Discussion paper

A smart Travel Survey What is the role of the respondent?

Daniëlle Remmerswaal
Barry Schouten
Jeldrik Bakker
Janelle van den Heuvel
Jonas Klingwort

January 2025

Summary

Travel Surveys are considered promising candidates to go ‘smart’. Respondents need to be both motivated and competent to correctly report all details of their travels for a specified time period. Location tracking offers options to remove burden on the respondent and to improve the quality of measurement. Adding contextual information, the collected location data may also be input to predictions of travel mode and purpose. However, location data are also subject to various types of error that in part can only be adjusted for with the help of respondents. Given that a reduction of burden and avoidance of recall errors and underreporting are arguments to go smart, asking the respondents to help checking and correcting data is contradictory. The resulting design decision is known as the active-passive data collection trade-off.

In 2022, Statistics Netherlands fielded a travel-app assisted experiment, in which respondents were invited to check and, if needed, impute or adjust daily stop-track segmentations. The amount of possible editing was randomized between a light and a heavy editing sample. Furthermore, the length of the tracking period was randomized between one day and a week.

In this paper, we evaluate the impact of respondent editing options and length of the reporting period on actual respondent editing behaviour. We concentrate on the in-app actions by respondents. A separate paper focusses on the impact of editing on stop-track segmentations and on data quality. We conclude that respondents use all editing options and that they, generally, remain active throughout a full week. We must note, however, that some brands and devices had low data quality. This has led to passive participation and drop-out for a subset of respondents.

Keywords

Smart surveys, Location tracking, Mobility, Push-to-smart data collection strategy, Active-passive data collection trade-off

Content

1. Introduction

Smart surveys employ the features of smart devices in collecting and/or processing data. They are particularly promising for surveys that are (cognitively) burdensome, demand for detailed knowledge or recall, or include topics for which questions provide weak proxies. Travel surveys are typical examples of surveys that satisfy these criteria. This has been recognized early on. Over the last decade a wide range of studies has been conducted on this subject (e.g. McCool, Lugtig and Schouten 2021, Harding et al 2022, Gillis, Lopez and Gautama 2023 and Lawson et al 2023). At Statistics Netherlands studies into the use of an app-assisted approach including location tracking started in 2017. The aim is to supplement data collection options in general population travel and time use surveys. This paper reports the anticipated role of respondents based on a second large-scale field test employing an app-assisted approach.

In smart travel and time use surveys, location tracking may serve four purposes: The first is the derivation of a daily frame of events, consisting of stops and tracks, that helps respondents fill in their diaries. The second is the prediction of travel modes that the respondents have used during their travels. The third is the prediction of the main purposes of stops that respondents have made. And specific to travel surveys, the fourth is the mapping of the travels on the mobility infrastructure (paths, roads, public transport lines, ferry). Location tracking data are, however, subject to missing data and imprecise measurements (e.g. Harding et al 2020, McCool, Lugtig and Schouten 2022, Klingwort et al 2024). Consequently, any derivation based on location data may be subject to error as well, leading to missed or spurious events and predictions with a low accuracy.

Some data errors may be partially resolved with the help of contextual data or historic data from the same respondent. Contextual data may consist of linked administrative data and points-of-interest data linked from external sources. Historic data may be travels and stops made by the same respondent on earlier days that had no or little error. See for example Chen et al (2016), Bantis & Haworth (2017), Krause & Zhang (2019), Smeets, Lugtig & Schouten (2019), McCool et al (2022 and 2024) and Zhou et al (2022).

Other, and in particular more influential errors require knowledge that only the respondents may have. Respondents may be asked to fill in larger gaps in location data, to check data sections with low accuracy, to check and revise stops and tracks, to check and supplement travel modes and to check and supplement travel purposes. The obvious questions are whether respondents are able to do such tasks and, subsequently, whether they are sufficiently motivated to do so adequately. Literature on this so-called active-passive trade-off in smart surveys is still thin. Schmidt (2014) and Hu et al (2017) provide results for diary studies, but without smart features. Wenz, Jäckle and Couper (2019) and Harding et al (2021) do present results on longitudinal studies in a smart context. All find that respondents have different preferences on the kinds of tasks they are willing to perform and for how long. Lunardelli et al (2024) performed a cross-country study

in Italy, the Netherlands and Slovenia into respondent perceptions on smart surveys and corresponding data collection tasks. Among the tasks was location tracking for time use surveys and for travel surveys. Again varying preferences were found for different tasks. Importantly, respondents in all countries indicated that they like to control what data are collected. In this paper, we focus on the active-passive trade-off travel surveys.

Between November 2022 and February 2023 a second large-scale field test was conducted by Statistics Netherlands. The cross-platform app had been completely redesigned. The new app ('CBS Onderweg in Nederland') included several options for respondents to edit automated stop-track segmentations that were shown to them during the diary reporting period. In the experiment, three design features were randomly varied: the length of the reporting period (one day or seven consecutive days), the amount of possible respondent editing (full editing or limited editing) and the offering of the web diary as alternative (direct at invitation, at first reminder or at second reminder). Here, we consider the editing options and the length of the reporting period.

In this paper, we describe the outcomes of the 2022-2023 experiment in terms of respondent involvement and respondent activity. Ultimately, the user interface and user experience (UI-UX)of the app-assisted Travel Survey should make explicit choices in what can be asked from respondents and what not. We, therefore, study three research questions:

1. What editing actions are performed by respondents?
2. To what extent can active editing be related to respondent characteristics?
3. How does length of the reporting period affect active participation?

We need to note that the 2022 travel survey app did suffer from some technical issues for specific brands/models. As a consequence, drop-out of the study was in part caused by low data quality. Again, we will make pragmatic choices when studying completion versus registration.

In parallel to this paper, other papers are produced based on the 2022-2023 study. Schouten et al (2024) describe the design of the 2022-2023 field test and evaluate the level and the representativeness of the response. Klingwort et al (2024) and Gootzen, Klingwort and Schouten (2024) evaluate, respectively, data quality and location data processing. Remmerswaal, Lugtig, Schouten and Struminskaya (2024) explore, in depth, the influence of the length of the reporting period on activity and data quality. A first study into respondent paradata in travel survey context was also conducted by Remmerswaal, Lugtig, Schouten and Struninskaya (2024) and elaborated by Giacobbe (2024).

This is the outline of the paper: We briefly revisit the design of the 2022-2023 field test in Section 2. In Section 3, we explore the editing actions performed by respondents. We then move to investigation of the types of respondents in Section 4. Next, we test whether the length of the reporting period affects activity during the study. We end with discussion and next steps in Section 6.

2. The 2022-2023 app-assisted field test

The field test consisted of two main parts, the app-assisted main mixed-mode travel survey (MMTS) and a follow-up online evaluation survey (ES). We briefly revisit the MMTS. Details can be found in Schouten et al (2024).

Three experimental conditions were randomized across the sample:

- Reporting period: An experiment with the number of participation days was conducted: Half of the respondents were invited to participate in the app for one day, the other half were invited to participate for one week. The one-day group was, however, told that they did not have to stop after the first full day and could continue up to a full week.
- Concurrent online questionnaire: An experiment with the timing of offering the online questionnaire was conducted: Respondents were offered to fill out the online questionnaire instead of using the app directly in the invitation letter, in the first reminder letter, or in the second reminder letter. Different invitation and reminder letters were used.
- Editing options: An experiment was added concerning the amount of editing respondents were invited to do (and were able to do). One half of the sample had full editing options and one half had limited editing options. Full editing included: adding stops or tracks, deleting stops or tracks, changing start and end times of stops and tracks, labeling travel modes and labeling travel purposes. Limited editing restricted options to deleting stops and tracks and labeling.

The three conditions were crossed, leaving us with 12 different subsamples. Table 2.1 presents all samples and experimental conditions including the follow-up sample as well. The follow-up sample had a seven-day reporting period and full editing.

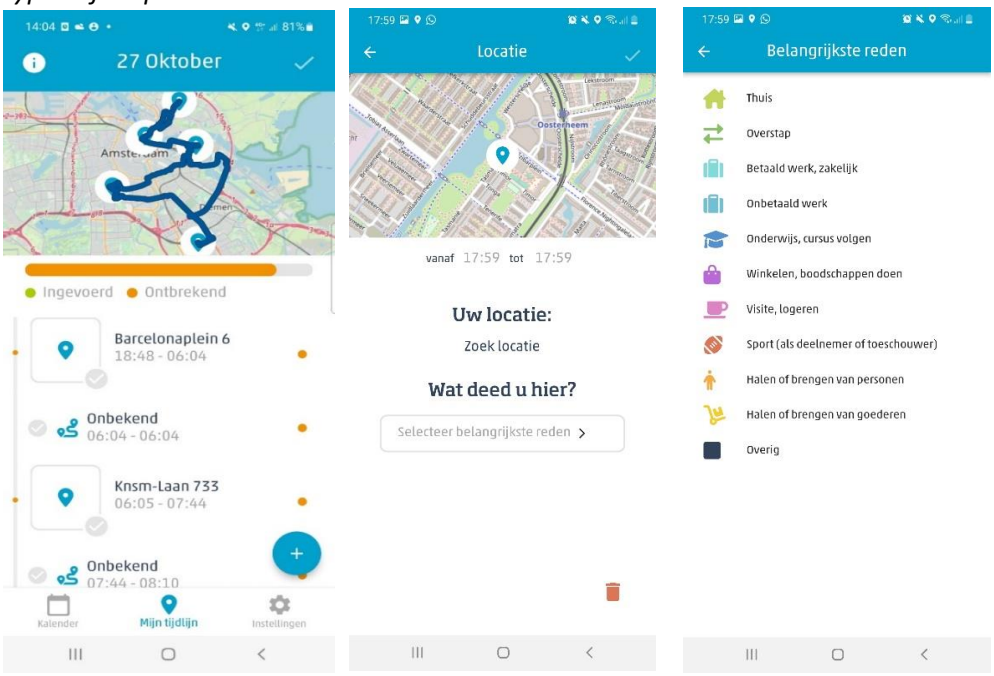
Table 2.1: Overview of samples and experimental conditions in MMTS.

Location tracking	Timing of questionnaire	Editing	
		Full	Limited
One day	Invitation	212	212
Seven days	Invitation	212	212
One day	1st reminder	212	212
Seven days	1st reminder	212	212
One day	2nd reminder	212	212
Seven days	2nd reminder	212	212

Figure 2.1 shows a number of screenshots of the MMTS app. Underlying to the app was an automated stop-track segmentation algorithm. Location points were clustered within stops based on radius and duration parameters. Missing location

data time slots were displayed in shaded colours. Details on stop-track decision rules can be found in Klingwort et al (2024).

Figure 2.1: Three screenshots of the MMTS app: Screen for daily stop-track segmentation, screen for adding a label to a stop location and screen listing the types of stops that could be selected.



3. What editing actions are performed by respondents?

In this section, we answer our first research question on in-app activity. Let us first discuss what involvement was asked from respondents.

After completing the in-app questionnaire, respondents needed to perform four types of actions:

- Check and, if needed, edit stops and tracks
- Label stops and tracks
- Provide contextual information on reporting days
- Submit/validate calendar days

In all cases, respondents could proceed without actually performing the actions, i.e. they could ignore labeling, leave stop-track decompositions unchecked, not answer day questions and leave calendar days non-validated. The app would keep

recording location data. Respondents are, however, pointed to uncompleted actions through the app UI.

The stop-track-decompositions and subsequent edits demand for more explanation: Stops and tracks are automatically derived by the app based on the collected location data through simple stop detection rules. Once a respondent resides in an area with a radius of 100 meters or less for at least three minutes, then the 'point of gravity' of the corresponding points is listed as a stop and shown as such in the day overview. Tracks follow from the time periods between stops. Time periods of more than 30 minutes without recorded location data were not included in stop-track decompositions and appeared as missing data in the daily overview. Respondents were asked to check, and if needed and allowed, to edit these periods. The app UI distinguished between the following actions a respondent can take:

- Deleting an existing stop or track
- Labeling the transport mode of tracks
- Labeling the purpose of stops
- Editing the begin and end times of an existing stop or track
- Creating a new stop or track by hand
- Supplementing stops and tracks for missing data periods

We tested two versions of the app: one with all edit options and one with limited edit options. Under limited editing, no new stops or tracks could be added, no editing of begin-end times of stops-tracks was possible and missing data could not be supplemented, leaving only options 1, 2 and 3

Next, we consider the following sub-questions:

- How much time on average do respondents spend in-app?
- What editing actions do respondents perform?
- What labeling actions do respondent perform?
- To what extent do respondents submit/validate days?

3.1 Time spent in the app

Based on the in-app navigation paradata, we summarized per user how many times they visit the app (number of sessions) and for how long (in minutes). It is, however, not trivial how to define a respondent session. This has two causes. The first is that a clear log message signaling the beginning of a session is not included in navigation audit trails. This has to be deduced from other log messages. The second is that a respondent can keep the app open on their phone while not being active. We, therefore, use the following definition: A new session is started when there is activity after ten minutes of inactivity, meaning that no log message has been collected in the past ten minutes. If the app is used within ten minutes after the last use of the app, this activity is combined in one session.

We distinguish two types of sessions: The first intro session and follow-up sessions. The first session takes longer on average, because respondents have to

log in and are asked to fill in a pre-questionnaire. Follow-up sessions are shorter on average.

In Table 3.1, we show the average in-app times for the different experimental conditions split into intro and follow-up sessions. For the intro session, there is a small difference between the apps with full editing and with limited editing. As may be anticipated, participants spend less time in the app with limited editing possibilities.

Table 3.1: Average time spent in-app in minutes per condition for the intro session and further sessions. Standard errors between brackets.

Condition	Intro session (in min)	Follow-up sessions (in min)
Full editing and one-day	3.9 (1.7)	1.1 (2.0)
Full editing and seven-days	4.3 (2.0)	1.6 (2.0)
Limited editing and one-day	3.7 (2.3)	1.1 (1.9)
Limited editing and seven-days	4.1 (2.3)	1.0 (1.6)

In Table 3.2, we display the average number of sessions and in-app times for the first full day excluding the intro session. Some care is needed because numbers per entry in the table are relatively small. As anticipated, limited editing decreases the average in-app time considerably and to a lesser extent also the number of sessions.

Table 3.2: Average number of sessions and in-app time on the first full day per condition. Standard errors between backets.

	First full day of 1 day		First full day of 7-day	
	Full	Limited	Full	Limited
Time in-app (in sec)	57 (41)	41 (51)	72 (48)	36 (27)
# in-app sessions	4.9 (2.9)	4.7 (3.6)	4.7 (2.7)	3.9 (2.3)

We conclude that in-app time is acceptable in total duration and that editing options strongly impact time spent.

3.2 Editing actions

We first look at actions respondents performed overall. Before we do, we must note that numbers of respondents are fairly small leading to fairly large standard errors. Table 3.3 shows the proportion of respondents' diary days that included each type of action at least once. Hence, respondents are included in the proportions as many times as they submitted days. Next, Table 3.4 splits results for the four experimental conditions limited to the first full day. So now denominators in the proportions are the number of respondents in a condition that participated at least for one day. For Table 3.4, standard errors are in the order of 6%.

Table 3.3: Proportions of submitted days with each type of action for the first day of tracking.

Type of action	Proportion of days
Deleted >0 stop-tracks	90%
Labeled stop-tracks	73%
Modified time stop-tracks	49%
Added >0 stop-tracks	31%
Number of diary days	Days = 1536

Table 3.4: Several indicators on in-app activity split by study duration and by editing conditions.

Indicators	One-day		Seven-days	
	Full editing (n= 61)	Limited editing (n= 72)	Full editing (n=84)	Limited editing (n=76)
Checked stop-tracks	100%	93%	90%	91%
Deleted >0 stop-tracks	79%	85%	76%	79%
Labeled stop-track	67%	63%	67%	60%
Added >0 stop-tracks	30%	Not applicable	50%	Not applicable

Some conclusions can be drawn, despite the small numbers. First, diary days almost always contain spurious events according to respondents; around 90% of days had at least one deleted event. This is in line with the choice to make stop-detection relatively sensitive. Second, around a quarter of diary days has no labels. This is a fairly large amount and may point at the added value of predictions. However, not labelling days may be related to influential data errors that discouraged respondents. Third, on around a third of diary days an event is missing according to respondents. It is not yet clear whether this is imputation of gaps in location data or whether events were missed by stop-detection.

Looking at the experimental conditions, some differences appear. However, except for the difference in adding days between one-day and seven-days samples, none of the differences would test as significant.

3.3 Labelling stops and tracks

In the previous evaluation, we have seen that respondents may label stops and tracks, but that also a fair amount of them leaves stops and tracks unlabeled. This may be the consequence of spurious stops and tracks and is evaluated in more

detail in Klingwort et al (2024). Here, we briefly consider what labels are given and, more importantly, whether sufficient diversity is present in the labels.

Figures 3.1 and 3.2 display proportions of labels for, respectively, stop purpose and travel mode. From these, for now, we conclude that all modes and all purposes appear with sufficient diversity.

Figure 3.1: Proportions labels stop purpose across categories in the MMTS app.

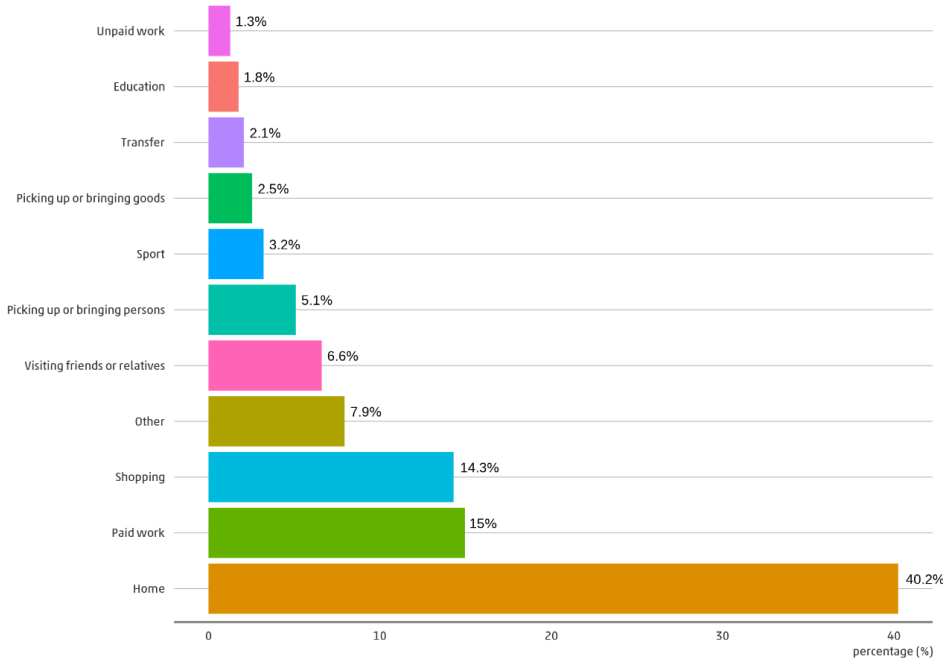
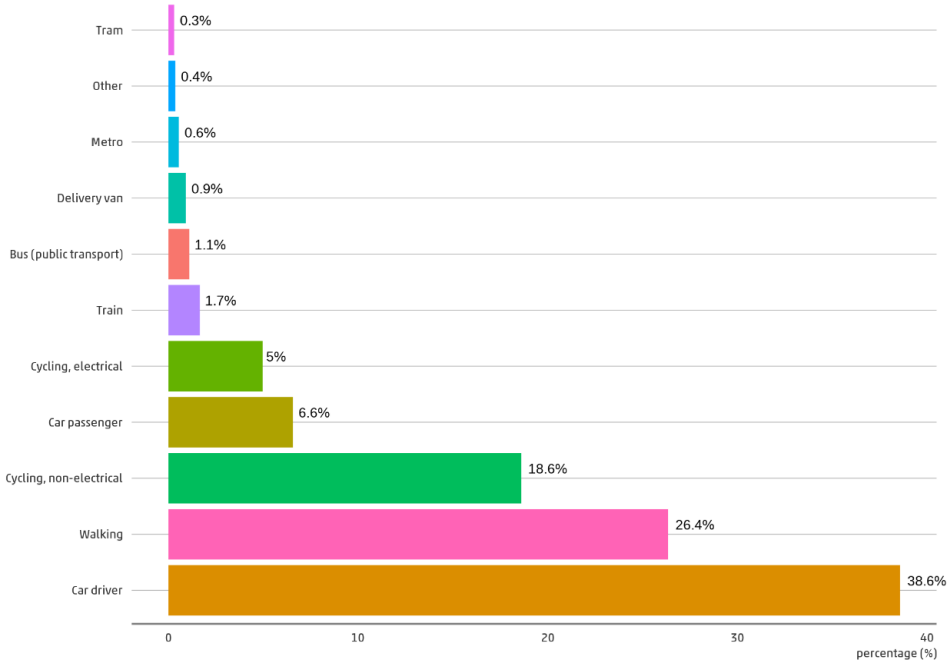


Figure 3.2: Proportions labels transport mode in the MMTS app.



3.4 Validating days and daily questionnaires

The final sub-question concerns the validation of study period calendar days. During the app survey, respondents were asked to check and validate each day of the survey. In addition, they could give comments about their behavior per diary day.

We split the evaluation between the one-day and seven-days samples since the one-day sample was only asked to participate for one day. Regardless of the instructions, participants had the possibility to validate all seven days. Of the one-day sample, 16% validated the requested one day, 39% validated more than one day, and the remaining 45% validated none of the days. Of the seven-days sample, 48% validated all days, 36% validated no days, and the remaining 16% of the respondents validated only a part of the days. Half of those who validated only a part of the week did start to validate but stopped after a while, the other half have validated days arbitrarily distributed over the week.

When validating days, the respondent gets an option to comment on the day they are validating and completing. If the device has not registered any travels that day, the user gets prompted an extra question: whether there are any specific reasons, they have not travelled that day. The respondent may provide a comment on why the day was different or not. Three-quarters of the respondents (77%) did not leave a comment. Of the respondents that did leave a comment, half of them answered the question we asked by commenting on why their day was different or why they did not travel that day. The other half did not answer the question but instead wrote a general statement about their day ('had a nice day') or wrote feedback on the workings of the app or the classification algorithm.

Summarizing the results in section 3, respondents can be roughly split into two groups: a group that is mostly consistent in labeling, validating and commenting and a group that ignores the requests. It is likely that the validation of the day depends on the validity of the stop-track decompositions and the amount of work editing may take to adjust for spurious stops and tracks and/or missing data.

4. To what extent can active editing be related to respondent characteristics?

In this section, we move from aggregate descriptives to individual respondent profiles in in-app editing behaviour. As a proxy for overall in-app activity, we focus on labelling. Labelling of stop purposes and transport modes should be done every

day by every respondent. Furthermore, providing labels is the last step and is preceded by some other form of editing. Any deficiency in checking and adjusting stops and tracks will often lead to problems in labelling as well. We address two subquestions:

- How consistent are respondents in labelling events during the reporting period?
- Can passive in-app behaviour be related to socio-economic characteristics?

4.1 Individual consistency in labelling as a proxy for activity

To explore consistency in behaviour, we first confront the number of labelled days with the number of location tracking days. Table 4.1 displays proportions of labelled days for respondents who had four or five tracking days and respondents who had six or seven tracking days. In the 4-5 days group 35% labelled all days and 20% missed just one day. In the 6-7 days group, the percentages are higher, they are 60% for all days and 18% for all days but one. We conclude that a large proportion of respondents labels all days or almost all days. Also there is a group of around 10% that never labels. Although numbers of respondents are small, this finding points at consistency in behaviour.

Table 4.1: Distribution of labelled days for respondents with four or five tracking and respondent with six or seven tracking days.

	No days labelled	All days labelled	One day not labelled	Other	Number of respondents
4 or 5 days	10%	35%	20%	30%	62
6 or 7 days	10%	60%	18%	13%	200

Next, we evaluate labelling as a function of editing actions. In Table 4.2, we confront labelling with respondent-made stops and tracks. For respondents who were asked to participate for one day the proportions labelled were 55% and 92% for no added events and at least one added event, respectively. For the seven-day group, the two proportions were 45% and 88%. We conclude that respondents who added at least one event also have a much higher proportion of labelled events. This finding confirms the general view that in-app activity clusters within respondents.

As a last step, we compared labelling to the accordance between online diary and unedited in-app diary. This analysis is restricted to the follow-up sample in the MMTS experiment. This sample was selected from former Travel Survey respondents and they were invited to use the app for a week and fill in the regular online diary for one day during this week. We followed the approach by Klingwort, Gootzen, Remmerswaal and Schouten (2024) in estimating the balanced prediction accuracy at the minute-level. The balanced accuracy is larger when app and online diary are better aligned, i.e. more minutes get the same stop-track assignment. Only 47 respondents were available for this evaluation, 36 of which always labelled. The mean balanced accuracy for those that always labelled was 0.87. For those respondents who missed labels, the balanced accuracy was 0.89. Numbers

of respondents are too small to draw strong conclusion. The results do tell us that larger deviations between app and online diary do not coincide with strong drops in labelling.

Table 4.2: Mean number of diary days with geolocations, mean number of diary days with labels, proportion of diary days which are labeled, and mean number of total hours with geolocations grouped by yes/no respondent-made stops or tracks and study duration (1 vs. 7 days).

Respondent-made stops or tracks?	Nr of days	Nr of resp's	Mean days (SE)	Mean Labeled days (SE)	Labeled days/diary days	Mean total hours with geolocations
No	1	43	2.9 (0.06)	1.6 (0.05)	0.55	33
	7	34	3.2 (0.08)	1.4 (0.07)	0.45	31
	All	77	2.8 (0.03)	1.5 (0.03)	0.52	32
Yes	1	17	4.7 (0.16)	4.3 (0.16)	0.92	60
	7	41	6.8 (0.04)	6.0 (0.05)	0.88	81
	All	58	5.5 (0.04)	4.9 (0.04)	0.90	75

4.2 Background characteristics and in-app behaviour

Given the relatively small number of respondents, we consider only three background characteristics: age (<25 years, 25 to 44 years, 45 to 65 years and >65 years), type of registered income (retired, employee/self-employed, other) and registered household income in quintiles. We must stress that the statistical power is insufficient to detect relatively small differences between subgroups. The lack of findings at common significance levels must, thus, not be interpreted as absence of interactions.

We compare the different categories of the three variables on the following types of behaviour:

- Total number of days with a deleted stop or track
- Total number of days with one or more labelled stops and tracks
- Total number of days with an added stop or track
- Total number of days that were validated and confirmed

As a benchmark for the four types of behaviour we evaluate the four types relative to the total number of days with location tracking data.

In Figures A.1 to A.3 in Appendix A, we display boxplots for the four types of behavior and the total number of location tracking days for age, type of income

and household income, respectively. In all cases, the differences between categories are too modest to be statistically significant.

5. How does length of reporting period affect active participation?

The last research question addresses the length of the reporting period. We have already seen in Section 3 that the one-day sample and the seven-day sample differ only mildly in what respondents do on the first diary day. We now investigate how the seven-day sample behaves as function of time in-study. In other words:

- Does average in-app time decrease during the reporting period?
- Do the editing actions change/decrease during the reporting period?

Again, we first look at in-app time for the follow-up sessions. In Table 5.1, we compare the average durations and numbers of session. The restriction is to respondents who completed all days, i.e. also the first full day is based only on respondents who did not drop out. The differences are relatively small, implying that respondent activity remained stable during the week.

Furthermore, we present the average numbers of events in Table 5.1. The number of events tend to get smaller averaging over all days, but standard errors are large and results only hint at a relation. If such an effect is really present, it may be the result of differences between week days and weekend days. Weekly averages, per definition, include all days of the week. More study is needed to see whether this difference is a true finding that results from such within-week variation.

Table 5.1: Average number of sessions and in-app time per condition for first full day and all days. Standard errors between brackets.

	First full day of 7-day		Mean day of 7-day	
	Full	Limited	Full	Limited
Time in-app (in sec)	70 (48)	39 (27)	59 (51)	41 (25)
# In-app sessions	4.7 (2.7)	4.3 (2.3)	3.9 (2.6)	3.8 (2.2)
# hours location data	16.1 (7.5)	16.8 (6.6)	13.0 (8.0)	13.0 (8.0)
# tracks	3.6 (2.8)	5.1 (4.3)	3.0 (2.8)	4.5 (5.2)
# stops	3.3 (2.7)	5.1 (4.2)	2.8 (2.8)	4.3 (5.5)

Table 5.2 takes the respondent viewpoint and displays the percentage of respondents performing each type of action as a function of day in field. Here, the limited editing sample is omitted. We note that only labeling was a necessary step

for all respondents. Editing was obviously only needed when the presented stops and tracks were perceived as incorrect. Contrary to Table 5.1, all respondents are included;, including those that dropped out. The proportions between the first day and later days differ substantially. We conclude that respondents who drop-out were less active than those that stayed. This could have multiple reasons (more technical issues, less able, less motivated) that cannot easily be disentangled given the low number of respondents.

Summarizing, we conclude that respondents who stay in the study use all edit actions and they remain to do so throughout the week.

Table 5.2: Proportions of respondents that performed each type of action on different days of field work.

Indicators (%)	Day 1	Day 2	Day 6	Day 7
Deleted >0 stop-tracks	80%	92%	96%	89%
Labeled stop-track	59%	77%	78%	72%
Modified time stop-tracks	31%	51%	56%	53%
Added >0 stop-tracks	26%	30%	35%	31%
Number of active persons	332	260	171	154

6. Discussion

In this discussion paper, we attempted to measure, understand and disentangle respondent in-app answering behaviour. Empirical knowledge on how motivated respondents are to do smart surveys and how competent they are to perform them accurately are imperative to the user interface design and to post-survey editing and adjustment procedures. Understanding behaviour can be done either through respondent evaluation or through paradata on in-app navigation and data entry. Given that respondent evaluation surveys suffer from nonresponse and detailed questions on behaviour will face recall errors, paradata seem to be a useful avenue. However, paradata analysis turned out more complicated than anticipated, because the app used for data collection suffered from technical issues and the number of respondents was relatively low. The response rate for the study was much lower than expected and, consequently, also the minimal observable differences. The results of the study are subject to sampling variance and caution was needed in drawing conclusions.

Some conclusions did, however, stand out quite clearly. One overall impression is that respondents employ all potential editing options they have available; they add and delete events, they change begin and end times, and they label events. A second overall impression is that once they do, they tend to do so throughout the entire study period. In addition, respondents that are relatively inactive, are so for the entire study period. Hence, respondents seem consistent and relatively early

on in data collection it can be determined whether a respondent will be (very) active or not.

The 2022-2023 MMTS study had a number of limitations that likely have affected the outcomes and that need to be resolved in future studies.
One limitation was that no user test was performed before the field test. The app UI-UX more or less was fielded based purely on experiences in prior smart survey experiments. This holds for the entire workflow and respondent journey. A future field test must be preceded by a user test of the full UI-UX starting from recruitment materials.
Another limitation was the measurement quality of the location data. Even when omitting some problematic brands and models, the drop-out was still sizeable and larger than in 2018. One cause may be that the MMTS app employed three parallel sensor measurements simultaneously and was, consequently, relatively heavy on the battery. The high battery load likely has lead to more missing data because mobile device batteries were depleted. A future test must have a smaller battery load.
In this respect, we point at other (discussion) papers linked to the MMTS study. These evaluate location tracking data quality, stop-track decision rules and respondent involvement, and further optimization of the corresponding methodology.

More study into the active-passive trade-off is definitely needed. One important research question is to what extent motivation and ability depend on technical issues that respondents may encounter. Another follow-up research question is the explanation of respondent behaviour through socio-demographic characteristics.

References

Bantis, T., Haworth, J. (2017), Who you are is how you travel: A framework for transportation mode detection using individual and environmental characteristics, Transportation Research Part C: Emerging Technologies 80, 286–309

Chen C, Ma J, Susilo Y, et al (2016) The promises of big data and small data for travel behavior (aka human mobility) analysis. Transportation Research Part C: Emerging Technologies 68:285{299. https://doi.org/10.1016/j.trc.2016.04.005.

Giacobbe, G. (2024), Unveiling User Dynamics. Examining User-Initiated Changes and Socio-Demographic Influences in app-based Travel Survey Data, EMOS Research Master Thesis, Leiden University, The Netherlands.

Gillis, D., A.J. Lopez, and S. Gautama (2023). “An Evaluation of Smartphone Tracking forTravel Behavior Studies”. In: ISPRS International Journal of Geo-Information 12.8, p. 335.

Gootzen, Y., Klingwort, J., Schouten, B. (2024), Data quality aspects, location tracking for smart travel and mobility surveys, CBS discussion paper, under review.

Harding, C., Faghih Imani, A., Srikukenthiran, S., Miller, E. J., & Nurul Habib, K. (2021). Are we there yet? Assessing smartphone apps as full-fledged tools for activity-travel surveys. Transportation, 48(5), 2433–2460. https://doi.org/10.1007/s11116-020-10135-7

Hu, M., Gremel, G. W., Kirlin, J. A., & West, B. T. (2017). Nonresponse and Underreporting Errors Increase over the Data Collection Week Based on Paradata from the National Household Food Acquisition and Purchase Survey. The Journal of Nutrition, 147(5), 964–975. https://doi.org/10.3945/jn.116.240697

Klingwort, J., Gootzen, Y., Remmerswaal, D., Schouten, B. (2024), Algorithms versus survey response. Validating a smart survey travel and mobility app, under review

Krause, C.M., Zhang, L. (2019), Short-term travel behavior prediction with GPS, land use, and point of interest data, Transportation Research Part B 123 (2019) 349–361.

Lawson, C.T., Krans, E., Rentz, E., Lynch, J. (2023), Emerging trends in household travel survey programs, Social Sciences and Humanities Open, 7, 100466.

Lunardelli, I., Heuvel, J. van den, Schouten, B., Nuccitelli, A., D'Amen, B., Lorè, B., Perez, M., Zgonec, M. (2024), Smart Perceptions Survey, Eurostat project Smart Survey Implementation Deliverable 1.2, available at https://cros.ec.europa.eu/book-page/coordination-and-integration-smart-baseline-stage-12

McCool, D., Lugtig, P., Schouten, B., Mussmann, O. (2021), Longitudinal smartphone data for general population mobility studies, Journal of Official Statistics, 37 (1), 149 – 170.

McCool, D.M., Lugtig, P., Schouten, B. (2022), Maximum interpolable gap length in missing smartphone-based GPS mobility data, Transportation, 51 (1), 297 – 327.

Remmerswaal, D., Lugtig, P., Schouten, B., Struminskaya, B. (2024), The effects of study duration on nonresponse and measurement quality in a smartphone app-based travel diary, under review.

Schmidt, T. (2014). Consumers' Recording Behaviour in Payment Diaries – Empirical Evidence from Germany. Survey Methods: Insights from the Field. https://doi.org/10.13094/SMIF-2014-00008

Schouten, B., Remmerswaal, D., Elevelt, A., De Groot, J., Klingwort, J., Schijvenaars, T., Schulte, M., Vollebregt, M. (2024), A smart Travel Survey Results of a push-to-smart field experiment in the Netherlands, CBS Discussion Paper, Statistics Netherlands.

Smeets, L., Lugtig, P., Schouten, B. (2019), Automatic Travel Mode Prediction in a National Travel Survey, CBS Discussion Paper, December 2019, available at

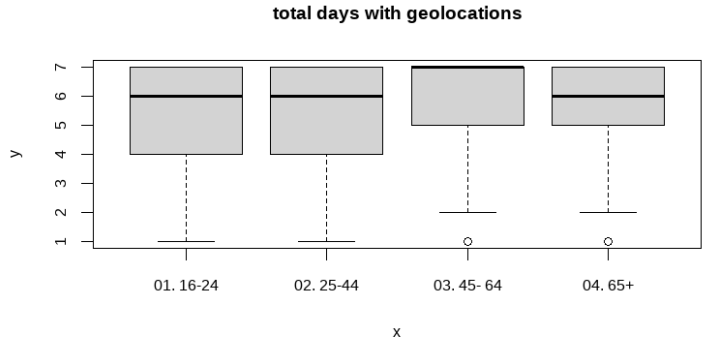
https://www.cbs.nl/-/media/_pdf/2019/51/dp-smeets-lugtig-schouten---vervoermiddelpredictie.pdf

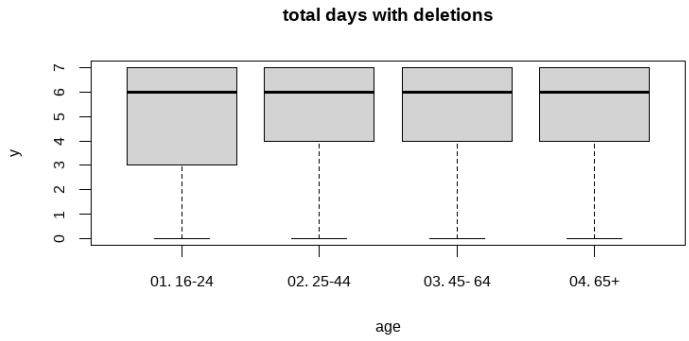
Wenz, A., Jäckle, A., Couper, M.P. (2019), Willingness to use mobile technologies for data collection in a probability-based household panel , Survey Research Methods, 13 (1), 1 – 22.

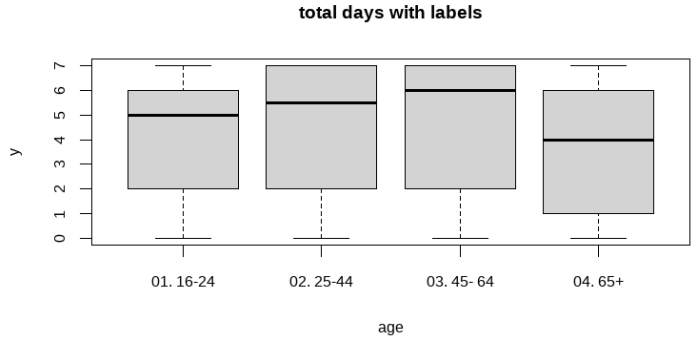
Zhou, Y., Zhang, Y., Yuan, Q., Yang, C., Guo, T., Wang, Y. (2022), The Smartphone-Based Person Travel Survey System: Data Collection, Trip Extraction, and Travel Mode Detection, IEEE Transactions On Intelligent Transportation Systems, 23 (12).

Appendix A - Types of in-app behaviour against socio-economic background

Figure A.1: Types of in-app behavior against age categories.







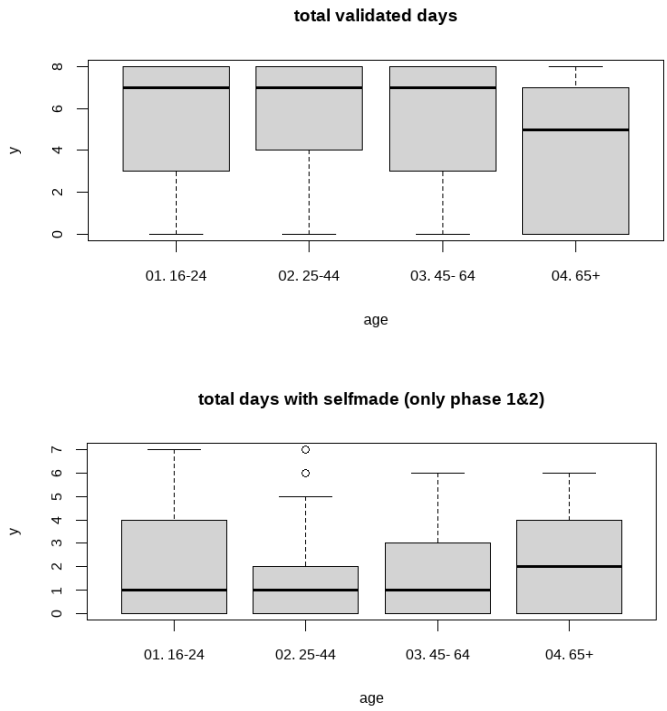
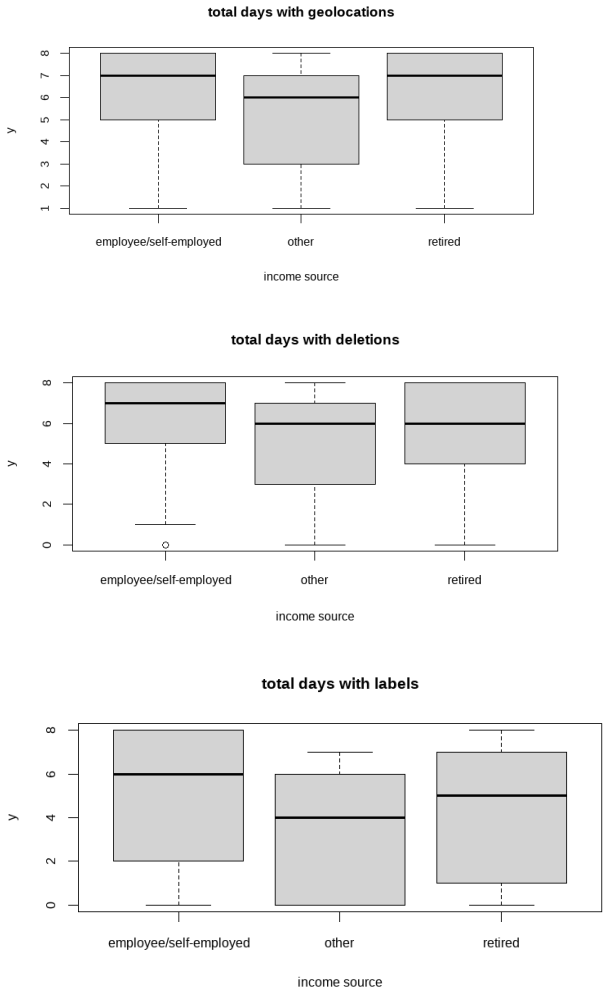


Figure A.2: Types of in-app behavior against type of registered income categories



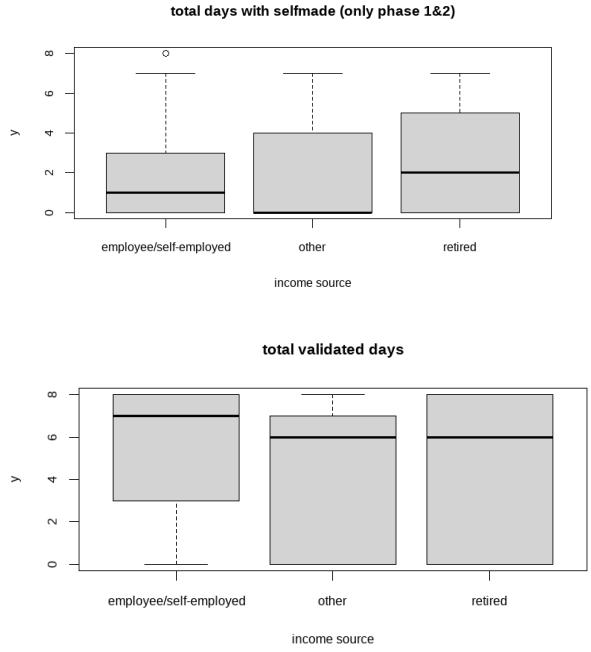
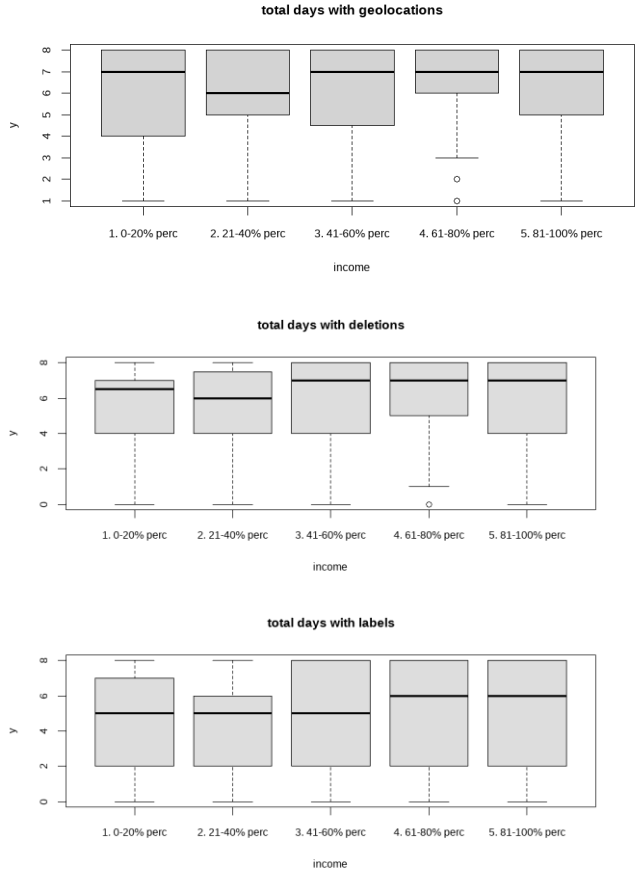
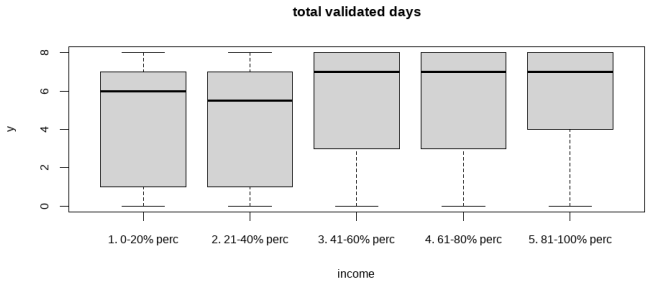
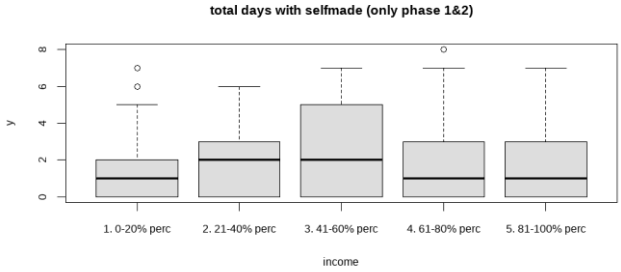


Figure A.3: Types of in-app behavior against household income quintiles.




total validated days
y
0
2
4
6
8
1. 0-20% perc
2. 21-40% perc
3. 41-60% perc
4. 61-80% perc
5. 81-100% perc
income


total days with selfmade (only phase 1&2)
y
0
2
4
6
8
1. 0-20% perc
2. 21-40% perc
3. 41-60% perc
4. 61-80% perc
5. 81-100% perc
income

Colophon

Publisher
Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Prepress
Statistics Netherlands, CCN Creation and visualisation

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contactform: www.cbs.nl/information