
Social Research Association Annual Conference 2024

Dear diary...

**The challenges and opportunities of
designing online diaries**

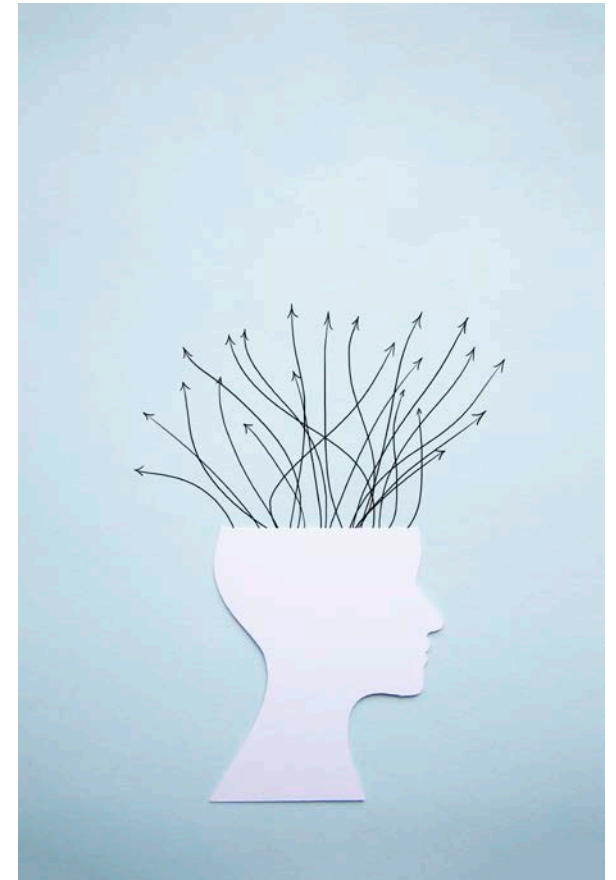
Eva Aizpurua, Peter Cornick & Shane Howe

Centre for Social Survey transformation

National Centre for Social Research

The Challenges of Recall-based Diaries

- With the increased shift towards self-administered and computerised data collection, a growing number of recall-based data collection tools (e.g., time-use, mobility, expenditure diaries) are moving online.
- Diaries pose unique challenges... They often span multiple days and require advanced survey features such as maps, summary screens, and complex loops, to effectively meet research needs.



The Wales National Travel Survey



Participants: Individuals 16 and over who reside in private households in Wales



Sampling frame: Postcode Address File (PAF)



Survey modes: Web + telephone + face-to-face



Survey topic: Travel attitudes and behaviours



Component: Questionnaire + travel diary (programmed within the survey software)

Diary Design Considerations: Recall Focus



Trip-based: Require respondents to enumerate every journey undertaken during the diary period (e.g., English National Travel survey)



Activity-based: Rather than asking respondents about the journeys they undertook, and then gather information about the trip's purpose and other attributes, this approach uses activities as the primary unit of analysis



Location-based: Diary keepers are asked where they went in chronological order, and then probed about how they got to each destination and the purpose of each visit (e.g., Netherlands Mobility Panel)

Diary Design Considerations: Recall Focus II

For help with filling in please unfold side flap for notes

JOURNEYS Please record each journey on a new row. Include very short ones and return journeys. Include walks if 1 mile or more

A	B	C	D	E	STAGES These columns are for entering details of each stage of your journey															
What was the purpose of your journey? See Note A	What time did you leave? See Note B	What time did you arrive? See Note C	Where did you start your journey? (Tick Home or give the name of the village, town or area) See Note D	Where did you go to? (Tick Home or give the name of the village, town or area) See Note E	Only fill in these columns if you used a CAR or OTHER MOTOR VEHICLE				Only fill in these columns if you used PUBLIC TRANSPORT											
					F	G	H	I	J	K	L	M	N	O						
					What method of travel did you use for each stage of your journey? See Note F	How far did you travel? (Miles) See Note G	How long did you spend travelling? (Minutes) See Note H	How many people travelled including you? See Note I	Which car or other motor vehicle did you use? See Note J	Were you the driver (D) or a passenger (P)? See Note K	How much did you pay for parking? See Note L	What type of ticket did you use? See Note M	How much did your ticket cost? See Note N	How many times did you board? See Note O						
1	Go to work	8:15	9:00	Home	Car	18	45	1	Fiesta	✓	£ 2.00									
2	Go food shopping	5:30	6:12	Home	Car	16	42	1	Fiesta	✓										
3	Go home	6:20	6:30	Home	Car	4	10	1	Fiesta	✓										
4	Go out for meal with friends	7:00	8:05	Home	walk	1	18	2												
					Train	8	27	2				Single	£ 2.90							
					Bus	1.5	8	2				Single	£ 1.00							
5	Go home	10:20	10:55	Home	Taxi	10	25	2												

Time	What were you doing? Record your main activity for each 10-minute period from 07.00 to 10.00!	What else were you doing? Record the most important parallel activity.	Where were you? Record the location or the mode of transport	Were you alone or together with somebody you know?					
	Only one main activity on each line! Distinguish between travel and the activity that is the reason for travelling.	Indicate if you used, in the main or parallel activity, a computer or internet	e.g. at home, at friends' home, at school, at workplace, in restaurant, in shop, on foot, on bicycle, in car, on motorbike, on bus ...	Mark "yes" by crossing					
				Alone	With other household members				Other persons that you know
					Partner	Parent	Household member up to 9 years	Other household member	
07.00-07.10	Woke up the children		At home	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
07.10-07.20	Had breakfast	Talked with my family		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
07.20-07.30	---	---		<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
07.30-07.40	Cleared the table	Listened to the radio		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
07.40-07.50	Helped the children dress	Talked with my children		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
07.50-08.00	Went to the day care centre	---	On foot	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

TNS Nipo

Home My addresses Daily Schedules Change password

Day 1 Day 2 Day 3

Daily schedule for Friday September 12 2014

Your next location is Office Change location

At what time did you arrive? 09:03 I arrived the next day

What did you do? Paid job at a fixed location

At what time did you leave? 17:30 I did not leave this location

Where did you go? Supermarket

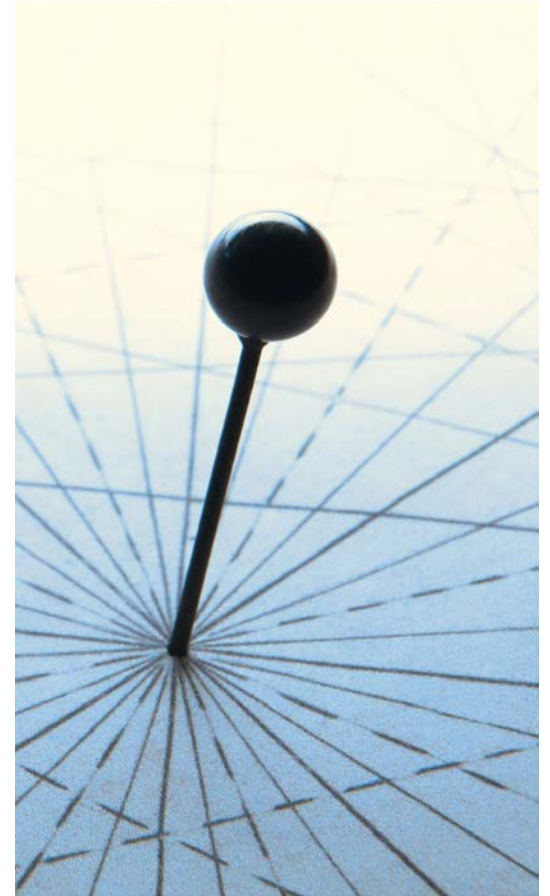
Save

The locations you have visited today

Time	Location	Time of arrival	Activity	Time of departure
00.00	Home			
08.11	School	08.11	Activities at home	07.45
08.12	Office	09.03	Dropping off people	08.25
08.25			Work outdoors - Paid job at a fixed location	17.30

Diary Design Considerations: Recall Focus III

- **Location-based** travel diaries tend to yield higher reported trips than activity- or trip-based diaries (Behrens & Masaoe, 2009).
- **The activity-oriented approach** aligns more closely with people's cognitive processing. Travel, in most cases, is a means to an end. However, activity-based diaries pose a greater burden because respondents need to retrieve information about all their activities, many of which may not involve travelling.
- **Location-based diaries offer a middle ground:** They focus on locations, which is a familiar concept, while only gathering trip information.



Diary Design Considerations: Retrieval of Location Information

You started at your home. From there, where did you go?

Please think about any journey, even if it was only a short walk or to run errands.

Address lookup

Pin drop

API Integration

I can't find the location on the map

Diary Design Considerations: Recall Focus IV

Insights from cognitive interviewing (Dec. 2023, $n = 12$) and usability testing (Jan. $n = 16$)

Comprehension: The location-based approach was understood by participants as intended, although some individuals were unsure of how small a trip could be (e.g., taking the bins out).

Usability: Participants tended to use the map integration as they would typically use Google Maps. Some, especially on smartphones, made use of zooming in/out.



Diary Design Considerations: Length of recall

	Advantages	Disadvantages
Single day	<ul style="list-style-type: none">- Lower respondent burden- Lower risk of recall errors/ underreporting- Lower risk of dropouts between the survey and the diary	<ul style="list-style-type: none">- Reduced accuracy due to lower volume of data- Does not capture intra-personal variation in travel behaviour- Less ability to capture infrequent travel activities/ modes (e.g., long-distance trips)- Less volume of trip data, which requires a larger sample
Multi-day	<ul style="list-style-type: none">- Higher level of precision (i.e. higher volume of data)- Captures infrequent travel behaviour/ modes- Captures intra-personal variability, allowing to study differences in individual behaviour- Requires a smaller sample size at the person level	<ul style="list-style-type: none">- Higher costs- Increased respondent burden- Increased attrition- Higher risk of underreporting- Lower response rates- Complex data management and analysis (independence of observations assumption is violated)

Diary Design Considerations: Length of recall II

Assumptions:

- Similar levels of journey recording per day as in the English NTS
- 40% drop-off between days 2 and 3, and days 3 and 4

Expected volume with 5,000 achieved responses	1 day diary	2-day diary	3-day diary	4-day diary
Total days of diary data recorded	5,000	10,000	13,000	16,000
Approx. days of diary data per day of week	714	1,429	1,857	2,286
Approx. days of diary data per day of week NEFF	621	1,242	1,615	1,988
Approx. total journeys recorded	9,450	18,650	24,230	29,630
Approx. total journeys NEFF	8,217	16,217	21,070	25,765
Approx. average journeys per day of week NEFF	1,174	2,317	3,010	3,681

Diary Design Considerations: Length of recall III

Next steps – Web pilot (Q2 2024):

- Examine drop-off rates between days 1 and 2
- Compare the number of trips reported on day 2 with respect to day 1
- Depending on the outcomes → Potential to incentivise day 2 completion separately



Diary Design Considerations: Approach to follow-up questions



Self-reported travel diaries are characterised by their use of follow-up questions to gather detailed information about travel times, distances, journey purposes, associated costs, etc. Main approaches:

- 1) **Interleaved:** Every question is immediately followed-up by a series of questions (e.g., English NTS)
- 2) **Grouped or roster:** All filter questions are presented in a block, before introducing any follow-up questions (e.g., Dutch Mobility Panel)

Diary Design Considerations: Approach to follow-up questions II

	Advantages	Disadvantages
Interleafed	<ul style="list-style-type: none">- Tends to be perceived as more conversational, with respondents being asked about each location once, rather than jumping between them- The focused nature of the interleafed design might ease information retrieval and impose less burden on respondents, improving recall accuracy (Clark-Fobia et al., 2018)	<ul style="list-style-type: none">- Increased risk of motivated misreporting (Kreuter, Eckman, & Tourangeau, 2020). In efforts to shorten the survey and reduce their burden, respondents might omit locations
Roster	<ul style="list-style-type: none">- Expected to reduce the risk of motivated misreporting, providing more accuracy for the filter questions (locations)	<ul style="list-style-type: none">- An additional consideration is the expected share of smartphone completions. Diary keepers using small devices face a higher break-off risk (Chung et al., 2021). In the grouped format, delaying follow-up questions could risk capturing detailed information about the journeys

Diary Design Considerations: Approach to follow-up questions III

Web test
Randomised test
(Q2 2024)

Interleaved

From [address], where did you go first? Please think about any trip, even if it was only a short walk or quick errand.



Follow-up questions

Roster

Please list, in order, everywhere you went between 5:00 am yesterday, [DD/MM/YYYY], and 5:00 am today, [DD/MM/YYYY]. Start with the place you visited first, and then list each subsequent place you went to until 5:00 am today.



Follow-up questions

Conclusions



Each design decision has trade-offs – there is no silver bullet



Testing the assumptions, performance, and usability of the instrument is crucial, especially when using complex features (e.g., maps)



Design decisions are interrelated, and they interact with each other



The success of many aspects ultimately depends on execution

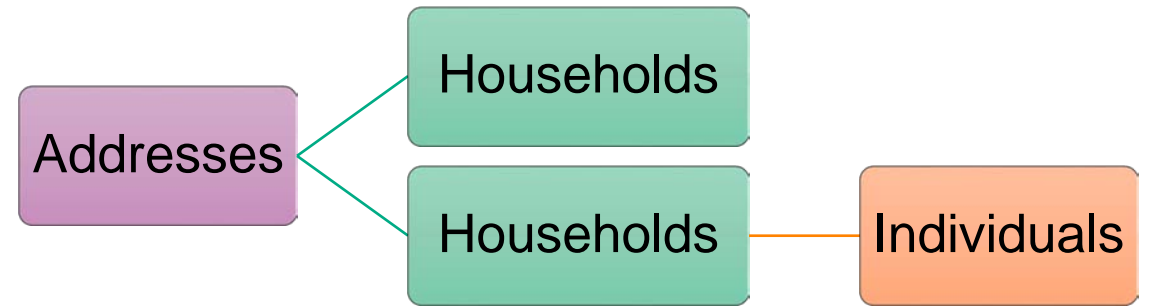
Thank you

E. eva.aizpurua@natcen.ac.uk



Diary Design Considerations: Household- vs. Individual-level surveys

- **Household surveys:** Every eligible member of the household is invited to participate in the survey (e.g., English NTS).
- **Individual surveys:** Extend invitations to specifically selected individuals.



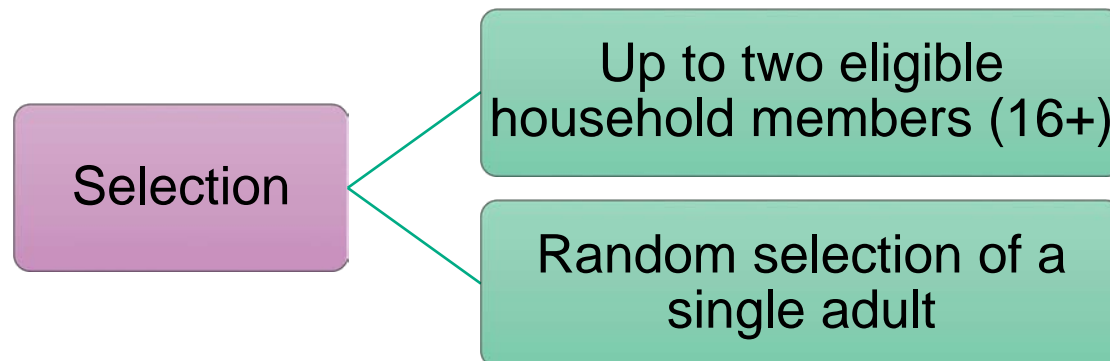
General information about the structure and composition of the household is collected in both cases. The choice between these two depends on several factors, including the available sampling frame or the mode of data collection.

Diary Design Considerations: Household- vs. Individual-level surveys II

	Advantages	Disadvantages
Household	<ul style="list-style-type: none"> - The amount of data is likely to be higher, even if response rates are lower - Provides a more holistic view of household travel behaviour/ dynamics - There is a lower risk of selection bias - Once a household is recruited, information for all members is obtained at minor cost 	<ul style="list-style-type: none"> - Cooperation from every household member is challenging - Requires relaxing proxy rules, reducing data quality and placing extra burden on households with children - Clustering requires more complex analysis (nested structure of the data) and leads to a loss of efficiency (variance increase) - Logistics can be challenging (e.g., distribution of access codes, following-up with partial responding households)
Individual	<ul style="list-style-type: none"> - Poses lower burden on respondents - Likely to yield higher response rates - New data collection tools for trip data are intrinsically individual based - If there is an interviewer-based component, there is extra flexibility in scheduling (only one person needs to be available) 	<ul style="list-style-type: none"> - High risk of selection bias, especially in self-administered environments - Household interactions/ dynamics are only partly reflected - The need to survey more people to get a comparable number of responses can lead to higher costs

Diary Design Considerations: Approach in the WNTS

- **Face-to-Face & Telephone Interviewing:**
Random selection of one household member (Kisch method)
- **Web survey:**
Randomised experiment (Q2 2024):



Experiment goals:

How does the selection process influence...

- 1) The composition of the sample (nonresponse error)?
- 2) The volume of data collected (precision of the estimates)?
- 3) Dropoff rates (nonresponse error)?