Survey Research Methods (2022)
Vol. 16, No. 1, pp. 13-27
doi:10.18148/srm/2022.v16i1.7835
European Survey Research Association

Nonresponse analysis in a longitudinal smartphone-based travel study

Peter Lugtig
Utrecht University
Netherlands

Katie Roth
Statistics Netherlands
The Hague, Netherlands

Barry Schouten
Statistics Netherlands, The Hague and
Utrecht University, Netherlands

Currently, travel diary surveys are the standard method for measuring mobility in official statistics. Nonresponse and measurement are problematic in travel surveys, due to the fact that respondents have to recall all their travels over the course of one or multiple days and have to derive distances for all these travels. To overcome these issues, new methods that rely on passive tracking of locations over time have emerged. The aim of this paper is to assess nonresponse in an experimental travel study carried out in the Netherlands. A smartphone application was developed that passively collects GPS coordinates and automatically populates a travel diary. Participants are then asked for additional information in the diary, such as travel mode. In the experiment, respondents from a random sample of the Dutch population participated in a 7-day study that varied how respondents were recruited into the study, as well as the level and timing of a monetary incentive. We study at what stage of the study respondents choose to participate and dropout, and study nonresponse bias across 13 variables from the Dutch population register. We find that respondents receiving lower incentives, respondents of higher age and respondents with lower levels of education are strongly underrepresented. The overall representativity of the study, as expressed in R-indicators and Coefficients of Variation are rather low because of this nonresponse. We found a similar bias in nonresponse for age going in opposite direction when we computed R-indicators for an earlier web-based travel-diary study. This implies that in the future, diary studies should focus on methods to successfully combine smartphone apps and diaries through the web or on paper if the goal is to limit nonresponse successfully.

Keywords: nonresponse; official statistics; mobile phone application; GPS tracking; R-indicator

1 Introduction

Measuring mobility is a challenging task. Traditionally, it is done by using travel diary surveys, which can be conducted face-to-face, via the phone, or on paper. Participants in these surveys are usually asked to report the start and end location, mode of transport, and provide extra information such as kilometers travelled for all their trips over a period of time.

Measurement is problematic in travel surveys, due to noncentrality of the requested information, and the fact that respondents rely on recall to reconstruct their day (Montini, Prost, Schrammel, Rieser-Schüssler, & Axhausen, 2015). Moreover, trip under-reporting is an issue; participants do not mention certain places they have visited or do not mention details about a certain trip (e.g. mode of transport) (Wolf, Oliveira, & Thompson, 2003). In order to reduce respondent burden, most travel surveys therefore only ask for one day of travel information (Statistics Netherlands, 2017). However, this severely limits the amount of data available for computing travel statistics and necessitates inviting large samples of respondents, in turn increasing costs.

Nonresponse is also an issue. In the Dutch National Travel Survey in 2017 for example, only 53,8% of the participants from the sample responded (Centraal Bureau Voor De Statistiek (CBS) & Rijkswaterstaat (RWS), 2018). For these reasons, a large range of studies has tried new data collection methods to improve trip reporting and response rates. Since most people have a mobile phone which can track geolocations continuously (Statistics Netherlands, 2019), smartphones apps offer the potential to track locations continuously with more precision and over a longer period than current travel diaries. Apps are already being developed with

Contact information: Peter Lugtig, Department of Methods and Statistics, Padualaan 14, 3584CH, Utrecht, The Netherlands (E-mail: p.lugtig@uu.nl)

the goal of improving the quality of diary studies (Elevelt, Lugtig, & Toepoel, 2019), to track behavior of people over time (Sugie, 2018), or in order to better stay in touch with sample members (Keusch, Leonard, Sajons, & Steiner, 2019; Luxton, McCann, Bush, Mishkind, & Reger, 2011).

Despite some obvious advantages that smartphone apps offer for measurement of travel behavior a big question is whether it is possible to use smartphone apps in a study among the general population. There are good reasons to believe that many respondents are not able or willing to participate in smartphone studies, and that nonresponse bias is a large.

The aim of this paper is to assess nonresponse rates and bias in a smartphone travel survey carried out by Statistics Netherlands in 2018. Within a large-scale pilot of a smartphone-app that aimed to measure travel behavior for a week, we used experiments on recruiting methods, and incentives to test which strategy works best to achieve good response rates, and to limit nonresponse bias. Because the goal of this study is to produce official statistics on travel behavior, it is important to understand potential selection biases. Variables like age, income and level of education correlate strongly with the choice of transport mode. Paper or web-based travel diary studies suffer from strong selection biases as well (eg. Forsman, Gustafsson, & Vadeby, 2007; Richardson, Ampt, & Meyburg, 1996), and even when smartphone-based travel studies do suffer from large rates of nonresponse, the question is whether nonresponse bias is higher or lower as compared to a more traditional diary survey. In this paper, we will therefore first study selection bias in the smartphone-based travel study in some detail, and then compare the representativity of this study to the representativity of a traditional paper-based diary, also carried out by Statistics Netherlands in 2016.

In the next section we will discuss recent developments in travel surveys. Then, we will describe the study set up and explain our statistical analyses. Following this we will present our findings. Finally, we will discuss limitations and future possibilities.

2 Background

Traditionally, travel behavior is measured via paper diary surveys. Travel studies have changed a lot in recent years. Several new methods exist: web surveys, GPS tracker studies, and smartphone apps. These new methods and their challenges will be discussed below. For a comprehensive literature reviews on developments and challenges in travel surveys, see Gadziński (2018) or Yue, Lan, Yeh, and Li (2014).

2.1 From paper-diaries to smartphone-based methods in travel studies

With the advent of web surveys in the 2000s, many (travel) diary studies, that were traditionally conducted using paper-and-pencil diaries moved to web (e.g. Bayart & Bonnell, 2007; Richard & Rabaud, 2018). Web surveys are cheaper and easier to conduct. Apart from effects on costs and logistics, the web can also help respondents to provide better quality data. In travel diaries, additional aids, such as a map can be used to put trips into context. A parallel test of the web- and paper-and-pencil version of a travel diary study showed that trip underreporting was lower and the average number of reported trips per day was higher in the web version (Hoogendoorn-Lanser, Schaap, & OldeKalter, 2015). A disadvantage of web diaries is that they still rely on respondents' recall. For travel diaries, this means that respondents have to recall details of every trip, their start- and endpoint location and times,the transport mode used, and whether the trip was taken alone or together with others. Zijlstra, Wijgergangs, and Hoogendoorn-Lanser (2018) discuss that it is generally very difficult for respondents to provide complete and correct information for all trips, and as a consequence, measurement error is considered to be a major problem in travel diary studies.

Instead of asking participants for their location, locations can be passively tracked using GPS tracker devices (e.g. Duncan, Badland, & Mummery, 2009; Shen & Stopher, 2014). In GPS studies a dedicated GPS tracker is provided to each participant. Sometimes the tracker is placed in the participant's car, in other cases the tracker has to be carried around by the participant. Some studies ask for only one day of data, others tested longer periods as well. Sometimes the GPS tracker only supplements an existing survey, while in other cases the trackers were replacing the survey altogether (Shen & Stopher, 2014).

Stopher, Fitzgerald, and Xu (2007) argue that a GPS tracker very well supplements a travel survey. It can be useful for producing correction factors for inaccurately reported travel times or travel distances. The researchers found that participants often misreport these in a survey, but a GPS tracker provided reliable measurements.

However, a study by Bricka, Sen, Paleti, and Bhat (2012) suggest that GPS tracker data are not always superior to the traditional diary. The researchers found that GPS data collection does not work well for the elderly, people who travel for leisure and people who are not familiar with the technology, as these people are less likely to carry the GPS tracker all the time. Such missing trips lead to an underreporting of trips for these specific groups. For younger respondents, and particularly for people who travel a lot the GPS studies were beneficial. Therefore, Bricka et al. (2012) advise that GPS tracking should only complement the travel survey, and not completely replace it.

In the last few years, there is a move away from GPS trackers towards using smartphones. Because so many people now own smartphones and carry those with them most of the time, these devices can potentially be used as a way to replace both the diaries and the GPS trackers. A key role here is played by smartphone apps, which serve as the software applications of smartphones, allowing the user to interact with the phone, and to record data. Many apps are currently being developed and improved (e.g. Geurs, Thomas, Bijlsma, & Douhou, 2015; Nitsche, Widhalm, Breuss, Brändle, & Maurer, 2014; Smoreda, Olteanu-Raimond, & Couronné, 2013; Vlassenroot, Gillis, Bellens, & Gautama, 2015). For a comprehensive overview of previous studies with mobile phones, see Wang, He, and Leung (2018).

Ohmori, Nakazato, and Harata (2005) provided a very early example of a smartphone app study to measure travel behaviour. The authors compared an app to a paper-and-pencil diary survey and found that the respondent burden was lower in the smartphone-app study, because participants could fill in the information in real-time. Data entry was also more frequent in the app than in the paper-and-pencil survey. However, battery consumption was affected by the app and respondent burden in the mobile phone survey depended on activity-travel behaviour and mobile phone literacy.

Due to technical advances, battery consumption is becoming less of an issue more recently. Nahmias-Biran et al. (2018) used a smartphone travel survey and compared its effectiveness to large-scale household surveys in Singapore and Tel-Aviv. The researchers found that the smartphone data had a higher resolution and accuracy of travel duration and start- and stop times. Also, it was able to better represent activity patterns that are often under reported in traditional surveys, such as sub-tracks that are part of a longer trip.

Greene, Flake, Hathaway, and Geilich (2016) reported on a proof-of-concept app that would replace a traditional one-day travel survey with a seven-day smartphone app. The researchers wanted to collect more data, while keeping in mind the respondent burden. Therefore, the app automatically collected GPS data, which was then used to populate a more traditional diary. Participants then would have to answer questions using a ‘prompted recall’ method about those trips

In a similar vein, Greaves et al. (2015) created a smartphone app and provided it as a supplement to the travel survey. The app would not only show the diary, but also show the tracks on a map in order to facilitate the respondent in recalling details about the trip. Respondents were given a choice between an app, and a traditional diary. Half of the participants decided to make use of the app. The study found that the app improved trip reporting, mostly so when participants also used the map function in the app. However, some individuals found the study too burdensome because of all the tasks and were not willing to participate. Also, the authors raised the issue of different mobile phones: it is hard (if not impossible) to create an app that works in the same way on all devices.

2.2 Predictors of nonresponse in smartphone and diary studies

The survey response process can be divided into a series of sequential steps that a researcher and respondent need to take in order to arrive at a complete response. In a traditional survey, contact has to be made first. If a participant cannot be reached, there is an issue of noncontact resulting in unit-nonresponse. Next, participants might refuse to participate. Some participants might not be able to participate (e.g. due to infirmity, language problems), which again will lead to unit nonresponse and might cause bias. Then, of course, participants have to answer all the questions in order to complete the survey (Bethlehem, Cobben, & Schouten, 2011). All these aspects of nonresponse may cause nonresponse bias when those not observed differ from the participants in the study.

When conducting a study in a smartphone app, the respondent has to go through several additional steps in the process of participating in the study. First, participants have to be able to download the app and run it on their phones. Then, they have to register in the app. Most importantly, they must interact with the app for the duration of the study. The researcher has very little control over interaction with the app after sending the invitations. Participants decide when and how much data they want to provide in the app, and often are asked to interact with the app multiple times a day. If respondents experience technical issues, it is hard to help them.

Existing findings from studies that have used smartphone apps in panel studies, have shown that nonresponse bias is a problem. For example, Elevelt et al. (2019) found that age was a strong predictor for participation in a smartphone-base time use diary in the Dutch LISS panel, with older age groups being less likely to participate. In the German IAB-smart panel (Keusch, Struminskaya, Antoun, Couper, & Kreuter, 2019), people with higher education and income levels, and younger age groups were more likely to participate, although differences here were not very large. Wenz, Jackle, and Couper (2019) found that gender, age, education, labour force status and housing tenure are significantly related to not participate in some tasks. Women, highly educated and younger members of the Understanding Society panel study in particular were more likely to participate in a smartphone spending study Jäckle, Burton, Couper, and Lessof (2019). Many studies point to the fact that some demographic variables do not cause the selection bias often present in smartphone-based app studies. Rather, these demographic groups often have better access to smartphones (Keusch, Bähr, Haas, Kreuter, & Trappmann, 2020), use these smartphones more often and are more willing generally to participate in smartphone-based studies (Revilla, Couper, & Ochoa, 2019; Struminskaya et al., 2021).

The relationship between these variables and nonresponse are not causal. Rather, familiarity and acceptance of smartphones as a general tool (Haan, Lugtig, & Toepoel, 2019), and a tool for research in particular (Struminskaya et al., 2021) are correlated to several socio-demographic variables. So are attitudes towards privacy in smartphone studies (Revilla et al., 2019).

Traditional travel diary studies conducted with web- or paper diaries also suffer from nonresponse bias. Younger people, individuals with a low income or lower educational level are less likely to participate (Bricka, 2008). Also, married people or individuals working for the government are less likely to respond (Zegras et al., 2018). In another study, Bricka and Bhat (2006) have found that men, unemployed, young, low educated individuals or individuals who make many trips and travel long distances are more often underreporting their travels. Overall, the picture from both travel studies and smartphone app studies is that there are often substantial biases related to socio-demographic variables.

Although it is likely that some nonresponse bias in smartphone-app travel studies and survey-diary based studies are similar (e.g lower educated are less likely to participate), there is ground to believe that some of the nonresponse predictors are different. Also, it remains a question whether findings about nonresponse bias in smartphone apps from panel studies hold in a cross-sectional setting, and when respondents are recruited from the general population.

In this paper we compare nonresponse bias for a new smartphone app study that automatically tracks people's location to a diary survey on travel behavior conducted a few years earlier. There are two major reasons why we concentrate on nonresponse biases in socio-demographic variables only in this article. First, we have register information for a range of socio-demographic variables, allowing us to assess nonresponse bias in detail. We don't have attitudinal register data for nonrespondents. Second, we can use findings from this study to correct for any nonresponse bias by using correlates of nonresponse in weighting models, or for future rounds of the study, to specifically target underrepresented groups in fieldwork. Weighting is especially important when the correlates of nonresponse are known to also correlate with our dependent variables. Several studies have documented how variables like age, educational level, employment situation and urbanicity relate to the modes of transport that are used, the number of trips and the distance traveled (Collia, Sharp, & Giesbrecht, 2003; Van den Berg, Arentze, & Timmermans, 2011). This makes it all the more important to understand whether these variables are also predictive of nonresponse.

In the remainder of this article we will explain how Statistics Netherlands collected the travel data using a smartphone-app. Then we will show how many people participate at every step of the smartphone study. We investigate nonresponse predictors and finally we explore different participation patterns and investigate if there is a relationship between the patterns and participant characteristics. Our main research question is: what demographic variables predict nonresponse in the app? A secondary question is whether the nonresponse biases we observe is worse than for a comparable paper-based travel diary survey.

3 Methods

3.1 Sample

The participants are recruited from two different samples. A distinction was made between previous OViN (Underway in the Netherlands; a mixed-mode diary only study) respondents, who already participated in the online OViN diary survey in September 2018 and newly drawn participants from the population register. 950 people were invited from the OViN survey and 946 people were randomly drawn from the population register using simple random sampling. Having two samples gives us an opportunity to see which method leads to a higher response: recruiting respondents straight into a smartphone app, or recruiting via a survey method that is familiar to most sample members.

Participants were also randomly divided into three incentive groups across both samples. All participants received €5 with their invitation letter, which is an amount that is used more often in experiments at Statistics Netherlands. Apart from the unconditional incentive, we also introduced conditional incentives. Pre-testing the app suggested that €10 or €20 seemed amounts that would potentially convince some respondents to participate. We here decided to also test splitting an incentive in two in order to encourage respondents to at least install the app. This resulted in three experimental groups. Incentive group one (5-5-5) received an extra €5 when they installed the app and again an extra €5 if they left the app active on their phone for seven days. Incentive group two (5-0-10) received €10 in one go, but only after seven days. Incentive group three (5-0-20) received €20 if they left the app run for seven days.

3.2 The Smartphone app

Data collection was done by Statistics Netherlands (CBS), through the "CBS verplaatsingen" app. The app was developed by CBS and is available for Android (5.0+) and iOS (9.0+) devices, covering about 98% of smartphones that were in use. The app was intended to work the same across operating systems, both in the location data it collected and the User Interface. The invitation letters were sent via postal mail on October 31st, 2018. In the letter, participants were asked to go to a specific CBS website. The website provided information about the study and asked participants to install the app on their phone. Then they had to open the app on their mobile phone and register with their personalized login from the letter (see Figure 1). After registration, they received instructions about how to use the app correctly, in the form of short videos in the app, and were asked for explicit consent for location tracking twice (once by the app, once the operating system). Participants were asked to run the app for at least seven days and provide extra information each day. A “do not track” button was provided in the app that would turn location tracking off. Participants who did not register by November 14th, 2018, got a reminder letter. Participants who did not provide seven days of data, got a reminder letter on November 21st, 2018. Data collection finished on December 15th, 2018.

The app automatically collects GPS data on the mobile phone. A combination of Wi-Fi detection and GPS sensors was used to determine the location. Measurements were taken every minute when a phone was stationary (not moving). Whenever location sensors detected that the phone was moving, measurements were taken every second, and it would switch back to 1-minute intervals after being stationary for about 3 minutes. The location measures were then used to populate a diary that splits the day into stops (periods of rest), and tracks (periods of travel) The participant would see the list of stops and tracks, and by clicking on them, would see the trip/track on a map (see Figure 2). Participants were here asked to provide extra information about their trips: the name and the purpose of the stops and the travel modes between the stops. In addition to the diary, there were also daily questions. The first question was an open question, asking if there were any comments for that day, followed by a multiple-choice question asking if the participant travelled differently than usual on that day. Finally, a Yes/No question asking if the participant carried his/her phone during the whole day. Participants were never prompted to annotate the diary. More details about the app itself, and how it worked for respondents can be found in McCool, Lugtig, Mussmann, and Schouten (2021).

Because of the fact that location measurements were taken dynamically, battery use in our app was not problematic. We tracked respondents’ battery levels throughout the study, and found only a handful of respondents in our study to experience an empty battery. An evaluation survey that was conducted among respondents about 2 weeks after the study also found no respondents commenting on battery consumption. Respondents did comment on the fact that the app sometimes crashed, or caused the phone to become ‘slow’. This was mainly caused by our data transfer protocol. Location data were always stored on the phone, and only transferred when a phone connected to WiFi, unless respondents manually sent data. Some respondents in our study connected to WiFi very infrequently, leading to a lot of data being stored within the app, making it slow. Respondents who experienced problems with the app could call or e-mail the helpline of Statistics Netherlands. Over the course of the study, the helpline got about 200 calls about the installation, use of the app, and completion of the study.

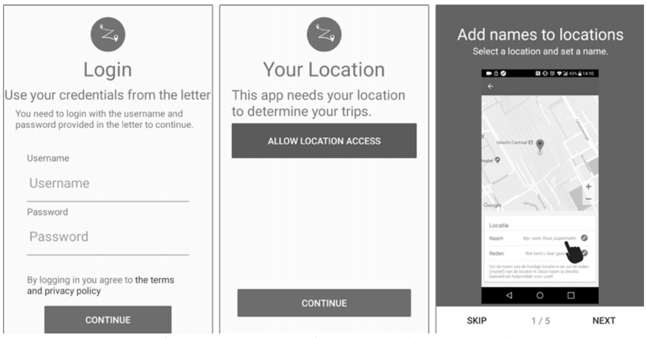

Figure 1. The first screen (left panel) is the login screen that was presented to participants upon installing and opening the app. Their personalized login consisted only of numbers and each participant was provided with a user id and password in their invitation letter. After logging in, respondents were asked to allow location access (middle panel), and were presented a series of 5 short video’s explaining how to interact with the app (right panel)

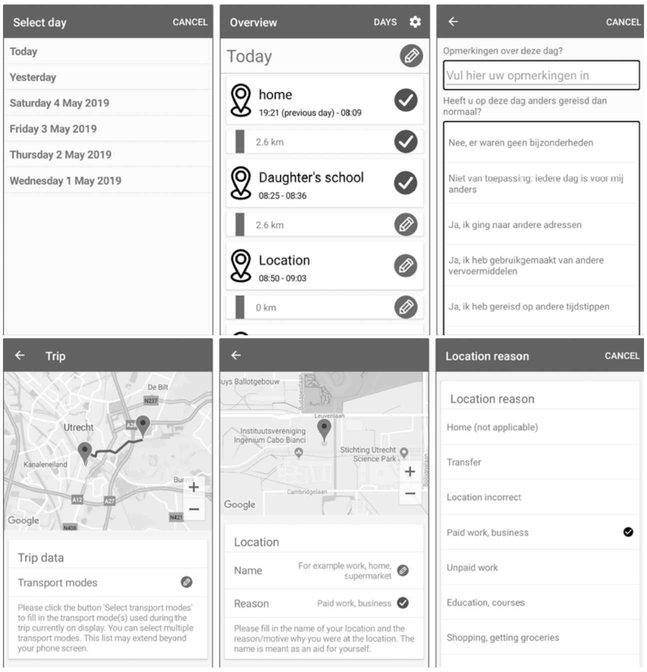

Figure 2. User interface of the app. Participants can (upper left panel) select a day of the week, after which they are presented with an automated travel diary separating the day into stops and trips (upper middle panel). Participants can annotate trips with the transport modes used (lower left panel) and give names to stops (lower middle panel), can explain the reason why stops were made (lower right panel), and finally answer questions about the day (“was it a normal day”) (upper right panel). After annotating a day, stop or trip, the pencil symbol In the overview of the day (upper middle panel) turns into a green tick mark to indicate that the task is complete.

3.3 The comparison study: OViN 2016

In order to put the nonresponse analyses for the smartphone-app study into context, we also compute R-indicators for the 2016 TDS Under way in the Netherlands Study (OViN), conducted annually by the Dutch Ministry of Infrastructure (Centraal Bureau Voor De Statistiek (CBS) & Rijkswaterstaat (RWS), 2018). OViN is designed primarily as a web-based diary study, where respondents keep track of all their trips, including start- and endpoints, and times for a specific day of the week. Participants are sampled from the Dutch population register. The sample is a two-stage stratified cluster sample of individuals within councils. Respondents from rural regions are somewhat oversampled. The invitation for the online survey is comparable to that of the app-based survey, with a mailed invitation letter and two reminder letters.

In order to limit nonresponse in the OvIN study, nonrespondents to the web diary are followed up with one of two modes. Nonrespondents with a registered phone number were contacted by phone and those without registered phone number by face-to-face and then asked again to conduct the survey. Because both the smartphone app study and OvIN are sampled from the same population register, we can use the same variables to compare the two studies in terms of nonresponse. In order to limit nonresponse in the OvIN study, nonrespondents to the web diary are followed up with one of two modes. Nonrespondents with a registered phone number were contacted by phone. Those without registered phone number by face-to-face and then asked again to conduct the survey. Because both the smartphone app study and OvIN are sampled from the same population register, we can use the same variables to compare the two studies in terms of nonresponse. Although the design and target variables of the surveys are largely the same, there are some small differences between OViN and our smartphone-app study that are worth pointing out: One main difference to the app-based study is that OViN asked for only one day of travel information, whereas the smartphone study rund for 7 days. Second, the target population for OViN is all ages, whereas the app-based survey was restricted to 16+ years. Third, the fieldwork year of OviN is different (2016 vs. 2018). Finally, fieldwork in OViN is conducted year-round, with the sample split into 52 weekly batches, whereas invitations for our smartphone study were all sent out at the same date (October 31st). Although these differences mean that nonresponse bias across the two studies may differ somewhat (e.g. because of year or month-specific nonresponse), the goal of the current study is mainly to put the findings from our smartphone study in context.

3.4 Statistical analyses

Key stages of nonresponse. The goal of this paper is to analyze nonresponse in the study. Nonresponse can occur at different stages of the study. Therefore, three key stages were defined that participants could reach/complete. These stages focus on registration, activation and completion.

Stage 1 indicates whether a participant downloaded and registered in the app using their credentials sent by mail. Registration is measured by checking if there was data delivered at least once. This stage is important, because in this stage we can investigate why some participants register and others do not.

Stage 2 indicates whether a participant interacted with the app after registration. This stage is defined as providing at least 1 day of data. This means that these participants delivered some data on the first day. Here, we are interested in why participants would discontinue to use the app after the first try.

We expect that participants who did not make it to stage 1 but dropping out at stage 2 dropped out due to technical issues or due to personal issues with the app, such as not understanding it or simply not liking it. Also, participants could choose not to share their GPS data. Thus, in stage 2 we look at the effect of the app itself.

Stage 3 indicates that a participant completed the study. In practice we find that in between activating the app and completing the study, respondents provide very different patterns of interaction with the app. For some respondents we find that data are collected continuously, but more respondents have missing location data due to switching their phone off, empty batteries or signal loss.

In order to analyze nonresponse in Stage 1, stage 2 and stage 3, we ran logistic regressions with stage completion (yes/no) as outcome and various background variables as predictors (see Table 1 for all variables). Other methods might have been also interesting to investigate, such as a survival analysis to see how far participants make it. However, we decided not to focus on those methods due to the complicated nature of dropout patterns and missing data. In this paper we concentrate on analyzing the potential biases that are introduced at the three major milestones of this study.

The variables we used as predictors for nonresponse were all taken from the Dutch population registry. We selected the variables to be able to compare them to predictors found in traditional travel surveys and other mobile phone apps, and focus our analyses on main effects of these predictors. We initially also included models with interaction effects, but found that the models without interaction effects fitted the data better (according to values of AIC). Apart from log-odds effects, we also report the Average Marginal Effects (AME) of the coefficients for easier interpretation and comparisons and focus our interpretation of the results on these. The AMEs were calculated with the "margins" package in R (Leeper, Arnold, & Arel-Bundock, 2018).

R-indicators. After computing logistic regression results, we further study how the variables that contribute towards nonresponse at different stages of the study, contribute to the overall representativity of the study. For this, we use so-called R-indicators, that are included in data collection monitoring dashboards at Statistics Netherlands. First, we monitor representativeness within the different stages of the study itself, i.e. from sample to registration to activity to submission of a complete diary. Second, we compare the representativeness of the smartphone-app study to the regular travel survey to evaluate whether the new study suffers from more or less nonresponse bias problems, and whether the nature of bias is different.

Representativeness indicators are introduced and discussed in Schouten, Cobben, Bethlehem, et al. (2009), Shlomo, Skinner, and Schouten (2012) and Schouten and Shlomo (2017). The indicators are functions of the sample variances of estimated response propensities from the logistic regression models. The more variance is found, the less representative the response is. R-indicators are standardized between 0 and 1, where values close to 1 indicate that there are no biases on any of the predictor variables. Apart from the overall R-indicators, we also evaluate partial indicators, which break down the overall indicator by predictor to illustrate what particular variable contributes to the (un)representativity of the data. There are two types of metrics that can be used to evaluate the relative variance in propensities in the dataset as whole, and within strata. One is to use R-indicators and partial R-indicators. Another is to use coefficients of variation (CV) and partial coefficients of variation (partial CV). The latter are more natural when the focus is on population means, and this option we choose here. SAS and R code and a manual can be found at http://www.risq-project.eu. All analyses were run in R (version 3.5.2) (R Core Team, 2018). The full code is available in the replication materials and on GitHub[1]

4 Results

In total 1896 sample members were invited for our smartphone-app. From these individuals, 674 participants (35.4%) downloaded and registered in the app. 541 (28.5%) people activated the app by giving permission to send GPS data, and provide at least 1 day of data. Of these, 450 participants (23.7%) completed the survey by providing at least seven days of data. Table 1 shows further sample characteristics, also stratified by recruitment method. There are no significant differences between the samples (not shown in table).

4.1 Registration, activation and completion

Using sample-level variables from the Dutch population register as covariates, we show logistic regression coefficients explaining participation in Stage 1 (registration), Stage 2 (activation) and Stage 3 (completion) in Table 2.

In stage 1 (registration), we find a significant effect of recruitment method (fresh or OViN respondents), incentive condition, age group, level of education, marital status and income. Some of these effects are strong, implying that there are relatively strong selection biases in the recruitment phase of the study. Most importantly we find that older participants are much less likely to register than younger participants. This effect increases as the age goes up. Conditional on other covariates in the model, 26-45 year old sample members are 13 percentage points less likely to register than 16-25 year olds. For 46-65 year old sample members, this percentage is 20 percentage points, whereas for sample members over the age of 65 it is 28 percentage points.

Highly educated participants are about 20 percentage points more likely to register than lower educated participants. Also, the probability of registering in the app is 16 percentage points higher for participants who previously participated in the OViN survey than for newly recruited participants. Furthermore, participants from a non-Western origin are 11 percentage points less likely to register than Dutch participants. The probability of registering is higher for incentive group two (5+0+10) (7 percentage points) and group three (5+0+20) (10 percentage points) than for incentive group one (5+5+5), implying that 1 conditional incentive awarded at the end of study leads to higher registration rates than a conditional incentives split in two, and a higher incentive (20 instead of 10) further leads to a registration rate that is about 3 percentage points higher. Further, there is a significant difference between the first two quantiles of income. Participants from quantile 20-40 are 10 percentage points less likely to register than participants from the lowest quantile (0–20). Finally, single participants are 6 percentage points less likely to register than married participants. We did test whether some of the design features (the sample and incentive groups) interacted with demographic characteristics, but found no interaction effects.

In stage 2 (registration), we find similar effects as in stage 1. We find that 133 of the 674 registered participants fail to activate the app and transmit data. The analyses in Table 2 are all conditional on sample membership, so in order to understand how selection bias may change over the course of the study, the change in coefficients between stages 1 and 2 are informative. We find here that although that are sometimes very small changes in the coefficients, differences are not meaningful. The effects of sample origin become a bit smaller, implying that respondents from the fresh sample are a bit more likely to activate the study conditional on registering the app. Similarly, the AME for level of education become about .05 smaller, meaning that higher educated respondents are less likely to register the app.

In stage 3 (completion), the change in the coefficients remain small in comparison to stage 1, and are negligible as compared to stage 2. This implies that although there is

[1]https://github.com/peterlugtig/Data-archive-SRM-TABI-nonresponse

Table 1
Descriptive statistics taken from register data for the sample, split by recruitment method

Sample	Overall	Split by recruitment method: Fresh	OVin 2018	Population 2018
Sample size (n)	1896	946	950	
Incentive (%)				
5-5-5	33.4	33.5	33.3	-
5-10	33.4	33.4	33.4	-
5-20	33.2	33.1	33.4	-
Age (%)				
16-25	13.9	14.2	13.7	14.6
26-45	30.0	29.1	30.9	29.3
46-65	36.3	35.7	36.8	32.9
>65	19.8	21.0	18.5	23.1
Drivers license = Yes (%)	79.8	77.0	82.6	77.6
Car owner = Yes	46.9	45.1	48.7	71.3
Moped owner = Yes	5.5	5.3	5.7	5.5
HIghest level of education[a]				
Primary	4.2	5.6	2.8	9.9
Lower secondary	11.3	11.6	10.9	21.0
Upper secondary/ vocational	24.7	23.2	26.3	37.3
Bachelor	14.9	13.4	16.3	19.5
Master	7.9	6.4	9.3	10.8
Unknown	37.0	39.7	34.3	1.4
Marital Status (%)				
Married	51.6	50.8	52.4	47.0
Single	48.4	49.2	47.6	
Origin (%)[a]				
Dutch	81.1	79.0	83.3	83.9
Not-western	9.0	10.8	7.2	7.6
Western	9.9	10.3	9.6	8.6
Income percentiles (%)[b]				
0-20	16.5	12.4	20.6	20.0
21-40	21.6	17.1	26.1	20.0
41-60	19.3	19.9	18.7	20.0
61-80	20.8	23.3	18.4	20.0
Unknown	1.8	2.5	1.1	
Gender (%)				
Male	49.3	47.4	51.3	47.9
Female	50.7	52.6	48.7	52.1
Urbanity (%)[b]				
Very strongly	21.3	21.9	20.6	23.6
Strongly	25.7	25.3	26.1	25.2
Moderate	19.8	19.8	19.8	17.1
Little	17.7	17.0	18.4	17.2
Not	15.6	16.1	15.1	17.0
Home Owner (%)				
Owner	67.6	66.3	68.9	66.2
Rent	29.5	31.1	27.9	31.2
Unknown	2.8	2.5	3.2	2.6

[a] For explanation about the Dutch school system. see https://www.nuffic.nl/en/subjects/education-in-the-netherlands/ [b] For the official cut-offs and calculations. see https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties

dropout over the course of the study (91 participants do not complete the study), this does not affect the selectivity of the study, beyond the bias that was introduced in stage 1 and 2.

A further question we turn to now is what these findings imply for the overall selectivity of the study. How representative is the smartphone-based travel app?

4.2 Representativeness and nonresponse bias

The smartphone-app. We again consider the three stages of response, i.e. registration of the app, activation and completion of the 7-day diary study. For each of the stages we will evaluate the R-indicator, and Coefficient of Variation (CV) to evaluate the overall representativeness of both the smartphone-app study, and compare our estimates to those to the OViN diary-study. In order to conform to the standard monitoring of OViN at Statistics Netherlands, we use a model with the following seven variables:

- Ethnicity: native, 1st generation non-western, 2nd generation non-western, 1st generation western, 2nd generation western
- AgeClasses: 12-17, 18-24, 25-29, ..., 65-69, 70-74, 75 and older
- IncomeClasses: No registered income, 1st quintile, 2nd quintile, ..., 5th quintile
- Urbanity: Not, little, moderate, strong, very strong
- CarHH: No cars in household, one car in household, two cars in household, three or more cars in household
- Moped: Yes or no owner of a moped
- License: Yes or no having a driver's license

Note that the variable coding differs slightly from the previous analyses: Ethnicity and age classes are more detailed and educational level, gender, home ownership and marital status are omitted. In the case of educational level this is a pity; educational level has only been administered fairly recently in the population register of the Netherlands, and as such is missing for most older respondents or respondents with a migration background. The reason that we here choose to use a different set of characteristics is that we want to be able to compare R-indicators for both the new smartphone-app study, and the OViN in diary study. Because we are not able to retroactively link other register data to 2016 OViN, we therefore adjust our covariates in this section.

Figure 3 presents a response-representativeness plot, or RR-plot, for the three stages. The vertical axis has the R-indicators and the horizontal axis the response rates. The diagonal lines represent constant values of the CV. The vertical bars around the points correspond to 95% confidence intervals. Obviously, the response rate is lowest for a complete diary and highest for registration. The R-indicator decreases slightly, but not significantly, when registered respondents drop-out. In general, representativeness is relatively low.

Next, we consider the variable-level partial CV for the three stages. Figure 4 gives the dashboard bar charts with unconditional and conditional partial CVs for each of the three stages. The unconditional values present the contribution of the variable to response propensity variance. The conditional values give the unique contribution of the variables adjusted for collinearity. Next, we consider the variable-level partial CV for the three stages.

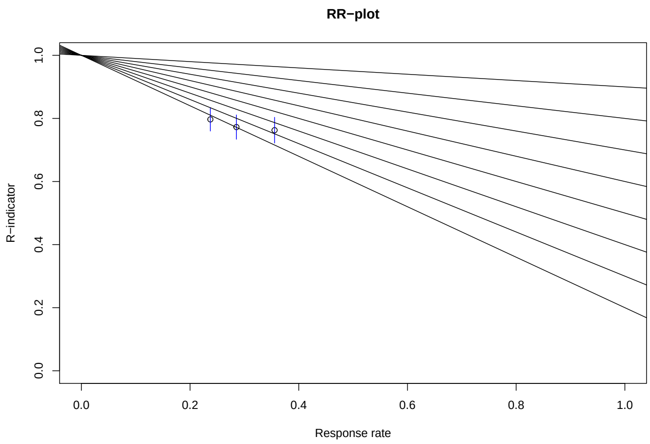

Figure 3. RR-plot for app registration, any app activity and complete app diary.

From Figure 4, we can conclude that representativeness changes only slightly from registration to completion of the study, confirming our earlier findings from Table 2. The strongest contributions come from age and income. The variables that are directly relevant for the travel survey; car and moped ownership and driver's license, give relatively small contributions, especially when adjusted for collinearity. The latter is a positive finding. It should be noted, however, that confidence intervals are still relatively wide for a sample size of 1900 and conclusions need to be drawn with care. We will consider category-level partial CVs in the next section when comparing representativeness to the regular survey.

Comparing variable level bias between app and diary. Response rates in OViN 2016 are 18.3% for the web survey, which then increases to 54.8% after CATI and CAPI fieldwork. We find that the web-response in the survey is thus about 12% points lower than in the smartphone-study, but after other modes are added, the response is much higher. Figure 5 gives the variable-level partial CV for the app-based survey, the online response for OViN 2016 and the total response for OViN 2016 (including telephone and face-to-face). We can make a number of observations: First, the regular online response in the 2016 OViN study shows strong similarity in representativeness on the variable level to the 2018 smartphone-app study. The same variables—age and income—stand out as most influential. Second, the partial CV's have a similar order of magnitude as well for these two sets of responses. Third, the full mixed-mode response is much more representative on all variables. We may conclude from these observations that an online questionnaire or

Table 2
Logistic regression results using registation, activation and completion as outcomes

	Registration (yes=674)			Activation (yes=541)			Completion (yes=450)		
	Coef.	Std. Err.	AME	Coef.	Std. Err.	AME	Coef.	Std. Err.	AME
Intercept	−1.11	0.41	-	−1.10	0.44	-	−2.04	0.51	-
Sample: ref= fresh	0.77*	0.10	0.16	0.66*	0.11	0.12	0.56*	0.12	0.12
Incentive (ref=5+5+5)									
5+10	0.38*	0.13	0.07	0.32*	0.14	0.06	0.31	0.15	0.06
5+20	0.52*	0.13	0.10	0.51*	0.13	0.09	0.55	0.14	0.09
Age (ref=18-25)									
26-45	−0.66*	0.19	−0.14	−0.60*	0.20	−0.12	−0.42*	0.21	−0.12
46-65	−0.94*	0.20	−0.20	−0.98*	0.21	−0.20	−0.89*	0.22	−0.20
>65	−1.39*	0.24	−0.28	−1.41*	0.25	−0.26	−1.35*	0.27	−0.26
Drivers license (ref=no)	0.24	0.17	0.05	0.33	0.18	0.06	0.36	0.19	0.06
Car owner (ref=no)	0.10	0.12	0.02	0.15	0.12	0.03	0.13	0.13	0.02
Moped owner (ref=no)	0.08	0.23	0.02	0.24	0.23	0.05	0.04	0.26	0.01
HIghest level of education (ref=primary school)									
Lower secondary	−0.05	0.33	−0.01	−0.04	0.35	−0.01	0.40	0.43	−0.01
Higher sec./ vocational	0.48	0.31	0.09	0.22	0.34	0.04	0.63	0.41	0.04
Bachelor	0.95*	0.33	0.20	0.77*	0.35	0.15	1.08*	0.42	0.15
Master	0.90*	0.36	0.19	0.80*	0.38	0.15	1.18*	0.45	0.15
Unknown	0.28	0.32	0.05	0.06	0.34	0.01	0.50	0.41	−0.06
Marital Status (ref=married)	−0.31*	0.12	−0.06	−0.35*	0.13	−0.06	−0.28	0.14	−0.06
Origin (ref=Dutch)									
Not-western	−0.60*	0.22	−0.11	−0.80*	0.24	−0.13	−0.89*	0.27	−0.13
Western	−0.25	0.18	−0.05	−0.46*	0.20	−0.08	−0.52*	0.22	−0.08
Income (ref=1-20)									
21-40	−0.53*	0.23	−0.10	−0.40	0.25	−0.06	−0.39	0.26	−0.06
41-60	0.15	0.21	0.03	0.24	0.23	0.04	0.22	0.25	0.04
61-80	0.38	0.21	0.08	0.38	0.23	0.07	0.34	0.24	0.07
81-100	0.35	0.22	0.07	0.27	0.23	0.05	0.27	0.25	0.05
Unknown	0.10	0.58	0.02	−0.39	0.80	−0.06	−0.93	1.07	−0.06
Female (ref=male)	0.15	0.10	0.03	0.17	0.11	0.03	0.34*	0.12	0.03
Urbanity (ref= very strong)									
Strong	−0.02	0.16	−0.01	−0.15	0.17	−0.03	−0.06	0.18	−0.03
Moderate	−0.03	0.17	−0.01	−0.13	0.18	−0.02	−0.09	0.19	−0.02
Little	−0.01	0.17	0.00	−0.16	0.18	−0.03	0.00	0.19	−0.03
Not	−0.31	0.18	−0.06	−0.41	0.19	−0.07	−0.27	0.21	−0.07
Homeownership (ref=owner)									
Rent	0.07	0.14	0.01	−0.09	0.15	−0.02	−0.04	0.16	−0.02
Unknown	−0.33	0.32	−0.06	−0.63	0.34	−0.10	−0.72	0.37	−0.10

Sample size is 1896.
* $p < 0.01$

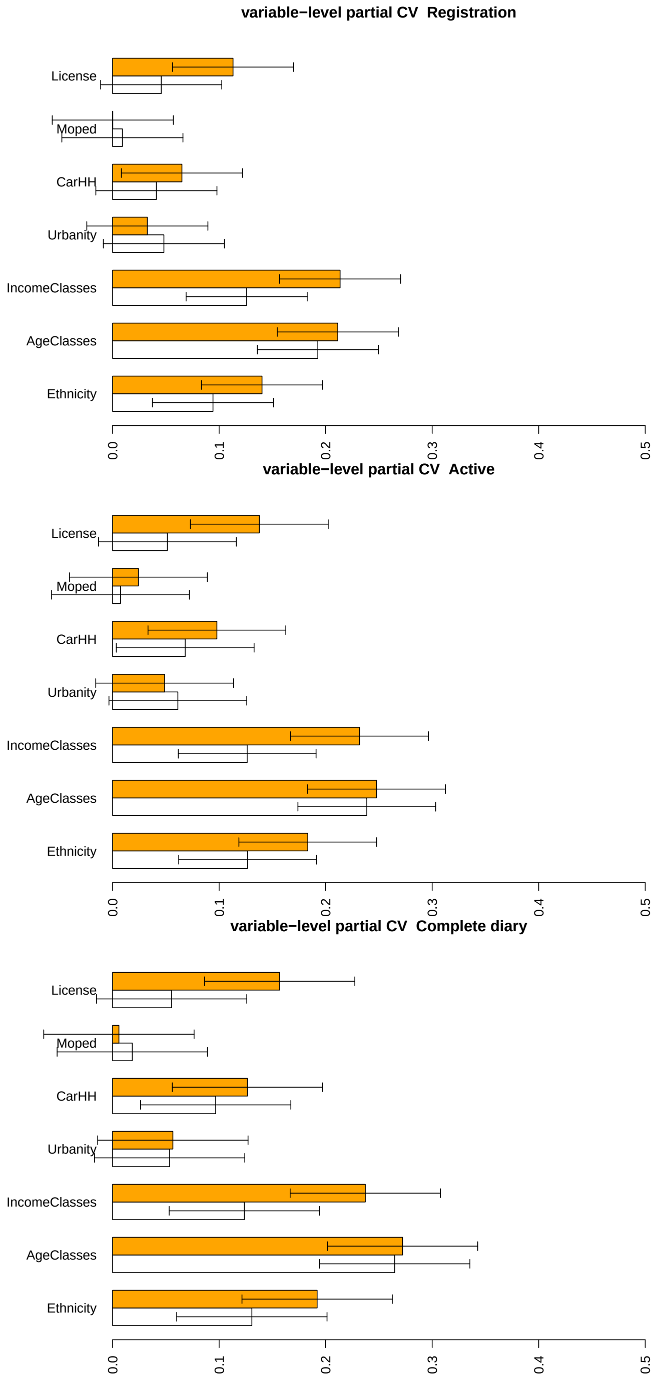

Figure 4. Unconditional and conditional variable-level partial CV for registration, activity and complete diary. Yellow and white bars represent, respectively, unconditional and conditional partial CV. 95% confidence intervals based on normal approximation are included

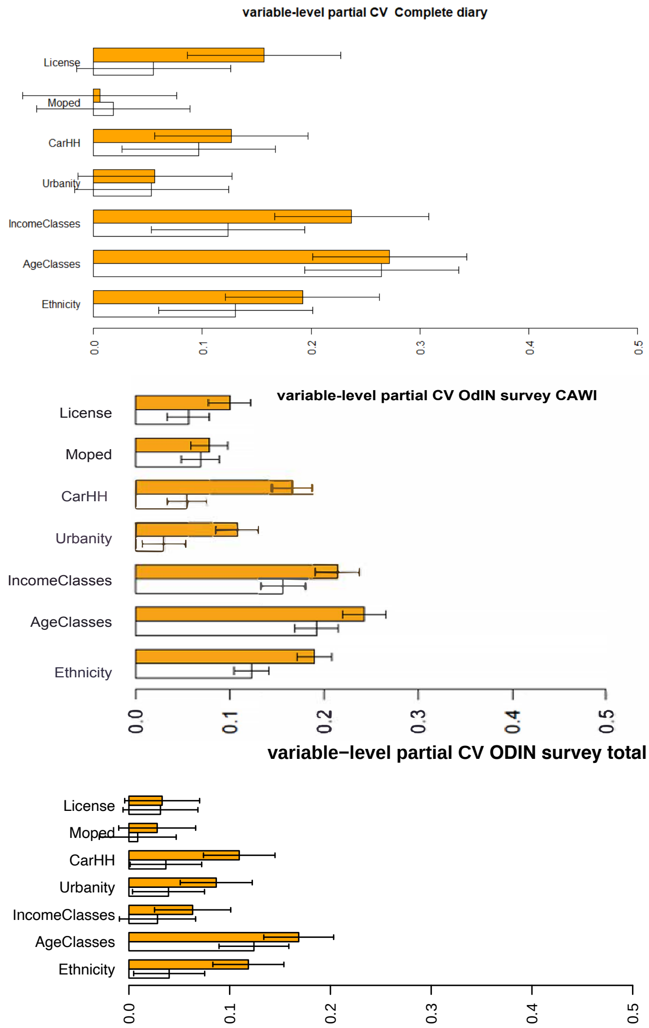

Figure 5. Unconditional and conditional variable-level partial CV for the app-based survey complete diary (upper panel), the online response to the regular survey (middle panel) and the total response to the regular survey (lower panel). Yellow and white bars represent, respectively, unconditional and conditional partial CV. 95% confidence intervals based on normal approximation are included.

an app result in similar representativeness overall. It would however still be possible that the different variables contribute differently towards the R-indicator. To see this, we need to look at the category-level partial CV.

Figure 6 shows the category-level partial CV for the variable age, which contributes most to the CV in both the smartphone-app and OViN. Category-level partial CVs include a positive-negative sign to indicate overrepresentation and underrepresentation, respectively. While the two variables have similar contributions to response propensity variance, the nature of the contributions is almost reverse. Younger age classes are overrepresented in the app-based

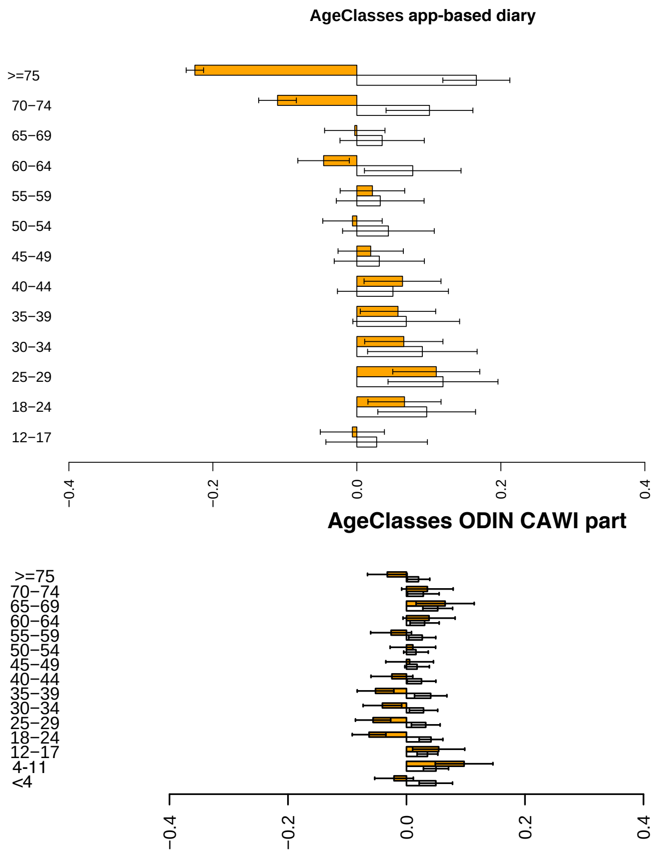

Figure 6. Unconditional and conditional category-level partial CV for the app-based survey complete diary and the online response to the regular survey. Yellow and white bars represent, respectively, unconditional and conditional values. 95% confidence intervals based on normal approximation are included.

survey, whereas older age classes are more present in the regular online survey. This finding may not be surprising, but is very relevant from a survey design point of view. If there are ways in which we can combine both types of data collection, we may be able to reduce nonresponse bias on age.

5 Conclusion and discussion

In this paper we investigated nonresponse in a new travel study conducted using a smartphone app. We analyzed nonresponse patterns in order to understand nonresponse bias in the study. Variables available for all sample members from the Dutch population register allowed us to establish that there are substantial nonresponse biases in the study. Most of these biases are introduced in the recruitment phase of the study. Despite the fact that our recruitment rate was relatively high (30%) we found that strong biases in especially age and level of education exist. Older and lower educated sample members were much less likely to participate. These biases persist throughout the study. After registering for the study, biases did not get worse or better in the next two stages we analyzed (activation and completion). We also found that higher incentives lead to a much better participation rate.

Overall, these biases lead to an R-indicator for the study that is about .70, which is rather low. When we computed the R-indicator using the same register variables for the OViN travel diary study which was the smartphone app study sought to replace and which was conducted through an 'old-fashioned' web diary, we found that the representativity of that diary-survey was similarly low. Interestingly, when partial indicators were evaluated at the variable level, we found that the biases in both designs are in opposite directions: whereas the response rate for people under the age of 25 in the smartphone app study is about 20% points higher than for people over 45, it is about 20% points lower in the earlier diary study. These findings may be a bit particular to the specific smartphone app we used and may vary somewhat with time. They are in line with findings from other studies however. Whereas in survey research there are problems with getting an adequate response rate among younger respondents, smartphone-based studies find it easier to attract young respondents.

Future studies could investigate further how the combination of a web or mixed-mode survey and mobile app works best. The smartphone app can be introduced within a broader mixed-mode fieldwork strategy, although that would lead to challenges in data harmonization across the different modes being used in the study.

We found that age and level of education were the main drivers of nonresponse bias. Because these variables are also related to various substantive outcome measures of the study (the number of trips, modes used and distance traveled), these variables would need to be included in post-survey adjustments for nonresponse. In our case, it would be possible to weight the data in order to correct to some degree for bias due to unit-nonresponse. Other travel studies often do not have the ability to sample from rich register data, and therefore also cannot correct for unit-nonresponse.

Weighting for unit-nonresponse alone does not solve the missing data problem however. For missing data within the study, a more sophisticated method is needed that currently does not exist. Even when respondents complete 7 days of the study, we often find that data are missing for short periods every day due to battery issues, technical problems, or respondents not taking their phone with them. Because of these missing episodes, estimates of for example the distance traveled are severely underestimated if one would use the location data naively. Before one can weight the data to deal with unit-nonresponse, it is necessary to deal with missing sections of trips, or missing hours or days within the app study. Data may be imputed at the level of individual location at particular times in case short sections of trips are missing, but probably have to be imputed at the aggregate track level when parts of a day are missing. Only after these steps can weighting be used to effectively correct for unit-nonresponse. Age and level of education would in such models be important covariates.

Two further findings stand out. The method to recruit respondents directly into an app yielded a response rate that is 16 percentage points lower than the response rate for respondents who earlier participated in OViN. However, one has to remember that OViN itself has a response rate of about 50%, and that conditional on the sample as a whole direct recruitment into the app was more successful. Similarly, we found that higher incentives lead to higher participation rates. However, we also find that the condition where conditional incentives are split across two moments (5+5+5) yielded about 6 percentage points lower response rates than the condition with only one conditional incentive (5+0+10). The total amount of the incentive is the same in these two conditions and it is easier to receive the incentives in the 5+5+5 condition. Yet, respondents prefer the 5+0+10 condition. We have to speculate why this is, but imagine that the simpler incentive structure with a higher conditional incentive appeals more. Although not tested in this study, the combination of an unconditional 5 euro incentive in combination with a conditional incentive seemed to work well. We did not find any interaction effects with the demographic variables we also tested, implying that the incentives we tested work similarly across demographic groups. Future rounds of the app should focus on methods to attract older and especially lower educated respondents. Different incentives structures may help to attract these underrepresented groups.

All in all, this new method to measure mobility is definitely promising. In a smartphone app a diary, GPS tracking and questions can be combined into one single study. The data resulting from this study is far more fine-grained and rich than could ever be achieved with a traditional diary study. We have also seen that it is possible to collect seven days of data, which goes beyond what diary studies normally collect. Nonresponse is however still an issue in doing a smartphone app study among the general population, and perhaps even more so than in traditional surveys or diary studies.

Acknowledgements

We are thankful to Danielle McCool and Laurent Smeets for help in the analyses for this paper. We also want to thank the department of infrastructure from the Ministry of Transport, Public Works and Water Management, and Statistics Netherlands for funding of the data collected for this project.

References

Bayart, C., & Bonnell, P. (2007). Web surveys: An opportunity to reduce non-response in traditional household travel survey. In *11th world conference on transport researchworld conference on transport research society*.

Bethlehem, J., Cobben, F., & Schouten, B. (2011). *Handbook of nonresponse in household surveys*. John Wiley & Sons.

Bricka, S. (2008). Non-response challenges in GPS-based surveys. In *8th international conference on survey methods in transport* (pp. 25–31).

Bricka, S., & Bhat, C. R. (2006). Comparative analysis of global positioning system–based and travel survey-based data. *Transportation Research Record*, *1972*(1), 9–20.

Bricka, S., Sen, S., Paleti, R., & Bhat, C. R. (2012). An analysis of the factors influencing differences in survey-reported and GPS-recorded trips. *Transportation research part C: emerging technologies*, *21*(1), 67–88.

Centraal Bureau Voor De Statistiek (CBS), & Rijkswaterstaat (RWS). (2018). Onderzoek verplaatsingen in nederland 2017—ovin 2017. doi:10.17026/DANS-XXT-9D28

Collia, D. V., Sharp, J., & Giesbrecht, L. (2003). The 2001 national household travel survey: A look into the travel patterns of older Americans. *Journal of safety research*, *34*(4), 461–470.

Duncan, M. J., Badland, H. M., & Mummery, W. K. (2009). Applying GPS to enhance understanding of transport-related physical activity. *Journal of Science and Medicine in Sport*, *12*(5), 549–556.

Elevelt, A., Lugtig, P., & Toepoel, V. (2019). Doing a time use survey on smartphones only: What factors predict nonresponse at different stages of the survey process? In *Survey research methods* (Vol. 13, pp. 195–213).

Forsman, Å., Gustafsson, S., & Vadeby, A. (2007). Impact of nonresponse and weighting in Swedish travel survey. *Transportation research record*, *1993*(1), 80–88.

Gadziński, J. (2018). Perspectives of the use of smartphones in travel behaviour studies: Findings from a literature review and a pilot study. *Transportation Research Part C: Emerging Technologies*, *88*, 74–86.

Geurs, K. T., Thomas, T., Bijlsma, M., & Douhou, S. (2015). Automatic trip and mode detection with move smarter: First results from the Dutch mobile mobility panel. *Transportation research procedia*, *11*, 247–262.

Greaves, S., Ellison, A., Ellison, R., Rance, D., Standen, C., Rissel, C., & Crane, M. (2015). A web-based diary and companion smartphone app for travel/activity surveys. *Transportation Research Procedia*, *11*, 297–310.

Greene, E., Flake, L., Hathaway, K., & Geilich, M. (2016). *A seven-day smartphone-based GPS household travel survey in Indiana*.

Haan, M., Lugtig, P., & Toepoel, V. (2019). Can we predict device use? An investigation into mobile device use in surveys. *International Journal of Social Research Methodology*, *22*(5), 517–531.

Hoogendoorn-Lanser, S., Schaap, N. T., & OldeKalter, M.-J. (2015). The Netherlands mobility panel: An innovative design approach for web-based longitudinal travel data collection. *Transportation Research Procedia*, *11*, 311–329.

Jäckle, A., Burton, J., Couper, M. P., & Lessof, C. (2019). Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: Coverage and participation rates and biases. In *Survey research methods* (Vol. 13, pp. 23–44).

Keusch, F., Bähr, S., Haas, G.-C., Kreuter, F., & Trappmann, M. (2020). Coverage error in data collection combining mobile surveys with passive measurement using apps: Data from a German national survey. *Sociological Methods & Research*, 0049124120914924.

Keusch, F., Leonard, M. M., Sajons, C., & Steiner, S. (2019). Using smartphone technology for research on refugees: Evidence from Germany. *Sociological Methods & Research*, *50*(4), 1863–1894. doi:10.1177/0049124119852377

Keusch, F., Struminskaya, B. [Bella], Antoun, C., Couper, M. P., & Kreuter, F. (2019). Willingness to participate in passive mobile data collection. *Public opinion quarterly*, *83*(S1), 210–235.

Leeper, T. J., Arnold, J., & Arel-Bundock, V. (2018). Margins: Marginal effects for model objects. *R package version 0.3*, *23*, 2018.

Luxton, D. D., McCann, R. A., Bush, N. E., Mishkind, M. C., & Reger, G. M. (2011). Mhealth for mental health: Integrating smartphone technology in behavioral healthcare. *Professional Psychology: Research and Practice*, *42*(6), 505.

McCool, D., Lugtig, P., Mussmann, O., & Schouten, B. (2021). An app-assisted travel survey in official statistics: Possibilities and challenges. *Journal of Official Statistics*, *37*(1), 149–170. doi:10.2478/jos-2021-0007

Montini, L., Prost, S., Schrammel, J., Rieser-Schüssler, N., & Axhausen, K. W. (2015). Comparison of travel diaries generated from smartphone data and dedicated GPS devices. *Transportation Research Procedia*, *11*, 227–241.

Nahmias-Biran, B.-h., Han, Y., Bekhor, S., Zhao, F., Zegras, C., & Ben-Akiva, M. (2018). Enriching activity-based models using smartphone-based travel surveys. *Transportation Research Record*, *2672*(42), 280–291.

Nitsche, P., Widhalm, P., Breuss, S., Brändle, N., & Maurer, P. (2014). Supporting large-scale travel surveys with smartphones—a practical approach. *Transportation Research Part C: Emerging Technologies*, *43*, 212–221.

Ohmori, N., Nakazato, M., & Harata, N. (2005). GPS mobile phone-based activity diary survey. In *Proceedings of the Eastern Asia Society for Transportation Studies* (Vol. 5, pp. 1104–1115).

R Core Team. (2018). R: A language and environment for statistical computing. R foundation for statistical computing, vienna, austria. 2012.

Revilla, M., Couper, M. P., & Ochoa, C. (2019). Willingness of online panelists to perform additional tasks. *methods, data, analyses*, *13*(2), 29.

Richard, O., & Rabaud, M. (2018). French household travel survey: The next generation. *Transportation Research Procedia*, *32*, 383–393.

Richardson, A., Ampt, E., & Meyburg, A. (1996). Nonresponse issues in household travel surveys. In *Conference proceedings* (Vol. 10, pp. 79–114).

Schouten, B., Cobben, F., Bethlehem, J., et al. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, *35*(1), 101–113.

Schouten, B., & Shlomo, N. (2017). Selecting adaptive survey design strata with partial r-indicators. *International Statistical Review*, *85*(1), 143–163.

Shen, L., & Stopher, P. R. (2014). Review of GPS travel survey and GPS data-processing methods. *Transport reviews*, *34*(3), 316–334.

Shlomo, N., Skinner, C., & Schouten, B. (2012). Estimation of an indicator of the representativeness of survey response. *Journal of Statistical Planning and Inference*, *142*(1), 201–211.

Smoreda, Z., Olteanu-Raimond, A.-M., & Couronné, T. (2013). Spatiotemporal data from mobile phones for personal mobility assessment. In *Transport survey methods: Best practice for decision making*. Emerald Group Publishing Limited.

Statistics Netherlands. (2017). Onderzoeksbeschrijving OViN 2017. Retrieved from https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/aanvullende%20onderzoeksbeschrijv%20ovin-2017

Statistics Netherlands. (2019). The Nethererlands on the European scale. Retrieved from https://longreads.cbs.nl/european-scale-2019/internet/

Stopher, P., Fitzgerald, C., & Xu, M. (2007). Assessing the accuracy of the sydney household travel survey with GPS. *Transportation*, *34*(6), 723–741.

Struminskaya, B. [B.], Lugtig, P., Toepoel, V., Schouten, B., Giesen, D., & Dolmans, R. (2021). Sharing data collected with smartphone sensors: Willingness, participation, and nonparticipation bias. *Public Opinion Quarterly*, *85*, 423–462. doi:10.1093/poq/nfab025

Sugie, N. F. (2018). Utilizing smartphones to study disadvantaged and hard-to-reach groups. *Sociological Methods & Research*, *47*(3), 458–491.

Van den Berg, P., Arentze, T., & Timmermans, H. (2011). Estimating social travel demand of senior citizens in the Netherlands. *Journal of Transport Geography*, *19*(2), 323–331.

Vlassenroot, S., Gillis, D., Bellens, R., & Gautama, S. (2015). The use of smartphone applications in the collection of travel behaviour data. *International Journal of Intelligent Transportation Systems Research*, *13*(1), 17–27.

Wang, Z., He, S. Y., & Leung, Y. (2018). Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society*, *11*, 141–155.

Wenz, A., Jackle, A., & Couper, M. P. (2019). Willingness to use mobile technologies for data collection in a probability household panel. In *Survey research methods* (Vol. 13, pp. 1–22). European Survey Research Association.

Wolf, J., Oliveira, M., & Thompson, M. (2003). The impact of trip underreporting on vmt and travel time estimates: Preliminary findings from the california statewide household travel survey GPS study. *Transportation Research Record*, *1854*, 189–198.

Yue, Y., Lan, T., Yeh, A. G., & Li, Q.-Q. (2014). Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. *Travel Behaviour and Society*, *1*(2), 69–78.

Zegras, P. C., Li, M., Kilic, T., Lozano-Gracia, N., Ghorpade, A., Tiberti, M., … Zhao, F. (2018). Assessing the representativeness of a smartphone-based household travel survey in Dar es Salaam, Tanzania. *Transportation*, *45*(2), 335–363.

Zijlstra, T., Wijgergangs, K., & Hoogendoorn-Lanser, S. (2018). Traditional and mobile devices in computer assisted web-interviews. *Transportation Research Procedia*, *32*, 184–194.