National Centre for Social Research

WNTS Stage 1b: Qualitative and Quantitative testing

Prepared for: Transport for Wales

NOVEMBER 2023

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
WNTS Stage 1b: Qualitative and Quantitative testing

Contents

Background
Development of the WNTS

1.1 Work completed to date

In April 2023 Transport for Wales commissioned the National Centre for Social Research (NatCen) to work as its delivery partner to design a suitable methodology for the new Wales National Travel Survey (WNTS). The WNTS will collect information on travel attitudes and behaviour in Wales and is expected to launch by December 2024.

The design of the WNTS is being guided by the National Centre for Social Research's REMoDEL approach – a systematic framework for designing or transforming social surveys and gathering robust evidence around the trade-offs associated with different design options (Cornick, 2021).

Figure 1: NatCen's REMoDEL Approach

Key Stages of NatCen's REMoDEL Approach:

Review the research and information needs, including any design parameters

Evaluate the feasibility of various methodological designs considering different quality dimensions

Model a prototype design for further development

Design and develop the new methodological approach and questionnaire

Experiment by testing the design in a quantitative context

Launch the new survey

To date, the Review, Evaluate and Model stages have been completed. Following the Review and Evaluate stages a mixed-mode, web-first survey (consisting of web, telephone and face-to-face) was identified as the optimal

National Centre for Social Research
WNTS Stage 1b: Qualitative and Quantitative testing

design for data collection, with the travel diary integrated into the survey software (for further details, refer to Cornick, Aizpurua, & Howe, 2023). Designing and integrating a more complex travel behaviour data collection tool, based around passive data collection, machine learning and prompted recall, was also identified as a longer-term objective.

During the Model stage NatCen outlined the options and recommendations for the operation of the data collection modes, the sampling design and fieldwork procedures (including communication and incentive strategies). These recommendations were drawn from a combination of survey literature, insights from similar projects and expert judgement.

Alongside these recommendations, the Model stage identified a series of evidence gaps and associated areas for quantitative experimentation. These are outlined in Table 1.

Table 1: Areas for experimentation identified during the Model stage

Area	Suggestions for experimentation
Data collection mode	a. Asses the efficacy of the telephone mode, with a particular focus on the operation of the diary in this context b. Assess the take-up of the telephone option to inform costings for Phase 2 c. Knock-to-nudge approach where interviewers offer all three modes concurrently, or promote the web d. Determine the optimal number of interviewer visits to addresses e. Passive data collection development for future collection of travel data f. Placement of the travel diary at the start or end of the survey, and the performance of complex data collection elements such as maps, location look-up and editable summary screens
Target population	a. Evaluate the feasibility of developing auxiliary sampling frames to include institutionalised populations b. Compare the performance of the 2-adult strategy and the 2-step Rizzo method for selecting adults within households in the web survey
Contact strategy	a. Test the performance of different envelope sizes and colours b. Evaluate the impact of including logos on the letters and/or envelopes c. Compare the effectiveness of a final reminder letter vs. a final reminder postcard d. Evaluate the impact of variations in how household members are addressed in the invitations and reminders e. Test the effectiveness of including a pre-notification letter before the survey invite is sent out f. Evaluate optimal approaches for letter dispatch to ensure uniform representation of travel behaviour across days of the week and months of the year
Incentives	a. Evaluate the need for future experimentation with the value of the conditional incentive

b. Evaluate whether there is drop-off in trip reporting on Day 2 of the diary and, if so, whether there is value in incentivising the diary separately to the survey

Source: Cornick, Aizpurua, & Howe (2023).

1.2 Design and Experimentation (Stage 1b)

Development of the WNTS is now moving into the Design and Experiment phases (Stage 1b).

During the Design phase the questionnaire and travel diary data collection tools will be designed, tested and programmed. This requires a series of qualitative-testing exercises encompassing both cognitive and 'cogability' testing.[1] We outline our proposed approach to this qualitative testing in the following section of this proposal.

Once the qualitative testing is complete the survey will progress to quantitative testing (the 'Experiment' phase). Our proposed approach to the quantitative testing is outlined later in this document.

---

[1] 'Cogability' testing combines cognitive and usability testing.

Qualitative testing

NatCen's recommended approach

1.3 Summary of our recommended approach

This section outlines our proposed plans for the qualitative testing of the questionnaire and travel diary data collection tools. Our proposed plan combines desk reviews with cogability testing of questions in both English and Welsh.

Cogability interviews combine elements of cognitive interviewing and user-testing. Cognitive interviewing methods are derived from cognitive psychology and allow researchers to examine the mental processes people go through when completing surveys. Cognitive interviewing techniques focus on four processes: whether participants comprehend the information provided, how they recall information necessary to assist with the task, the judgments (or shortcuts) they make as to what information to provide and how they respond to the request. User-testing interviews are used to establish whether a product meets end user requirements. For this project we want to establish whether the questionnaire templates developed for the survey are user-friendly in terms of their design interface, and that they work well on both small and large screen devices (i.e. smartphones and desktop computers).

At NatCen, we routinely conduct cogability interviews to ensure that respondents understand the questions and can provide the information requested with minimal barriers. If issues are identified, we are then able to make improvements and suggest solutions based on the feedback received. By doing this we can ensure that survey questions are fit for purpose from a user's perspective and can be used to collect valid survey data.

1.4 Initial survey and diary assesment and re-design

During earlier stages of the design of the WNTS NatCen undertook an intital assesment of the draft WNTS questionniare. This included two reviews: i) a general review of best practice design for creating questions; and ii) identification of measurement issues due to mixed modes of data collection.

In addition to item-level assessment, macro level issues and design conventions were also considered. These included pagination, branching questions, and universal or optimal mode design. Based on this initial assessment the following conventions were agreed between NatCen and TfW:

- Inclusion of a refusable CASI

- Addition of showcards
- Universal questionniare design
- Incorporation of non-substantive answer options (Refusal and Dont know)
- Questionnaire to be reordered to maximise engagement
- Implement branching questions were appropriate to limit burden
- Batteries of scales to be permitted
- Pagination to be optimised to enhance user experience

NatCen also reviewed the proposed diary element of the questionnaire. The diary itself is a key component for WNTS and will ultimately provide information to meet a number of TfW's aims and objectives. Following this review it was agreed with TfW the diary would take a location-based approach to data collection, would be based on 48 hour recall and would be developed within the Blaise 5 software to enable a seamless transiton from the survey to the diary. The inclusion of complex data collection elements, such as location look-up, interactive maps and editiable summary screens were also highlighted as areas for experimentation.

Based on this, NatCen will first re-design the existing quesitonniare and diary in-line with these conventions.

As part of this process, where required NatCen will meet with colleagues from TfW to understand the underlying research questions for any items where the initial intent or concept was assessed to be unclear.

Following this redesign a further independent review will be undertaken by our Questionaire Design Testing (QDT) hub before embarking on cognitive and user testing. This extra review is to ensure we effectively prioritise which questions should be subject to pre-testing. We will present our recommendation on test aims and priorities to TfW before embarking on user testing.

At the end of the proposed work packages the final output will be a fully-programmed quesitonnaire and diary in both Welsh and English that has undergone usability and cognitve testing revisions. This final programmed version will then be suitable for piloting and any quantiative testing.

1.5 Overview of work packages

To develop the Welsh National Travel Survey, we are proposing a development process consisting of four work packages. These are as follows:

1. Implementing agreed question recommendations following our initial questionniare assessment before a further independent **desk review** by the Questionnaire Development and Testing (QDT) hub of the adjusted questions to identify which questions we believe would benefit most from undergoing cognitive testing or user-testing.

2. A **first round of cognitive interviews in English** to test the question wording identified as the highest priority through the desk review. If the review highlights questions where visual rendering may be an issue there will be an option to test mock-ups for non-standard screen designs.

3. A **workshop to identify questions in Welsh which would benefit from additional Welsh language testing** (e.g. items where translation was problematic or where text length/ optimal screen design may be impacted as a result of the translation).

4. A **second round of cogability interviews in both English and Welsh** will allow us to re-test any substantial wording changes newly programmed questions, and lower priority questions which we were

unable to include in the first round of interviews. Cogability interviews combine cognitive interviewing techniques with user-testing. These interviews would involve testing programmed web instruments. This allows for more potential to explore user experience issues with screen designs and layouts. Testing in Welsh will allow us to examine whether any questions may be problematic for Welsh speakers.

We will now describe each of these work packages in further detail.

1.6 Work package 1: Desk review

Work package 1 will involve a desk review of the adjusted questionnaire undertaken by experienced members of NatCen's Questionnaire Development and Testing (QDT) Hub. Approximately 120 questions from the survey and diary will be reviewed using a tailored version of the Questionnaire Appraisal System (QAS-04). The QAS-04 is a detailed checklist used for systematically reviewing survey questions. The desk review process will involve questions being reviewed individually to assess question wording, response options, question order and any instructions. To minimise the risks of coder bias and ensure all issues are identified, the review would be undertaken independently by two researchers, before meeting to discuss their findings.

Output: Following the desk review a Word document will be provided. This document will outline the findings from the desk review and will provide recommendations on which questions we feel would benefit most from undergoing further testing.

1.7 Work package 2: Cognitive interviews in English

To ensure the survey is comprehensively tested and that questions can be easily understood, interpreted and completed by participants we would recommend conducting two rounds of test interviews. Testing iteratively is considered best practice within questionnaire design, as it allows for questions to be refined and then re-tested where necessary (for example in the event of major changes to wording).

In reflection of the findings from the desk review, and in consultation with TfW and Welsh Government, 20-30 questions will be selected to undergo cognitive testing within the first round of interviews. In the first round of cognitive interviews, questions will be interviewer administered using question 'mock-ups' rather than programmed into a web instrument. The focus of the first round of testing will be to refine source (English) question wording prior to instrument translation and programming. Cognitive interviews will look at question comprehension, recall, editing and response options. We will conduct interviews with up to **16 participants** for this round.

We will use a recruitment agency to recruit participants for both rounds of interviews. When recruiting participants quotas will be set to ensure a diverse range of people are interviewed to adequately test the questions asked. These quotas will be agreed with TfW and Welsh Government once we have agreed test priorities. At this point we propose the use of **soft quotas** around age, sex, highest educational qualification and modes of transport used other criteria such as ethnicity can be explored. These factors can affect participants' understanding and ability to answer the questions, so testing the questions with a diverse mix of people can expose findings which might not have been picked up otherwise. Final quotas will be set once the priority questions for cognitive testing are agreed.

During interviews we will use a mixture of **observations, think aloud and probing techniques.** This mix of techniques is used, starting with participant-led methods, to minimize the risks of the interviewers' probes biasing the participants' answers. Interviewers are trained to observe respondents' behaviours and reactions to survey

questions, for example hesitations or frustrations, and will record these during the interviews. Interviewers will also ask participants to verbalise their thought process whilst answering each question, a process called 'thinking aloud' and collect further information using **probing techniques** after each set of questions to understand how the participants go about answering the questions. TfW and Welsh Government representatives will be invited to provide comments on the protocols before testing begins for all rounds. Questions within the protocols can be ordered in terms of priority, which will allow a greater volume of material to be tested, such that lower-priority items would be tested time-permitting.

All interviews will be conducted remotely via Zoom. Each cognitive interview would last approximately one hour, with participants receiving a £30 voucher to thank them for taking part. Interviews will be recorded with the participants' consent. A summary of each interview will be written up into a structured Excel pro-forma as soon as possible after the interview is completed. A summary of participants' characteristics (such as age, gender, educational level and answers to the other screening questions) will be included in the first column, thus allowing for anonymisation of the raw data. A summary of each area of investigation will be recorded in subsequent columns. This approach allows for a systematic analysis, whereby data can be read horizontally as a complete case record for an individual, or vertically by looking at a particular area under investigation across all cases.

After the first round of cognitive interviews, we will have a debriefing with all interviewers involved in the project which TfW and Welsh Government will be invited to attend. At this debriefing, interviewers will provide a summary of the main issues that have arisen with the questions tested

> **Output**: As a result of this meeting, we will have the opportunity to make suggested revisions to the questions tested and discuss which questions should be tested again in the second round. A **short interim report** of findings and recommendations will be provided before the second round of interviews commences. TfW and Welsh Government will be able to comment on the proposed changes.

1.8 Work package 3: Workshop on priorities for Welsh Language testing

Prior to further interviews, we will need to translate the Welsh National Travel Survey into the Welsh language (following sign-off from TfW and Welsh Government), and programme a Welsh protype version of the web instrument. Following the translation of the entire questionnaire into Welsh, we will hold a workshop with the research team and the Welsh translators. The aim of this workshop is to allow us to prioritise which questions should be the subject of cogability testing in the Welsh language. For example, we will review whether there were any questions that were more problematic to translate or where multiple wording options have been considered. We will also check whether translation of the questionnaire into Welsh has had any noticeable impact on questionnaire rendering or usability (for example by making questions longer or screens more cluttered).

> **Output:** Post this workshop we will circulate a **short report** listing priority questions and/or screens to test in the Welsh language interviews and any extra aims for the next round of testing.

1.9 Work package 4: Cogability interviews in English and Welsh
In the second round of testing we will conduct 'cogability interviews' in both English and Welsh. Cogability interviews combine techniques used in cognitive interviews (to test question wording) with techniques used in user-testing (to test appropriateness of the web instrument in terms of user needs including interface design and navigation).

The aims of the second round of testing will be:

- Re-test a selection of questions from Round 1 to ensure that the revisions made to questions following testing are working.
- Test a programmed version of the questions (survey and diary), especially any items that rely on a non-standard screen design.
- Test Welsh question translations for selected items identified in work package 3
- [Space permitting) test any lower priority questions which could not be tested in the first round.

Interviews for round two will be conducted in English and Welsh. For this round we will programme questions within a web survey template, as this will allow us to explore some usability issues, such as any non-standard screen templates e.g. looped questions, error messages and look-up functions. As with round 1 we will set soft quotas around age, sex, highest educational qualification and modes of transport used, other items such as ethnicity can be considered. In addition, we will set quotas to ensure we get participants trialling the questions on different types of device (both smartphones and desktop devices) and we will set quotas on self-rated digital literacy. We plan to conduct up to **24 interviews for this work package** (12 in English and 12 in Welsh).

All interviews will be conducted remotely via Zoom. We will use screen sharing so we can make detailed observations of how people interact with the web instrument on their own device. Each cogability interview will last approximately one hour, with participants receiving a £30 voucher to thank them for taking part.

Welsh language testing
Selected Welsh language questions would be programmed into a web survey template. The aim of this round of testing will be to check how well the Welsh translations retain the meaning of the equivalent English questions and how well the Welsh language is visually presented on a screen. The testing for this round would have as much parity with the English testing round as possible (i.e. so we can ensure that questions have the same meaning in English and Welsh). However, we may also identify new items to test as a result of the translation priorities workshop.

Welsh language cogability interviews will be conducted by our preferred partner organisation. NatCen will develop the cogability interview protocols that will be used by our preferred partner and we will provide them with full training on how to use these. As part of this training our QDT experts will observe partner interviewers complete a test cogability interview (in English) and provide feedback on this. Our preferred partner will then complete 12 cogability interviews (in Welsh) to test Welsh Language versions of the web instruments. We will set quotas to ensure we get Welsh speaking participants trialling taking part on different types of device (both smartphones and desktop devices) and who have different levels of self-rated digital literacy.

All interviews will be conducted remotely via Zoom. Screen-share recordings will be shared between NatCen and our preferred partner.

1.10 Joint Analysis Meeting
After all Welsh cogability interviews have taken place we will conduct a Joint Analysis Meeting (JAM) with our preferred partner. The aim of this meeting is to identify what issues have arisen for Welsh speakers and to determine whether these are different to the issues that arose in the English language testing.

NatCen has extensive experience of leading Joint Analysis Meetings after cognitive interviews have been conducted in different languages. For example, we have co-ordinated similar activities for the last four waves of

the European Social Survey (ESS) after working with partners to conduct interviews in 4 to 5 different European languages.

During the JAM we will discuss any issues arising in the Welsh language cogability interview. For each issue that arises we will determine whether:

- The issue detected also applied to the 'source question' (i.e. if the same issue applies in both English and Welsh)
- The issue detected was due to simple translation error; or
- The issue detected was due to source material being problematic to translate (i.e., if it contained multiple clauses)

By classifying the cause of the issues, we can determine whether changes only need to be applied to translations or whether changes should also be made to source English questions (i.e., make translation easier). Where appropriate we will ask our Welsh language speakers to provide updated recommendations for translating as part of this meeting.

> **Output:** Post the Joint Analysis Meeting we will circulate a **short report** summarising the main issues detected in round 2 and recommendations for final question changes. This report will include recommendations on question wording changes, screen design and translation edits. In addition, we will produce a final synthesis report. This report will summarise the methods used in all work packages and will document the testing process undertaken as a whole.

The final output of the qualitative testing phase will be a full tested and programmed questionnaire, ready for quantitative testing.

1.11 Additional round of diary usability testing

The proposed approach to the design of the diary element of the survey involves complex data collection tools, such as interactive maps, autocomplete open fields and editable summary screens. As such, it would benefit from a third round of user-testing.

The proposed work packages allow for the minimal amount of testing needed to explore the unique features of the diary:

- The usability of the proposed design for respondents with various backgrounds and travel behaviours.

- The ability of participants to enter addresses using interactive features and any functions surrounding repeated trips.

- The ability of participants to remove and add trips and trip legs (i.e., trip stages).

- The ability of participants to review travel entries before submitting their answers.

- The transition from the survey component to the travel diary.

The risks associated with only one round of user testing are that respondents are unlikely to coalesce and provide a unified voice on improvements or corrections to the diary features. This is potentially problematic for the project as the performance and usability of these revised changes will not have undergone further testing before administering this live. As such we cannot guarantee these problems are fully resolved or have not produced new ones. This could lead to increased instances of missing data, break-off or lower quality data due to burden or misinterpretation of the task. This risk would be further limited with this additional round of testing.

However, a further stage of testing poses a number of challenges to the timetable. If conducted in advance of the quantitative testing it would add approximately an additional six weeks to the development time allowing for additional recruitment, instrument adjustment, interviewer briefing and interviewing. An ideal schedule would be to fully complete all forms of user testing and development before commencing any piloting or live data collection as this allows greater confidence in the instruments used to generate data. However, while it would be beneficial to complete this before entering large scale quantitative testing, this would delay the launch of the live survey into 2025.

An alternative option is to undertake this additional testing concurrently with our first pilot to limit timetable disruptions. This would limit analytical possibilities of the pilot if notable adjustments were made as a result of this additional round of testing, but we believe that the potential benefits of the additional testing outweigh this potential drawback. As such, we propose that this testing is undertaken alongside the first quantitative pilot.

For this round of testing, we recommend conducting 12 interviews using a recruitment approach similar to the previous rounds (quotas based on sex, age, and travel behaviours) or any other characteristics prior testing suggest would be relevant. Testing would be completed on both PCs and smartphones to evaluate the usability of diary features across devices.

> **Additional Output:** At the end of this additional stage, we will circulate a short report summarising the main issues detected and recommendations for final question changes.

Quantitative testing
NatCen's recommended approach

1.12 Quantitative testing prioritisation

Following the qualitative testing phase we will progress to the quantitative testing phase. Due to time-constraints around the quantitative testing – it is an essential requirement that the survey is made live by the end of 2024 – it is not possible to address every area for experimentation in quantitative testing prior to the launch of the survey. In light of this, NatCen has reviewed the suggested areas for quantitative experimentation (see Table 1) and plotted them on a decision priority matrix. The aim of this exercise was to i: identify the areas which should be prioritised for testing; and ii: inform the design of the quantitative experimentation phase.[2]

Figure 2: WNTS Experimentation decision prioritisation matrix

---
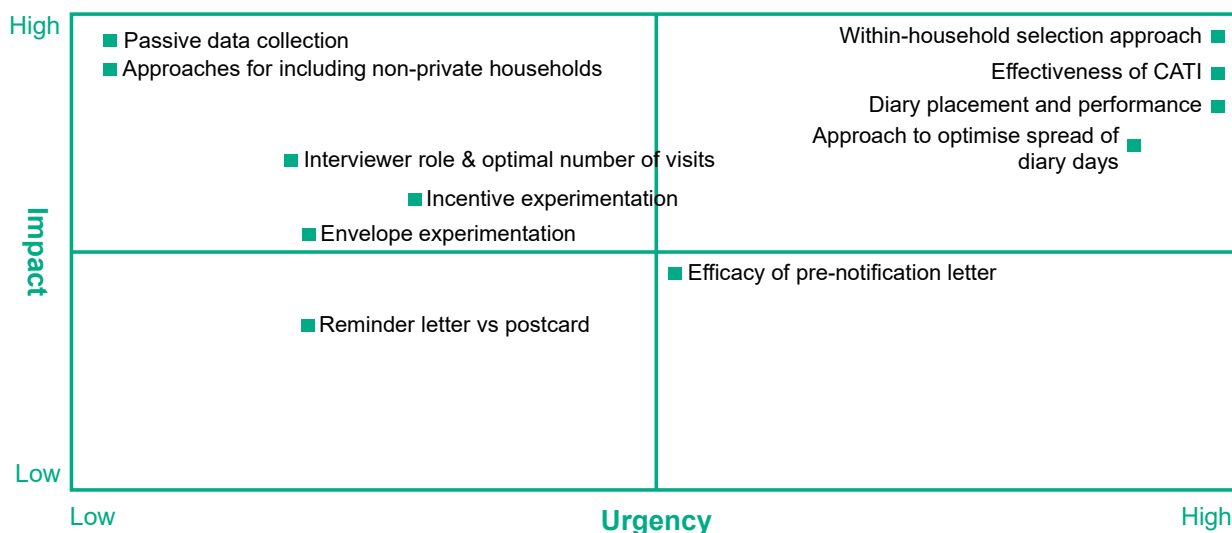
[2] Please note that in the matrix some potential areas for experimentation have been grouped together based on similar themes for ease of presentation.

Areas which are considered to be urgent – those in the right-hand side of the matrix – are rated as such because they will play a key role in determining the cost of the survey. As such, they are necessary to finalise the design of the survey ahead of its launch. These are:

- The within-household selection approach ('b' in the 'Target population' category in Table 1)
- The efficacy of administering the diary during the Computer Assisted Telephone Interviewing stage, and understanding take-up of the telephone option ('a' and 'b' in the 'Data collection mode' category in Table 1)
- Diary placement at the start or end of the survey, testing the roster vs. interleaved approaches, and performance of complex data collection elements ('f' in the 'Data collection mode' category in Table 1)
- The optimal approach to ensuring a spread of travel diary days ('f' in the 'Contact strategy' category in Table 1)

Areas which have a low-to-middle expected impact and a low-to-middle urgency are considered likely to have an impact on survey response, but are not essential to resolve prior to survey launch. Instead, these areas can be experimented on during the live survey with the aim of developing the optimal strategy over time. These are:

- Incentive experimentation, if the testing phase identifies the requirement for further testing ('a' and 'b' in the 'Incentives' category in Table 1)
- Experimentation with the contact strategy ('a', 'b', 'c' and 'd' in the 'Contact strategy' category in Table 1)
- The efficacy of using a pre-notification letter ('e' in the 'Contact strategy' category in Table 1)
- Experimentation with the interviewer role ('c' and 'd' in the 'Data collection mode' category in Table 1)

Areas which have a very high anticipated impact but low urgency are those which have the potential to make a substantial difference to the survey but require a longer development period. These are longer-term objectives that can be addressed alongside the live survey and introduced into the WNTS when considered ready. These are:

- Passive data collection ('e' in the 'Data collection mode' category in Table 1)
- The inclusion of those living in institutions and non-private households ('a' in the 'Target population' category in Table 1)

Based on the decision priority matrix, and within the known time-constraints, NatCen recommends that quantitative testing is conducted over four different periods.

We recommend that two separate periods of testing are conducted prior to the survey launch. These are to address areas for experimentation which will play a key role in determining the cost of the main survey and to conduct sufficient checks on both the design of the questionnaire/diary and the survey operation process. In particular, we recommend conducting two separate pilots, each of which will test specific aspects of the survey design/operation. The second pilot will also serve as a dress rehearsal, which will check the end-to-end survey process.

Less urgent experimentation can be conducted during the live survey – this includes aspects of the communication and incentivisation strategy. Longer-term experimentation – around the inclusion of non-private households and of passive data collection, can be conducted alongside, but separate from, the live survey.

Table 2 outlines how NatCen proposes to allocate each of the areas for experimentation to these different testing periods.

Table 2: Proposed allocation of areas for experimentation to different testing phases

Area	Pilot 1	Pilot 2 / Dress rehearsal	Live survey	Longer term
Effectiveness of CATI		✓		
F2F Interviewer role		✓	✓	
Optimal number of visits			✓	
Passive data collection				✓
Approaches for including non-private households				✓
Within-household selection approach	✓			
Envelope size/colour & logos			✓	
Reminder letter vs postcard			✓	
Use of pre-notification letter			✓	
Approach to optimise spread of diary days		✓		
Diary placement and performance	✓	✓		
Incentivisation	✓[3]		✓	
End-to-end process & formatting of data output		✓		
Interleaved vs. roster format	✓			

[3] Please note that we do not propose to conduct an incentive experiment during Pilot 1. Instead, it will be used to inform whether an incentive experiment is required at a later stage and, if so, what that experimentation should aim to achieve.

National Centre for Social Research
WNTS Stage 1b: Qualitative and Quantitative testing

This rest of this proposal focuses on NatCen's recommended approach for the testing prior to the launch of the live survey. Testing during the live survey, and longer-term development, will be discussed with Transport for Wales and the Welsh Government separately.

1.13 Summary of our recommended approach
As outlined above, this proposal contains NatCen's recommendation for quantitative testing prior to the survey launch. In particular, we recommend conducting two pilots, with the second pilot also acting as a dress rehearsal. These two stages will each have specific aims and objectives and will build upon the work completed to date. The final output of this testing will be a fully designed and tested survey ready to launch by December 2024.

We recommend that Pilot 1 is conducted via web only. It will have two broad aims: i) to test two separate approaches to within-household selection; and ii) evaluate the performance of the questionnaire and diary in a quantitative setting. By generating quantitative data on the questionnaire and diary for the first time we will be able to see how the data collection tools perform, assess the distribution of responses, and identify any areas for refinement ahead of further testing. This will also inform the data processing strategy, identifying the extent of coding requirements and any need for imputation as well as informing the data structure. The outcome of this stage will be a firm recommendation on the optimal within-household selection approach to adopt in the web-element of the survey and a review of the questionnaire and diary, including any recommendations for refinement.

Following Pilot 1, we recommend conducting a second pilot which includes all survey modes (web, telephone and face-to-face). This pilot will be used to test the performance and value of the CATI element, specifically in regard to operating the diary element in a non-visual setting. It will also be used to explore different strategies for optimising the spread of diary days and to generate estimates of the trip rate – these estimates will be vital in informing the sampling strategy for the main survey (that is, ensuring that sufficient sample is generated to meet TfW/WG's information needs). The outcome of this stage will be a recommendation on whether or not the diary is operable in the telephone stage (and, if not, a summary of the trade-offs associated with collecting survey data through this mode) and a full cost for delivering the main survey. To save time, we also recommend treating Pilot 2 as a dress rehearsal. This would check the whole survey process from end-to-end.

We now outline each of these stages in more detail.

1.14 Pilot 1
We recommend that the first pilot is conducted via web-only. There are a number of reasons for this. Firstly, a key aim of the testing phase is to determine the most efficient within-household selection approach for the web element of the survey. Therefore, it is unnecessary to test this for the face-to-face mode, where the selection approach is well-established. Secondly, following the design and qualitative testing of the questionnaire and diary it will be necessary to conduct a quantitative test to assess the performance of the data collection tools with a live sample. This will be particularly important for the diary which may include a number of complex elements such as location look-up, mapping and editable summary screens. These elements will have an unknown impact on data quality and respondent experience, so will need to be fully tested prior to the launch of the main survey. The test will also generate robust datafiles which will be used to interrogate the data and inform how it should best be processed.

Conducting a web-only test will be quicker and more cost-effective than conducting a test using all three survey modes.

The aims of Pilot 1 would be as follows:

- Test the within-household selection approach.
- Evaluate the performance of the questionnaire and the performance and placement of diary in a quantitative setting and make any recommendations for amendments necessary ahead of Pilot 2.
    - Examine the impact of diary placement (before vs. after the questionnaire) in survey completion (drop-offs, partial responses) and trip estimates.
    - Explore the effect of adopting an interleaved vs. roster approach in the design of diary questions in survey completion (drop-offs, partial responses) and trip estimates.
- Inform the data processing strategy and provide a rough first dataset for assessment by NatCen and TfW/WG

In addition, Pilot 1 will allow us to produce a first rough estimate of the trip rate recorded in the travel diary. This will inform whether or not a two-day travel diary is likely to generate sufficient data to meet TfW's information needs, or if it is necessary to explore options for collecting travel behaviour over a longer timeframe.

Testing the within-household selection approach

To test the within-household selection approach we propose to use a split-sample design. The sample would be randomly split into two sub-samples. One sample would follow **the 'all adults' approach**. Under this approach we would allow up to two adults per household to complete the online element of the survey. There would be no specific selection of which adults take part. This is the approach which is commonly used across a range of push-to-web surveys, including the British Social Attitudes survey (BSA).

The second sample would follow a two-stage selection approach, the aim of which is to select one eligible respondent at random from the household to complete the web element of the survey. To do this, we would follow the **Rizzo-Brick-Park method** (Rizzo, Brick & Park, 2004) which is as follows:

- Any household member completes the 'household-element' of the questionnaire and is asked about the total number of eligible people living in the household. (This would likely require some minor adjustment to the questionnaire for the experimental group.)
- If there is only one adult in the household, that adult is chosen for the individual-level interview.
- If there are two or more adults in a household, one adult is randomly selected with a probability equal to the inverse of the total number of adults. If the selected respondent is the same one who has completed the household-element of the questionnaire, the selection process ends, and they continue to complete the survey. If the person who completed the household-element is not selected and there are two adults in the household, the informant is told that the other adult is selected for the interview and the selection process ends.
- If the household informant is not selected and there are more than two adults in the household, another selection method (such as the Kish method or the last birthday method) is used to select an adult after excluding the household contact.

This method takes the advantage of the fact that more than eight in ten households in Wales have two or one adults. As a result, this method has the potential to significantly reduce the effort required to make a selection, while still generating a known and non-zero probability of selection.

Following the pilot, we would assess whether there is any benefit in adopting the Rizzo-Brick-Park method in the WNTS. This would be based on an assessment of the relative response rates and sample profiles achieved by each selection approach.

Evaluate the performance of the questionnaire and diary in a quantitative setting

Following the design and qualitative testing of the questionnaire and the diary, Pilot 1 would be used to evaluate the performance of the data collection tools in a quantitative setting. In particular, the Pilot would assess:

- The distribution of survey responses for items in the questionnaire. This will provide a quantitative measure of how well the question performs (for instance, if there is a high proportion of non-substantive answers); whether it captures the concept that it intends (for instance, if respondents are able to give an answer or tend to satisfice); and if the question is sufficiently discriminatory of views to elicit usable data.
- The level of dropouts from the survey and where they fall. This will identify whether there are any particular questions or set of questions, which cause respondents to close the survey early, and if the survey length is too burdensome
- The levels of consent for any recontact request (such as to join a panel), or for any data linkage[4]
- Responses to the diary element of the survey.
    - Whether the data produced meets TfW's information needs
    - Whether the roster or interleaved approach is most appropriate
    - Whether the pre-coded lists used for mode and purpose are sufficient, or if there is a high-level of respondents selecting 'other' responses
    - Whether the approach for location selection/coding works correctly, or if there is any need for post-hoc coding
    - The level of dropouts during the diary element
    - Whether there is a drop-off in trip reporting for the second day
    - Whether the survey suffers from any non-specificity in response that needs to be corrected

Following this assessment, NatCen would outline any recommendations for amending the data collection tools ahead of further testing.

Inform the data processing strategy

A key element of the success of the WNTS is producing a dataset (or datasets) in a structure and format that is easy to navigate for data users. Pilot 1 will be used to assess the raw data output and inform discussions around the optimal structure and format of the data. Questions to answer will include:

- At what level the data is produced – that is, whether there are separate datasets for household, individual and diary elements
- How trip and stage data is presented – whether it is necessary to produce separate datasets for trips and stages
- How locations are presented in the data
- Whether any auxiliary data needs to be appended to location data
- Whether any imputation is required
- What level of data cleaning is necessary
- Whether any derived variables are necessary

---

[4] We note that TfW has an interest in exploring consent for linking responses to the SAIL Databank. We would work with TfW to understand the linking process and test consent among respondents

- What format the data should be in when delivered to TfW

Following this assessment, NatCen would work with Transport for Wales and the Welsh Government to produce a data protocol: a comprehensive document with specifications and procedures to be used in the production of WNTS data files.

Proposed design

In order to meet these aims, NatCen proposes the following design:

- Mode: Push-to-web (no alternative modes)
- Target population: Adults (16+) living in private accommodation in Wales
- Target completes: 1,000
- Languages: English and Welsh
- Sample frame: Postal Address File (PAF)
- Sampling approach: Stratified random probability selection of addresses in Wales
- Survey software: Blaise 5
- Contact strategy: 1 x invitation letter; 2 x reminder letters.
    - Mailing 1: white C5 envelope; 2-sided letter; 2nd class postage. Sent on a Wednesday
    - Mailing 2: manilla C5 envelope; 2-sided letter; 2nd class postage. Sent one week after mailing 1
    - Mailing 3: white C5 envelope; 2-sided letter; 2nd class postage. Sent one week after mailing 2
- Assumed deadwood/ineligibility rate: 10%
- Respondent selection: Split sample. Half to follow 'all adult' approach (limited to two adults); half to follow Rizzo-Brick-Park approach
- Response rate estimate: 15% (Please note that this is an estimate for sampling purposes. A key aim of the pilot is to determine the response rate. NatCen cannot guarantee this level of response.)
- Issued sample size: 7,500
- Incentive: £10 conditional incentive (administered by Love2Shop e-voucher)
- Fieldwork period: 3 weeks
- Coding: Assumption that no coding/back coding is required for this pilot

NatCen would provide TfW with a raw data file, in SPSS format (.sav), after the completion of this fieldwork. At this stage the data will not be cleaned and processed. Instead, it will be used to inform the strategies for cleaning and processing the data. Following the production of the first draft of the data protocol, the pilot data will be cleaned and processed as a proof-of-concept and to inform revisions to the data protocol.

NatCen's responsibilities would be:

- Design a probability-based sampling approach;
- Draw sample from the PAF;
- Develop an initial contact letter and two reminder mailings;
- Administer print/postage of invitation and reminder letters;
- Develop a project-specific area of the NatCen website to host key information about the survey;
- Host the online survey;
- Manage incentive administration;
- Securely store survey data in accordance with GDPR requirements;

- Conduct analysis of within-household selection approach and questionnaire and diary performance and provide recommendations to TfW/WG;
- Produce initial draft of data protocol (to be refined following further testing).

1.15 Pilot 2 / Dress rehearsal

Following Pilot 1 we propose to conduct a second pilot. This pilot will be conducted using the full proposed selection of modes (CAWI, CATI and CAPI). This will be used to inform the transfer of sample from mode-to-mode and to assess the take-up of the CATI option and how it should be included within the main survey. This will be particularly important to consider in the context of the diary element of the survey – whether or not it is possible to complete it in a non-visual medium.

The aims of Pilot 2 would be as follows:

- Test the take-up and efficacy of the CATI element, in particular in relation to the administration of the diary element
- Test the transfer of the sample from mode-to-mode, including the testing of draft principles for the survey administration
- Test approaches to maximise the spread of travel diary days

For Pilot 2 we propose to use a non-random sample. This is for two main reasons: i) to facilitate the inclusion of the face-to-face element it is necessary to select areas purposively based on interviewer availability for the pilot; and ii) the design features being tested in this pilot do not require a random sample.

Testing the efficiency of the CATI element

The timeframe available for this pilot, and the associated size and scope of the sample that is possible, means that it is not viable to conduct an experiment where a proportion of the sample is offered the CATI element and another proportion is not. The expected take-up of the CATI option is low, so it would not be possible to generate sufficient sample to explore the impact on the sample profile or survey estimates. Indeed, following discussions with TfW/WG, it is apparent that the inclusion of the CATI element is primarily driven by accessibility requirements rather than methodological ones. Instead, we propose to allow the whole sample to have the CATI option. This will allow us to test and assess the following:

- The likely take-up of the CATI element for the mainstage and preferred language of CATI respondents. This will allow the creation of a cost-effectiveness assessment
- The process of transferring sample from the CAWI mode and to the CAPI mode (as necessary)
- An indication of the profile of those who choose to take part via CATI
- Any issues in administering the diary element of the questionnaire in a non-visual setting

Following this assessment NatCen would report on the findings and give recommendations for how the CATI element is most efficiently included in the survey design. There will also be options to refine the approach over time if necessary.

Test the transfer of the sample from mode-to-mode

In any mixed-mode survey a key challenge will be how to efficiently transfer sample from mode-to-mode without allowing multiple completions from the same individual/household and without negatively impacting on the

experience of the respondent. To do this, NatCen proposes to use Blaise 5 software, which is developed by Statistics Netherlands for Official Statistics and complex surveys. We also propose to use a bespoke sample control system to transfer the sample. This software is relatively new and requires full testing ahead of use in a live survey environment. This pilot would be used to develop protocols for moving the sample across modes and assess the performance of those protocols.

Following this assessment, NatCen would produce recommendations for any changes required to the protocols used for moving sample.

Test approaches to maximise the spread of travel diary days

A key requirement for the WNTS is to capture seasonality in travel behaviour. This means that it is necessary to ensure as even a spread of travel diary days across the week, month and year as possible. In interviewer-administered surveys the survey contractor has a good degree of control over the spread of diary days – it is possible to implement complex rules to allocate specific days to specific addresses. However, in self-administered surveys we do not enjoy this level of control. It is not possible to allocate specific days to each address because there is no control over when the household/respondent will open the invitation and complete the survey (it could be after any specific day) and there is a high chance of drop-out if the allocated diary day is in the future.

Instead, it is necessary to control completion by controlling the dispatch of invitation letters. For this pilot we propose to test two different approaches to dispatch to explore which creates the best spread of diary days (depending on the outcome of Pilot 1). For half the sample we will issue the invitation letter and reminders in single batches, but with each mailing on a different day of the week. For the other half sample, we will issue the letters in three batches on different days of the week.

Following this assessment, NatCen would analyse which approach elicits the best spread of travel diary days and the associated costs with the different approaches (the higher the number of mailings, the higher the cost). We would then work with TfW/WG to decide which approach is used for the mainstage survey.

Dress rehearsal

The optimal approach would be to conduct a separate dress rehearsal following the completion of Pilot 2. However, this is not feasible within the available timeframe. Instead, we propose to treat Pilot 2 as a combined pilot/dress rehearsal.

The dress rehearsal would test the whole survey process from end-to-end. It would be designed to replicate the mainstage as closely as possible.

The aims of the dress rehearsal element would be:

- Test end-to-end process of delivering the survey: sampling process, interviewer training, fieldwork administration, and the transition between interview and diary
- Test data processing procedures from end-to-end, including the production of data to test the format of the output

The combined pilot/dress rehearsal is not planned as a statistical test and therefore the data will not be weighted, but the data will be reviewed by NatCen statisticians to confirm that it includes the required information for it to be

National Centre for Social Research
WNTS Stage 1b: Qualitative and Quantitative testing

weighted. Because the data is not representative it will not be suitable for any internal analysis of travel attitudes or behaviour.

Proposed design

In order to meet these aims, NatCen proposes the following design:

- Mode: Push-to-web with CATI opt-in, followed by CAPI for non-responding addresses
- Target population: Adults (16+) living in private accommodation in Wales
- Target completes: 300
- Languages: English and Welsh
- Sample frame: PAF
- Sampling approach: Points selected purposively based on interviewer availability, but addresses selected at random
- Survey software: Blaise 5
- Contact strategy: 1 x invitation letter; 2 x reminder letters.
    - Mailing 1: white C5 envelope; 2-sided letter; 2nd class postage
    - Mailing 2: manilla C5 envelope; 2-sided letter; 2nd class postage
    - Mailing 3: white C5 envelope; 2-sided letter; 2nd class postage
    - Trial of alternative mailing strategies. Half sample to be sent in one batch; half the sample to be sent in three batches
- Assumed deadwood/ineligibility rate: 10%
- Respondent selection: Based on outcome of Pilot 1.
- Response rate estimate: 15% for web-CATI element; 25% for face-to-face element; 37% overall. Please note that this is a pre-pilot estimate. The response rate estimate will be refined following the completion of these pilots.
- Issued sample size: 910
- Incentive: £10 conditional incentive (administered by Love2Shop e-voucher)
- Fieldwork period: 7 weeks. All modes will be available at weekends
- Data processing: Full data processing (excluding weighting), based on data protocols developed during Pilot 1. Data processing to be further refined during the production of data.

NatCen would provide TfW with a data file, in SPSS format, after the completion of this fieldwork. This would be a first draft based on the data protocol developed following Pilot 1. It would be intended to be used to assess the structure and format, rather than the content.

NatCen's responsibilities would be:

- Design a purposive sampling approach;
- Draw an address sample from the PAF;
- Develop an initial contact letter and two reminder mailings;
- Administer print/postage of invitation and reminder letters;
- Develop a project-specific area of the NatCen website to host key information about the survey;
- Host the online survey;
- Manage request for CATI completion;
- Conduct face-to-face interviewing;
- Manage incentive administration;

- Securely store survey data in accordance with GDPR requirements;
- Conduct analysis of efficacy of CATI element; transfer of sample from mode-to-mode; and optimal spread of travel diary days;
- Conduct assessment of end-to-end process, including any recommendations for amendment to be considered and agreed by TfW and Welsh Government;
- Produce initial draft of data format based on data protocol designed following Pilot 1.

Following the dress rehearsal NatCen would assess how it went and make any agreed adjustments necessary to ensure the smooth delivery of the mainstage fieldwork.

Progression to Phase 2

Initial steps to progress to mainstage

It is a key requirement that Phase 2 of the WNTS delivery programme – the roll out of the mainstage data collection across Wales – begins before the end of 2024.

The recommended WNTS design requires a large volume of face-to-face fieldwork in Wales. To cover this volume the National Centre for Social Research would need to scale-up its interviewer resource capacity in Wales. We estimate that scaling-up to the required number of interviewers would take six months. In effect, this means that in order to launch the mainstage survey in December 2024 it would need to be formally commissioned by 31 May 2024.

This lead-in time means that it will be necessary for TfW to seek approval to progress to Phase 2 before all of the testing and piloting in Stage 1b has been completed. Specifically, Pilot 1 will have been fully completed and reported and any agreed changes for Pilot 2 will be in the process of being implemented, but fieldwork for Pilot 2 will not have started.

This means that the backlog of approved changes required to be made to the methodology, systems and processes for Phase 2 will not be fully known.

However, Pilot 2 is primarily focused on testing and refining the survey process, rather than the methodology. In particular, it looks at the efficacy of the CATI element, the process of transferring sample across modes, and how best to maximise the spread of travel diary days. While these are key aspects to resolve ahead of Phase 2 we are confident that there will be ample time to make any changes and that the potential impact on the costing of Phase 2 is low.

That said, it will be necessary to base some of the costing variables for Phase 2 on assumptions. These will specifically be around the face-to-face response rate, the number of Welsh-language face-to-face interviews, the number of CATI interviews (both in English & Welsh), the level of drop-off in the CATI element, the full scope of the weighting requirement; and the scale of any changes to the programme as a result of Pilot 2.

We would work closely with TfW to define these assumptions and (where possible) to provide an evidence base behind them. This will ensure that the costing provided is fully transparent. We would also work closely with TfW to agree on a costing strategy which will most efficiently allow any adjustments which Pilot 2 identifies to be made prior to the launch of Phase 2.

Our proposal for Phase 2 would clearly outline the assumptions made and any associated risks with commissioning Phase 2 at that point.

References

Cornick, P. (2021). *The NatCen Remodel Approach. A systematic approach to transforming social surveys*. Briefing Paper. Centre for Social Survey transformation.

Cornick, P., Aizpurua, E., & Howe, S. (2023). *Wales National Travel Survey: Review Report*. Prepared for Transport for Wales.

Rizzo, L., Brick, J. M., & Park, I. (2004). A Minimally Intrusive Method for Sampling Persons in Random Digit Dial Surveys. *The Public Opinion Quarterly, 68*(2), 267–274. http://www.jstor.org/stable/3521592