ScienceDirect

Transportation Research Interdisciplinary Perspectives
Volume 12, December 2021, 100501

App-based automatic collection of travel behaviour: A field study comparison with self-reported behaviour

Ingeborg Storesund Hesjevoll, Aslak Fyhri, Alice Ciccone

Institute of Transport Economics, Gaustadalléen 21, 0349 Oslo, Norway

Received 14 April 2021, Revised 17 July 2021, Accepted 31 October 2021, Available online 4 December 2021, Version of Record 4 December 2021.

Check for updates

Show less

Outline | Share Cite

https://doi.org/10.1016/j.trip.2021.100501
Get rights and content

Highlights

- Travel behaviour measured by survey and by automated app simultaneously.
- More registered activity by app than survey for all modes except public transport.
- App-survey differences vary with modes and level of aggregation.
- Concordance for modes registered varies between 61 and 87.8 %.
- Active modes diverge more for weekly than for one-day data.

Abstract

Smart phone apps hold great promise for travel behaviour research, but their performance relative to traditional methods is still not well understood. The aim of this study is to evaluate the magnitude and direction of differences between travel behaviour from a completely automatic travel mode detection mobile app and a traditional travel behaviour survey. We present data from n=230 participants who used the app (sense.dat) for four weeks. Participants also completed a one-day travel diary and one-week retrospective account of cycling and walking in the same period. Correspondence between app and survey varied across levels of aggregation and modalities. Overall, the app recorded substantially more km, minutes and non-zero trip days than the one-day survey, but when split up by mode this was not true for public transport. On the individual level there was a tendency for the app to register modes not self-reported by the respondents for all modes except public transport, possibly indicating that the app captures trips that the user may have forgot or intentionally left out. For bike, car and foot, the Spearman correlations between app and survey registered (one-day) distances and durations were moderate (r>0.5) or strong (r>0.8) when based on observations that were non-zero in both data sources, and moderate or weak when based on all observations. For one-week reports of active transport modes, app-survey correlations were lower than for the one-day data, especially for foot.

Previous Next

Keywords

Travel behavior; Smartphone application; Self-report; Travel survey data collection

1. Introduction

1.1. Motivation

The collection of accurate travel behaviour data is often a prerequisite for making informed decisions about transport policy, and for evaluating outcomes of transport related investments and incentive schemes. Recent and ongoing technological advancements have increased the number of tools available for collecting travel behaviour data. Knowledge on the benefits and weaknesses of different methods, and how their results compare is needed. In particular, smart phone apps have seen a large increase in use, and there is a need to better understand their potential and performance compared to traditional data collection methods.

Two of the main challenges reported in the literature are scaling up of data collection (in number of respondents and for longer periods of time) and recording the most accurate travel behaviour with the lowest possible respondent burden. This study addresses both issues using a smartphone app to automatically measure travel behaviour, and comparing these data for a specific day with a web survey where participants were asked to report a travel diary during the fall of 2018.

1.2. Background

Traditionally, travel behaviour data has been collected using self-reported, i.e. by asking participants to fill out travel diaries on paper, over phone or online. The accuracy of the resulting data relies heavily on respondents' ability and willingness to provide correct and precise information, and concerns about participant fatigue or attrition typically limit such surveys to one or very few days, and few or broad trip characteristics.

Replacing or supplementing traditional surveys with passive, automated data collection methods enables the collection of more detailed travel behaviour data over longer time periods while limiting the reliance on respondents' knowledge and memory (Harrison et al., 2020, Lee et al., 2016). Based on GPS tracking devices or smartphone apps, these methods typically combine GPS trajectories with other data to generate travel diaries either automatically (i.e. as a standalone approach) or semi-automatically (Prelipcean et al., 2018b). In semi-automatic approaches, users correct or validate the automatically collected data in prompted recall surveys. The richness, detail and potential for longitudinal data possible from GPS devices or travel apps hold potential utility for a wide range of applications, including transport modelling (Harrison et al., 2020), and research focused on specific transport modes such as infrastructure use or route choice for cycling (Romanillos et al., 2016), or longitudinal tracking of travel behaviour to evaluate the effects of interventions targeting mode change.

Reported accuracies for mode classifications for automated travel diaries appear to be relatively high. In a review of GPS travel survey methods, Shen & Stopher (2014) note that reported accuracy for mode detection is often 90 to 96 % in studies comparing data from dedicated GPS devices with traditional travel diaries or prompted recall surveys. However, some reports of travel mode accuracy refer only to trips that are successfully matched, which may exclude a substantial share of recorded trips (see e.g. Lee et al., 2016, Pronello and Kumawat, 2020) and mode detection accuracy may differ between modes and trip characteristics (Chen et al., 2010, Bohte and Maat, 2009, Geurs et al., 2015). Results may also differ based on the method used for inferring trip characteristics (e.g. mode, start and end times) from the raw data (Gong et al., 2014, Prelipcean et al., 2017, Sadeghian et al., 2021, Wu et al., 2016).

Studies based on automatic or semi-automatic travel diaries tend to record more trips than traditional self-report, which is interpreted as under-reporting of trips in surveys. This is found when comparing both app- and GPS-derived trips to both independent and concurrent surveys (Bohte and Maat, 2009, Flake et al., 2017, Faghih Imani et al., 2020, Kang et al., 2018, Lynch et al., 2019, McCool et al., 2021, Nguyen et al., 2017, Schüssler, 2010, Thomas et al., 2018, Zhao et al., 2015). However, the implied underreporting of survey trips varies between studies. For instance, Thomas et al. (2018) found a trip rate of 3.9 using an app combined with prompted recall, and 3.3 from regular self-report, whereas the app used by Faghih Imani et al. (2020) showed an average trip rate of 5.1 per day compared with 2.6 in a regional travel survey.

Automatic travel diary collections also appear to influence estimates of travel durations. Self-report is prone to rounding of travel times (e.g. to the nearest 5 or 10 min; Sato and Maruyama, 2020, Zhao et al., 2015), which is limited by automatic tracking. A systematic review of GPS-derived and self-reported trip durations found that self-report consistently results in higher trip duration estimates (+2.2 to +13.5 min; 9–75% longer than GPS recorded durations), possibly reflecting respondents rounding, and including other travel related activities when they report trip durations (Kelly et al., 2013). Similarly, Houston et al. (2014) found an average difference of 11.2 daily minutes between self-reported and GPS-derived walking. In contrast, the app used by McCool et al. (2021) found longer travel times (mean 1.1 h/day) compared to a self-reported survey (0.6 h/day). However, as McCool et al. (2021) used an independent census survey for comparison, the extent to which finding reflects the high level of aggregation or the use of an independent survey for this comparison is unclear.

The degree to which self-reported and automatically or semi-automatically derived travel diaries correspond varies based on the characteristics of respondents and trips. For instance, the degree of self-reported travel time rounding may differ between modes or trip purposes (Sato & Maruyama, 2020), and automatic and self-reported trip correspondence is associated with trip length (e.g. less "under reporting" of short trips (e.g. Faghih Imani et al., 2020, Thomas et al., 2018, Zhao et al., 2015, Schüssler, 2010), and with modality (Kang et al., 2018, Houston et al., 2014). Characteristics of respondents that have been found to influence the correspondence between survey reported and automatically recorded trips include age, employment status, travel frequency, education and income level, although results are not entirely consistent across studies (e.g. Bricka et al., 2012, Houston et al., 2014, Nguyen et al., 2017, Thomas et al., 2018). This indicates both that direct comparisons between studies should be done with caution, and that results differ based on the level of aggregation, such as whether indicators of correspondence (e.g. trip rates, durations) are split up by mode.

While automated and semi-automated travel diaries can reduce several of the drawbacks associated with self-report, there are challenges associated with these methods too. Many existing studies interpret the automatic or semi-automatic travel diaries as a ground truth, but accuracy can be limited by GPS signal loss or cold starts, and according to Stopher et al. (2015) prompted recall surveys are subject to several of the same biases as regular self-report surveys. Additionally, GPS devices require the users to bring and charge them, which may be problematic: Nguyen et al. (2017) found that 44% of survey-recorded trips were not recorded by GPS, and for 60 % of these trips, the reason for missingness was respondents not wearing the GPS unit. The potential for large-scale data collection with GPS devices is also limited by the cost associated with acquiring and distributing these devices (Rojas et al., 2016).

In contrast, the marginal cost of scaling data collections using dedicated smartphone apps is lower due to the high penetration rates of smartphones (Prelipcean et al., 2018b), and the risk of respondents forgetting to bring or charge their phone may be lower. Dedicated smartphone apps typically combine mobile based GPS data with other phone sensors such as accelerometers, which can improve detection of location and movement patterns, especially when GPS signal coverage is an issue (see e.g. Wang et al., 2018). In a recent review of smartphone apps to collect mobility data, Pronello and Kumawat (2020) identified 81 different such apps, noting that associated publications, when available, were often proof of concept studies or pilot studies.

Medium to large scale data collections with dedicated apps have been conducted in e.g. Sweden (Prelipcean et al., 2018a, Allström et al., 2016), USA (Lynch et al., 2019, Flake et al., 2017), Singapore (Zhao et al., 2015), Canada (Faghih Imani et al., 2020), and the Netherlands (Geurs et al., 2015, Thomas et al., 2018, McCool et al., 2021), most in combination with prompted recall surveys or validation prompts. As these studies differ in e.g. specific apps, samples, geographic contexts and validation approaches (when available), direct comparison of their results is not feasible. For instance, Zhao et al., 2015, McCool et al., 2021 compare findings with (independent) census data, the app evaluated by Geurs et al., 2015, Thomas et al., 2018 is analysed based on the prompted recall survey and the main focus of the study by Faghih Imani et al. (2020) is differences between two similar apps. Still, all studies illustrate the feasibility of recruiting relatively large samples to use such apps over multiple days, and that the resulting data is more detailed and extensive than what is available from survey approaches to studying travel behaviour. For instance, app-based trip detection may be especially helpful for limiting under-reporting of short trips (Faghih Imani et al., 2020, Thomas et al., 2018, Zhao et al., 2015) and irregular trips (Thomas et al., 2018). Validation results are also promising, but varying. The app/platform applied in the Netherlands was used by 615 respondents and had mode classification success rates between 78 % for the shortest trips and 95 % for the longest trips (Thomas et al., 2018), and in the Swedish case studies report 54 % accuracy in mode detection (Allström et al, 2016). In a 2016 experiment, Harding et al. (2021) used 13 different apps concurrently while also recording a ground truth. They found highly varying accuracy, with some apps showing substantial errors and others producing highly reliable travel diaries despite the experimental set of trips being highly random and apps not being calibrated to the city's public transport system.

Despite the large number of travel tracking apps available, knowledge is lacking on their performance both relative to a "ground truth" and to traditional methods (Pronello and Kumawat, 2020). The wide range of potential applications for automated travel diaries, and previous research indicating that correspondence between (semi) automated and self-reported data vary across levels of aggregation and analysis outcomes (e.g. Houston et al., 2014, Kang et al., 2018) indicate that different outcomes and levels of analyses in comparisons are needed as apps become increasingly popular.

Validating app-recorded behaviour with independent or prompted recall surveys is likely to provide the highest quality data (Harding et al., 2021, Pronello and Kumawat, 2020), but this approach implies a relatively high respondent burden, especially for longer data collection periods. Such a high degree of user involvement may not be feasible or desirable in all contexts, e.g. due to concerns about data quality (Safi et al., 2017), sample size, attrition and/or selection bias.

A possible solution is further improvement of apps that allow for completely automatic data collection, i.e. with limited or no required user interaction. While completely automated tracking may be less precise than combining it with prompted recall surveys, knowledge on how automated apps as a standalone approach compare to standalone surveys is limited. Prompted recall approaches are not necessarily comparable to independent surveys, and so the implications of using apps with minimal user interaction versus traditional surveys are not well understood.

1.3. Objectives

The current study compares travel behaviour data from an automatic app with minimal user interaction requirements to an independent self-reported travel behaviour survey for the same respondents on the same day(s). Our main aim is to assess the differences that would result from choosing an automated app as the main method of data collection, as compared to an independent travel behaviour survey. The current study contributes to the knowledge gap on how travel apps and traditional methods compare by assessing correspondence on different levels of aggregation, and for different travel behaviour indicators. Hence, the results can help researchers assess the magnitude and direction of differences that may be expected due to differences in these two measurement methods.

Specifically, this paper explores the following research question: What is the level of congruence between the app and the survey for mode detection, distance and duration at different levels of aggregation?

2. Materials and methods

The data used in the current study result from a large data collection starting in August 2018 in which a baseline survey was used to recruit respondents to three different research projects, and to use a smartphone app that automatically tracked travel behaviour over the course of five weeks. Each respondent was recruited to no more than one research project, and those recruited to a lottery field experiment (Ciccone et al., 2021) received a follow-up survey at the end of the period of app data collection, which included a travel diary. Respondents in the current study are thus a sub sample of participants in the lottery experiment.

2.1. Recruitment

The initial recruitment survey was distributed through Facebook, among members of the Norwegian Automobile Association, to previous survey respondents who had agreed to be contacted anew, and to a list of university students. Lottery participants were recruited among the respondents who agreed to use a smartphone app to track travel behaviour for the next few weeks. At the end of the recruitment survey, respondents received instructions on how to download the app from App Store and Google Play. They were also informed that among respondents who used the app for the whole 5-week period, one randomly selected person would win 1000 NOK (approx. € 96).

Then, the lottery was implemented as a 3-week field study, where participants' travel behaviour was continuously measured with the app (for details, see Ciccone et al., 2021). Lottery participants used the app from late August to late September. The follow-up survey containing the travel diary and items about weekly walking and biking was sent out towards the end of the 5-week period, after the lottery experiment. The current study compares responses to this travel diary with the app data registered from the same respondents on the corresponding date(s). The study was approved by the Norwegian Social Science Data Services, and all participants gave their informed consent before completing the questionnaire, and again before downloading the app.

2.2. Survey

The surveys were administered online, and the recruitment survey contained questions related to daily travel and factors explaining choice of travel mode, as well as demographics. The follow-up survey was shorter, and contained a travel diary section and a section on weekly trips by bike and foot which are used in the current study.

The travel diary asked respondents if they had any trips outside their home the previous day. The name of the pertaining weekday was displayed to avoid confusion about what "yesterday" referred to. Trips were defined by purpose, and the instructions for the initial question explicitly prompted the inclusion of short trips by foot or bicycle. Respondents who had at least one trip were then asked to fill in, for up to 6 trips: The main purpose of the trip (12 categories in a drop- down list, meant to aid recall), the mode of transport used (7 categories in a drop-down list), and the duration (minutes) and distance (km) of each trip.

For foot and bike, weekly trips were also included using items similar to the IPAQ methodology (Craig et al, 2003): Respondents were first asked which of the past seven days they had a bike trip that lasted at least 10 min. Weekdays were provided in reverse chronological order, starting with the weekday for the previous date. Next, respondents who chose at least one day were asked about the total number of minutes and km they biked for the selected days (with selected days listed). Afterwards, respondents indicated which of the past seven days they had walked for at least 10 min (if any). If one or more day(s) were selected, they were asked for how many minutes, and how many trips of at least 10 min, they typically made on the selected days (selected days listed in question).

2.3. App

The smartphone app used, Sense.DAT, uses GPS, Wi-Fi signals and sensors in the smartphone to automatically register travel behaviour. Once installed, travel is registered without requiring any user action, but users may view their automatically registered trips (in a list view, and in a map), and can approve and/or edit aspects of their data (e.g. the modality). Sense.DAT is an implementation of the Mobidot platform, which was also used by Geurs et al., 2015, Thomas et al., 2018, who describe the platform and processing approach in depth (but note that the prompted recall aspect was not used in the current study). In contrast to some earlier versions, the app integrated real-time timetables from the local public transport providers to aid mode detection. The data resulting from the app includes estimates per trip for e.g. duration (seconds), distance (meters), start time and date, and modality. For simplicity we refer to unimodal trip(legs) as trips, even though some registrations might be trip legs in multimodal trips. For instance, a trip by foot to catch the train is described as one trip by foot and one by train, not two trip legs of the same trip.

2.4. Data processing

Certain differences between the survey and the app could lead to systematic differences in their results even if there had been no measurement error in either. To limit the influence of known method-specific differences we aimed to exclude from each data source that which could in principle not be detected by the other. App registrations for tram, metro, bus, train, and ferry were collapsed to match the survey category “Public transport”, and two survey categories for car (passenger and driver) were collapsed to match the app. Trips by other modalities (“unknown” and boat in the app, motorcycle and unreported/missing in the survey) were excluded due to low prevalence; ten respondents had one such trip each in the app and/or survey. Note that ferry is part of the public transport system for which real-time timetables were used to aid modality classifications.

In the travel diary, automatic answer control in the survey defined the maximal allowed trip distance as 99 km. Thus, respondents likely set any longer trips to 99 km or refrained from reporting them. Accordingly, trips longer than 98 km were excluded from both data sources, and to better match the precision level from the survey, the app data were rounded to the nearest minute (from seconds) and 100 m (from meters). This data processing was conducted on the level of trips, but for the one-day data we use person-day as the unit of analysis: For each respondent, trips by the same modality were summarised to daily trips, km and minutes for both the app and the survey. This, too, was done to ensure comparable units between data sources: On the trip level there was insufficient information to match the data, with the app lacking information on trip purpose, and the survey lacking information on time and location. Additionally, differences in trip definition between sources (by purpose in survey and by changes in distinct movement patterns/modality in app) lead us to expect limited comparability at the level of trips. Note that while trips exceeding 98 km were removed from the data, person-days may exceed this limit for respondents with several long trips.

For the weekly survey data, respondents with impossible or highly implausible values were removed for each modality. This included reporting a daily number of (10+minute) trips higher than 10 (n=4, foot), reporting a higher number of km than minutes (n=4, bike), or a combination of minutes and km that indicated a speed higher than 35 km/h (n=8 bike). Next, the weekly survey data were summarized, for foot and bike, as number of days with trips of at least 10 min duration (0–7), as well as minutes. Minutes were reported for the sum of all days for bike, and for the “typical day” for foot. To calculate weekly minutes for foot, we multiplied the number of days with the minutes for the “typical day”, as per the IPAQ methodology (Craig et al, 2003). In addition to minutes travelled, respondents reported the weekly number of km by bike, and the number of trips by foot. Descriptive statistics for weekly km and trips are reported in Appendix B.

From the app, the seven days prior to each respondent’s survey response date (i.e. the days corresponding to the weekly survey items) were screened for trips of at least 10 min by foot or bike. Based on this, the number of days with at least one 10-minute trip by each modality was calculated, as well as the sum of minutes, km (bike only) and trips (foot only) recorded in trips of at least 10 min for these days.

With the app, short trips can be more error prone and the degree to which they are correctly registered can differ systematically between respondents, depending on factors such as regularity of travel, interactions with the app, and GPS coverage. This relates in part to the app activating from a battery-saving state when the phone is moving, which might not reliably occur on shorter trips, especially if they are irregular. To explore the implications of removing short trips from app, the analyses for the daily data were repeated on a dataset excluding trips shorter than 10 min/1 km. These results are available in Appendix A.

2.5. Analytical approach

Based on respondents’ reported trips in the survey and app registrations on the corresponding dates, we compare travel behaviour from the two data sources for both a one-day travel diary and a weekly survey. Note that for weekly data we only asked to recall trips by foot and bike. To assess the degree of correspondence between app and survey, we first compare the total sum of the data registered for each modality (aggregate level) for both one-day diary and weekly data. For the daily data, we also compare the mode split resulting from each data collection method.

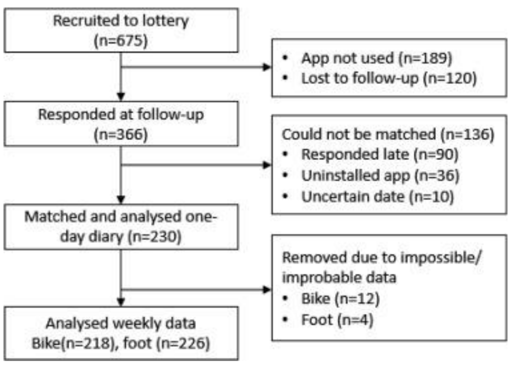
To examine any differences found on the aggregate level, we assess the overlap in mode classifications between the two methods, as well as correspondence for distance and duration (and trips for weekly walking). In this part of the analysis we use person-day or person-week unit of analysis, unless stated otherwise. Modality detections were classified as corresponding if, for a given participant and modality, a) at least one trip was registered both data sources, or b) no trip was registered in either data source. As some degree of correspondence would be expected due to mere chance, we report Cohen’s Kappa ($\boldsymbol{\kappa}$ unweighted; Cohen, 1960). Cohen’s $\boldsymbol{\kappa}$ is an indicator of inter-rater agreement on categorical variables, that accounts for the «agreement» that would be expected due to randomness. A $\boldsymbol{\kappa}$ of 0 represents the agreement expected due to mere chance and 1 indicates perfect agreement.

The degree of correspondence between app and survey registered durations and distances was assessed by correlations and mixed model regression analyses. Spearman correlations were interpreted according to the guidelines suggested by Ferguson (2016), with 0.8 representing a strong association, and 0.5 a moderate association. In the current context we consider correlations below 0.2 as not being practically significant. The regression analyses use reported minutes (and km) as the dependent variable, and data source (app or survey) and modality as predictors. To account for the non-independence of observations from the same respondent, we fitted (linear) mixed effect models, with random intercepts for participants using the R-package lme4 (Bates et al., 2014). To test whether any differences in app and survey registered km and min differed between modalities, we ran these analyses first for each modality and finally for all modalities jointly, with modality as a predictor. Both minutes and km were log-transformed before analyses. Marginal means were estimated used to interpret interaction effects in the regression analyses (Lenth, 2021). All data processing, analyses and data visualization were conducted with R version 4.0.2 (R Core Team, 2020, Wickham et al., 2019).

3. Results

3.1. Recruitment and sample

Among the n=675 recruited to the lottery and invited to use the app, 72 % downloaded the app. These lottery participants used the app for on average 28.3 days, and registered approximately seven trips per day (Ciccone et al., 2021). As seen in Fig. 1, n=366 both used the app and responded to the follow-up survey. Respondents were included in the analysis if they still had the app installed at the date corresponding to their survey travel diary response (which was also the last day of the weekly measure). This excluded n=136 respondents who completed the survey two or more days after the end of app data collection period, uninstalled the app early, or completed the survey across several days such that the date corresponding to the survey diary could not be determined.



Download : Download high-res image (182KB)

Download : Download full-size image

Fig. 1. Participant flowchart.

The respondents for which we could match app and survey data (n=230) all installed the app more than two weeks prior to the analysed date(s) and had, on average, the app installed for a total of 29.7 days (SD=3.6). The average number of trips registered across the period of app use was 153 per person (SD=54.9), and on average, respondents validated 24 % of their app-logged trips for the full period. The travel diary date was a Thursday for the majority of the respondents (n=134). Respodents' average age was 40.5 (SD=10), and n=113 (49%) were female. Respondents were recruited from three major Norwegian cities, and nearly 60% lived in the city of Oslo. As travel patterns and education levels differ between urban and rural areas of Norway, Table 1, Table 2 display age and education (Table 1) and mode shares (Table 2) for the sample and for the population of Oslo. The sample is highly educated, and older and younger age groups are underrepresented relative to the Oslo population.

Table 1. Distribution of age and education level for sample and for adults in Oslo in 2018. Percent.

	Sample	**Oslo**[1]
Age group		
18–25	7	13
26–35	22	26
36–45	41	19
46–55	21	15

	Sample	Oslo[1]
56–65	9	12
Higher education		
None	15	46
less than 5 years	31	32
5+ years	54	22

Note: [1]Statistics Norway (2018), refers to adults (>17 years). As only age groups present in the study sample are included, Oslo age groups do not sum to 100.

Table 2. Mode share (trips) for app, survey and the population of Oslo. Percent.

Mode	Survey	App	Oslo[1]
Bike	38	32	5
Foot	21	25	32
Car	22	29	37
Public transport	18	12	26
Other	0	1	0

Note: Bike shares in Norway vary by season, and are larger in the summer. [1] Hjorthol et. al (2014).

As a result of inclusion criteria for the lottery, all respondents owned or had access to a bicycle, and were either active in the workforce (92%) or students. In Norway, 75 % of the adult population owned or had access to a bicycle in 2014 (Hjorthol et al., 2014). A total of n=142 (62%) reported that they usually cycled at least 4 times a week, and an additional n=64 (28%) said they normally did so 1–4 times a week. More than three out of four (n=174, 77%) reported that their most common mode of transport to work was bike. Hence, based on the demographic composition of the sample we would not expect the travel behaviour of this sample to reflect that of the general population.

3.2. Correspondence for one-day travel diary

On the travel diary date, most respondents (n=183) registered at least one trip in both the app and the survey, while n=8 registered no trips in either. Activity registered only in the app was found for n=27 respondents, and the remaining n=12 registered trips in the survey only.

3.2.1. Aggregate correspondence for travel diary

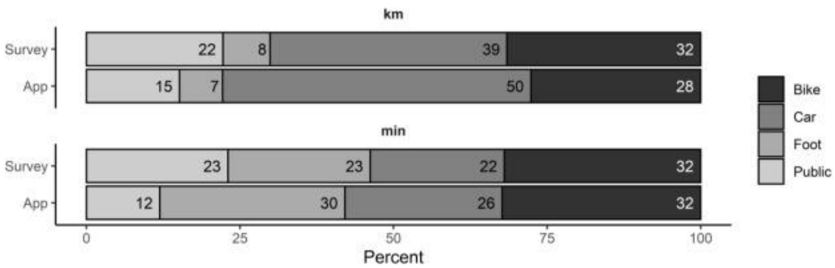
As displayed in Table 3, the median values for registered km and minutes per person per day are similar across data sources whereas mean values are higher for the app than the survey. In the survey, the prevalence of non-zero registrations is lower for km (n=182) than for minutes (n=195) due to some respondents not reporting km for trips for which they did report duration. Overall, 15 respondents registered at least one trip without reporting the associated distance, whereas no respondents registered trips without reporting duration. Trips where neither distance or duration were reported are not included in this material.

Table 3. Descriptive statistics for outcomes from app and survey, disregarding modality.

Data source	Outcome	Non-zero registrations						All registrations		
		Median	*Mean*	*SD*	*Min*	*Max*	*n*	*Mean*	*Median*	*SD*
App	Km	16	25.8	33.5	1	338	210	23.4	15	32.8
	Min	68	83.7	61.8	6	460	210	76.4	64	63.6
Survey	Km	16	20.8	18.9	1	141	182	16.5	13	18.8
	Min	60	68.3	43.8	2	280	195	57.9	52	47.2

Note: For «all registrations», n is 230 and minimum observations are zero for all outcomes.

Fig. 2 displays modality distributions for all registered km and minutes. At this aggregate level, the rank-order of modality distributions are the same for distance: the highest share of travelled km is by car, followed by bike, public transport, and lastly foot. For duration, the app captures lower shares of public transport activity than car driving, while in the survey users report a higher share of public transport activity than car driving. In both cases, there are clear differences between the modality distributions.

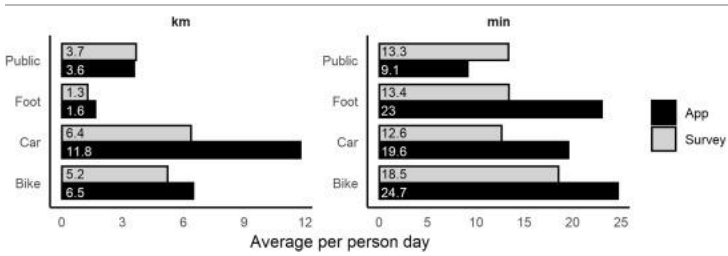


Download : Download high-res image (105KB)

Download : Download full-size image

Fig. 2. Modality distributions for app and survey. Shares across full sample.

Fig. 3 illustrates the average distances and durations registered by mode, indicating that the mode share differences observed in Fig. 2 largely reflect the app registering fewer minutes and km relative to the survey for public transport, but more for all other modes. The distances registered by car in the app are nearly twice as high as those registered by the survey. When short trips are excluded, the overall patterns do not change notably, with the exception that the difference between all app- registered and all survey-registered minutes by foot is smaller (Fig. A.2 in appendix).



Download : Download high-res image (98KB)

Download : Download full-size image

Fig. 3. Average km and minutes registered in each data source by modality. N=230.

3.2.2. Modality correspondence for travel diary

The degree to which the differences observed in section 3.2.1 occur due to the two data sources recording different distances and durations for the same activities, or different activities, cannot be assessed on the aggregate level. To explore the nature of the observed aggregate level differences, we investigate correspondence on the individual level in terms of modality classifications, and differences between app and survey registered distances and durations for respondents who have registered the same modality in both data sources.

Modality correspondence was assessed by the share of respondents with any activity registered in app only, survey only, both data sources or neither data source for each modality. Cohen's Kappa (κ unweighted; Cohen, 1960) was calculated to account for the degree of concordance that may be expected due to chance. As seen in Table 4, the number of concordant cases (i.e. the sum of observations registered in both or neither data source) make up between 61 and 88% of the observations, depending on modality. Foot has the lowest total concordance and κ, while public transport has the highest κ, with a substantial share of zero-concordant observations.

Table 4. Registrations of activity per modality and data source for travel diary data. Cohen's κ and percent for n=230. Row sums of "total" (or not total) columns are 100%.

		Percent concordant			Percent discordant		
		Total	*Of which*		Total	*Of which*	
Modality	κ		*Non-zero*	*zero*		*In survey*	*In app*
Bike	**0.52**	**75.7**	*38.7*	*37.0*	**24.3**	*4.8*	*19.6*
Car	**0.54**	**78.7**	*23.9*	*54.8*	**21.3**	*4.8*	*16.5*

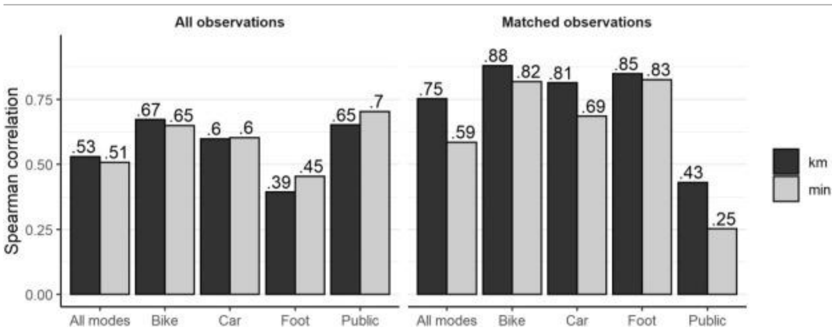
		Percent concordant			Percent discordant		
		Total	***Of which***		**Total**	***Of which***	
Modality	**κ**		***Non-zero***	***zero***		***In survey***	***In app***
Foot	**0.26**	**61.3**	*26.5*	*34.8*	**38.7**	*7.4*	*31.3*
Public	**0.70**	**87.8**	*21.7*	*66.1*	**12.2**	*5.2*	*7.0*
All modes	**0.20**	**83.1**	*79.6*	*3.5*	**16.9**	*5.2*	*11.7*

Among the discordant observations, registrations in the app only is 3–4 times higher than survey only for bike, car and foot. Excluding short trips (Table A.4) has the most notable effect for foot, indicating that the large share of respondents with trips by foot in the app only in Table 3 is partly explained by short trips. When these are removed the total concordance for foot increases to 80.4 % (κ= 0.52), mostly due to an increase in respondents with no trips by foot in either data source.

The results of Table 4 indicate that the higher registrations by the app relative to the survey (section 3.2.1) are explained at least in part by the app recording activity by modalities that the respondents do not report in the survey, and that this occurs to a lesser extent for public transport than for other modalities. To assess whether there are systematic differences in the recorded distances and durations for modes that are registered in both data sources, we assess the correspondence between app- and survey reported km and minutes for each modality separately both for all observations and for the concordant non-zero observations.

3.2.3. Correspondence for distance and duration in travel diary

The concordant non-zero observations (from here on "matched observations") make up a smaller share of app-registered data (69 % of all km) than survey registered data (88% of all km), with some differences between modalities (see Table A.1 for details). Hence, focusing on observations detected in both data sources excludes a larger share of app-registered than survey registered data. For most modes, this means that the differences between the total registered minutes and km for the two data sources are smaller; For car, where 1.9 times as many km were registered in app as in the survey for the full data (section 3.2.1), this ratio is 1.4 for the matched data. Similarly, the sum of minutes recorded by foot were 1.7 times higher in the app than in the survey for the full data, and 1.2 times higher in the matched data. Ratios of the total activity registered in the app to the survey by mode are available in the appendix (Table A.2). Fig. 4 displays Spearman correlation coefficients for the association between app and survey distance (kilometres) and duration (minutes) for matched observations and all observations.



Download : Download high-res image (161KB)
Download : Download full-size image

Fig. 4. Spearman correlations by outcome and modality. Matched observations represent person-modality combinations that are non-zero for both app and survey.

For bike, car and foot the associations between app and survey data are moderate (>0.5) or strong (>0.8) for the matched observations, and moderate or weak for all observations. For public transport, however, associations are moderate for all observations and weak for the matched observations. For all modalities jointly (sum of all registered km and minutes per person-day), the overall spearman correlations between app and survey data are moderate for both outcomes and datasets. When short trips are excluded, the correlations for foot for all observations are above the threshold for moderate for both minutes and km (Fig. A.3).

The mixed model regression analyses conducted to assess the influence of data source and modality on the registered durations and distances for matched data are presented in Tables 5 (distance) and 6 (duration). The control variables included age, gender, city of residence (reference: Oslo), travel behaviour assessed in the recruitment survey (distance travelled and number of modes used), respondents' app approvals (share of all registered trips for the full period) and the average number of app-registered trips per day for the full app use period. The control variables were added in steps to assess whether they influenced the estimates for mode and data source. Finally, subgroup regression analyses were conducted for each modality separately. As the number of observations were low for these subgroup analyses, the models included only a predictor for data source (reference level survey) and a random intercept for each participant.

Table 5. Regression models for (log) km. One-day matched data.

		Model 1		Model 2		Model 3	
	Predictor	**Estimate**	**95 % CI**	**Estimate**	**95 % CI**	**Estimate**	**95 % CI**
	Intercept	2.74	**2.51;2.98**	2.91	**2.62;3.19**	2.87	**2.59;3.16**
Data source	**App**	0.23	−0.05;0.52	0.23	−0.06;0.52	0.23	−0.05;0.52
Mode	**Bike**	−0.42	**−0.70;-0.13**	−0.44	**−0.73;-0.15**	−0.41	**−0.70;-0.12**
	Foot	−1.74	**−2.05;-1.42**	−1.76	**−2.07;-1.44**	−1.74	**−2.06;-1.43**
	Public	−0.07	−0.04;0.27	−0.08	−0.42;0.26	−0.07	−0.41;0.27
Mode* data source	**Bike: App**	−0.16	−0.51;0.20	−0.16	−0.52;0.20	−0.16	−0.52;0.20
	Foot: App	−0.17	−0.56;0.23	−0.17	−0.56;0.23	−0.17	−0.56;0.23
	Public: App	−0.78	**−1.20;-0.36**	−0.78	**−1.21;-0.36**	−0.78	**−1.21;-0.36**
Demographics	**Age[1]**			−0.01	−0.02;0.00	−0.01	−0.02;0.00
	Female			−0.25	**−0.46;-0.03**	−0.21	−0.42;0.01
	City B			−0.01	−0.3;0.28	−0.07	−0.37;0.23
	City T			−0.05	−0.31;0.21	−0.01	−0.27;0.26
Travel beh	**Distance[2]**					0.00	0.00;0.01
	Modes[2]					0.04	−0.11;0.18
App use	Approvals[2]					0.00	0.00;0.00
	Avg. trips[2]					0.05	−0.01;0.11
Random effects							
	σ^2	0.54		0.54		0.54	
	τ_{00}	0.28		0.26		0.26	
	ICC	0.34		0.33		0.33	

Note: Confidence intervals in bold are significant at $p<.05$. [1]Age centered at 40years old. [2]Variables mean scaled before analyses. Each model included 164 respondents with a total of 472 observations.

Table 6 displays the regression results for minutes. Estimates for modality and data source were largely unaffected by controlling for background variables. Estimated marginal means for mode and data source, based on model 3 (Table 6), are displayed in Fig. 6.

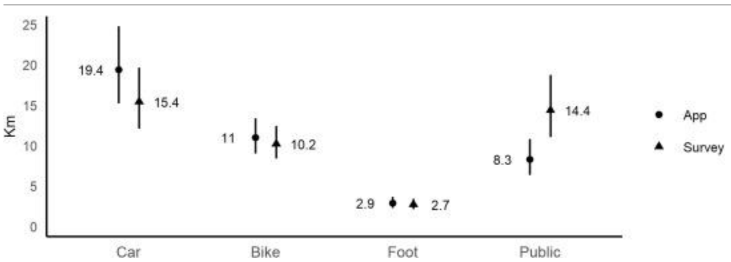
Table 6. Regression models for (log) daily minutes. One-day matched data.

		Model 1		Model 2		Model 3	
	Predictor	**Estimate**	**95 % CI**	**Estimate**	**95 % CI**	**Estimate**	**95 % CI**
	Intercept	3.48	**3.28; 3.68**	3.52	**3.28; 3.76**	3.49	**3.25; 3.73**
Data source	**App**	0.19	−0.05; 0.43	0.19	−0.05; 0.43	0.19	−0.05; 0.43
Mode	**Bike**	0.18	−0.06; 0.42	0.18	−0.07; 0.42	0.21	−0.04; 0.45
	Foot	−0.08	−0.34; 0.17	−0.09	−0.35; 0.17	−0.07	−0.33; 0.19
	Public	0.30	**0.03; 0.57**	0.29	**0.02; 0.57**	0.31	**0.03; 0.58**
Mode* data source	**Bike:App**	−0.10	−0.41; 0.2	−0.10	−0.41; 0.20	−0.10	−0.41; 0.20

		Model 1		Model 2		Model 3	
	Predictor	**Estimate**	**95 % CI**	**Estimate**	**95 % CI**	**Estimate**	**95 % CI**
	Foot:App	0.05	−0.28; 0.37	0.05	−0.28; 0.37	0.05	−0.28; 0.37
	Public:App	−0.74	**−1.09; -0.40**	−0.74	**−1.09; -0.4**	−0.74	**−1.08; -0.4**
Demographics	**Age[1]**			0.00	−0.01; 0.01	0.00	−0.01; 0.01
	Female			−0.06	−0.24; 0.12	−0.04	−0.23; 0.15
	City B			0.09	−0.16; 0.34	0.05	−0.20; 0.31
	City T			−0.09	−0.31; 0.14	−0.05	−0.28; 0.18
Travel beh.	**Distance[2]**					0.00	0.00; 0.00
	Modes[2]					0.02	−0.1; 0.14
App use	Approvals[2]					0.00	0.00; 0.00
	Avg. trips[2]					0.05	0.00; 0.11
Random effects							
	σ^2	0.40		0.41		0.40	
	τ_{00}	0.21		0.21		0.22	
	ICC	0.34		0.34		0.35	

Note: Confidence intervals in bold are significant at p<.05. [1]Centered at 40 years old. [2]Variables mean scaled before analyses. Each model includes 174 respondents with 510 observations.

The main effects of modality for km (Table 5) indicate that longer distances were registered by car (reference level) than by bike and foot. The main effect of data source indicates that the app registers approximately 26 % more km than the survey (i.e. exp(0.23)-1=0.26) for car (reference level modality), but this effect is not statistically significant. The mode by data source interaction term is significant only for public transport, and both estimates and confidence intervals for modality and data source are stable when background variables are controlled for. Fig. 5 displays the estimated marginal means for modality and data source based on Model 3 (Table 5). The patterns observed in Fig. 5 were also found in the subgroup analyses; App registered km were lower than those registered in the survey for public transport, whereas the opposite was true for the remaining modes. In the subgroup analyses for km, the difference between app- and survey registered distances were statistically significant for public transport and car (Table A.10).

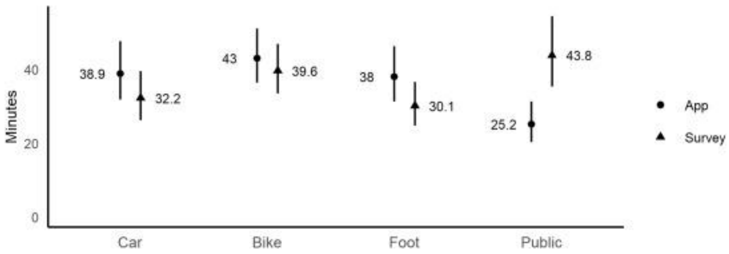


Download : Download high-res image (51KB)

Download : Download full-size image

Fig. 5. Estimated marginal means and 95% CI for daily travel distance by mode and data source. Matched observations.

For minutes, the interaction between data source and transport mode was only significant for public transport, where more activity was registered in the survey than in the app. For the remaining modes, there is a tendency for the app to register more than the survey. The subgroup analyses showed the same pattern as dispayed in Fig. 6, although the difference between the app and survey registered minutes were statistically significant for both foot and public transport (Table A.11)



Download : Download high-res image (54KB)

Download : Download full-size image

Fig. 6. Estimated marginal means and 95% CI for daily travel duration by mode and data source. Matched observations.

3.3. Correspondence for weekly data

3.3.1. Aggregate correspondence

For the weekly survey data, 78 % (n=170) reported at least one day with trips (of at least 10min duration) by bike, and 87 % (n=197) reported at least one day of walking. From the app, the number of respondents with registrations on at least one day was 88% for bike and 84% for foot. As displayed in Table 7, the average registered activity higher for the survey than for the app for all outcomes.

Table 7. Descriptive statistics for non-zero weekly registrations for days and minutes by bike and foot.

Data source		Bike					Foot				
	Outcome	**Mean**	**SD**	**Median**	**Range**	**n**	**Mean**	**SD**	**Median**	**Range**	**n**
App	Days	3.5	1.6	4	1–7	192	2.8	1.6	2	1–7	190
	Minutes	156	130	135	12–1009	192	138	127.0	102	11–850	190
Survey	Days	4.0	1.7	4	1–7	166	4.4	2.2	4	1–7	180
	Minutes	175	146	150	10–999	166	165	203	105	10–1400	180

3.3.2. Modality correspondence

With regard to modality detection, three out of four respondents reported travelling by bike for one or more days in both the survey and the app, and these shares are fairly similar for foot (Table 8). For bike, the pattern of non-corresponding cases is similar to the travel diary: the share of respondents with trips registered only by the app is more than three times higher than the share of respondents with trips only registered in the survey. For foot, the share of respondents with trips in the survey only is higher than the opposite. Hence, the survey registering a higher total number of days, minutes and trips by foot than the app (Table 7) is probably in part explained by 11 % of respondents reporting walking in the survey but not in the app. Cohen's Kappa was 0.23 for foot and 0.36 for bike.

Table 8. Registrations of (any) activity per modality and data source for weekly data. Percent for n=230. Row sums of "total" (or not total) columns are 100%.

	Percent concordant			**Percent discordant**		
	Total	***Of which***		**Total**	***Of which***	
Modality		***Non-zero***	***zero***		***In survey***	***In app***
Bike (n=218)	**81.7**	*73.9*	*7.8*	**18.3**	*4.1*	*14.2*
Foot (n=226)	**81.0**	*76.1*	*4.9*	**19.1**	*11.1*	*8.0*

3.3.3. Correspondence for distance and duration.

The app-survey correlations displayed in Table 9 show that the associations are moderate for days and minutes by bike, whereas those for foot are smaller. In contrast to the one-day diary this is true for both matched and all observations.

Table 9. Spearman correlations for app and survey for weekly data by modality.

	Matched observations				All observations			
	Bike		Foot		Bike		Foot	
	days	min	days	min	days	min	days	min
rsp	0.66	0.68	0.34	0.29	0.68	0.72	0.46	0.45
n	161	161	172	172	218	218	226	226

Regression models similar to those for the one-day data (Table 5, Table 6) were run for the wekly minutes and trip days by foot and bike. For brevity, the full regression results are presented in the appendix (Table A.11, Table B1). For both minutes and trip days, results were unaffected by controlling for background variables. For minutes, the difference between survey and app-registered data was not statistically significant, and neither was the interaction between modality and data source. For trip days there was a significant effect of modality, as well as a modality by data source interaction. The estimated marginal means for minutes and trip days for the one-week data are presented in Fig. 7. For minutes, the difference in estimated marginal means are less than 5 min for bike and approximately 19 min for foot. For trip days, the app-registered number of days is notably lower than that registered in the survey for foot, whereas the difference for bike is smaller.

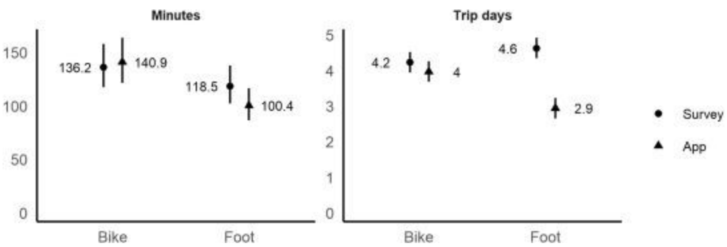


Fig. 7. Estimated marginal means and 95% CI for weekly minutes and trip days by modality and data source.

3.4. Summary of results

The current study presents findings from a large-scale automated collection of travel behaviour data, based on a subset of users (n=230) who completed an independent travel behaviour survey while using the smartphone app. Our analyses compare survey and app data for one day where users responded to a travel diary, and for one week where they responded to walking and biking activity. On average, respondents used the app for about four weeks.

For the one-day data, the minutes and km registered by the app greatly exceeded that registered in the survey, both overall and for each transport mode except public transport. Regarding modality detection, we found concordance (a given mode registered in both or neither data source) for between 61 and 88% of respondents, depending on the mode. For foot, car and bike we found a clear tendency for the app to register activity for more respondents than the survey, whereas this was not found for public transport.

For bike, car and foot the correlations between app and survey data are moderate (>0.5) or strong (>0.8) for respondents who had the mode registered in both data sources, and moderate or weak when based on all observations. For respondents who have the same mode registered in both data sources, the differences in estimated marginal means for app and survey are below 1 km and 10 min for foot and bike. Public transport had the highest level of correspondence for modality detection, largely explained by the low prevalence of such trips in both data sources. However, for matched public transport observations, correlations between app and survey registered minutes and km were far lower than for other modalities, and the survey reported distances and durations exceeded those recorded by the app.

The exclusion of short trips has the largest influence for foot; the relative “overreporting” of the app becomes smaller, and the shares of respondents with walking only in the app is reduced, leading to larger overall correspondence in modality detection. In the weekly data, more trips days and minutes were registered in the survey than in the app for foot and bike. For respondents with registrations in both data sources, the app-survey correlations for the weekly data were moderate (about 0.07) for bike and weak (about 0.3) for foot.

4. Discussion

There are several well-known trade-offs in choosing data collection methods for travel research. Detailed knowledge of the consequences of choosing different methods are important both for the planning of studies, and for the synthesis of studies using different methods. In comparing results from two different data collection methods used for the same respondents at the same time the current study highlights some interesting aspects of how the measures correspond.

The current study differs from earlier related work in ways that prevents direct comparisons, such as the use of an independent (vs. prompted recall) survey, the specific app (/version) used, the level of analysis (person-days versus trips), sample, and geographical area, which can all influence the association between self-reported and automatically recorded travel behaviour. Still, some general similarities are worth noting. The correspondence found for modality detection for the daily data in this study is slightly lower than that found for trip-rate correspondence for previous versions of the Mobidot platform when validated data was compared to prompted recall surveys (Geurs et al., 2015, Thomas et al., 2018). The higher prevalence of app only (vs. survey only) registrations for most modalities, and the app recording more daily minutes and longer daily distances overall, is in line with previous research indicating that automated travel diary collection results in registration of more trips than traditional self-report (e.g. Sammer et al., 2018). There are several plausible reasons why this may occur, including recall issues (e.g. forgetting a trip, or reporting a trip on the wrong day). Furthermore, since the app potentially tracks all movement, activity that users might not classify as trips in the survey, such as walking a dog or biking for exercise purposes could contribute to these tendencies, especially for walking and cycling.

One interesting finding is that across outcomes and analyses, results for public transport deviate from those of the other modalities for the daily data (less commonly registered in the app only, and lower durations and distances in the app than the survey for matched observations). The sample size is limited when data are split up by modality, limiting the certainty by which we can draw substantial conclusions. Still, there are several plausible reasons for why we would observe this. First, some respondents may have misreported the public transport part of a chained trip, or over reported durations to include delays, waiting times, access and egress trips. Second, respondents may have less precise knowledge of the distance travelled as duration is typically stated by trip/route planning tools. Third, challenges with mode classifications specific to the app could be important (Geurs et al., 2015), however the app version used in the current study incorporates real-time public transport timetables to improve accuracy. While it may be that the app misclassifying modes contributes to the diverging pattern, the results when not splitting up the data by modality follow the pattern of the app registering activity for more respondents than the survey, possibly indicating that misclassified modes cannot fully explain our results. In future research aiming to track all transport modes (e.g. Pronello and Kumawat, 2020, Winkler et al., 2020) it is important to ensure that measurement are equally valid across transport modes.

Weekly trips by foot and bike was assessed with items similar to the IPAQ methodology (Craig et al, 2003). The app-survey correlations for foot were small in size, which is in line with studies that compare IPAQ-assessed walking with pedometer/accelerometer data (Lee et al. 2011): Spearman correlations range between 0.08 and 0.57. Lee et al. (2011) also found tendencies for IPAQ to result in higher estimates of physical activity relative to other "objective" measures. Although such comparisons were not available for walking specifically, they seem to correspond to our finding that the weekly data for walking show longer durations reported in the survey than registered in the app, both in the aggregate and on the individual level. The lower degree of correspondence for the weekly measures than for the one-day diary are unsurprising, and plausibly related to the greater difficulty of recalling behaviour for a week than for the past day.

For (weekly) days and minutes by bike, however, the aggregate registrations were more similar for app and survey. This higher degree of correspondence for weekly bicycling in our study is likely related to the large share of bicycle commuters in our sample; regular trips might be easier for the respondents to recall, and possibly for the app to register correctly. Furthermore, the lottery incentivised reporting cycling in the app, which may have prompted participants to pay closer attention to how often and how long they went by bike.

The degree of correspondence between GPS-derived and self-reported travel behaviours may be associated with demographics (e.g. Houston et al., 2014, Geurs et al., 2015), and the accuracy of automatically generated trip diaries is associated with trip characteristics (Thomas et al., 2018). Although we found no influence of participant characteristics on the associations between matched app and survey recorded distances and durations, our sample was representative. Accordingly, our results might not generalise to samples with different demographic compositions or travel patterns. Specifically, the high share of bicycle commuters, and the previous participation in an intervention focused on bicycling could be relevant: It may be that the observed differences between modes in part reflect their frequency of use in the current sample. Additionally, older age groups and unemployed people were underrepresented in the current study. These groups have previously been shown to display lower correspondence between automated and self-reported travel diaries (Bricka et al., 2012, Houston et al., 2014, Thomas et al., 2018), potentially indicating that lower correspondence may be expected in representative samples. Recruiting large-scale representative samples is a general challenge for collection of app-based travel behaviour (Silvano et al., 2020, Verzosa et al., 2021, Lee and Sener, 2020), and exploring how recruitment rates can be increased or results weighted to account for underrepresented groups is an important focus for future research.

It is important to note that this study employed a particular app with a specific technology, and may not generalize to other tools. Moreover, the app relies on many types of input and is sensitive to a series of parameters such as GPS signal coverage and public transport information availability. In all automated data collection methods, raw GPS traces require substantial processing to derive relevant parameters such as modes, ends and purposes of trips. As noted in the introduction, several different methods for doing so have been proposed and used in the literature, and the choice of data processing methods likely influence results. As the app used in the current study is proprietary, details on the pre-processing algorithms used are not available.

The association between automated and self-reported travel behavior is also likely to vary depending on how the survey is conducted, which varies across studies and sub-fields of transportation research. For instance, national surveys typically rely on geolocation and start and end times, whereas reviews of interventions of travel behavior commonly include studies where participants explicitly report numbers of trips, durations or distances (e.g. Doğru et al., 2021, Arnott et al., 2014, Semenescu et al., 2020). In the current study, trip distances and durations were asked directly, and results may not transfer to other self-report methods.

The large share of app-recorded distances and durations that could not be matched to the survey on the person-modality level (approx. 30 % of all registered minutes and km) clearly indicates that the measurement error in either data source is substantial. Without assuming that either data source reflects respondents' true travel behaviour, we cannot reasonably conclude to which degree observed differences between the data sources are due to measurement errors in one or the other. Presumably, measurement error is present in both the app and the survey, and potential sources of systematic bias may differ between them and have different effects. Previous research indicates that automated travel diaries can capture short trips that respondents fail to self-report, but also that short trips are more error prone when using the platform implemented in sense.Dat (Thomas et al., 2018). In our study, excluding short trips increased correspondence for foot, and also reduced the share of app-registered distances and durations that could not be matched (to 11 % of app km, and 17 % of app minutes). Unfortunately, the degree to which this reflects short trips being forgotten by the respondents versus short trips having high measurement error cannot be determined based on the current data.

Even in the absence of measurement error perfect correspondence would not be expected due the previously noted tendency for respondents to underreport trips as well as differences in e.g. how trips are defined (as patterns of movement vs. by purpose). This is, however, not a challenge unique to the current study, as previous studies comparing self-reported and device-recorded travel behaviour have used similar trips definitions (see e.g. Kelly et al., 2013), and our research interest was in comparing the output of automated data collection with that which would otherwise have been the preferred data collection method, i.e. a self-reported travel behaviour survey.

5. Conclusions

Our results indicate that replacing a one-day travel survey with an automated travel behaviour app results in registering more activity (km and minutes) both overall and for most transport modes. At the individual level, the app and the survey had a concurrence rate between seventy-five and ninety percent for car, public transport and cycling, and just above sixty percent for walking. For weekly travel by foot and bike, correspondence between self-reported and app- registered duration was comparable to previous research based on other "objective" tracking devices, but lower than the correspondence we observed with the one-day travel diary.

Perfectly corresponding measurements between app and survey data are not expected, and what constitutes an *acceptable* level of correspondence will vary depending on the area of application. The current study illustrates that the correspondence between self-reported and automatically registered travel behaviour data can differ based on the level of analysis, modalities, and data processing (i.e. inclusion of short trips). For the matched data, the app reports fairly similar results to the traditional survey. Results for public transport did, however, not align with the pattern of results found for other modes. This discrepancy may stem from recollection problems or may indicate that trip registration by the app for public transport remains problematic even though the real-time API of the public transport providers were included in the modality classification algorithm. Special attention should be placed on this modality in future research.

The choice of data collection method is a complicated issue, and while fully automated data collection may be less precise than the additional use of independent or prompted recall surveys, there are other clear benefits of the automated approach that might justify its use in some cases. This includes e.g. continuous collection of multi-week travel behaviour data, the collection of location data and minimising respondent burden. The current study illustrates that multi-week data collection can be feasible with a high participation rate and low rate of drop out. This is important for any study as it reduces the issues related to self-selection bias. However, for results to be comparable with self-reported data, some degree of respondent involvement, and careful consideration of short trips appears to be required.

Some important caveats should be considered before replacing travel diaries with an automated app, relating to the main goal of the data collection. First, using the app to measure interventions directed towards specific modes may require another setup than using it for a general travel behaviour study, as the app can apparently be tweaked to better collect data for given modes. Second, the differences between modes in the agreement between app and survey implies that choosing one or the other data collection method might be more consequential for studies focusing on mode shares. Third, even though such apps may attempt to infer trip purpose, this information is not always reliable and requires validation from users. This is the reason why it was not included in the current study, and we suggest it should be a focus in future research. To improve our understanding of the implications of using automated data collection methods versus traditional surveys, more research is needed to assess whether these findings generalise to other automatic travel behaviour apps, and other versions of travel behaviour surveys.

CRediT authorship contribution statement

Ingeborg Storesund Hesjevoll: Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. **Aslak Fyhri:** Conceptualization, Funding acquisition, Methodology, Investigation, Project administration, Writing – review & editing. **Alice Ciccone:** Conceptualization, Methodology, Investigation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Funding: This research was funded from the Norwegian Research Council's Transport 2025.

research program (project number 267855). The funding source(s) had no involvement in study design, in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Appendix A.

One-day matched data

Table A.1 displays, for each modality and outcome, the share of the total observations that are included in the “matched” data. Note that the matched data refers to concordant non-zero observations, i.e. person-modality combinations with non-zero registrations in both the survey and the app. For bike, 91% of survey-registered km are included in the matched data, whereas the same is true for 76 % of the app-registered km.

Table A.1. Data included in matched subset. Percent of total sum of registered km and min by data source and modality.

	Km		**Min**	
Mode	**app**	**survey**	**app**	**survey**
Bike	76	91	74	91
Car	69	91	72	84
Foot	56	72	58	82
Public	56	85	74	85
All modes	69	88	68	86

Table A.2 displays the ratio of the sum of app-registered to survey-registered data for each modality based on the full and the matched data. Again, data are considered to be matched when a mode is registered (non-zero) in both data sources for the same respondent. For car, the sum of all recorded km were 1.85 times higher in the app than in the survey, while for the matched data the app-recorded km were 1.4 times higher than the survey-recorded km. Hence, even for respondents who had car recorded in both the app and the survey, the app-recorded distances are higher than those recorded in the survey.

Table A.2. Ratios of all registered activity in app to all registered activity in survey for all and matched data.

	All data		**Matched data**	
Mode	**km**	**min**	**km**	**min**
Bike	1.25	1.33	1.04	1.08
Car	1.85	1.55	1.40	1.32
Foot	1.30	1.72	1.01	1.22
Public	0.97	0.68	0.65	0.59
All modes	1.42	1.32	1.10	1.05

One-day results excluding short trips

This section repeats the analyses of the one-day data based on a dataset filtered by the same criteria as the weekly data, i.e. excluding trips shorter than 10 min and/or 1 km from either data source, as well as any impossible or highly improbable trips. The specific criteria are outlined in section 2.4 (data processing) in the main text.

The shares of the total data set included in the appendix is illustrated in Table A.3. These data make up approximately 92 % of the survey-registered minutes and km, and a somewhat smaller share of app data, especially for minutes. The modality most affected by removing short trips is foot, where approximately 70 % of the app-registered walking remains (Table A.3).

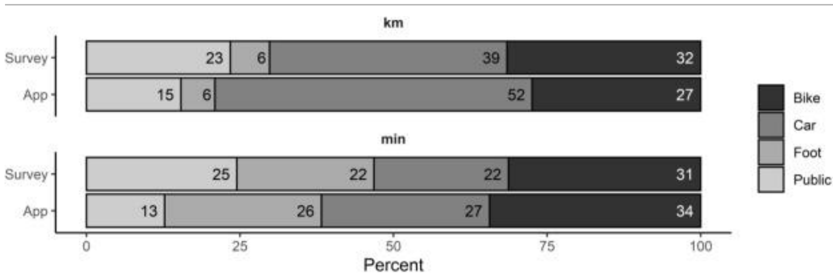
Table A.3. Data retained when excluding short trips, by modality and data source. Percent of total km and min registered in the main dataset.

	Km		Min	
Mode	**app**	**survey**	**app**	**survey**
Bike	88.8	92.4	88.9	90.5
Car	92.0	92.7	89.4	92.8
Foot	70.3	76.8	70.8	89.2
Public	90.7	97.5	89.2	98.3
All modes	89.4	92.4	83.6	92.5

Table A.4. Descriptive statistics for minutes and km from app and survey, disregarding modality. Short trips excluded.

		Non-zero registrations						***All registrations***		
Data source	**Outcome**	***Median***	***Mean***	***SD***	***Min***	***Max***	***n***	**Mean**	**Median**	**SD**
App	**Km**	14	24.0	32.5	1	322	201	20.9	13	31.4
	Min	58	73.1	56.8	10	432	201	63.9	51	58.4
Survey	**Km**	16	20.2	18.5	1	141	173	15.2	11	18.3
	Min	60	66.6	43.0	10	280	185	53.5	45	46.7

Note: For «all registrations», n is 230 and minimum observations are zero for all outcomes.



Download : Download high-res image (103KB)

Download : Download full-size image

Fig. A.1. Modality distributions app and survey. Shares across full sample.

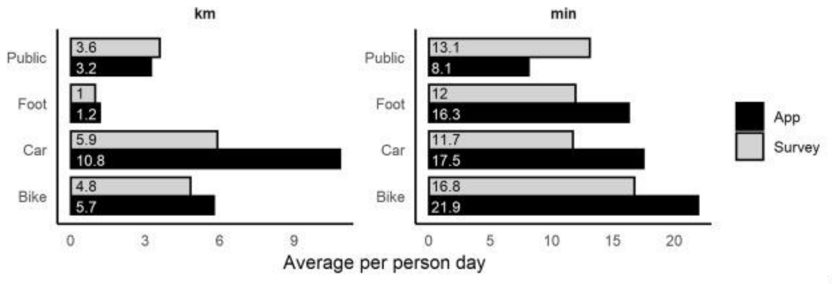


Fig. A.2. Average km and minutes registered in each data source by modality, excluding short trips. N=230 for each modality and data source.

Table A.5. Registrations of activity per modality and data source for travel diary data. Percent for n=230. Row sums of "total" (or not total) columns are 100%.

		Percent concordant			**Percent discordant**		
		Total	***Of which***		**Total**	***Of which***	
Modality	**κ**		***Non-zero***	***zero***		***In survey***	***In app***
Bike	**0.59**	**79.5**	*35.2*	*44.3*	**20.4**	*4.3*	*16.1*
Car	**0.50**	**79.1**	*18.7*	*60.4*	**20.9**	*6.1*	*14.8*
Foot	**0.52**	**80.4**	*18.7*	*61.7*	**19.6**	*7.0*	*12.6*
Public	**0.62**	**85.6**	*17.8*	*67.8*	**14.4**	*8.3*	*6.1*
All modes	**0.36**	**82.6**	*75.2*	*7.4*	**17.4**	*5.2*	*12.2*

Table A.6. Ratio of all registered activity in app to all registered activity in survey for all and matched data.

	All data		**Matched data**	
Mode	**km**	**min**	**km**	**min**
Bike	1.20	1.31	0.99	1.05
Car	1.84	1.49	1.28	1.20
Foot	1.19	1.36	0.93	1.09
Public	0.90	0.62	0.62	0.57
All modes	1.37	1.19	1.03	0.98

Note: both matched and total data here refer to data excluding short trips.

Table A.7. Data included in matched subset. Percent of total sum of registered km and min by data source and modality.

	Km		**Min**	
Mode	**app**	**survey**	**app**	**survey**
Bike	75	91	72	90
Car	62	90	63	78
Foot	60	78	63	79
Public	54	77	71	78
All modes	64	86	67	82

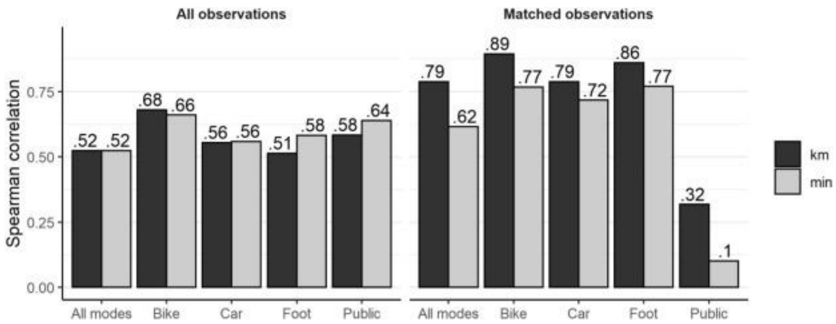


Fig. A.3. Spearman correlation coefficients for the association between app and survey, by modality and outcome.

Table A.8. Regression results for log(minutes) for matched one-day data excluding short trips.

		Model 1		Model 2		Model 3	
	Predictor	**Estimate**	**95 % CI**	**Estimate**	**95 % CI**	**Estimate**	**95 % CI**
	(Intercept)	3.67	**3.50;3.85**	3.72	**3.51;3.93**	3.72	**3.51;3.94**
Data source	**App**	0.09	−0.12;0.29	0.09	−0.12;0.29	0.09	−0.11;0.29
Mode	**Bike**	−0.00	−0.21;0.21	0.00	−0.21;0.22	0.02	−0.19;0.23
	Foot	0.12	−0.11;0.36	0.14	−0.10;0.37	0.14	−0.09;0.38
	Public	0.22	−0.02;0.46	0.23	−0.01;0.47	0.21	−0.03;0.46
Mode* data source	**Bike:App**	−0.09	−0.34;0.16	−0.09	−0.34;0.16	−0.09	−0.34;0.16
	Foot:App	−0.02	−0.3;0.27	−0.02	−0.30;0.27	−0.02	−0.30;0.27
	Public:App	−0.65	**−0.94;-0.35**	−0.65	**−0.94;-0.35**	−0.65	**−0.94;-0.36**
Demographics	**Age***			0.01	0.00;0.01	0.01	0.00;0.01
	Female			−0.11	−0.26;0.05	−0.11	−0.27;0.05
	City B			−0.03	−0.24;0.19	−0.01	−0.23;0.20
	City T			0.00	−0.20;0.19	0.02	−0.18;0.21
Travel beh	**Distance***					0.00	0.00;0.00
	Modes*					−0.07	−0.18;0.04
App use	Share approvals*					0.00	0.00;0.00
	Trips per day*					0.05	**0.00;0.10**
Random effects							
	σ^2	0.23		0.23		0.23	
	τ_{00}	0.15		0.16		0.15	
	ICC	0.4		0.4		0.4	

Note: n 159, 416 observations

Table A.9. Regression results for log(km) for matched one-day data excluding short trips.

	Predictor	Model 1 Estimate	Model 1 95 % CI	Model 2 Estimate	Model 2 95 % CI	Model 3 Estimate	Model 3 95 % CI
	(Intercept)	3.04	**2.83;3.25**	3.17	**2.92;3.42**	3.16	**2.91;3.41**
Data source	**App**	0.14	−0.09;0.37	0.14	−0.10;0.38	0.14	−0.09;0.37
Mode	**Bike**	−0.67	**−0.91;-0.42**	−0.67	**−0.92;-0.43**	−0.66	**−0.91;-0.41**
	Foot	−1.60	**−1.88;-1.31**	−1.60	**−1.88;-1.31**	−1.59	**−1.88;-1.31**
	Public	−0.30	**−0.59;0.00**	−0.29	−0.58;0.00	−0.30	**−0.59;-0.01**
Mode* data source	**Bike:App**	−0.17	−0.46;0.11	−0.17	−0.46;0.11	−0.17	−0.46;0.11
	Foot:App	−0.25	−0.59;0.09	−0.25	−0.59;0.09	−0.25	−0.59;0.09
	Public:App	−0.60	**−0.95;-0.26**	−0.60	**−0.95;-0.26**	−0.60	**−0.95;-0.26**
Demographics	**Age***			0.00	−0.01;0.01	0.00	−0.01;0.01
	Female			−0.26	**−0.45;-0.07**	−0.25	**−0.44;-0.05**
	City B			−0.07	−0.33;0.20	−0.06	−0.33;0.22
	City T			0.04	−0.20;0.27	0.06	−0.17;0.30
Travel beh	**Distance***					0.00	0.00;0.00
	Modes*					−0.06	−0.19;0.07
App use	Share approvals*					0.00	0.00;0.00
	Trips per day*					0.06	0.00;0.12
Random effects							
	σ^2	0.29		0.29		0.29	
	τ_{00}	0.24		0.22		0.23	
	ICC	0.45		0.44		0.44	

Note: n 149, 382 observations.

Table A.10. Regression analyses and estimated marginal means by mode for km.

	Bike		Car		Public transport		Foot	
	Est.	95% CI	Est.	95% CI	Est.	95% CI	Est.	95% CI
Intercept	2.29	2.14;2.44	2.77	2.46;3.08	2.59	2.34;2.85	0.95	0.7;1.2
Data: app	0.07	−0.03; 0.18	0.23	0.05; 0.41	−0.55	−0.84; 0.26	0.06	−0.09; 0.22
Random effects								
σ2	0.13		0.22		0.46		0.18	
T00	0.39		1 0.05		0.25		0.72	
ICC	0.75		0.83		0.35		0.80	
respondents	87		51		43		55	
observations	174		102		86		110	
Survey	9.9(8.5;1.5)	15.9(11.6;21.8)	13.4(10.4;17.3)	2.6(2.0;3.3)				
App	10.6(9.1;12.4)	20(14.6;27.5)	7.7(6.0;10)	2.8(2.1;3.6)				

Note: Each column represents a separate analysis. Confidence intervals in bold are significant at p<.05.

Table A.11. Regression analyses and estimated marginal means by mode for minutes.

	Bike		Car		Public transport		Foot	
	Estimate	**95% CI**	**Estimate**	**95% CI**	**Estimate**	**95% CI**	**Estimate**	**95% CI**
Intercept	3,6	**3.47;3.73**	3,49	**3.26;3.72**	3,75	**3.55;3.95**	3,35	**3.12;3.58**
Data: app	0,08	**−0.02;0.19**	0,19	**−0.01;0.39**	−0,55	**−0.79;-0.31**	0,23	**0.08;0.38**
Random effects								
σ2	0,12		0,28		0,38		0,18	
T00	0,27		0,5		0,14		0,65	
ICC	0,69		0,64		0,27		0,79	
respondents	87		55		50		61	
observations	174		110		100		122	
Survey	36.7(32.2;41.9)		32.7(25.8;41.5)		42.5(34.7;52)		28.5(22.6;36)	
App	39.9(35;45.5)		39.5(31.2;50.1)		24.5(20;30)		36(28.6;45.5)	

Note: Each column represents a separate analysis. Confidence intervals in bold are significant at p<.05.

Appendix B. . Weekly data.

Regression analyses

Subgroup analyses

Table B1. Mixed model regression analysis for trip days, matched data.

		Model 1		Model 2		Model 3	
Predictors		**Estimate**	**95 % CI**	**Estimate**	**95 % CI**	**Estimate**	**95 % CI**
	Intercept	**4.12**	**3.85;4.39**	**4.04**	**3.69;4.38**	**4.03**	**3.68;4.37**
Modality	**Foot**	0.48	**0.12;0.83**	0.42	**0.07;0.78**	0.40	**0.04;0.75**
Data source	**App**	−0.26	−0.61;0.08	−0.26	−0.61;0.08	−0.26	−0.61;0.08
Modality*data	**Foot:App**	−1.45	**−1.93;-0.96**	−1.42	**−1.91;-0.94**	−1.42	**−1.91;-0.94**
Demographics	**Age[1]**			−0.01	−0.03;0.01	−0.01	−0.03;0.01
	Female			0.13	−0.19;0.45	0.16	−0.17;0.49
	City B			0.42	−0.03;0.87	0.37	−0.09;0.83
	City T			−0.04	−0.43;0.35	0.00	−0.40;0.41
Travel beh.	**Distance[2]**					0.00	−0.01;0.01
	Modes[2]					0.10	−0.11;0.32
App use	**Approvals[2]**					0.00	−0.01;0.00
	Avg. trips[2]					0.07	−0.03;0.17
Random effects							
	σ^2	2.42		2.39		2.39	
	τ_{00}	0.55		0.56		0.56	
	ICC	0.18		0.19		0.19	

Note: Estimates in bold are significant at p<.05 [1]Centered at 40years old. [2]Variables mean scaled before analyses. Each model includes 211 respondents with 624 observations.

Table B2. Mixed model regression analysis for one-week log(minutes), matched data. Estimates and CI back-transformed to original scale.

		Model 1		Model 2		Model 3	
Predictors		**Estimate**	**95 % CI**	**Estimate**	**95 % CI**	**Estimate**	**95 % CI**
	Intercept	**4.87**	**4.37;5.01**	**4.84**	**4.66;5.02**	**4.82**	**4.64;5.00**
Mode	**Foot**	−0.11	−0.29;0.07	−0.13	−0.31;0.05	−0.14	−0.32;0.04
Data source	**App**	0.03	−0.14;0.21	0.03	−0.14;0.21	0.03	−0.14;0.21
Mode*data	**Foot:App**	−0.21	−0.45;0.04	−0.20	−0.45;0.05	−0.20	−0.45;0.05
Demographics	**Age[1]**			0.00	0.00;0.01	0,00	0.00;0.01
	Female			0.03	−0.14;0.21	0.07	−0.11;0.24
	City B			0.19	−0.05;0.43	0.15	−0.09;0.40
	City T			−0.04	−0.25;0.17	−0.01	−0.22;0.21
Travel beh.	**Distance[2]**					0,00	0.00;0.00
	Modes[2]					0,00	−0.12;0.11
App use	**Approvals[2]**					0,00	0.00;0.00
	Avg. trips[2]					0.05	−0.01;0.1
Random effects							
	σ2	0.62		0.61		0.61	
	τ00	0.19		0.18		0.18	
	ICC	0.23		0.23		0.23	

Note: Estimates in bold are significant at p<.05 [1]Centered at 40years old. [2]Variables mean scaled before analyses. Each model includes 212 respondents with 628 observations.

Weekly trips and km

Descriptive results for weekly km by bike and weekly trips by foot are presented in Table B3, Table B4. Note that the one-week recall items did not include number of trips by bike, nor km by foot.

Table B3. Descriptive statistics for non-zero weekly registrations for km by bike and trips by foot.

	Bike (*n*=218), km		Foot (*n*=226), trips	
	app	survey	app	survey
Mean	36.6	40.2	3.8	7.2
SD	37.7	50.6	3.6	7.6
Median	25	25	3	5
Range	0–203	0–400	0–18	0–40

Table B.4 presents spearman correlations and difference scores for weekly trips and km. Difference scores were created for each person-modality unit by subtracting the survey-registered trips (or km) from those registered by the app. Hence, a negative difference score implies that the survey registered more than the app.

Table B4. Spearman correlation for app and survey. All observations.

	Matched observations		All observations	
Statistic	**Km, bike**	**Trips, foot**	**Km, bike**	**Trips, foot**
rsp	0.74	0.27	0.76	0.41

Statistic	Matched observations		All observations	
	Km, bike	Trips, foot	Km, bike	Trips, foot
n	161	172	218	226

References

Allström et al., 2016 A. Allström, G. Gidofalvi, I. Kristoffersson, A.C. Prelipcean, C. Rydergren, Y.O. Susilo, J. Widell
Experiences from smartphone based travel data collection: system development and evaluation
Accessed from (2016)
https://fudinfo.trafikverket.se/fudinfoexternwebb/Publikationer/Publikationer_003001_003100/Publikation_003043/SPOT_final_report_v1.0.pdf ↗
Google Scholar ↗

Arnott et al., 2014 B. Arnott, L. Rehackova, L. Errington, F.F. Sniehotta, J. Roberts, V. Araujo-Soares
Efficacy of behavioural interventions for transport behaviour change: systematic review, meta-analysis and intervention coding
International journal of behavioral nutrition and physical activity, 11 (1) (2014), pp. 1-23, 10.1186/s12966-014-0133-9 ↗
View PDF ↗ Google Scholar ↗

Bates et al., 2014 D. Bates, M. Mächler, B. Bolker, S. Walker
Fitting linear mixed-effects models using lme4
Journal of Statistical Software, 67 (2014), pp. 1-48
https://doi.org/10.18637/jss.v067.i01 ↗
View in Scopus ↗ Google Scholar ↗

Bricka et al., 2012 S.G. Bricka, S. Sen, R. Paleti, C.R. Bhat
An analysis of the factors influencing differences in survey-reported and gps-recorded trips
Transportation Research Part C: Emerging Technologies, 21 (1) (2012), pp. 67-88, 10.1016/j.trc.2011.09.005 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Bohte and Maat, 2009 W. Bohte, K. Maat
Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands
Transportation Research Part C: Emerging Technologies, 17 (3) (2009), pp. 285-297, 10.1016/j.trc.2008.11.004 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Chen et al., 2010 C. Chen, H. Gong, C. Lawson, E. Bialostozky
Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study
Transportation Research Part A: Policy and Practice, 44 (10) (2010), pp. 830-840, 10.1016/j.tra.2010.08.004 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Ciccone et al., 2021 A. Ciccone, A. Fyhri, H.B. Sundfør
Using Behavioral Insights to Incentivize Cycling: Results from a Field Experiment
Journal of Economic Behavior & Organization, 188 (2021), pp. 1035-1058, 10.1016/j.jebo.2021.06.011 ↗
View PDF View article Google Scholar ↗

Cohen, 1960 J. Cohen
A coefficient of agreement for nominal scales
Educational and psychological measurement, 20 (1) (1960), pp. 37-46, 10.1177/001316446002000104 ↗
View article ↗ Google Scholar ↗

CRAIG et al., 2003 CORA.L. CRAIG, ALISON.L. MARSHALL, MICHAEL SJ??STR??M, ADRIAN.E. BAUMAN, MICHAEL.L. BOOTH, BARBARA.E. AINSWORTH, MICHAEL PRATT, ULF EKELUND, AGNETA YNGVE, JAMES.F. SALLIS, PEKKA OJA

International Physical Activity Questionnaire: 12-country reliability and validity
Medicine and Science in Sports and Exercise., 35 (8) (2003), pp. 1381-1395, 10.1249/01.MSS.0000078924.61453.FB ↗
View article ↗ Google Scholar ↗

Doğru et al., 2021 O.C. Doğru, T.L. Webb, P. Norman
What is the best way to promote cycling? A systematic review and meta-analysis
Transportation Research Part F: Traffic Psychology and Behaviour, 81 (2021), pp. 144-157, 10.1016/j.trf.2021.06.002 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Ferguson, 2016 C.J. Ferguson
An effect size primer: A guide for clinicians and researchers
A.E. Kazdin (Ed.), Methodological issues and strategies in clinical research, American Psychological Association (2016), pp. 301-310, 10.1037/14805-020 ↗
View article ↗ Google Scholar ↗

Flake et al., 2017 L. Flake, M. Lee, K. Hathaway, E. Greene
Use of smartphone panels for viable and cost-effective GPS data collection for small and medium planning agencies
Transportation Research Record, 2643 (1) (2017), pp. 160-165, 10.3141/2643-17 ↗
View article ↗ View in Scopus ↗ Google Scholar ↗

Geurs et al., 2015 K.T. Geurs, T. Thomas, M. Bijlsma, S. Douhou
Automatic trip and mode detection with move smarter: First results from the Dutch mobile mobility panel
Transportation Research Procedia, 11 (2015), pp. 247-262, 10.1016/j.trpro.2015.12.022 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Gong et al., 2014 L. Gong, T. Morikawa, T. Yamamoto, H. Sato
Deriving personal trip data from GPS data: A literature review on the existing methodologies
Procedia-Social and Behavioral Sciences, 138 (2014), pp. 557-565, 10.1016/j.sbspro.2014.07.239 ↗
View PDF View article Google Scholar ↗

Harding et al., 2021 C. Harding, A. Faghih Imani, S. Srikukenthiran, E.J. Miller, K. Nurul Habib
Are we there yet? Assessing smartphone apps as full-fledged tools for activity-travel surveys
Transportation, 48 (5) (2021), pp. 2433-2460, 10.1007/s11116-020-10135-7 ↗
View PDF ↗ View in Scopus ↗ Google Scholar ↗

Harrison et al., 2020 G. Harrison, S.M. Grant-Muller, F.C. Hodgson
New and emerging data forms in transportation planning and policy: Opportunities and challenges for “Track and Trace” data
Transportation Research Part C: Emerging Technologies, 117 (2020), p. 102672, 10.1016/j.trc.2020.102672 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Hjorthol et al., 2014 R. Hjorthol, Ø. Engebretsen, T.P. Uteng
2013/14 National travel survey – key results
Institute of Transport Economics, Oslo, Norway (2014)
Google Scholar ↗

Houston et al., 2014 D. Houston, T.T. Luong, M.G. Boarnet
Tracking daily travel; assessing discrepancies between gps-derived and self-reported travel patterns
Transportation Research Part C: Emerging Technologies, 48 (2014), pp. 97-108, 10.1016/j.trc.2014.08.013 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Faghih Imani et al., 2020 A. Faghih Imani, C. Harding, S. Srikukenthiran, E.J. Miller, K. Nurul Habib
Lessons from a Large-Scale Experiment on the Use of Smartphone Apps to Collect Travel Diary Data: The “City Logger” for the Greater Golden Horseshoe Area
Transportation research record, 2674 (7) (2020), pp. 299-311
View article ↗ CrossRef ↗ View in Scopus ↗ Google Scholar ↗

Lynch et al., 2019 J. Lynch, J. Dumont, E. Greene, J. Ehrlich
Use of a Smartphone GPS Application for Recurrent Travel Behavior Data Collection

Transportation Research Record, 2673 (7) (2019), pp. 89-98, 10.1177/0361198119848708 ↗
View article ↗ View in Scopus ↗ Google Scholar ↗

Kang et al., 2018 M. Kang, A.V. Moudon, P.M. Hurvitz, B.E. Saelens
Capturing fine-scale travel behaviors: a comparative analysis between personal activity location measurement system (PALMS) and travel diary
Int J Health Geogr, 17 (2018), p. 40, 10.1186/s12942-018-0161-9 ↗
View PDF View article Google Scholar ↗

Kelly et al., 2013 P. Kelly, P. Krenn, S. Titze, P. Stopher, C. Foster
Quantifying the difference between self-reported and global positioning systems-measured journey durations: A systematic review
Transport Reviews, 33 (4) (2013), pp. 443-459, 10.1080/01441647.2013.815288 ↗
View PDF ↗ View in Scopus ↗ Google Scholar ↗

Lee et al., 2011 P.H. Lee, D.J. Macfarlane, T.H. Lam, S.M. Stewart
Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review
International Journal of Behavioral Nutrition and Physical Activity, 8 (1) (2011), p. 115, 10.1186/1479-5868-8-115 ↗
View PDF View article Google Scholar ↗

Lee and Sener, 2020 K. Lee, I.N. Sener
Emerging data for pedestrian and bicycle monitoring: Sources and applications
Transportation research interdisciplinary perspectives, 4 (2020), p. 100095, 10.1016/j.trip.2020.100095 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Lee et al., 2016 R.J. Lee, I.N. Sener, J.A. Mullins
An evaluation of emerging data collection technologies for travel demand modeling: From research to practice
Transportation Letters, 8 (4) (2016), pp. 181-193, 10.1080/19427867.2015.1106787 ↗
View article View in Scopus ↗ Google Scholar ↗

Lenth, 2021 R.V. Lenth
emmeans: Estimated Marginal Means, aka Least-Squares Means
R package version, 1 (6) (2021), p. 1
https://CRAN.R-project.org/package=emmeans ↗
Google Scholar ↗

McCool et al., 2021 D. McCool, P. Lugtig, O. Mussmann, B. Schouten
An App-Assisted Travel Survey in Official Statistics: Possibilities and Challenges
Journal of Official Statistics, 37 (1) (2021), pp. 149-170, 10.2478/jos-2021-0007 ↗
View article ↗ View in Scopus ↗ Google Scholar ↗

Nguyen et al., 2017 Nguyen, T. T., Armoogum, J., Madre, J. L., & Pham, T. H. T. 2017. GPS and travel diary: Two recordings of the same mobility. In ISCTSC, 11th International Conference on Transport Survey Methods (p. 13p). https://hal.archives-ouvertes.fr/hal-01588420/ Accessed July 16, 2021.
Google Scholar ↗

Prelipcean et al., 2017 A.C. Prelipcean, G. Gidófalvi, Y.O. Susilo
Transportation mode detection–an in-depth review of applicability and reliability
Transport reviews, 37 (4) (2017), pp. 442-464, 10.1080/01441647.2016.1246489 ↗
View article View in Scopus ↗ Google Scholar ↗

Prelipcean et al., 2018a A.C. Prelipcean, G. Gidófalvi, Y.O. Susilo
MEILI: A travel diary collection, annotation and automation system
Computers, Environment and Urban Systems, 70 (2018), pp. 24-34, 10.1016/j.compenvurbsys.2018.01.011 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Prelipcean et al., 2018b A.C. Prelipcean, Y.O. Susilo, G. Gidófalvi
Collecting travel diaries: Current state of the art, best practices, and future research directions

Transportation Research Procedia, 32 (2018), pp. 155-166, 10.1016/j.trpro.2018.10.029 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Pronello and Kumawat, 2020 C. Pronello, P. Kumawat

K. Arai, S. Kapoor, R. Bhatia (Eds.), Intelligent Systems and Applications. IntelliSys 2020. Advances in Intelligent Systems and Computing, 1251, Springer, Cham (2020), 10.1007/978-3-030-55187-2_35 ↗

View article ↗

R Core Team, 2020 R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Google Scholar ↗

Rojas et al., 2016 M.B. Rojas, E. Sadeghvaziri, X. Jin

Comprehensive review of travel behavior and mobility pattern studies that used mobile phone data

Transportation Research Record, 2563 (1) (2016), pp. 71-79, 10.3141/2563-11 ↗

View article ↗ View in Scopus ↗ Google Scholar ↗

Romanillos et al., 2016 G. Romanillos, M. Zaltz Austwick, D. Ettema, J. De Kruijf

Big data and cycling

Transport Reviews, 36 (1) (2016), pp. 114-133, 10.1080/01441647.2015.1084067 ↗

View article View in Scopus ↗ Google Scholar ↗

Safi et al., 2017 H. Safi, B. Assemi, M. Mesbah, L. Ferreira

An empirical comparison of four technology-mediated travel survey methods

Journal of Traffic and Transportation Engineering (English edition), 4 (1) (2017), pp. 80-87, 10.1016/j.jtte.2015.12.003 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Sadeghian et al., 2021 P. Sadeghian, J. Håkansson, X. Zhao

Review and evaluation of methods in transport mode detection based on GPS tracking data

Journal of Traffic and Transportation Engineering (English Edition), 8 (4) (2021), pp. 467-482, 10.1016/j.jtte.2021.04.004 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Sammer et al., 2018 G. Sammer, C. Gruber, G. Roeschel, R. Tomschy, M. Herry

The dilemma of systematic underreporting of travel behavior when conducting travel diary surveys – a meta- analysis and methodological considerations to solve the problem

Transportation Research Procedia, 32 (2018), pp. 649-658, 10.1016/j.trpro.2018.10.006 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Sato and Maruyama, 2020 Y. Sato, T. Maruyama

Modeling the Rounding of Departure Times in Travel Surveys: Comparing the Effect of Trip Purposes and Travel Modes

Transportation Research Record., 2674 (10) (2020), pp. 628-637, 10.1177/0361198120935435 ↗

View article ↗ View in Scopus ↗ Google Scholar ↗

Schüssler, 2010 N. Schüssler

Accounting for similarities between alternatives in discrete choice models based on high-resolution observations of transport behaviour (Doctoral thesis)

ETH Zurich (2010), 10.3929/ethz-b-000263353 ↗

View article ↗ Google Scholar ↗

Semenescu et al., 2020 A. Semenescu, A. Gavreliuc, P. Sârbescu

30 Years of soft interventions to reduce car use–A systematic review and meta-analysis

Transportation research part D: transport and environment, 85 (2020), p. 102397, 10.1016/j.trd.2020.102397 ↗

View PDF View article View in Scopus ↗ Google Scholar ↗

Shen and Stopher, 2014 L. Shen, P.R. Stopher

Review of gps travel survey and gps data-processing methods

Transport Reviews, 34 (3) (2014), pp. 316-334, 10.1080/01441647.2014.903530 ↗

View PDF ↗ View in Scopus ↗ Google Scholar ↗

Silvano et al., 2020 A.P. Silvano, J. Eriksson, P. Henriksson
Comparing respondent characteristics based on different travel survey data collection and respondent recruitment methods
Case Studies on Transport Policy, 8 (3) (2020), pp. 870-877, 10.1016/j.cstp.2020.05.015 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Statistics Norway, 2018 Statistics Norway
Population by sex and age groups
Accessed July 17 2021
https://www.ssb.no/en/statbank/table/07459/ ↗ (2018)
Google Scholar ↗

Stopher et al., 2015 P.R. Stopher, L. Shen, W. Liu, A. Ahmed
The challenge of obtaining ground truth for GPS processing
Transportation Research Procedia, 11 (2015), pp. 206-217, 10.1016/j.trpro.2015.12.018 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Thomas et al., 2018 T. Thomas, K.T. Geurs, J. Koolwaaij, M. Bijlsma
Automatic trip detection with the Dutch mobile mobility panel: Towards reliable multiple-week trip registration for large samples
Journal of Urban Technology, 25 (2) (2018), pp. 143-161, 10.1080/10630732.2018.1471874 ↗
View PDF ↗ View in Scopus ↗ Google Scholar ↗

Verzosa et al., 2021 N. Verzosa, S. Greaves, C. Ho, M. Davis
Stated willingness to participate in travel surveys: a cross-country and cross-methods comparison
Transportation, 48 (3) (2021), pp. 1311-1327, 10.1007/s11116-020-10096-x ↗
View PDF ↗ View in Scopus ↗ Google Scholar ↗

Wang et al., 2018 Z. Wang, S.Y. He, Y. Leung
Applying mobile phone data to travel behaviour research: A literature review
Travel Behaviour and Society, 11 (2018), pp. 141-155, 10.1016/j.tbs.2017.02.005 ↗
View PDF View article View in Scopus ↗ Google Scholar ↗

Wickham et al., 2019 H. Wickham, M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani
Welcome to the tidyverse
Journal of Open Source Software, 4 (43) (2019), p. 1686, 10.21105/joss10.21105/joss.01686 ↗
View article ↗ Google Scholar ↗

Winkler et al., 2020 Winkler, C., Meister, A., Schmid, B., & Axhausen, K. W. (2020). TimeUse+: A comprehensive app-based longitudinal study of travel, time use, and expenditures. Working Paper, 1542, Institute for Transport Planning and Systems (IVT), ETH Zurich, Zurich.
Google Scholar ↗

Wu et al., 2016 L. Wu, B. Yang, P. Jing
Travel mode detection based on GPS raw data collected by smartphones: a systematic review of the existing methodologies
Information, 7 (4) (2016), p. 67, 10.3390/info7040067 ↗
View article ↗ Google Scholar ↗

Zhao et al., 2015 F. Zhao, F.C. Pereira, R. Ball, Y. Kim, Y. Han, C. Zegras, M. Ben-Akiva
Exploratory analysis of a smartphone-based travel survey in Singapore
Transportation Research Record, 2494 (1) (2015), pp. 45-56, 10.3141/2494-06 ↗
View article ↗ View in Scopus ↗ Google Scholar ↗

Emerging trends in household travel survey programs

2023, Social Sciences and Humanities Open

Show abstract

The effects of a subvention scheme for e-bikes on mode share and active mobility

2022, Journal of Transport and Health

> *Citation Excerpt :*
>
> …By contrast to the argumentation that self-reported measures often lead to underreporting of walking and cycling (Handy et al., 2014). A recent study by Storesund Hesjevoll, Fyhri, and Ciccone (2021), found that the app (Sense.DAT) overall recoded substantially more kilometres for cycling than a one-day self-report travel survey (i.e. when comparing the same days). However, these participants were motivated to correct wrongly identified cycling trips in the app.…

Show abstract

Image-based research methods for mapping tourist behaviour: smart photos

2023, Tourism Recreation Research

The Effects of a Subvention Scheme for E-Bikes on Mode Share and Overall Bicycle Usage

2021, SSRN