of respondents dropping out of the survey during the selection process.

Table 23. Respondent selection questions prompting disengagement.

Question	N
Kish	2
KishSel	1
RespCont	1
RespName	2
RespOut	2
RespRef	2
RespSel	1
RespSel2	2
RespSel3	2
RespSes2	1
Total	16

A final hypothesis is that the added burden and complexity provided by the selection questions would result in a lower response rate due to the possibility of excluding the most interested parties. However, when comparing the household-level response rate (that is, the proportion of households where at least one survey was completed), the response rate was slightly higher in the random selection group than in the ‘up to two adults’ group. In other words, while fewer people took part in the random selection group, more households participated. However, the relatively small sample size means that this difference is not statistically significant.

It is likely that any difference is caused by the letter design. The random approach only requires one single log-in on the letter and fewer instructions, which may lead to higher levels of engagement.

Table 24. Response outcomes.

	Respondent selection approach	
	Up to two	Random
Issued sample (addresses)	3,750	3,750
Issued sample (cases)	7,500	3,750
Assumed eligible addresses	3,375	3,375
Assumed eligible respondents	6,379	3,375
Productive cases	562	446
Productive households	412	446
Household level response rate	12.2%	13.2%
Productive cases per household	1.36	1.00
Individual level response rate	8.8%	13.2%

Indicators of Data Quality from the Questionnaire

In addition to examining the volume of journey data and the impact of the experiments tested on it, we analysed a series of indicators focusing on the diary component of the WNTS to evaluate the extent to which the proposed design met the study requirements.

10.2 Consent to follow-up studies

Achieving a high level of consent to follow-up studies was a key requirement set in the review stage of the process (Cornick, Aizpurua, & Howe, 2023). The rationale was to have a high number of respondents that could be invited for future research (e.g., administering a travel diary, focusing on particular topics or subpopulations).

At first instance, 72% of respondents provided consent, with an additional 8% of respondents indicating they were not sure and required more information. A follow up question was presented to respondents who indicated uncertainty, offering additional details about what participation in future entails. This question converted just under half of the uncertain respondents into agreeing to be recontacted (45%). In total, 75% of respondents consented to be recontacted.

Since the consent question was placed after the incentive block, an additional 2% of respondents who had completed the survey (i.e., they had reached the incentive block) did not answer this question. If increasing consent is an objective for the main stage, reordering the questions to place this item before the incentive block should be considered, regardless of diary placement. This recommendation aligns with recent research indicating that asking for consent to data linkage earlier in the questionnaire increases consent rates (Jäckle et al., 2023).

Because diary placement can impact task burden, which can reduce respondent cooperation, we explored whether consent rates were affected by diary placement. We found no evidence of this, with comparable levels of consent achieved when the diary preceded the questionnaire and when it followed it ($\chi^2(2) = 0.57, p = 0.751$).

Table 25. Consent for further research, overall and by placement condition.

Question	Overall % (n)	Early placement % (n)	Late placement % (n)
Initial consent			
Yes	71.7% (723)	70.3% (319)	72.9% (404)
No	17.7% (178)	17.8% (81)	17.5% (97)
Not sure – Need further information	8.4% (85)	9.0% (41)	7.9% (44)
Not answered	2.2% (22)	2.9% (13)	1.6% (9)
Follow-up consent question			
Yes	84.4% (38)	51.2% (21)	38.6% (17)
No	31.8% (27)	24.4% (10)	38.6% (17)
Not answered	23.5% (20)	24.4% (10)	22.7% (10)

10.3 Questionnaire length

The time it takes to complete a survey is commonly regarded as an indicator of task difficulty and effort required. While completion times for the WNTS were expected to vary widely depending on respondents' travel behaviours, it was anticipated that, on average, it would take around 40 minutes to complete the survey.

Most people completed the survey in the same day (97%), but there were examples of respondents who took longer to do so (ranging from 1 to 22 days). This dispersion made the mean a less reliable measure of central tendency, and instead the median was used (a measure that is less affected by outliers). Across conditions, the median time taken to complete the survey was 23:49.

While the difference in completion time between versions (early - late) and placement conditions is not significant ($W = 121329$, $p = .225$ and $W = 132169$, $p = .163$), Table 26 reveals an interaction between the two experiments. While the median time and interquartile range are very similar for the roster and interleaved versions in the late placement condition, the median time is over 6 minutes longer in the roster version than in the interleaved version when the diary preceded the questionnaires. In this version, the dispersion of completion times, as measured by the interquartile range, was also larger than in the other conditions. This might be attributed to respondents in this condition entering, on average, a larger number of journeys than in any other condition (3.5 journeys versus 2.4-2.6, for further details, see Table 12).

Table 26. Completion time (minutes: seconds) by experimental condition.

Version	Placement	Median	IQR	n
Roster	Early	27.8 (27:45)	22.4 (22:24)	219
Interleaved	Early	21.2 (21:13)	18.8 (18:50)	235
Roster	Late	23.2 (23:13)	18.3 (18:21)	273
Interleaved	Late	23.8 (23:49)	18.5 (18:33)	281

There were several extreme cases where the completion time for the full WNTS was unusually short. For example, three individuals, all assigned to the 'up to two adults' and interleaved design, submitted their survey in less than five minutes (with completion times of 4 minutes and 30 seconds, 3 minutes and 48 seconds, and 2 minutes and 11 seconds). These times are extremely fast given the number of questions included in the questionnaire and the request to complete two days' worth of travel data. Since speeding is considered an indicator of poor data quality and low engagement with survey questions, processes should be developed to identify and exclude instances of unusually low completion times.

Drop-offs

The proportion of people who finish the survey once started is a key indicator of data quality in survey research. While early abandonment is relatively frequent in self-administered surveys, it has important implications for available sample size, non-response error, fieldwork efforts, and processing and analysis (e.g., imputation of missing data). For this reason, examining drop-off rates and identifying sections where drop-offs are more likely to occur are helpful to understanding and improving respondents' experience.

In

Table 27, we display the number of valid responses for the first question in each block asked to everyone, as well as the proportion of missing answers compared to the first question in the block immediately before.

The largest drop in substantive answers occurred between the "Consent" and "Contact" (-23.2%) modules at the end of the questionnaire. This was followed by moderate drops in substantive responses between the "Travel tickets/ Discounted travel" and "Travel frequency" modules (-8.4%) and between "CASI start" and "Impairment" modules (-3.1%).

Table 27. Valid responses to the first question in each block.

Block	Valid responses	Dropout %
Noise	1249	NA
Household	1245	-0.3
Welsh	1209	-2.9
Travel tickets/ Discounted travel	1199	-0.8
Travel Frequency	1098	-8.4
Community Transport	1095	-0.3
Taxi	1097	0.2
Car	1096	-0.1
Cycle	1064	-2.9
Ferry	1100	3.4
Employment	1072	-2.5
CASI Start	1072	0.0
Impairment	1039	-3.1
Accessibility	1038	-0.1
Equality measures	1035	-0.3
Survey assessment	1007	-2.7
Incentives	1010	0.3
Consent	991	-1.9
Contact	761	-23.2

11. Recommendations

11.1 Volume of Data and Approach to Follow-Up Questions

The achieved number of journeys in Pilot 1 was substantially lower than expected (-26.5%), and the data suggests underreporting of trips on day 2 of the diary. Following this finding, there are two potential options:

- 1) Shortening the diary recall period**, from 2 days to 1, whilst keeping a similar level of detail in the follow-up questions. If a single day is used, we suggest using the roster version, which maximised the number of journeys on day 1.
- 2) Maintaining the recall period (2 days)** whilst substantially reducing the number of follow-up questions in the travel diary. While the number of journeys across both days was higher in the roster version (2.9 journeys on average vs. 2.6), the difference with the interleaved version did not reach statistical significance due to the drop in reported journeys on day 2 (1.9 vs. 1.0), which was less pronounced in the interleaved version (1.5 vs. 1.1). Therefore, the roster design might be preferred to maximise journeys, although the interleaved version allows for easier correction of wrongly entered journeys (by clicking back). Because wrongly entered journeys increase measurement error and reduce the amount of available data (they often result in NA or uncodable responses), the difference in usable journeys between interleaved and roster might further decrease when accounting for this.

Shortening the diary recall period from two days to one day would require a significantly larger sample size to meet the WNTS's precision requirements. This, in turn, significantly increases the costs. Our cost modelling shows that such an increase in price is not cost-effective. Therefore, we recommend retaining the recall period at two days and reducing the level of detail collected.

Regarding the diary design, since the difference in journeys across both days is not statistically significant and amending journeys is substantially more laborious in the roster version without editable summary screens, we recommend proceeding with the interleaved approach. This approach is expected to perform better than the roster version in the interviewer-administered context, as it is more conversational, gathering all details about a journey before moving on to the next, rather than jumping back and forth between journeys for additional details.

11.2 Diary Placement

Overall, there is evidence that early diary placement results in a higher volume of journey data (on average, 3.0 journeys across both days vs. 2.5). However, there is also evidence that drop-offs increase when respondents are presented with the diary first, resulting in a smaller sample of people with a greater number of trips.

In addition, the moderation analysis suggests that the effect of diary placement depends on the diary design. With an interleaved design, the placement of the diary resulted in comparable number of journeys. For this reason, if the interleaved version is used, we recommend placing the diary after the questionnaire, since its early placement does not maximise journey volumes and can increase dropouts.

11.3 Within-Household Selection

The decision to employ either a random selection of one individual per household or allow up to two adults per household in the web element of the WNTS is not straightforward. Each approach has advantages and disadvantages.

Our experimentation in Pilot 1 showed that employing an ‘up to two adults’ approach would generate a larger sample size in the web element, which would result in lower survey costs when aiming to reach a target number of complete surveys. However, this method is potentially more complex, as it introduces intra-household clustering. The recommended WNTS design involves random selection of an individual in the face-to-face element, and ideally, the two modes would aim to recruit the same number of individuals per household.

There is also some evidence to suggest that the ‘up to two adults’ approach might lead to more individuals but fewer households participating in the web element of the WNTS, although larger-scale testing would be required to confirm this. Finally, to our knowledge, no survey using an ‘up to two adults’ or ‘all adult’ approach has been granted Accredited Official Statistics status (previously National Statistics status). The Office for Statistics Regulation, the regulatory arm of the UK Statistics Authority, is the relevant accreditation body for Accredited Official Statistics.

The random selection approach used in Pilot 1 would generate a lower sample of individuals than the ‘up to two adults’ approach from a set sample. It would, therefore, be more expensive to run. The Pilot 1 analysis also showed that the random selection approach does not generate a more representative sample than the ‘up to two adults’ approach. This suggests that there is some level of non-compliance with the selection criteria. However, there is also evidence that the household-level response rate is higher, which may be caused by a simpler letter. Further, there is no evidence that the selection approach itself drives respondents to drop out of the survey in substantial numbers.

The use of the random selection approach would create greater consistency with the face-to-face element and theoretical compliance with the principles of random selection, which may be required to acquire Accredited Official Statistics status. With the random selection approach, there is lower risk of falsification (that is, creating fictitious additional household members to trigger an incentive payment), although it should be noted that evidence of falsification in the ‘all adult’ approach is limited and small scale. Finally, it would be possible to further refine the selection process over time with the aim of improving compliance.

Table 28. Advantages and disadvantages for selection type.

	Advantages	Disadvantages
Up to two eligible respondents	<ul style="list-style-type: none"> - Increased amount of data generated per productive household (journeys and questionnaire items) - Smaller sample required to achieve required effective samples size - Lower surveys costs due to a smaller sample size - Extensive examples of the method working effectively on other surveys - Lower carbon footprint 	<ul style="list-style-type: none"> - Potentially fewer overall households completing the web element - Inconsistency with the face-to-face element - Clustering within household could require more complex analytical approach - The clustering at the household level can lead to a loss of statistical efficiency

	Advantages	Disadvantages
		<ul style="list-style-type: none"> - Would require use of quasi-selection weights - Unknown whether a survey using a non-random selection technique would be granted Accredited Official Statistics status - Greater opportunity for falsification due to conditional incentives
Random respondent	<ul style="list-style-type: none"> - Some initial evidence that the simplified approach could yield higher response rates and provide more productive households - Consistency from respondent and data users' perspectives across survey modes - Likely to support accreditation of Accredited Official Statistics status - No additional risk of falsifying a completed interview for incentive 	<ul style="list-style-type: none"> - Sample profile achieved is no more representative of the inferential population than the 'up to two adults' approach - Evidence demonstrates a lack of compliance with the selection process - further work is needed to increase compliance - Lower quantity of trip and questionnaire gather per household - Higher administration costs as larger sample required - Higher carbon footprint

The decision of which selection approach to use is, therefore, a finely balanced one. If cost-effectiveness is the key criteria, we recommend using the 'up to two adults' approach: it is slightly lower cost to operate, and the random selection approach does not provide clear quality benefits through a more representative sample. However, if cost is not the key driver, we recommend the random selection approach: there is a simplicity and consistency to using the same approach across modes, it does not appear to drive dropouts and adheres more closely to the theory of random probability sampling. We believe that more work could be done over time to understand and lower non-compliance with the selection approach, meaning that there could be longer-term quality benefits.

11.4 Summary Screens

In 168 instances, respondents indicated that the locations or journeys they had entered were incorrect, with an additional 40 cases where the modes of transport were incorrect. While this is a relatively small number of cases compared to the total number of journeys (2,743), it indicates the benefits of offering options for respondents to rectify their answers. This is further supported by responses to the semi- and open-ended questions, where respondents indicated that they had entered journeys accidentally and were not able to correct them, forcing them, in some cases, to complete entire loops of diary questions about non-existent or erroneous journeys.

Taking into account the potential risks of disengagement caused by this and the increased measurement error introduced by these erroneous journeys, it is important to evaluate the feasibility of offering respondents ways to amend their entries and ensure that the questions they receive are relevant. This ideally includes testing the performance of editable summary screens where respondents can confirm the accuracy of the information or amend any details (e.g., incorrect start or end locations) in a quantitative setting. This is particularly important if a roster design is adopted for mainstage, since amending locations often entails going back

multiple screens. With an interleaved design, this is less problematic, given that follow-up questions are presented immediately after the filter and could more easily be amended.

While summary screens are desirable features to correct errors, they are not needed for most journeys where the information is accurate. Their inclusion, therefore, can have unintended consequences resulting from the increased burden of reviewing, editing, and/or confirming location and transport mode information (e.g., drop-offs). For this reason, we recommend developing and testing the performance of editable summary screens and other approaches to amend journey details in the future. This testing should include qualitative insights to guide the design but will require quantitative data to evaluate the risk of respondents dropping off.

11.5 Map Usability

In most cases, respondents were able to find the locations they travelled to using the search functionality of the map or dropping a pin. However, there were 278 instances where they indicated not being able to find the desired location.

In the majority of these cases, they used the free-text option to indicate their origin or destination. The usability of the entered locations varied widely, ranging from places that could be pinpointed manually during processing stages to uncodable answers. The most common category of responses consisted of town, city, or area names where the exact location of the respondent could not be identified, but an approximate location could be assigned (e.g., city centre). To make the most of the data collected, we recommend using existing information whenever possible, even if it reduces the level of location precision. This requires developing processes for coding unspecific locations entered by respondents (e.g., assigning them a specified location mid-point) to facilitate the inclusion of other journey information, such as transport mode.

Since some respondents reported feeling uncomfortable providing full addresses for privacy reasons, we suggest experimenting with privacy and confidentiality reassurances in the diary. We also recommend inputting the house address to avoid ambiguous instances where it was unclear whether the respondent referred to their own home address or someone else's.

11.6 Summary of Recommendations

Table 29 provides a consolidated overview of the recommended approaches, based on findings from Pilot 1 analysis. These recommendations are considered optimal for the WNTS, balancing data quality, cost-effectiveness, and respondent engagement. However, all design options evaluated in this report are deemed viable and capable of achieving the required data standards for the WNTS.

Table 29. Summary of Recommendations from Pilot 1.

Dimension	Recommendation
Recall period	Retain a two-day recall period and reduce the level of detail collected
Diary design	Interleaved
Diary placement	Late (after the questionnaire)

Dimension	Recommendation
Within-household selection approach	“Up to two adults” to promote cost-effectiveness; random selection to maximise consistency across modes and adhere more closely to the theories of probability-based sampling
Summary screens	Develop and test editable summary screens to amend journey details
Map usability	Implement approximate coding for vague locations, test privacy assurances, and automatically input known locations (i.e., home)

12. Analytical Approach

Given the complexity of the data, characterised by the high volume of variables resulting from the instrument structure (with 20/21 journey loops per diary day) and experiments embedded in Pilot 1 (three fully crossed split-ballot experiments), all analyses presented in this report, except for the sample profile and compliance, were conducted independently by two researchers. They performed the analyses separately, each developing their own code, and then compared the results to ensure consistency. The outputs were fully consistent in all cases except for one, where there was a discrepancy: One analyst recorded that a respondent had reported 6 journeys while the other recorded 7 journeys. This discrepancy did not alter the substantive analysis and conclusions drafted in the report.

The items used in the analysis – except those listed in the sample composition section, which already include item names - are listed in Table 30.

Table 30. Items used in the report by metric.

Metric	Source	Items
Journey count	Diary	<p>D7 How did you travel from [your home/address] to [address]? If you used more than one mode of transport, please select the one you used to travel the longest distance.</p> <ol style="list-style-type: none"> 1) Bus or coach 2) Cycle 3) Car or other motor vehicle (e.g., light van, campervan, motorcycle, moped) 4) Community transport (e.g., Community Buses, Volunteer Car Schemes) 5) Taxi 6) Train 7) Walking or wheeling 8) Other
Trip summary screen	Diary	<p>QTripSumCheck Please review the information below about your journeys from yesterday. You will not be asked to correct any entries, only to specify any errors. If more than one issue applies to a single journey, you can select all relevant options.</p> <ol style="list-style-type: none"> 1) Everything is correct 2) I need to modify a location 3) I need to modify a journey/ part of a journey 4) This journey did not take place
Mode summary screen	Diary	<p>QmodeSumCheck Please review the information below about the modes of transport you used yesterday. You will not be asked to correct any entries, only to specify any errors. If more than one issue applies to a single journey, you can select all relevant options.</p> <ol style="list-style-type: none"> 1) This information is correct 2) The main mode of transport for this journey is incorrect 3) Any other mode of transport for this journey is incorrect
Map issues	Diary	<p>QDRosDay2_OE QEDIL_2_ENDD2 QLDRosDay2_OE QLDIL_2_ENDD2 Thank you for recording your journeys. We are interested in understanding your experience of entering your travel information and using the maps. Did you experience any difficulties while recording your journeys, or do you have suggestions for improvements? Feedback from this question will inform the development of this part of the survey. [Open-ended]</p>

Metric	Source	Items
Journey purpose	Diary	<p>D5 What was the purpose of this journey to [address]? If your journey had more than one purpose, please select all that apply.</p> <ol style="list-style-type: none"> 1) Exercise or play sports 2) Go to work 3) Holiday or day trip 4) Medical consultation or treatment 5) Personal appointment (e.g., bank, hairdresser, launderette) 6) Personal journey during work 7) Shopping, even if there was no intention to buy 8) Education (e.g., go to school, college) 9) Social or entertainment (e.g., meet friends or family, voluntary work) 10) Informal caring responsibilities (e.g., care for an elderly relative) 11) Accompany someone (e.g., drop someone off, pick someone up) 12) Return home 13) Other
Perceived burden and difficulty	Questionnaire	<p>Sur01 How difficult was it for you to understand and answer the questions in this survey?</p> <ol style="list-style-type: none"> 1) Not at all difficult 2) Slightly difficult 3) Moderately difficult 4) Very difficult 5) Extremely difficult
	Questionnaire	<p>Sur02 Approximately how many minutes did it take you to answer all the questions in this survey? Please enter an amount in the box below using numbers. [Numeric entry] Don't know</p>
	Questionnaire	<p>Sur03 Was today's survey longer than you expected, shorter than you expected, or about as long as you expected?</p> <ol style="list-style-type: none"> 1) Much longer 2) Somewhat longer 3) About as long as I expected

Metric	Source	Items
ConsFutr	Questionnaire	4) Somewhat shorter 5) Much shorter Thank you again for contributing to this important piece of research. Would it be okay for us to contact you again in the future for similar projects such? If you agree, you may be contacted in the future by the Transport for Wales or another research organisation to take part. When you are contacted, you can decide if you still want to participate at that time. 1) Yes – I would like to be given the option to take part in the future 2) No – I would not like to be invited to take part 3) I'm not sure – I would like more information
ConsInfo	Questionnaire	Please click on the links below for some more information about the study

[HELPTXT: What will taking part involve?

You will be invited to take part in the future for a survey similar to the one you have just completed or this may involve other research methods such as focus groups. You can choose not to take part in the future when you are contacted again and it will always be made clear what taking part involves.]

[HELPTXT: Will I receive a voucher for taking part?

Future invitations will explain what you will receive for taking part.]

[HELPTXT: What will you do with the information I give?

The information you give will be used for research purposes only and held securely in accordance with current data protection legislation. We may analyse your answers together with answers you gave previously to help us understand how people's views are changing over time.]

- 1) Yes – I would like to be given the option to take part in the future
- 2) No – I would not like to be invited to take part

