

An App-Assisted Travel Survey in Official Statistics: Possibilities and Challenges

Danielle McCool¹, Peter Lugtig¹, Ole Mussmann², and Barry Schouten³

Advances in smartphone technology have allowed for individuals to have access to near-continuous location tracking at a very precise level. As the backbone of mobility research, the Travel Diary Study, has continued to offer decreasing response rates over the years, researchers are looking to these mobile devices to bridge the gap between self-report recall studies and a person's underlying travel behavior. This article details an open-source application that collects real-time location data which respondents may then annotate to provide a detailed travel diary. Results of the field test involving 674 participants are discussed, including technical performance, data quality and response rate.

Key words: Non-response; travel diary; sensor data for surveys; app design; android background restriction.

1. Introduction

Understanding the true underlying movement behavior of persons in a given geographic area is a key component in the foundation of national infrastructure decisions. Institutions responsible for generating official statistics have designed streamlined instruments to enable the collection of important travel behavior metrics. Most organizations currently implement some form of travel diary survey (TDS), in which participants record a series of trips and stops over a specified time period. When these diaries are completed within probabilistic samples, the aggregate results can be used to model travel demand between regions, generate statistics on transportation modes, or monitor the uptake of green incentives such as telecommuting.

Usage of TDS in official statistics to create a granular picture of individual travel behavior over time has spanned more than half a century. Modes of administration have evolved with the times, from face-to-face interviews in the 1950s transitioning gradually into mail and telephone survey instruments in the 1980s and 1990s, followed by a transition to web-based methodology in the early twenty-first century (Adler et al. 2002; Arentze et al. 2005; Axhausen 1995). Although this mode evolution has led to both reduced costs as well as increased ease of administration, reliance on respondent recall for generation of the diary has remained constant.

¹ University of Utrecht, Department of Methodology and Statistics, Padualaan 14, 3584 CH Utrecht, the Netherlands. Emails: d.m.mccool@uu.nl and p.lugtig@uu.nl

² Statistics Netherlands, P.O. Box 4000, 2270 JM Voorburg, the Netherlands. Email: bo.mussman@cbs.nl

³ Statistics Netherlands, Research and Development, P.O. Box 24500, Den Haag, 2490HA, the Netherlands. Email: jg.schouten@cbs.nl

Researchers have long been aware of the tendency of this method to lead to trip under-reporting (Clarke et al. 1981; Richardson et al. 1995). Recent studies comparing concurrent GPS and recall methodologies have demonstrated that reliance on recall methodology produces under-reporting of short trips, differences in reported trip departure times, overestimation of trip length, and underestimation of vehicle miles of travel (Bricka et al. 2009; Carrion et al. 2014; Forrest and Pearson 2005; Kelly et al. 2013; Stopher and Shen 2011; Wolf et al. 2003). More generally, the response rates for TDS have been decreasing steadily over the decades. Bricka et al. (2009) noted disproportionate nonresponse for large households, low-income households and younger adults, and Ogle et al. (2005) showed similar nonresponse trends within households making the fewest and most trips. Although presumably less burdensome, studies using standalone GPS devices have presented similar nonresponse challenges, but do seem to increase uptake among younger participants (Bricka 2008). Managing lowering response rates requires increasing the cost per respondent of an already expensive design, which has prompted researchers to find more cost-effective ways to access the information.

Smartphone-based travel studies have been proposed as a solution for addressing issues of cost and decreasing response rate among younger households. Smartphone penetration over the last decade has neared saturation, with recent numbers from Statistics Netherlands indicating that over 90% of the Dutch population owns a mobile device (Centraal Bureau voor de Statistiek 2019). Additionally, Roddis et al. (2019) found that respondents rated interaction with an app more enjoyable than either a traditional user-completed travel diary or a personal log, rating it as less burdensome, and both Roddis et al. (2019) and Safi et al. (2017) demonstrated that smartphone-based apps provided higher-quality data in comparison with recall-based TDS. These potential advantages have led to the introduction of multiple app-based travel diaries (Berger and Platzer 2015; Cottrill et al. 2013; Greaves et al. 2015; Lynch et al. 2019; Patterson and Fitzsimmons 2016; Prelijpean et al. 2018). To date, however, these studies have yet to address the unique challenges involved with fieldwork within the general population, nor has the impact of a large-scale implementation been assessed. To that end, the primary goal of this research is to present a realistic assessment of a smartphone-based travel study within a national sample, showcasing problems at the different component levels of Total Survey Error. For fitness in general population surveys, important requirements are acceptable recruitment rates, low drop-out, low in-app missing data and high in-app data quality. This article explores these features. Apart from these more methodological requirements, there are complex logistical and procedural requirements. Although these are referenced here, they are not focal in this exploration.

Introduction of the app-based TDS is not a silver bullet. There are known issues arising from Global Navigational Satellite System (GNSS) measurements themselves, ranging from the problems of urban canyons to the length of time required to establish an initial signal (Park et al. 2014). Additionally, although processed Android location data can be as accurate as $\pm 10\text{m}$, this accuracy varies across different devices (Liu et al. 2017; Menard et al. 2011). Even when the accuracy is acceptable, technical issues with the applications themselves can lead to completely missed travel behaviors (Roddis et al. 2019). Independent of technical issues, semantic issues present an additional hurdle. The transition to automated stop identification from respondent's subjective interpretation of

their travel behavior has proven a difficult task, and the reduction in burden expected from travel mode identification is yet to surface (Prelicean et al. 2016; Yang et al. 2016; Zhao et al. 2015). Importantly, apps must also remain user-friendly or risk early dropout (Assemi et al. 2018). In this study, we aim to investigate the feasibility of an app-based TDS to be robust enough against these errors to be usable in official general population surveys of travel behavior.

2. Statistics Netherlands Travel Application Overview

The initial objective of this study was to develop an application that would be able to assess the potential of smartphone technology in mobility research. To this end, the app was designed both to collect the data of interest as well as metadata and user input that would allow assessment of the data quality. In order to collect and represent back to the participant their mobility data, the app needed to provide latitude and longitude updates frequently enough to reliably determine location, separate these measurements into a series of moving periods (tracks) and stationary periods (stops). This would then allow the user to enrich the data with auxiliary day-level and trip-level information.

The Travel App System is comprised of a front end and a back end. The front end consists of the Statistics Netherlands (SN) Travel App, which collects the location data, resolves stops and tracks, and exposes these to the user for annotation purposes. Both the raw location data and the resolved data are stored locally in a SQLite database on the mobile device. The back end consists of an API written in GO that performs the data ingestion and transformation into a PostgreSQL database that ultimately receives and stores the data (see Figure 1).

2.1. CBS Travel App

In order to be able to deploy equivalent algorithms to both Android and iOS versions, the client was developed in C# using the Xamarin framework. This framework provides compilation to Intermediate Language which is Just-in-Time compiled to native assembly on Android devices, and Ahead-of-Time compilation into native ARM assembly code for the iOS build. The application was developed Open Source and hosted on a publicly accessible collaborative code repository in order to facilitate distribution and address potential privacy concerns. Full code for back end and front end is available at <https://gitlab.com/tabi/archive>.

2.1.1. User Interface

Users who download the application to their smartphone are requested to log in using the credentials received in the invitation letter received in the mail. Upon successful registration of their device on the server, users are asked to enable location permissions on their device. Specifics of this permission request differ across devices, although the UI prompt does not. Following this, regardless of permission status, users are provided with a brief video tutorial explaining proper use of the application. Users are shown how to navigate between stops and tracks within a day, adding annotative information, as well as how to pause and resume the application's location-tracking behavior (see Figure 2).

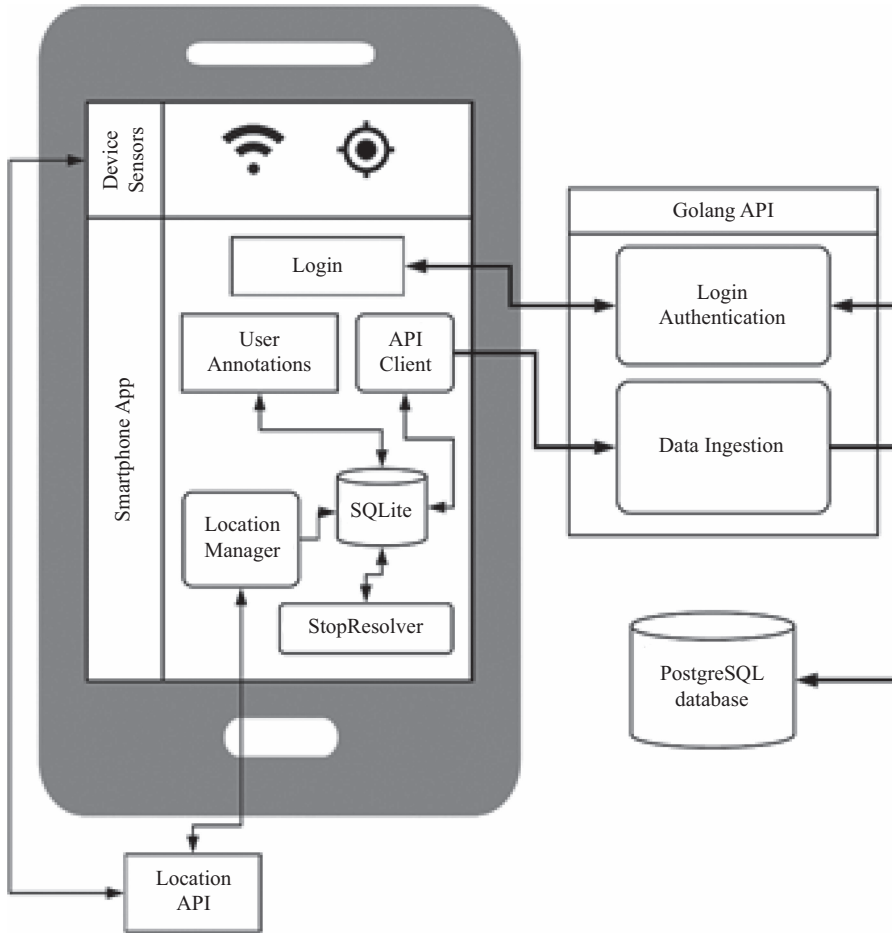


Fig. 1. Technical implementation and integration of back end and front end.

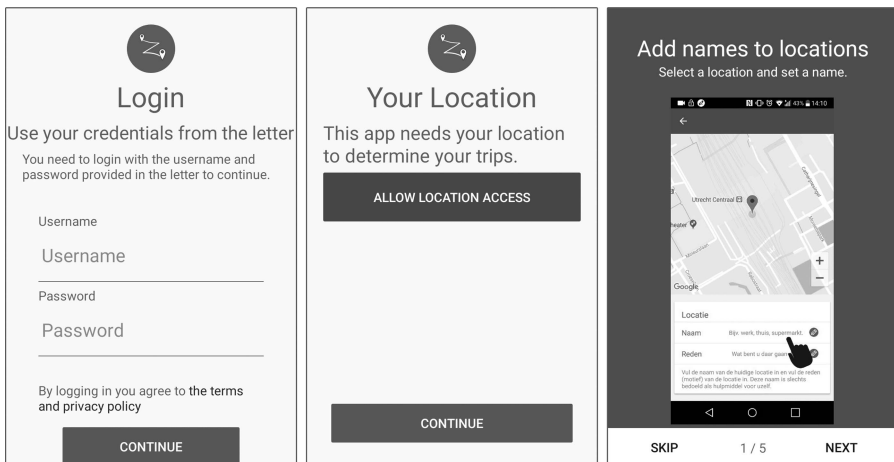


Fig. 2. Interface for initial login, permissions, and use instructions.

While the application is running, an icon is visible in the system bar. When the user accesses the notification drawer, a message is visible alerting the user that the app is running in the background to track movement.

The user interface for adding annotations is organized by day, with days beginning at midnight. A user opening the app while it is running in background mode will be taken to a list of the current day’s stops and tracks, called the Day Overview. From the Day Overview, it is possible to return to the calendar to see a list of available days, to drill down into each stop or track, or to answer day-specific questions (see Figure 3).

Clicking on a stop opens a map with a point and surrounding radius representing the user’s registered location for that time period. Clicking on a track will provide a map with two points representing start and stop locations and a blue line representing their movement trajectory. This map can be manipulated by zooming or panning in order for the user to determine with greater accuracy where they were. The stop menu requests users to record a name and reason for the stop. Figure 4 shows the track menu, which requests that users enter the mode of transportation used in the trip leg.

The user is presented with differential icons indicating whether the annotation item has been completed. A user who completes a day by assigning motives and mode annotations to each stop and track on a day as well as completes the day-level annotation questions will see a check mark for the day.

2.2. Algorithms and Implementation

The application collects and saves raw data from an operating-system-specific location API implementation. For iOS, this is Core Location. For Android, Google Play services’ Google Location Services API is preferred. When it is not available, the native Android location API is called in its place. All three APIs return similar information. Although the methods that produce the location information are proprietary, it is known that these location APIs aim to produce accurate location results by combining information from GNSS, local Wi-Fi signals and cell phone tower signals. Android devices offer a setting

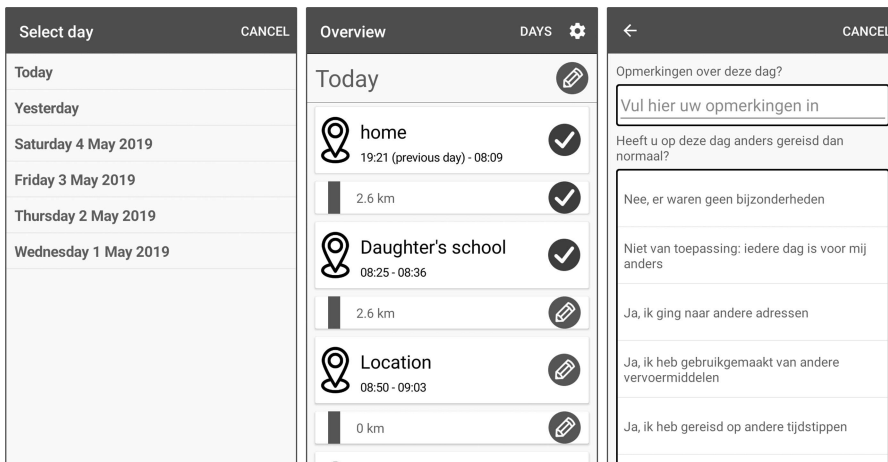


Fig. 3. User Interface for travel log display and day-level questions.

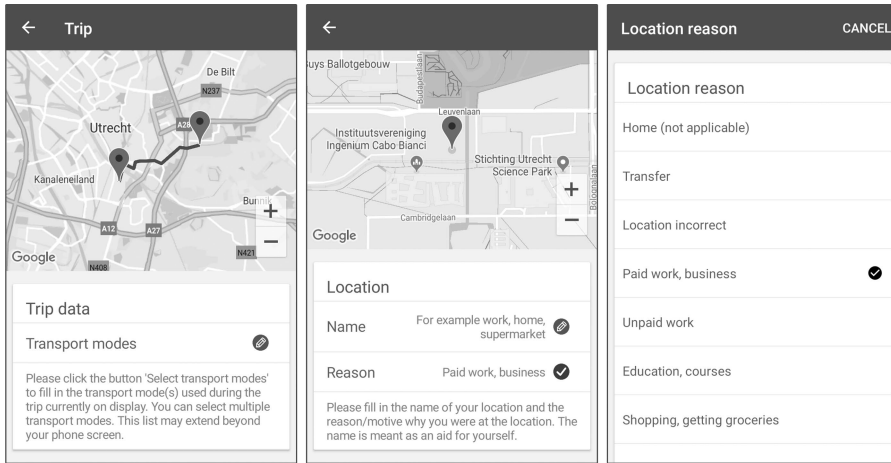


Fig. 4. User Interface for trip and stop detail annotation.

external to our application where the user may prefer to use only GNSS signals, or to turn on ‘high-accuracy mode’ which provides a location derived from the combination of Wi-Fi, Bluetooth, cell tower and GNSS signal information. Devices running iOS as an operating system also offer a setting to disable GPS, although it is accessible only through a series of menus. Beginning with iOS 11, Apple devices offer the user the option to share location information with the requesting application either only when the app is open and running in the foreground, or at all times.

The format and content of the coordinates returned by the location APIs vary by the polled system. Latitude and longitude are universally returned, although the number of significant digits varies. Altitude is reported if and only if a connection is established between the device and a GNSS system. The accuracy returned by the Google Location Services API is defined as the radius of 68% confidence in meters. The interpretation of the horizontal accuracy variable returned by CoreLocation is currently undocumented by Apple.

2.2.1. High/Low Tracking

The application running on the mobile phone requests information from the OS-specific API at regular intervals in order to generate a pattern of movement. When the device is determined to be not in motion, the application requests a location update once per minute in order to preserve battery life. When the device is in motion, the location request is submitted once per second. On Android, the application also accepts other location updates that were requested by different applications with the goal of providing the highest possible accuracy at no increased cost to battery life.

An algorithm was developed in order to determine when a device should move between the two tracking profiles. This algorithm has two parameters: time and distance. The implementation is similar in the Android and iOS implementations. A listener is engaged to receive location updates, all of which are saved in the location repository on the device. Each updated location is checked against the time parameter to determine whether a

distance check should be performed. Once the time parameter is exceeded, the distance from the immediately preceding location is calculated using the Haversine formula. This calculated distance is compared to the distance parameter to determine whether or not the user is moving. In the event that the reported accuracy of the location is larger than the distance parameter, this number is used instead of the distance parameter in order to offer adjustment for the potentially erratic behavior in a situation in which multiple low-accuracy locations are returned because of poor signal availability. If the calculated distance exceeds this number, the application activates high tracking mode and begins requesting location updates at the rate of one per second. If it does not, low tracking mode is activated and location updates are requested once per minute.

2.2.2. Stop Detection/Resolution

The stop detection algorithm developed within this application functions similarly to the high/low-tracking algorithm. The algorithm identifies a stop when the device reported location is within a given radius for a particular length of time. By altering these two parameters, respectively radius and duration, it is possible to adjust the sensitivity with which a set of locations can be consolidated into a single stop. In Section 3 Field Test, we describe the design of our study in which we varied these attributes across participants. A simulation study conducted post-hoc on the field-test data in which we clustered participants' locations in each combination of radius and duration parameters identified minimal differences in the number of identified stops (Killaars et al. 2019).

While all location updates were stored in the repository, only those locations where the reported accuracy was less than 80 meters were used for the separate step of stop determination. The stop detection algorithm consists of four steps. First, the timestamp of the last known stop is requested from the internal location repository. Second, the set of locations with timestamps greater than or equal to the last known stop are returned. Third, the set of all positions are divided into groups based on the distance between the positions; this step involves calculating the distance between each new proposed location and the set of all locations identified as belonging to a stop. If this distance is less than the radius parameter, the location is incorporated into the existing stop. If the distance is greater than the radius parameter, it becomes the first location in a proposed new stop. Fourth, the time elapsed between the first and last location in the proposed stop is calculated and compared against the duration parameter. If the elapsed time exceeds this value, the mean latitude, longitude and the beginning and ending timestamp are registered in the local database.

Adjacent stops are merged when the distance between the two stops' average position is less than one hundred meters. Following the stop resolution phase, stop visits can be processed. Stops may be places to which users return, such as a house or workplace. A stop visit is therefore a singular instance of having been at a stop.

2.2.3. Tracks

Tracks were defined as the set of locations between stops. Start time, beginning latitude/longitude and ending latitude/longitude are saved in the local repository. Track length was calculated as a summation over individual distances between consecutive locations after filtering out coordinates with an accuracy greater than 100 meters.

2.2.4. Additional Data Collected

Data were collected at various intervals. Upon registration with the server, device information was recorded in the database, including phone manufacturer, model, OS and version number. As described in the preceding section, location data were collected at time intervals of varying length and processed client-side into tracks and stops.

Users were requested to annotate all tracks with the mode(s) of transportation utilized in the movement. Various common options were provided for the user in a drop-down select menu in addition to an open text field to accommodate less common forms of transportation. Users were requested to annotate stops with the motive of the stop. Users were provided with a list of many common motives, such as home, work and shopping. Additionally, users could submit their own motive or mark the stop as incorrect. Stops could be given a name for ease of recall.

To facilitate detection of unexpected events, users were requested to provide feedback on three day-level questions: ‘Did you have your phone with you at all times today?’, ‘Was today a normal day for you?’ and ‘Do you have comments on the day?’.

3. Field Test

A field test was designed and conducted in late 2018 in order to address two primary research questions:

1. To what extent are persons willing to register an app and provide seven days of time-location sensor data?
2. What is the quality of the resulting sensor data and additional survey data?

To facilitate comparison with existing mobility research, the TDS Underway in the Netherlands (ODiN), conducted by the Dutch Ministry of Infrastructure, was selected as the basis for the pilot study ([Statistics Netherlands 2019](#)). ODiN is designed 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.

3.1. Participants

The field-test survey was sent to 1902 sample persons, half of whom were randomly sampled from the Dutch population register and half of whom were sampled from the pool of previous ODiN respondents whose surveys had been conducted in the months of September and October 2018. The target population consists of people 16 years and older living in non-institutionalized households.

3.2. Methods

3.2.1. Incentive Stratification

The sample was split into three and randomly assigned to one of three incentive conditions. All incentives were paid in the form of gift cards mailed to recipients. All sample persons received an unconditional EUR 5. One third of the sample was promised a split sum of EUR 10, with EUR 5 conditional on registration and EUR 5 conditional on

seven days of recorded travel data. One third was promised EUR 10 conditional on seven days of travel data. One third received EUR 20 conditional on seven days of recorded travel data. In the following, we will omit the reference to the unconditional EUR 5 and denote the incentive conditions by $5 + 5$, $0 + 10$ and $0 + 20$.

3.2.2. Stop Detection Stratification

Sampled persons were additionally allocated different stop detection parameters at random. Participants were assigned a duration, d , of either 2-, 3-, 4- or 5-minute intervals and a radius, r , between 60 and 100 meters inclusive, representing the maximum distance from a central point of a stop. These parameters were implemented to influence the relative strictness of the algorithm to automatically identify a stop, with lower values representing looser criteria and higher values representing more stringent requirements. These parameters were implemented in the back end of the app and were unknown to participants.

3.2.3. Materials

The sample was sent an invitation letter by mail with app login information and both a QR code and URL leading to the SN Travel App landing page. The invitation letter arrived on November 2nd and persons were given the opportunity to respond until December 15th. On the landing page, persons were given background information and a brief explanation of the study and the app, and were directed to the appropriate application store based on device operating system (Android or iOS). On November 16th, those persons yet to register with the app were sent a reminder letter in the mail containing the same login information, QR code and URL. On November 23rd, respondents who had not reached seven days of travel data were sent a motivation letter.

All letters, the app, and the landing page contained information on how to contact Statistics Netherlands if there were questions or difficulties.

3.2.4. Analysis Plan

As the primary goal was to establish the feasibility for future widespread implementation, we aimed to investigate initial uptake, study dropout and data quality. Initial uptake analyses were stratified by the demographics available in the Dutch population register, as well as by incentive condition and previous participation in Travel Diary Study status. Dropout analyses were stratified by these same variables and additionally included information on the type of mobile device used by the participant and its operating system. Quality of the data was judged by frequency of collection and alignment of summary measures for field test data and ODiN responses on the subset of participants who were involved in the September 2018 data collection.

4. Results

4.1. Registration and Response

Of particular interest was whether use of an app-based TDS could lead to acceptable participation rates, and to determine which factors would lead to either nonresponse or participation. Additionally, this study sought to determine the extent to which

non-response and dropout were selective by using administrative data available on all Dutch residents of all potential respondents.

The following sequential stages were required for full participation in the study: receiving and reading the invitation letter, downloading the application, registering within the app, accepting location permissions on the device, not closing the app for the full seven days, and providing annotative data. Unit response was considered to coincide with the third step, device registration.

Of the 1902 respondents who were sent a letter, 674 registered a device using the login information they received, leading to a total unit response rate of 35.4%. Previous ODiN participants had a response rate of 44.4% compared with the newly-acquired sample's response rate of 26.5%. This reflects a slightly lower response rate than the 31.0% and 27.9% obtained for ODiN in 2018 and 2019 respectively (Statistics Netherlands 2019; CBS/RWS-WVL 2020). As shown in Table 1, the three incentive conditions, 0+20, 0+10 and 5+5 achieved response rates of 39.7%, 36.4% and 30.1% respectively.

The measure of non-registration represents an overestimation of nonresponse. However, distinguishing non-contact, refusal and non-eligibility from other factors was possible only for those persons who independently contacted Statistics Netherlands. Contact was received from people who attempted to participate in the study, but were unable to download and install the application for various reasons. Unregistered is, within this study, a close analogue of nonresponse. Most who fail to register a device are likely to be traditional nonresponders.

4.2. Nonresponse

Device registration status varied across known demographic variables from the Dutch population register, which is a governmental database containing administrative information on all persons registered as living in the Netherlands. Younger persons were more likely to register a device than older persons. Immigrants were less likely to register a device than those originally of Dutch origin, and first generation immigrants were less likely to register a device than second-generation immigrants. Persons with college degrees or higher were more likely to register a device than those with elementary or vocational school degrees.

Divorced and widowed persons were less likely to register a device in comparison with persons who were never married. However, single-person households were less likely to

Table 1. Device registration (unit response) rate by sample source and incentive condition.

		Unregistered		Registered	
		n	%	n	%
Sample source	ODiN respondents	529	55.6	422	44.4
	Newly acquired	699	73.5	252	26.5
Incentive	5+5	443	69.9	191	30.1
	0+10	403	63.6	231	36.4
	0+20	382	60.3	252	39.7
Total		1228	64.6	674	35.4

respond than family households with children. Home-owners were more likely to register a device than were renters. People with higher household incomes were also more likely to register their device. See [Table 2](#).

Most transportation-related characteristics that could be drawn from the Dutch population register, including possession of a car, moped or lease car, were not significantly related to registration status. Possession of a driver’s license, however, was significantly related to registration status, with those in possession of a driver’s license more likely to register a device than those not in possession of a license.

Geographic variables, including address density and province- or city-related variables were not significantly related to device registration status. A full table of response across all available variables can be found in [Table 7](#) in Appendix (Section 6).

Table 2. Device registration by sample characteristics.

		Unregistered	Registered
Age categories	[15,30]	240 (58.0%)	174 (42.0%)
	(30,50]	343 (58.7%)	241 (41.3%)
	(50,70]	437 (66.5%)	220 (33.5%)
	(70,96]	202 (83.8%)	39 (16.2%)
Origin	Dutch	961 (62.5%)	577 (37.5%)
	Non-western	130 (76.5%)	40 (23.5%)
	Western	131 (69.7%)	57 (30.3%)
Generation	Dutch	961 (62.5%)	577 (37.5%)
	First	149 (78.8%)	40 (21.2%)
	Second	112 (66.3%)	57 (33.7%)
Marital status	Married	624 (63.7%)	355 (36.3%)
	Never married	410 (60.8%)	264 (39.2%)
	Divorced	121 (71.2%)	49 (28.8%)
	Widow/widower	67 (91.8%)	6 (8.2%)
Education	Vocational	155 (72.4%)	59 (27.6%)
	Elementary	63 (78.8%)	17 (21.2%)
	Secondary	282 (60.1%)	187 (39.9%)
	Graduate	74 (49.7%)	75 (50.3%)
	University	136 (48.2%)	146 (51.8%)
Household type	NA	512 (72.9%)	190 (27.1%)
	Single	244 (72.6%)	92 (27.4%)
	Partners	420 (64.2%)	234 (35.8%)
	Partners, child	473 (60.4%)	310 (39.6%)
	Single parent	79 (69.3%)	35 (30.7%)
Has drivers license	Other household	6 (66.7%)	3 (33.3%)
	No	281 (73.4%)	102 (26.6%)
Home ownership	Yes	941 (62.2%)	572 (37.8%)
	Own	790 (61.6%)	492 (38.4%)
	Rent, corporation	275 (73.1%)	101 (26.9%)
	Rent, other	123 (66.8%)	61 (33.2%)
	Unknown	4 (80.0%)	1 (20.0%)
Total	NA	30 (61.2%)	19 (38.8%)
		1222 (64.5%)	674 (35.5%)

^a All χ^2 differences significant, $p < .01$

^b Table omits 6 non-responders lacking register data

4.3. Drop-out

Technical difficulties contributed to ambiguity in the identification of dropout as participants reported an inability to send data following successful device registration. Of the 674 participants who registered a device, 98 were never able to send GPS data. Participants who contacted Statistics Netherlands with this problem were instructed to upgrade their OS version, reinstall the application, or install on a different device. Each additional registration was recorded separately. In total, the 674 participants completed 748 registrations, of which 706 were unique configurations (differing model or OS version), and the remaining 42 were reinstallations. In total, 136 unique-to-user device configurations produced no GPS data. Table 3 shows the distribution of lack of GPS data across OS and OS version. The likelihood that an Android device would send GPS location data at least once generally increased across operating system versions. Although we found that iOS devices were more likely never to send data, we did not observe differences across versions of iOS. Distinguishing in a determinative way those users who expressly denied location permission when prompted in the OS from those who experienced technical issues with location provision or for whom other settings in their mobile device disallowed location provision was not possible with the available data.

In addition to dropout over time within the seven day period, we identified a pattern of dropout within a day in which the app would not send location reports during the full 24 hours, but would continue to send data either later in the day or on a subsequent day. While it is not possible to distinguish intentional closing and reopening of the application from the data, the patterns identified in the data lend support to the idea that this effect comes from the behavior of the OS. In order to investigate this phenomenon and determine its effects, we created a measure called gap time. Gap time refers to the length of time between two subsequent location reports from a single device. A device functioning properly and without additional restrictions imposed by the operating system of the device should have gap times of approximately one second while in motion, and of one minute while stationary. After removing duplicate records, on average, we identified a mean gap time of 18.8 seconds during trips and a mean gap time of 47.5 seconds during contiguous non-trip activity.

Table 3. OS and version by data status.

		GPS	No GPS
Android version	< 6.0	22 (81.5%)	5 (18.5%)
	6	42 (89.4%)	5 (10.6%)
	7	83 (96.5%)	3 (3.5%)
	8.0.0	170 (95.0%)	9 (5.0%)
	8.1.0	20 (95.2%)	1 (4.8%)
	9	2 (100.0%)	0 (0.0%)
iOS version	< 11.4.1	30 (75.0%)	10 (25.0%)
	11.4.1	24 (77.4%)	7 (22.6%)
	12.0	14 (77.8%)	4 (22.2%)
	12.0.1	124 (75.6%)	40 (24.4%)
	12.1	70 (76.9%)	21 (23.1%)
Total		601 (85.1%)	105 (14.9%)

At face value, receiving an update during trips at half the rate and updates during stationary periods more often than requested seems to indicate that very little data is lost. However, if we instead consider the maximum gap time per day per user, a different picture emerges. Figure 5 shows that approximately 30% of trips and 65% of non-trip activity contains gap times greater than one hour. Any statistics calculated on the basis of available data, therefore, must consider this fact, as aggregate summative statistics will likely be underestimated if we assume that gap times are likely to cover periods of meaningful data. Although determining the length of an acceptable gap time to consider coverage to be complete may depend on end goals, choosing a maximum gap time no larger than fifteen minutes yielded an average of 12.8 hours covered per user-day.

Complicating the problem, the gap times are not evenly distributed throughout the day. In fact, there is evidence that these gap times represent app or device fall-off, leading to a greater proportion of missingness later in the day than in the beginning of the day. Figure 6 shows the hour at which contact is lost and the hour at which we again start receiving data from the device, following from a gap time of at least thirty minutes.

4.4. Comparison to Traditional Diary

Data from the ODIN travel diary were made available to facilitate comparison between the app-based and travel diary methods. Only the data for respondents who participated in both this study as well as the ODIN study were used for comparison. Three measures were selected for comparison between the two methodologies: number of trips within a day, trip length and trip distance. ODIN respondents were asked to self-report information for a single, specific day that was assigned randomly. Trip-level measures were summed per day to establish number of trips per day, total trip length and total trip distance within a day. Trip-level characteristics were calculated per trip and then averaged for the user.

The total number of trips within a day were compared between the ODIN data and this study. As shown in Table 4, the median number of trips as determined by the SN App mechanism was four as compared to the median of three in the ODIN data. Additionally, we see a distribution with a much longer tail from the SN App data in Figure 7. Some

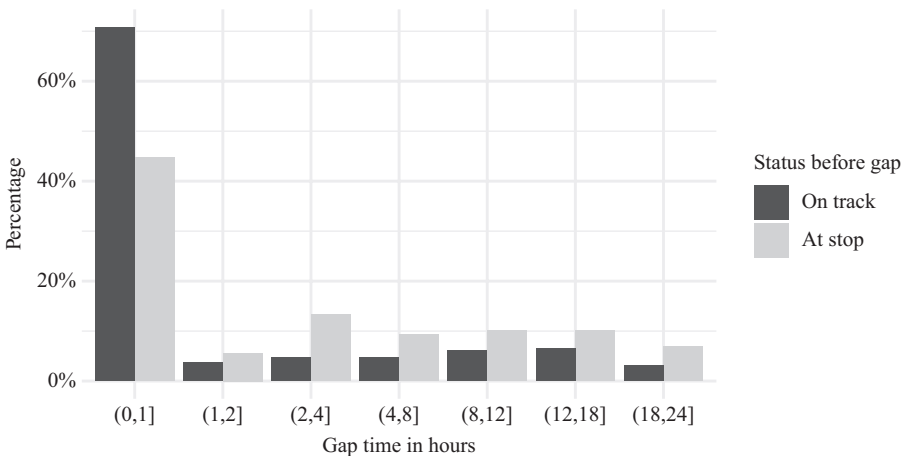


Fig. 5. Maximum gap time per user-day.

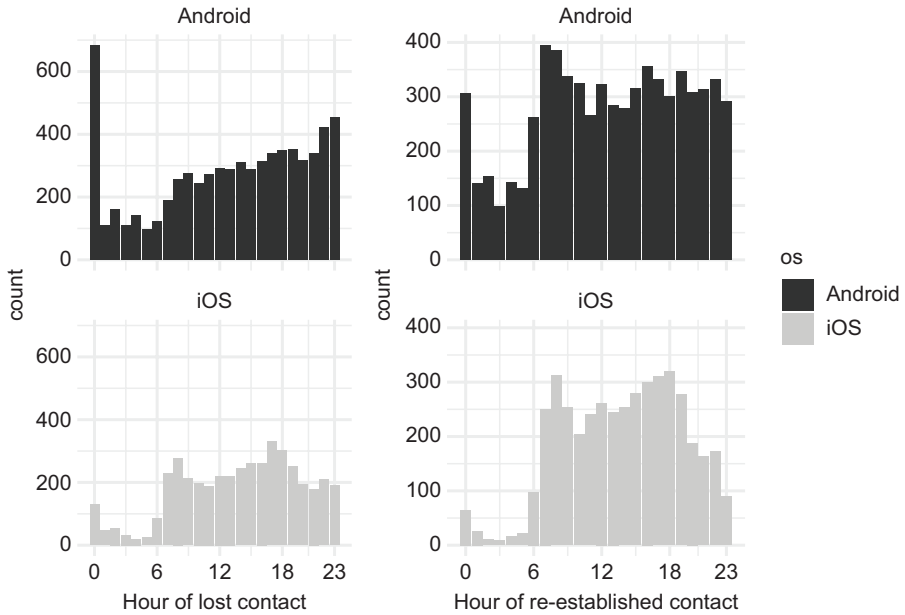


Fig. 6. Hour of lost and re-established contact, by OS for situations for > 60 minutes of no device activity.

Table 4. Descriptive statistics of trips per day by survey.

Survey	Respondent days	Mean	Median	Max
ODiN	321	3.54	3	12
Current study	1353	5.13	4	33

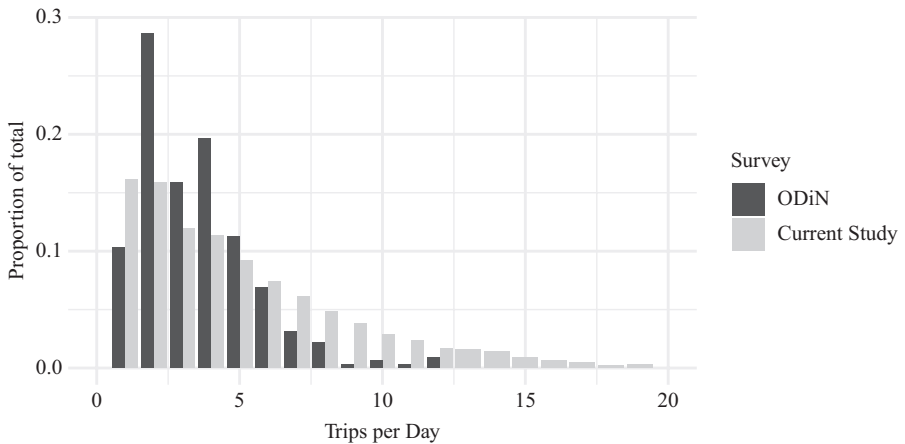


Fig. 7. Comparison of number of trips per user day.

portion of this increase likely represents a desirable outcome for our study, in which we are capturing short trips known to pose problems in self-report measures.

Additionally, our study records a longer time spent in transit per day than did ODiN. The average time spent in travel is higher for the SN app data, as shown in Table 5. Figure 8 shows recorded total travel time per day within both the current study and ODiN. The SN App recorded more time spent in all categories above one hour, which is in alignment with our hypothesis that automation would capture travel behavior at a more granular level. However, the combination of an increased number of trips per day and an increased time spent in travel could additionally be due to mechanisms within the tracking application that are too sensitive to movement, inflating both counts by including trips that would generally fall outside the purview of travel behavior, such as trips from an office workplace to a canteen in the same building. These differences may also be related to natural month-to-month discrepancies in travel behavior.

Curiously, although we demonstrate both more trips per day and more time spent traveling, aggregated distance within a day is notably shorter in this current study as compared to the ODiN data. Table 6 shows descriptive statistics for total distance traveled in a day within both studies. ODiN respondents reported a median distance per person-day of 32.2 kilometers compared to the median of 8.47 kilometers as tracked within this study. Figure 9 shows the differential distribution, with over 40% of person-days within the current study summing to fewer than 5 kilometers in comparison to approximately 15% within ODiN. It is likely that the missing data problem contributes to this difference. For example, consider a situation in which the app successfully initiates tracking a commuter successfully in the morning, but which loses contact in the course of the day, dropping the likely return trip in the evening.

Table 5. Descriptive statistics for travel time (hours) per day by survey.

Survey	Respondent days	Mean	Median	Max
ODiN	321	0.59	0.33	8.00
Current study	2327	1.14	0.51	20.26

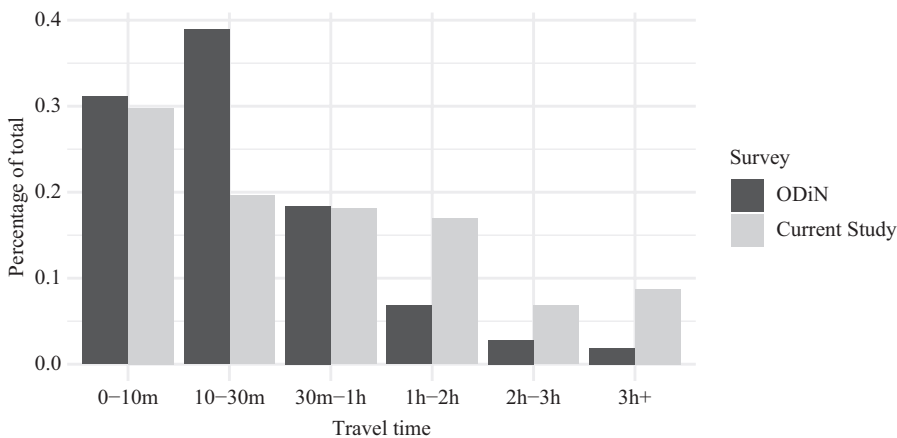


Fig. 8. Travel time per day by survey.

Table 6. Descriptive statistics for travel distance (km) per day by survey.

Survey	Respondent days	Mean	Median	Max
Current study	2290	34.68	8.47	1366.41
ODiN	319	288.22	32.30	12884.00

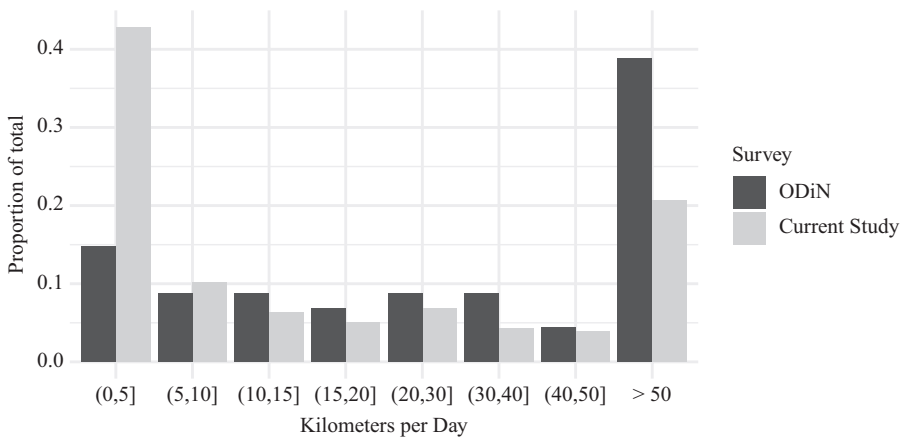


Fig. 9. Kilometers per day by study.

5. Conclusions

The primary goal of this research was to determine whether or not direct implementation of smartphone travel diaries was feasible in large scale own-device studies. In order to be considered feasible, we outlined requirements that we must have an acceptable response rate and the data collected from the device must be of high enough quality to allow for important metrics to be calculated reliably.

Response rates for this study were similar to those of the travel diary study ODiN. Although prior to 2018, response rates were upwards of 50%, this was largely a function of the CATI and CAPI contact initiatives following initial nonresponse, which were very costly. Since 2018, these modes have been phased out, leaving only the CAWI mode, which lowered response rates to 31.0% and 27.9% respectively for 2018 and 2019. The 26.5% response rate for this study's new responders is similar, but in place of a single day's data, they are responding to our request for a week's worth of data.

Respondents also proved willing to provide annotative data to the passive traces. Although this currently takes the form of labeling stops for purpose and tracks for transportation mode, these are variables which could conceivably be calculated from the data itself as the accuracy of mobile-device GPS units approaches centimeter-level accuracy (Dabov and Di Pietra 2019; Humphreys 2018). This frees researchers to reduce respondent burden by either relying on the passive data itself or adapting a verification approach where respondents either confirm the predicted mode or correct it.

Importantly, this study identifies a major hurdle for researchers wishing to transition to app-based TDS implementations. There is ample evidence within our data to suggest that concerns over dropout and missing data are inherently device-related. A recent survey of the state of the art of current smartphone-based travel apps demonstrated that across 22

different apps, device characteristics such as operating system and phone manufacturer impacted both quantity and quality of collected data (Harding 2019). Researchers in this field have remained hopeful over the years that improved smartphone technology would address current limitations with missing data, low accuracy, urban canyons and battery life (Berger and Platzer 2015; Verzosa et al. 2017; Cottrill et al. 2013; Greaves et al. 2015; Allström et al. 2017; Gong et al. 2014; Geurs et al. 2015). Although these predictions are likely warranted as they pertain to the inexorable march of technology, it does not necessarily follow that the resultant data will be any better as new technological issues will arise to replace the old. Consider battery life: in order to keep pace with increasing battery requirements, Android has implemented mechanisms to close apps without informing the user, leading to gaps and unintentional drop-out in the data (Petter et al. 2019). For the same battery concerns, iOS has taken over strict control of the location management system, restricting the frequency that locations can be polled, leading to the problem of cold starts in the data.

Researchers ultimately have little control over the device upon which the application will be installed. Google, Apple and the various device manufacturers are unlikely to be open in releasing details describing the precise functioning of their location systems. Additionally, as versions move quickly and much changes between iterations, a system that functions well one year may require an intensive change in the following year in order to continue to function, and the data that are generated may ultimately be quite different as well. It may be that researchers must, at least for the time being, continue to involve participants in assessing their passive data with active control questions.

It is easy to view this as a negative if we view it in the context of the relatively clean and complete data that comes back from written survey instruments. Researchers have been promised higher-quality data that strictly improve upon the TDS, but the current study demonstrates that any increase in quality is by no means free. Instead, these data must be judged on their own terms. The richness is not only useful, but key to its use. Identification of the issues underlying missingness and measurement is an important first step. The next step must be to find ways to compensate for the technological issues as we do for other methodological issues, and here the unique characteristics of the data itself may prove invaluable: the longitudinal nature of the data collected may allow recovery of missing track segments within a person's own travel log, the spatial nature of the data may allow us to correct for measurement error and the sheer size of the data may open up new methods of inference to us. Technological challenges will persist, necessitating development of robust methodology for identification of the true underlying behaviors, but the data are sufficient to provide grounds for good inference if researchers can move past these initial steps.

6. Appendix

Table 7. Device registration by full sample characteristics.

	Unregistered (N = 1228)	Registered (N = 674)	p value
Age			< 0.01 ¹
N-miss	6	0	
Mean (SD)	50.4 (19.0)	44.3 (17.1)	
Range	15.0 – 96.0	15.0 – 90.0	
Agecat			< 0.01 ²
N-miss	6	0	
15-20	90 (7.4%)	63 (9.3%)	
21-30	150 (12.3%)	111 (16.5%)	
31-40	157 (12.8%)	117 (17.4%)	
41-50	186 (15.2%)	124 (18.4%)	
51-60	229 (18.7%)	118 (17.5%)	
61-70	208 (17.0%)	102 (15.1%)	
>70	202 (16.5%)	39 (5.8%)	
Gender			0.77 ³
N-miss	6	0	
Male	606 (49.6%)	329 (48.8%)	
Female	616 (50.4%)	345 (51.2%)	
Origin			< 0.01 ²
N-miss	6	0	
Dutch	961 (78.6%)	577 (85.6%)	
Non-western	130 (10.6%)	40 (5.9%)	
Western	131 (10.7%)	57 (8.5%)	
Generation			< 0.01 ²
N-miss	6	0	
Dutch	961 (78.6%)	577 (85.6%)	
First	149 (12.2%)	40 (5.9%)	
Second	112 (9.2%)	57 (8.5%)	
Marital status			< 0.01 ²
N-miss	6	0	
Married	624 (51.1%)	355 (52.7%)	
Never married	410 (33.6%)	264 (39.2%)	
Divorced	121 (9.9%)	49 (7.3%)	
Widow/widower	67 (5.5%)	6 (0.9%)	
Education			< 0.01 ⁴
N-miss	518	190	
Elementary school	63 (8.9%)	17 (3.5%)	
High school	282 (39.7%)	187 (38.6%)	
Vocational school	155 (21.8%)	59 (12.2%)	
University	136 (19.2%)	146 (30.2%)	
Graduate school	74 (10.4%)	75 (15.5%)	
Household type			< 0.01 ²
N-miss	6	0	
One-person household	244 (20.0%)	92 (13.6%)	
Partners, no child	420 (34.4%)	234 (34.7%)	
Partners, child	473 (38.7%)	310 (46.0%)	
Single parent	79 (6.5%)	35 (5.2%)	
Other household	6 (0.5%)	3 (0.4%)	

Table 7. Continued

	Unregistered (N = 1228)	Registered (N = 674)	p value
Urbanicity			0.55 ⁴
Very high	259 (21.2%)	144 (21.4%)	
High	316 (25.9%)	171 (25.4%)	
Moderate	239 (19.6%)	136 (20.2%)	
Slight	208 (17.0%)	128 (19.0%)	
Rural	200 (16.4%)	95 (14.1%)	
Province			0.24 ²
N-miss	6	0	
Groningen	37 (3.0%)	28 (4.2%)	
Friesland	46 (3.8%)	26 (3.9%)	
Drenthe	47 (3.8%)	12 (1.8%)	
Overijssel	75 (6.1%)	45 (6.7%)	
Flevoland	31 (2.5%)	13 (1.9%)	
Gelderland	149 (12.2%)	88 (13.1%)	
Utrecht	85 (7.0%)	59 (8.8%)	
Noord-Holland	206 (16.9%)	103 (15.3%)	
Zuid-Holland	256 (20.9%)	127 (18.8%)	
Zeeland	27 (2.2%)	13 (1.9%)	
Noord-Brabant	179 (14.6%)	115 (17.1%)	
Limburg	84 (6.9%)	45 (6.7%)	
Home ownership			< 0.01 ²
N-miss	36	19	
Own	790 (66.3%)	492 (75.1%)	
Rent, corporation	275 (23.1%)	101 (15.4%)	
Rent, other	123 (10.3%)	61 (9.3%)	
Unknown	4 (0.3%)	1 (0.2%)	
Std. income pct.			< 0.01 ¹
N-miss	20	5	
Mean (SD)	55.4 (27.7)	64.3 (25.6)	
Range	0.0 – 100.0	0.0 – 100.0	
Has car			0.11 ³
N-miss	6	0	
No	665 (54.4%)	341 (50.6%)	
Yes	557 (45.6%)	333 (49.4%)	
Has moped			1.00 ³
N-miss	6	0	
No	1155 (94.5%)	637 (94.5%)	
Yes	67 (5.5%)	37 (5.5%)	
Has drivers license			< 0.01 ³
N-miss	6	0	
No	281 (23.0%)	102 (15.1%)	
Yes	941 (77.0%)	572 (84.9%)	

1. Linear Model ANOVA

2. Fisher’s Exact Test for Count Data with simulated p-value (based on 500 replicates)

3. Fisher’s Exact Test for Count Data

4. Trend test for ordinal variables

7. References

- Adler, T., L. Rimmer, and D. Carpenter. 2002. "Use of Internet-Based Household Travel Diary Survey Instrument." *Transp Res Rec* 1804(1): 134–143. DOI: <https://doi.org/10.3141/1804-18>.
- Allström, A., I. Kristoffersson, and Y. Susilo. 2017. "Smartphone Based Travel Diary Collection: Experiences from a Field Trial in Stockholm". *Transp. Res. Procedia* 26: 32–38. DOI: <https://doi.org/10.1016/j.trpro.2017.07.006>.
- Arentze, T., I. Bos, E. Molin, and H. Timmermans. 2005. "INTERNET-BASED TRAVEL SURVEYS: SELECTED EVIDENCE ON RESPONSE RATES, SAMPLING BIAS AND RELIABILITY". *Transportmetrica* 1(3): 193–207. DOI: <https://doi.org/10.1080/18128600508685648>.
- Assemi, B., H. Jafarzadeh, M. Mesbah, and M. Hickman. 2018. "Participants' Perceptions of Smartphone Travel Surveys". *Transportation Research Part F: Traffic Psychology and Behaviour* 54: 338–348. DOI: <https://doi.org/https://doi.org/10.1016/j.trf.2018.02.005>.
- Axhausen, K.W. 1995. "Travel Diaries: An Annotated Catalog". *Institut für Strassenbau und Verkehrsplanung*. DOI: <https://doi.org/10.3929/ethz-b-000024848>.
- Berger, M., and M. Platzer. 2015. "Field Evaluation of the Smartphone-Based Travel Behaviour Data Collection App 'SmartMo'". *Transp. Res. Procedia* 11: 263–79. DOI: <https://doi.org/10.1016/j.trpro.2015.12.023>.
- Bricka, S. 2008. "Non-Response Challenges in GPS-Based Surveys." 8th International Conference on Survey Methods in Transport, Annecy, France, May 25–31, 2008.
- Bricka, S.G., J. Zmud, J. Wolf, and J. Freedman. 2009. "Household Travel Surveys with GPS: An Experiment". *Transp Res. Rec.* 2105(1): 51–56. DOI: <https://doi.org/10.3141/2105-07>.
- Carrion, C., F. Pereira, R. Ball, F. Zhao, Y. Kim, K. Nawarathne, N. Zheng, C. Zegras, and M. Ben-Akiva. 2014. "Evaluating FMS: A Preliminary Comparison with a Traditional Travel Survey." Available at: <https://trid.trb.org/view/1289999> (accessed February 2021).
- CBS/RWS-WVL. 2020. Centraal Bureau voor de Statistiek (CBS), and Rijkswaterstaat (RWS-WVL). "ODiN 2019" Centraal Bureau voor de Statistiek (CBS). DOI: <https://doi.org/10.17026/DANS-XPV-MWPG>.
- Centraal Bureau voor de Statistiek. 2019. "Internet; Toegang, Gebruik En Faciliteiten." October. Available at: <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83429NED/table?fromstatweb> (accessed February 2021).
- Clarke, M., M. Dix, and P. Jones. 1981. "Error and Uncertainty in Travel Surveys". *Transportation* 10(2): 105–126. DOI: <https://doi.org/10.1007/BF00165261>.
- Cottrill, C.D., F. Cãmara Pereira, F. Zhao, I. Ferreira Dias, H. Beng Lim, M.E. Ben-Akiva, and P.C. Zegras. 2013. "Future Mobility Survey: Experience in Developing a Smartphone-Based Travel Survey in Singapore". *Transp Res. Rec.* 2354(1): 59–67. DOI: <https://doi.org/10.3141/2354-07>.
- Dabove, P., and V. Di Pietra. 2019. "Towards High Accuracy GNSS Real-Time Positioning with Smartphones". *Adv Space Res* 63(1): 94–102. DOI: <https://doi.org/10.1016/j.asr.2018.08.025>.
- Forrest, T.L., and D.F. Pearson. 2005. "Comparison of Trip Determination Methods in Household Travel Surveys Enhanced by a Global Positioning System". *Transp Res Rec* 1917(1): 63–71. DOI: <https://doi.org/10.1177/0361198105191700108>.

- Geurs, K.T., T. Thomas, M. Bijlsma, and S. Douhou. 2015. “Automatic Trip and Mode Detection with Move Smarter: First Results from the Dutch Mobile Mobility Panel”. *Transp. Res. Procedia* 11: 247–262. DOI: <https://doi.org/10.1016/j.trpro.2015.12.022>.
- Gong, L., T. Morikawa, T. Yamamoto, and H. Sato. 2014. “Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies”. *Procedia – Soc. Behav. Sci.* 138: 557–565. DOI: <https://doi.org/10.1016/j.sbspro.2014.07.239>.
- Greaves, S., A. Ellison, R. Ellison, D. Rance, C. Standen, C. Rissel, and M. Crane. 2015. “A Web-Based Diary and Companion Smartphone App for Travel/Activity Surveys”. *Transp. Res. Procedia* 11: 297–310. DOI: <https://doi.org/10.1016/j.trpro.2015.12.026>.
- Harding, C. 2019. “From Smartphone Apps to in-Person Data Collection: Modern and Cost-Effective Multimodal Travel Data Collection for Evidence-Based Planning”. Doctoral dissertation, University of Toronto, Canada. Available at: <http://hdl.handle.net/1807/97472> (accessed February 2021).
- Humphreys, T.E. 2018. “Centimeter Positioning with a Smartphone-Quality GNSS Antenna.” DOI: <https://doi.org/10.15781/T2HT2GV0R>.
- Kelly, P., P. Krenn, S. Titze, P. Stopher, and C. Foster. 2013. “Quantifying the Difference Between Self-Reported and Global Positioning Systems-Measured Journey Durations: A Systematic Review”. *Transp Rev* 33(4): 443–459. DOI: <https://doi.org/10.1080/01441647.2013.815288>.
- Killaars, L., B. Schouten, and O. Mussmann. 2019. “Stop and Go Detection in Gps-Position Data.” Centraal Bureau voor de Statistiek. Available at: https://www.researchgate.net/publication/338402776_Stop_and_Go_detection_in_GPS-position_data_-_Discussion_Paper.
- Liu, J., Y. Hu, D. Zhang, and H. Liu. 2017. “Performance Assessment of GNSS Measurements from Android Platform.” In 6th International Conference on Computer Science and Network Technology (ICCSNT) 21 October, 2017, Dalian, China: 472–76. DOI: <https://doi.org/10.1109/ICCSNT.2017.8343742>.
- Lynch, J., J. Dumont, E. Greene, and J. Ehrlich. 2019. “Use of a Smartphone GPS Application for Recurrent Travel Behavior Data Collection”. *Transp Res Rec* 2673(7): 89–98. DOI: <https://doi.org/10.1177/03611981198119848708>.
- Menard, T., J. Miller, M. Nowak, and D. Norris. 2011. “Comparing the GPS Capabilities of the Samsung Galaxy S, Motorola Droid X, and the Apple iPhone for Vehicle Tracking Using FreeSim_Mobile.” In 14th International IEEE Conference on Intelligent Transportation Systems (ITSC) 5–7 October 2011, Washington, D.C., U.S.A.: 985–90. IEEE / IEEE. DOI: <https://doi.org/10.1109/ITSC.2011.6083141>.
- Ogle, J., R. Guensler, and V. Elango. 2005. “Georgia’s Commute Atlanta Value Pricing Program: Recruitment Methods and Travel Diary Response Rates”. *Transp Res Rec* 1931(1): 28–37. DOI: <https://doi.org/10.1177/0361198105193100104>.
- Park, M.H., H.C. Kim, S.J. Lee, and K.S. Bae. 2014. “Performance Evaluation of Android Location Service at the Urban Canyon.” In 16th International Conference on Advanced Communication Technology, 16–19 February 2014, Pyeongchang, Republic of Korea. 662–65. DOI: <https://doi.org/10.1109/ICACT.2014.6779045>.
- Patterson, Z., and K. Fitzsimmons. 2016. “DataMobile: Smartphone Travel Survey Experiment”. *Transportation Research Record* 2594(1): 35–43. DOI: <https://doi.org/10.3141/2594-07>.

- Petter, O., M. Hirsch, E. Mushtaq, P. Hevesi, and P. Lukowicz. 2019. "Crowdsensing Under Recent Mobile Platform Background Service Restrictions: A Practical Approach." In Adjunct Proceedings of the 2019 Acm International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 Acm International Symposium on Wearable Computers September, 2019, London UK: 793–797. DOI: <https://doi.org/10.1145/3341162.3344867>.
- Prelipcean, A.C., G. Gidofalvi, and Y.O. Susilo. 2016. "Measures of Transport Mode Segmentation of Trajectories". *Int J Geogr Inf Sci* 30(9): 1763–1784. DOI: <https://doi.org/10.1080/13658816.2015.1137297>.
- Prelipcean, A.C., G. Gidófalvi, and Y.O. Susilo. 2018. "MEILI: A Travel Diary Collection, Annotation and Automation System". *Comput Env. Urban Syst* 70: 24–34. DOI: <https://doi.org/10.1016/j.compenvurbsys.2018.01.011>.
- Richardson, A.J., E.S. Ampt, and A.H. Meyburg. 1995. *Survey Methods for Transport Planning*. Eucalyptus Press Melbourne.
- Roddis, S., S. Winter, F. Zhao, and R. Kutadinata. 2019. "Respondent Preferences in Travel Survey Design: An Initial Comparison of Narrative, Structured and Technology-Based Travel Survey Instruments". *Travel Behav. Soc.* 16: 1–12. DOI: <https://doi.org/10.1016/j.tbs.2019.03.003>.
- Safi, H., B. Assemi, M. Mesbah, and L. Ferreira. 2017. "An Empirical Comparison of Four Technology-Mediated Travel Survey Methods". *J. Traffic Transp. Eng. Engl. Ed.* 4(1): 80–87. DOI: <https://doi.org/10.1016/j.jtte.2015.12.003>.
- Statistics Netherlands. 2019. "Onderweg in Nederland (Odin) 2018." Statistics Netherlands. Available at: <https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/aanvullende-onderzoeksbeschrijvingen/onderweg-in-nederland-odin—onderzoeksbeschrijving-2018> (accessed September 2020).
- Stopher, P., and L. Shen. 2011. "In-Depth Comparison of Global Positioning System and Diary Records". *Transp Res Rec* 2246(1): 32–37. DOI: <https://doi.org/10.3141/2246-05>.
- Verzosa, N., S. Greaves, and R. Ellison. 2017. "Smartphone-Based Travel Surveys: A Review." DOI: <https://doi.org/10.3141/2246-05>.
- Wolf, J., M. Oliveira, and M. Thompson. 2003. "The Impact of Trip Underreporting on VMT and Travel Time Estimates: Preliminary Findings from the California Statewide Household Travel Survey GPS Study". *Transp Res Rec* 1854: 189–198. DOI: <https://doi.org/10.3141/1854-21>.
- Yang, F., Z. Yao, Y. Cheng, B. Ran, and D. Yang. 2016. "Multimode Trip Information Detection Using Personal Trajectory Data". *J Intell Transp Syst* 20(5): 449–460. DOI: <https://doi.org/10.1080/15472450.2016.1151791>.
- Zhao, F., A. Ghorpade, F. Câmara Pereira, C. Zegras, and M. Ben-Akiva. 2015. "Stop Detection in Smartphone-Based Travel Surveys". *Transp. Res. Procedia* 11(January): 218–226. DOI: <https://doi.org/10.1016/j.trpro.2015.12.019>.

Received January 2020

Revised September 2020

Accepted December 2020