A national-scale mobility pricing experiment using GPS tracking and online surveys in Switzerland

Response rates and survey method results

Working Paper

Author(s):
Molloy, Joseph; Castro Fernández, Alberto; Götschi, Thomas; Schoeman, Beaumont; Tchervenkov, Christopher; Tomic, Uros; Hintermann, Beat; Axhausen, Kay W.

Publication date:
2020-08

Permanent link:
https://doi.org/10.3929/ethz-b-000441958

Originally published in:
Arbeitsberichte Verkehrs- und Raumplanung 1555

A National-Scale Mobility Pricing Experiment using GPS Tracking and Online Surveys in Switzerland: Response Rates and Survey Method Results

Joseph Molloy
IVT, ETH Zurich, Switzerland
Email: joseph.molloy@ivt.baug.ethz.ch

Alberto Castro
WWZ, University of Basel, Switzerland
Email: alberto.castrofernandez@unibas.ch

Thomas Götschi
WWZ, University of Basel, Switzerland
Email: thomas.goetschi@unibas.ch

Beaumont Schoeman
WWZ, University of Basel, Switzerland
Email: beaumont.schoeman@unibas.ch

Christopher Tchervenkov
IVT, ETH Zurich, Switzerland
Email: christopher.tchervenkov@ivt.baug.ethz.ch

Uros Tomic
INE, Zurich University of Applied Sciences, Switzerland
Email: uros.tomic@zhaw.ch

Beat Hintermann
WWZ, University of Basel, Switzerland
Email: beat.hintermann@unibas.ch

Kay W. Axhausen
IVT, ETH Zurich, Switzerland
Email: axhausen@ivt.baug.ethz.ch

Word Count: 5941 words + 6 table(s) $\times$ 250 = 7441 words

Submission Date: August 1, 2020

ABSTRACT
This article presents the first results and observations from the MOBIS Study, a nation-wide mobility pricing field experiment in Switzerland. Mobility pricing is widely regarded as a promising policy measure to combat congestion, internalize external costs of transport, and offset decreasing fuel tax revenues. However, the implementation of mobility pricing in Switzerland is hindered by a lack of empirical evidence, among other things. In the field experiment participants participated through the use of a GPS tracking app, Catch-my-Day, which logged their daily travel on different transport modes and imputed the trip segments and modes. The experiment lasted 8 weeks, bookended by online surveys. After the first 4 week control phase, participants were split into three treatment groups. The first continued as a control. The second received information on their external costs, and the third received a real monetary budget, from which their external costs were deducted. The first results show that the technology is capable of supporting such an experiment on both Android and iOS, the two main mobile platforms. Significant differences in the engagement and attrition were observed between iOS and Android participants over the 8 week period. Finally, the attrition rate did not vary between treatment groups.

Keywords: GPS tracking, mobility pricing, external costs, response rates, mobility behaviour

INTRODUCTION

Mobility pricing is widely regarded as a promising policy measure to combat congestion, internalize external costs of transport, and offset decreasing fuel tax revenues. The concept of mobility pricing was first proposed in the 1920's as an example of a corrective tax to internalize congestion externalities (1). Since then, there has been much study of the topic, including mathematical theory (2, 3) and simulation experiments (4, 5, 6). Most of the research and practical implementations have focused specifically on road pricing, which is a limited form of mobility pricing that focuses on drivers. Despite the theoretical capabilities to maximise infrastructure utilisation, mobility pricing has only been sparsely implemented in practice as it is typically viewed as a 'new tax' and is thus associated with strong political resistance. Schemes in London (7, 8) and Stockholm (9) are two well-known examples where limited mobility pricing has been implemented in the form of congestion charges: Cars entering the central business district during certain hours have to pay a fee. These 'congestion charges' don't reflect all the external costs from all modes of transportation. Schemes have also been implemented in a number of cities including Milan, Paris, Rome, Stuttgart and Singapore.

Although there is evidence on the success of congestion pricing (7, 8, 9), understanding the effects of broader mobility pricing schemes remains a challenge. A key challenge is understanding the potential impacts of the proposed policy. Multiple studies have looked at route, mode and destination choice within the context of various pricing schemes using stated-preference experiments (10, 11, 12). Work on the acceptance of pricing schemes includes Vrtic et al. (13), Jakobsson et al. (14). More recently, the proliferation of affordable GPS tracking and mobile connectivity has opened up the possibilities to do field experiments exploring transport users' behavioural responses under a pricing scheme, which would have been financially and logically infeasible in the pre cell-phone era. In one of the first examples, Nielsen (15) equipped 500 cars with a GPS-based device, and monitored participants for a control period before exposing them to a pricing scheme for the Copenhagen region. This study was in the pre-smartphone era and hence limited to a small sample size and no control group. A similar study using car-based GPS loggers was performed in Melbourne, in which 1,400 toll road users experienced different types of congestion charges (16, 17). A period of several months was used to monitor baseline behaviour before the pricing schemes were introduced for three quarters of the sample. In both these experiments, only car trips with the primary household vehicle were tracked. Public transport and active modes were not recorded. The Melbourne study did investigate possible modal shifts to rail commuting, by identifying car trips and subsequent parking at railway stations. The study reportd that 30% of participants reported changing their road travel use under the pricing scheme. Until now there have been no studies that have attempted to use smartphone-based GPS tracking to look at road or mobility pricing, limiting the opportunity to understand modal shifts.

The use of GPS tracking for mobility research is now widespread. Multiple studies have identified how traditional travel diaries under-report the number of trips, due to, among other reasons, response burden and memory recall (18, 19, 20). Passive tracking mostly mitigates these issues, although the collecting of trip metadata such as detailed trip purpose, fellow passengers and travel expenses mostly still requires more traditional survey methods. Furthermore, the performance of GPS tracking depends on the quality of the GPS traces, and the algorithms used to identify trips, stages and activities, as well as the mode and purpose of travel. Here there has been significant advances in recent years (21, 22). For two comprehensive reviews on the processing of GPS tracking data, the reader is referred to Shen and Stopher (23) and Nikolic and Bierlaire (24).

Other studies note that the performance of the algorithms is highly dependent on the quality of the GPS data (25, 26, 27).

One of the key factors influencing the quality of GPS data is the device used. This can be either a dedicated GPS logger, or a smartphone, where the data is collected through an app. The quality of the data can vary between devices, in particular between iOS and Android devices, depending for example on battery saving settings.

Few studies have explored the implications of this iOS/Android dichotomy and the implications for mobility studies using app-based tracking. Harding (26) compared the performance of trip identification and mode detection by different apps and found that iOS-based apps tended to have a higher accuracy. However, not only is the quality of the recorded data important, but also the attrition rate throughout the study, as this ultimately determines the sample size. This is an open question that has not been widely explored. The market penetration rates of iOS and Android - and even different Android-based manufacturers - varies across regions and, possibly, segments of the population. For studies requiring a representative sample, for example official national travel surveys, an understanding of these factors is important. As the MOBIS study aims to analyze societal impacts of mobility pricing to inform policy and decision making, obtaining a representative sample was a key objective.

We report our experiences undertaking a tri-lingual, national-scale mobility pricing survey and randomized controlled trial in Switzerland, combining traditional survey methods and app-based GPS tracking. MOBIS aims to understand the effects on travel behaviour of a) informing subjects about congestion, health effects, and carbon emissions of their mobiliity, and b) actually charging subjects the external costs associated with these 3 factors under a mobility pricing experiment. To do this, we examine two different treatments - information and pricing. In the current political discourse it is of interest to understand if information measures are found to have a similar impact as mobility pricing. On the other hand, evidence for pricing would support calls to restructure current mobility taxes and subsidies. In this paper, we focus on the survey method and the role of app-based tracking. In particular, contributions include a detailed analysis of the response rate over the duration of the study, and how it was impacted by the differences between iOS and Android devices.

METHODOLOGY

The 8-week study consisted of two consecutive 4-week phases, a control and treatment phase respectively, book-ended by introductory and concluding online surveys. A pretest with a mail-out sample of 1,500 letters was undertaken to estimate the expected response rate for the main study and test the surveys and GPS tracking.

Initial Recruitment

For the main study, a representative list of 60,000 addresses randomly selected across the major agglomerations (in the German and French speaking parts) of Switzerland from the Swiss Federal Office of Statistics was used. Based on the response rate in the pretest, this address sample was skewed to account for under-represented groups. Additionally, to achieve the desired sample size of 3,500 study participants, a second wave of around 30,000 persons were contacted using addresses from a private vendor, yielding a total of a little over 90,000 invitations. Only people living in an agglomeration area of Switzerland (excluding the Italian-speaking canton Ticino) were invited to participate in the study.

The letter invited the recipients to fill in a screening-survey with transport-related questions and, if they met the inclusion criteria, to participate in a smartphone-based mobility experiment where they would receive 100 CHF ($100 USD) for participating for the entire 8 weeks. Neither the 'mobility pricing' nature of the study nor the focus on the external costs of transport was shared with the participants.

Two reminder letters were also sent in the first wave, 4 and 7 weeks after the invitation letter was received, to those who had not responded to previous letters. No reminders were sent in the second wave as the target number of 3,500 participants had already been achieved.

Introduction and final surveys

The initial survey was designed to determine a respondent's eligibility for the main tracking study and collect data that would be needed in the calculation of external costs (such as mobility tool ownership, car type and age, and some general attitudes towards transport policies). The final survey included a series of stated-choice experiments and lifestyle and values questions, as well as awareness questions to gauge if participants understood the experiment and were therefore 'knowledgeable' participants. Completion of the final survey was a condition for receiving the incentive.

Recruitment for the field experiment

The participants who completed the introduction survey were assessed against the eligibility criteria for the field experiment. Specifically, participants

- had to use a car at least two days a week
- were restricted to the age of 18 to 65
- must be able to walk without assistance
- must own a smartphone
- were not allowed to drive in a professional capacity - i.e. postman/woman or taxi driver.

Those who met the requirements for the study and gave consent to participate were sent an email with a unique registration code and a link to download the Catch-My-Day app and participate in the tracking study.

Tracking app

The Catch-My-Day app is a location tracker for iOS and Android, which uses the location services of the respective operating system. GPS tracks are stored on the phone and uploaded to the MotionTag analytics platform, where stages, travel modes and activities are imputed. The following modes are detected the by Catch-my-Day app.

- Airplane
- Bicycle
- Bus
- Car
- Ferry
- S-Bahn (Local train)
- Regional train
- Subway
- Train (other)
- Tram
- Walk

Those marked with an asterisk are not automatically detected, but selectable by the user as a correction

- Boat*
- Carsharing*
- Motorbike/Scooter*
- Taxi/Uber*

Users can view their daily travel patterns on their phone in the form of a logbook, validate the travel mode and activity purpose or indicate if a trip or activity did not take place. The database stores both their correction and the original algorithmic imputation. There are some user-interface differences between the iOS and Android versions, which are most noticeable in the trip validation interface.

Users of could view their daily travel log in the app, and correct any incorrect travel mode imputations. Validation in the treatment phase was still allowed, even for the pricing group. Disabling validation in the treatment phase would have disadvantaged those affected by mis-detection, especially if they had made corrections in the control phase, due to the lower external costs of public transport. To counter any possible 'gaming' of the experiment, an outlier analysis was performed before transferring the incentive to the participants. No clearly suspicious behaviour was observed, except for one participant who seemed to switch to riding his e-bike for the entire second phase of the study. Figure 1 presents the validation interface of the app for the respective operating systems.

Users were required to activate the app by creating an account, which requires the provision of an email address and the choice of a password, along with the unique registration code provided. Participants are not required to validate their trips and activities, but were informed that this was possible and would be appreciated.

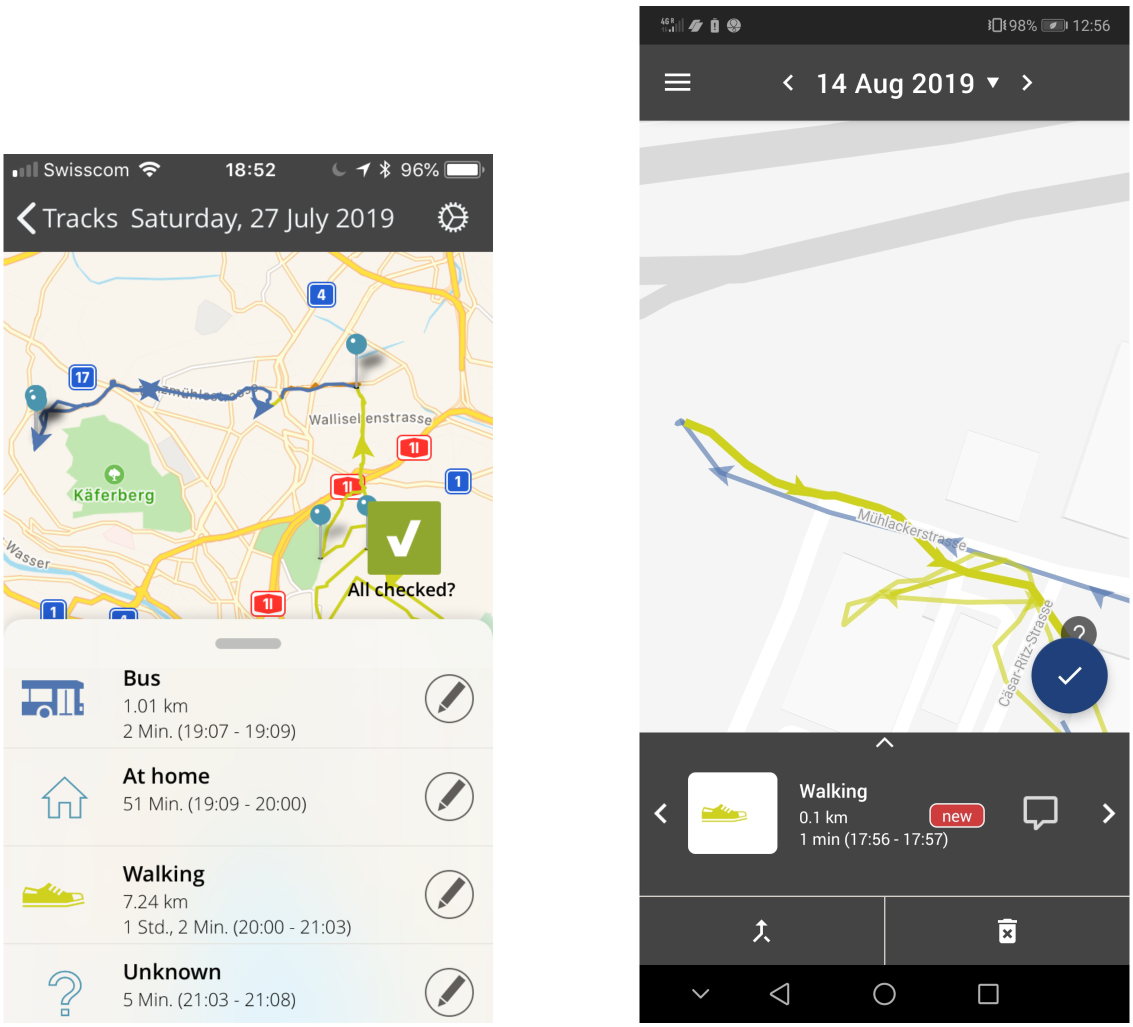
To increase the retention rate, automated reminder emails were sent to participants when they had not activated the app, or no data was recorded for a certain number of days. A helpdesk was set up for participants experiencing difficulties. User-guides on correctly configure one's smartphone for the app were provided. Additionally, participants who did not record data on at least 12 of the first 28 days were removed from the study, and notified by email.

Treatment groups

The 8-week study period was divided into two 4-week phases. In the first phase, participants were tracked using the app, and received weekly reports on the kilometers traveled per mode. At the beginning of the second phase, participants were randomly assigned to either the control group, or one of the two treatment groups. The control group continued to receive the same basic information on their behaviour, whereas both the information and pricing groups received additional information on the externalities they caused. Furthermore, participants in the pricing group were provided with a mobility budget, equal to 120% of their external costs in the first phase, from which their external costs in phase 2 were subtracted; with the balance remaining transferred to them, as an incentive to reduce their externalities. An example of the weekly reports is provided in Figure 2

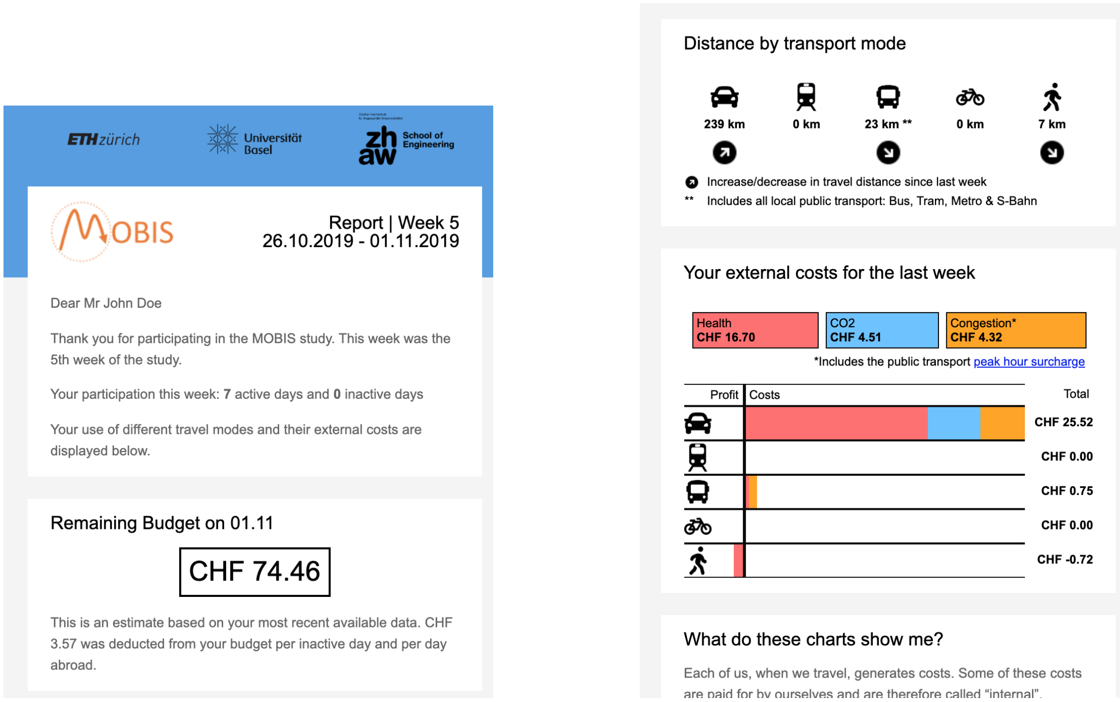
These externalities were separated into health, environmental and congestion costs, which were computed using a data pipeline run every evening. For more details on the externality computation, please refer to Tchervenkov et al. (28). The calculations are based on the HBEFA (Handbook for emissions analysis), relevant Swiss norms and the IVT MATSim scenario for Switzerland

FIGURE 1: Trip/validation interface



(a) iPhone (b) Android

FIGURE 2: Weekly report to participants in the pricing group



(a) First sections (b) Further sections

(29). Additionally, data collected from the introduction survey was incorporated into the data processing pipeline to improve the computation: Information on the participant's main vehicle was used to calculate individualized external costs.

RESULTS AND DISCUSSION

In this paper, we present the results in terms of participation and the collection of tracking data. The analysis of the field experiment is still ongoing and will be presented elsewhere.

Response rates

Invitations to the study were sent by post to 90,090 persons. From this sample, 23.70% completed the initial survey. This response rate was likely elevated by the prospect of the 100 CHF incentive for the tracking experiment, mentioned in the invitation letter (even though no incentive was provided for participation in the introductory survey on its own). Only 31.89% of those who completed the introduction survey met the criteria for the field experiment. This was predominately due to the minimal car-use requirement. Many people (age 16 and over) in Switzerland neither have access to a car (22%), nor a drivers license (18%) (30).

The two reminder letters were also effective in the first wave. of the 5320 who registered, 2397 (45%) did so before a reminder letter was sent, and 1793 (34%) and 1245 (23%) did so after the first and second reminder respectively.

Of those who qualified, 78.06% agreed to participate. This compares similarly to the other studies in Table 1. At the next stage, out of the remaining 5364 participants, 1146 (21.4%) did not start tracking. Either they either never installed the app, removed it before data was recorded, or were unable to get it to work successfully. Of those who did track, the share with an iOS device was 61%, much higher than the reported 44,4% national market share in 2019 (31), indicating that relatively more Android users were unable or unwilling to use app. Anecdotal evidence from the staff on the study help desk also indicated that more participants had issues installing the app for Android than iOS, and required assistance from the help desk in doing so (32).

Finally, 3690 participants successfully completed the 8-week tracking period, giving a completion rate of 69.4% for those that registered, and 4.06% overall. This is somewhere in the middle of the results from previous studies, with the high incentive appropriately offsetting the long tracking period.

TABLE 1: Response rates in various tracking studies

Project name	MOBIS	SPOT (33)	IN-THE-MOMENT (34)	ATLAS (35)	AKTA (15)	Cincinnati (36)	Atlanta (37)	Reno (38)	Tel Aviv HTS (39)
Tracker	MotionTag	MEILI	rMove	SITSS	Device	Device	Device	Device/App	App
Country	Switzerland	Sweden	USA	NZ	Denmark	USA	USA	USA	Israel
Year	2019	2015	2015	2014	2001-2003	2010	2011	2015-2016	2016-2017
Tracking days	56	7	7	3	112	3	7	7	2
Min. incentive (USD)	$100		$25		variable	$25	$25	$25	
Validation/annotation	Optional	Yes	Yes	yes	no	yes	yes	yes	yes
Invited persons (N)	90,909	130,000	1,427	186	25,000	11,118	16,374	25,817	67,199
Intro survey (N)	21,571								
% of invited	23.73%								
Qualified (N)	6895	1159	478						38,500
% of invited	7.58%	0.89%	33.50%						57.29%
% of intro completed	31.96%								
Registered (N)	5375	495	295	77	500	4656	1938	602	27,415
% of invited	5.91%	0.38%	20.67%	41.40%			11.84%		40.80%
% of qualified	77.96%	42.71%	61.72%						71.21%
Started tracking (N)	4218	293	295	73	500	4656		602	25,201
% of invited	4.64%	0.23%	20.67%	39.25%	2.00%	41.88%			37.50%
% of qualified	61.17%	25.28%	61.72%						65.46%
Completed tracking (N)	3690	51	240	65	500	3849	1061	312	23,240
% of invited	4.06%	0.04%	16.82%	34.95%	2.00%	34.62%	6.48%	1.21%	34.58%
% of qualified	53.52%	4.40%	50.21%						60.36%
% of registered	68.65%	10.30%	81.36%	84.42%	100.00%	82.67%	54.75%	51.83%	84.77%

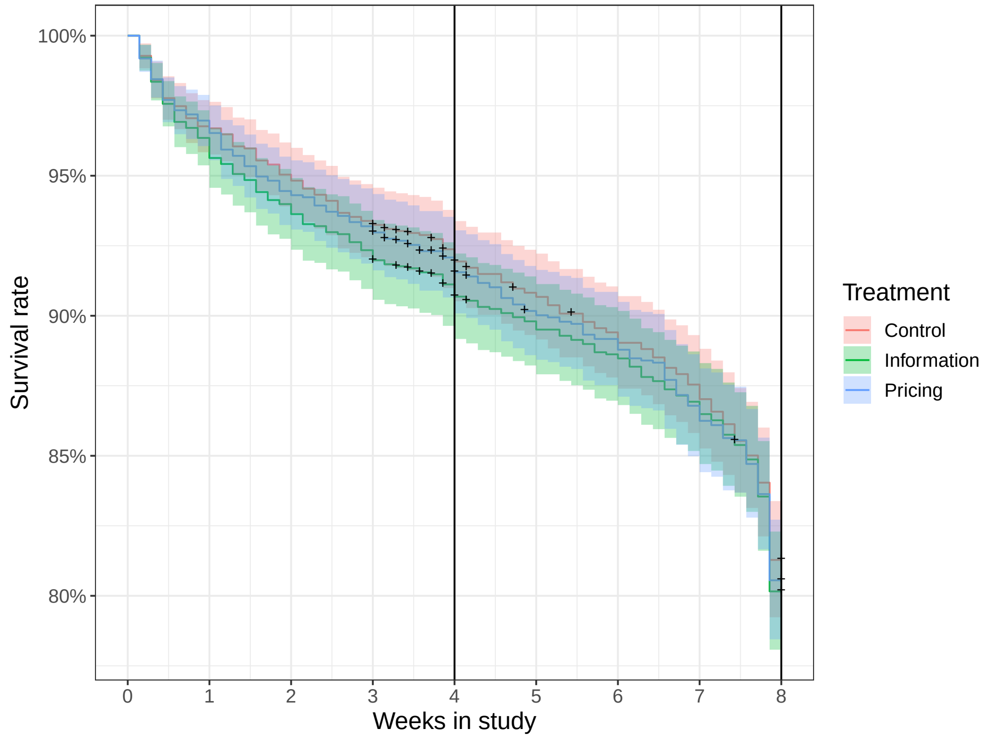


FIGURE 3: Kaplan-Meier survival curve by treatment group. The cross indicates censoring of participants

Participant retention

To explore the retention rate of participants in the tracking phase, we performed a survival analysis on the duration of tracking in the study. First, a Kaplan-Meier approach (see Figure 3) shows the impact of the treatment on the length of time which participants would track. Participants who were automatically dropped out after phase 1 due to poor tracking compliance but were still tracking at the end of phase 1 were censored (marked by a cross). There is no significant difference between the three treatment groups in their survival curves. A sharp decrease in survival is evident in the last study week. As participants were informed at the end of the study that they could delete the app, the last few days of tracking were sometimes not collected before the app was deleted.

Although the participants in the study had a clear participation goal of 8 weeks, after which they would receive the incentive, the survival curve is extremely linear. One would intuitively expect that the attrition rate would be highest early on in the study, and flatten out as participants neared the 8-week goal. This appears to only slightly be the case, with the dropout rate remaining constant throughout the study, even in the second phase. Furthermore, Figure 3) shows that the treatment didn't affect the attrition rate in the second phase.

A time-variant Cox proportional hazards model is to investigate the impact of different factors on the participation duration (see Table 2 for the model results). To account for time-dependent effects, the study period was stratified into fortnightly windows. Those in high-income brackets (more than 12,000 CHF/year) were more likely to stop tracking. Conversely, those from larger households and those with tertiary education were more likely to track for longer. A significant gender-based difference was only observed in the final fortnight, where females were more likely to remain in the study.

	Beta (SE)	HR (95% CI)	p
Income $>$ 12,000 CHF	0.28 (0.09)	1.32 (1.10, 1.58)	0.003 **
Household size	-0.07 (0.03)	0.93 (0.87, 1.00)	0.038 *
Age (decades)	0.00 (0.03)	1.00 (0.95, 1.06)	0.883
Tertiary education	-0.19 (0.08)	0.83 (0.70, 0.97)	0.022 *
German speaking	0.03 (0.09)	1.03 (0.87, 1.22)	0.752
Female			
fortnight=1	0.02 (0.15)	1.02 (0.77, 1.35)	0.895
fortnight=2	-0.07 (0.20)	0.93 (0.62, 1.39)	0.721
fortnight=3	-0.04 (0.22)	0.96 (0.62, 1.48)	0.841
fortnight=4	-0.28 (0.12)	0.76 (0.60, 0.96)	0.022 *
Android			
fortnight=1	0.87 (0.16)	2.38 (1.73, 3.26)	0.000 ***
fortnight=2	0.46 (0.22)	1.58 (1.02, 2.45)	0.040 *
fortnight=3	-0.01 (0.25)	0.99 (0.60, 1.62)	0.960
fortnight=4	0.41 (0.13)	1.51 (1.17, 1.94)	0.002 **
Huawei			
fortnight=1	0.38 (0.20)	1.47 (0.99, 2.18)	0.057 .
fortnight=2	0.37 (0.32)	1.45 (0.78, 2.70)	0.239
fortnight=3	0.29 (0.41)	1.33 (0.59, 2.98)	0.487
fortnight=4	0.15 (0.21)	1.16 (0.77, 1.75)	0.465
Employed			
fortnight=1	-0.33 (0.16)	0.72 (0.53, 0.97)	0.033 *
fortnight=2	-0.07 (0.23)	0.94 (0.60, 1.47)	0.775
fortnight=3	0.24 (0.27)	1.27 (0.75, 2.15)	0.369
fortnight=4	0.05 (0.14)	1.05 (0.80, 1.38)	0.718
AIC		10484.33	
Coordance		0.602	
Num. events		655	
PH test		0.76	
Note:			$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

TABLE 2: Cox porportional-hazard model

Contrary to expectations, there was no significant effect of age on the hazard rate. This suggests that common concern about the feasibility of tracking studies for older age groups is unfounded, at least up to the age of 65, the age limit in this study.

The coefficient on employment is also time-dependent. Those in the workforce (i.e. excluding students, homeworkers and retirees) were more likely to remain in the study throughout the first fortnight.

The participant's mobile device played a much larger role. Having an Android phone of any model increased the hazard drastically. However, this effect was strongest in the first week. The effects were even larger for Huawei models. The incompatibility of GPS loggers with Android

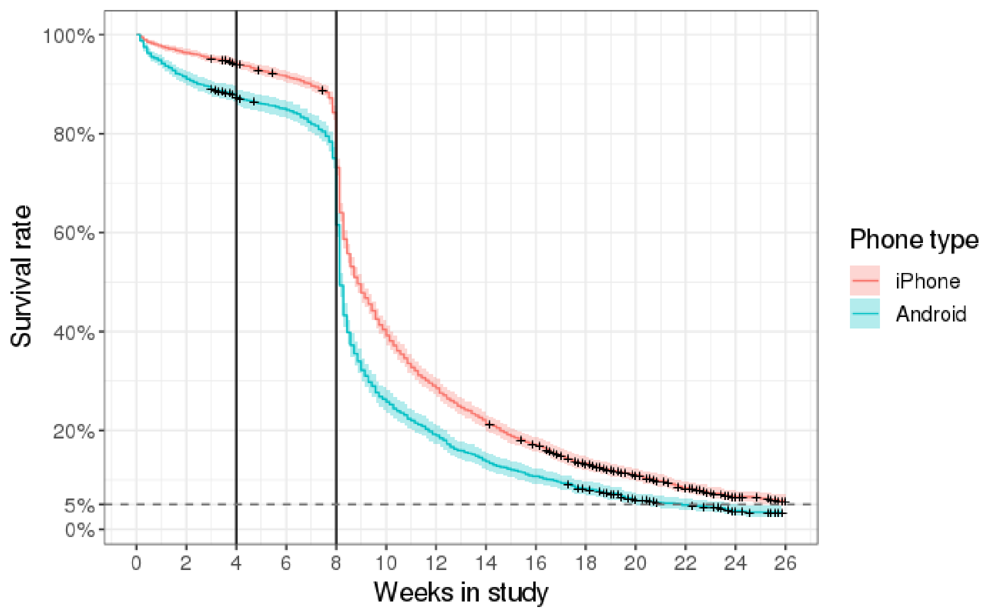


FIGURE 4: Post-study participation survival curve

(and particularly Huawei devices) is already well known; however, here the effect is quantified, and seen to be dramatic. The effect was also time-dependent, with the most significant hazard in the first fortnight. At the end of the second fortnight, participants who tracked insufficiently were removed from the study - this explains the reduction in the Android hazard coefficient for the third fortnight, when many of them could have been expected to stop tracking, had they not been removed from the study.

Post-study retention

At the end of the tracking study, participants were told that they could delete the app, but were also encouraged to continue using it if they wished. Figure 4 shows the dropout rate for the whole study, including the post-study period. The majority of the participants dropped out soon after the study, but even 6 months after the study was completed, around 5% of participants continued to use the app. Anecdotal reports from participants indicated that they enjoyed having an overview of their travel, and that it even continued to inform their mobility decisions. The impacts of the mobile operating system continued even after the study, with the post-study retention rate falling faster for Android users.

Participant engagement

Participants in the information and pricing groups were effectively treated through information provided in a weekly email detailing their externalities and the costs incurred. Interactions with the emails were recorded using standard email tacking techniques. Emails that remained unopened were effectively missed treatments. Table 3 presents a overview of the engagement with the email communications. The open rate did not change drastically over the duration of the study. Participants in the pricing group viewed their emails much more often than the control or information groups. The information group also opened their emails repeatedly in the first two weeks of phase two, before returning to a pattern similar to the control group, whereas the pricing group continued to repeatedly open their emails.

Participants in the treatment groups likely repeatedly reopened the emails to check their externalities and remaining budget. We suggest that this ‘repeat opening’ behaviour is a useful indicator to measure the level of engagement with the treatment.

TABLE 3: Engagement with various emails through the study

Email & Treatment	n	% Opened	Times opened (mean)	Time to open (h) median (IQR)
Welcome				
-	5475	82.36	2.78	8.50 (2.88 - 20.33)
Report 1				
	4168	84.88	2.13	7.37 (2.53 - 19.22)
Report 2				
	4132	81.03	1.87	6.66 (2.59 - 18.37)
Report 3				
	4105	78.59	1.83	6.19 (2.51 - 17.85)
Report 4				
Control	1247	79.23	1.62	5.40 (2.30 - 14.65)
Info	1262	83.68	1.99	5.40 (2.40 - 16.83)
Pricing	1222	82.90	2.64	6.06 (2.35 - 17.57)
Halfway				
Control	1250	76.80	1.60	5.60 (2.41 - 15.54)
Info	1263	83.29	1.72	5.50 (2.53 - 17.35)
Pricing	1222	80.93	2.17	5.51 (2.24 - 17.15)
Report 5				
Control	1243	76.43	1.55	5.96 (2.42 - 15.37)
Info	1255	80.80	1.90	6.28 (2.42 - 17.29)
Pricing	1213	80.54	2.24	6.94 (2.66 - 19.82)
Report 6				
Control	1238	77.06	1.87	5.78 (2.35 - 16.89)
Info	1252	78.12	1.87	5.87 (2.57 - 17.32)
Pricing	1208	79.22	2.09	6.24 (2.41 - 17.87)
Report 7				
Control	1235	74.98	1.61	5.83 (2.35 - 15.83)
Info	1248	77.64	1.66	6.08 (2.44 - 18.16)
Pricing	1205	80.25	2.02	6.07 (2.33 - 17.49)
Report 8				
Control	1231	79.69	1.50	6.11 (2.55 - 17.01)
Info	1246	78.33	1.46	6.41 (2.49 - 18.85)
Pricing	1200	81.50	2.01	6.55 (2.49 - 18.80)

Trip mode and purpose validation

Participants were invited to use the validation interface to confirm the detected mode and purpose of their trips and activities. This was optional, but they were encouraged in the weekly email reports to do so. Even in the second phase, participants were trusted to correct the mode detected by the app. As the mode is crucial in determining the external costs deducted from the mobility budget for the pricing group, this consequently gave them the opportunity to 'game' the experiment, by for example 'correcting' car trips to another transport mode. To test for this, a regression analysis using a zero-inflated negative binomial model was performed with the number of corrections for a day as the dependent variable (see Table 4). A zero-inflated model was used to accommodate the large number of participants who did not correct any trips. While a significant increase in the number of corrections was observed in phase 2, no increase in the number of corrected trips specific to the pricing group was observed. Conversely, the parameters are insignificant but negative. In fact, the information group saw a significant reduction in the corrections in phase 2. One hypothesis is that by receiving more information on their externalities in the weekly reports in the second phase, participants felt discouraged from correcting their trips in the app. Also, no indication was given to participants that they would be penalised for any suspicious behaviour. The fact that no significant change in the average correction rate was seen between treatment groups, suggests that the trust in the participants was justified.

In recent years, state-of-the-art machine learning algorithms for mode and activity detection have achieved accuracy rates of over 90%, depending on the approach (40, 24). Hence, we made validation of the trip purpose and mode optional for participants, in order to ensure a minimal response burden over the 8 weeks. 85.7% of participants confirmed at least 1 of their trips; however, of those who did use the validation functionality, 20.4% of iPhone users and 44.1% of Android users did not make a single correction over the 8 weeks, respectively. Even with state-of-the-art accuracy rates, such a validation behaviour is extremely unlikely. As such, we can assume that these participants did not use or understand the validation interface correctly, and these participants are therefore removed from the following analysis on the mode detection performance. It also indicates that the iPhone validation interface was much more intuitive.

Mode detection performance

The mode detection provided by the tracking app was a key component of the MOBIS study. As far as the authors aware, this is the first study to incentivise changes in mobility behaviour based on the output of a mode detection algorithm. As seen in Table 5, the algorithm worked exceptionally well on location data from both operating systems. There is small difference in accuracy between iOS and Android, with iOS being on average slightly better (92.23% vs 92.10%) with a p-value of 0.01, test of equal proportions). However, the differences in accuracy are more observable at the categorical level. The iOS performs better on car, local rail, regional rail, tram and walk. However, the differences are only 1-3% in accuracy. Note that 'Rail' groups all rail modes together for conciseness. It is also worth noting that while the accuracy of some individual rail modes is quite low, the overall rail accuracy is very good. The main confusion was between different rail mode types.

Table 6 presents the confusion matrix between the modes. Here we can see that the algorithm often mis-detected car travel as bus travel. For conciseness, the category 'Other *' includes those modes which could be manually selected by the participant, but which were not automatically detected. These included: Carsharing, Taxi/Uber, Motorbike/Mopeds, and Gondolas. Most

TABLE 4: Zero inflated negative binomial model of the validation behaviour

	Count model (1) Corrections/day		Zeros model (2) Correction/day > 0	
Constant	0.744	(0.032)***	1.504	(0.046)***
Phase 2	0.047	(0.014)**	0.050	(0.020)*
Age (decades)	−0.024	(0.003)***	−0.014	(0.005)**
Male	0.074	(0.012)***	0.047	(0.017)**
Treatment				
Control	-		-	
Information	−0.029	(0.022)	−0.053	(0.032)
Pricing	−0.083	(0.069)	−0.335	(0.103)**
Education				
Mandatory	-		-	
Trade/traineeship (baseline	−0.098	(0.023)***	−0.220	(0.033)***
Higher education	−0.014	(0.023)	−0.321	(0.033)***
Income (CHF per month)				
Less than 4000	-		-	
4000 <= 8000	−0.134	(0.022)***	−0.208	(0.032)***
8000 <= 12,000	−0.203	(0.022)***	−0.324	(0.032)***
12,000 <= 16,000	−0.230	(0.024)***	−0.429	(0.035)***
More than 16,000	−0.124	(0.025)***	−0.360	(0.038)***
Interactions				
Control * male	-		-	
Information * male	−0.027	(0.028)	0.139	(0.040)***
Pricing * male	−0.004	(0.027)	−0.001	(0.040)
pricing * mandatory	-		-	
pricing * trade/traineeship	−0.113	(0.057)	0.099	(0.081)
pricing * higher education	−0.166	(0.057)**	−0.023	(0.082)
pricing * less than 4000	-		-	
pricing * 4000 <= 8000	0.174	(0.059)**	0.278	(0.084)***
pricing * 8000 <= 12,000	0.285	(0.058)***	0.354	(0.083)***
pricing * 12,000 <= 16,000	0.187	(0.065)**	0.456	(0.092)***
pricing * more than 16,000	0.128	(0.068)	0.368	(0.099)***
Observations	147,450			
Log Likelihood	−127,206.400			
Note:	$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$			

TABLE 5: Comparison of the MotionTag mode detection performance bewteen iOS and Android

	% Correct	
Mode	Android	iOS
Airplane	99.48%	98.86%
Bicycle	81.59%	79.14%
Bus	66.98%	66.82%
Car	92.98%	93.15%
Rail	89.50%	91.05%
Local train	88.67%	90.18%
Regional train	71.35%	73.40%
Subway	93.56%	92.53%
Train	63.13%	63.78%
Tram	95.01%	96.64%
Walk	95.56%	97.21%

of these were detected as car travel, and the 1,500 'Bicycle' trips which were corrected to 'Other' were predominately trips by motorbike or moped.

These mode detection results confirmed the indications of our pretest that the automatic detection could indeed be used to calculate the external costs of travel with sufficient accuracy and determine the phase 2 budget and deductions based on these. If the accuracy had been too low, more participants would have dropped out of the study, seeing it as 'unfair' if the budget and deductions did not match their travel behaviour.

Identified mode detection issues

As previously mentioned, the quality of the mode detection was key to the mobility pricing field experiment. A few issues were identified which are worth considering in future studies that apply algorithmic mode detection.

The first consideration concerns those leisure activities that are movement based over a larger area, such as a bike tour, hiking and skiing. Skiing is especially important in alpine areas: In Switzerland, the percentage of the population that ski regularly is 37% (41). Gondolas and chairlifts move at between 15 and 50km/h, meaning that these trips are often confused for car travel unless the algorithm has been specifically calibrated. On the downhill, skiers reach similar speeds. Taking a strict definition of a transport trip, such movement-based activities should be excluded from the calculation of external costs. If they were to be included, a person could end up being charged for a long hike in the wilderness on the weekend - which would arguably not be in the spirit of a mobility pricing scheme.

The second consideration is trip chaining. Shen and Stopher (23) note that all methods to date (albeit in 2014) did not consider trip chains when detecting the transport mode, and only considered each individual stage. While the mode detection provided by the app was sufficient for the purpose of the mobility pricing field experiment, anecdotal evidence indicates that considering trip chains could further improve the performance of the algorithm.

TABLE 6: Confusion matrix of mode detection accuracy

	Confirmed mode									
	Airplane	Bicycle	Boat	Bus	Car	Rail	Tram	Walk	Other	**Total**
Predicted										
Airplane	**2,113**	-	-	-	22	-	-	-	-	**2,135**
Bicycle	4	**26,201**	136	438	1,499	177	149	2,771	1,500	**32,875**
Bus	1	435	2	**35,713**	15,085	140	280	889	865	**53,410**
Car	372	2,495	741	8,028	**366,649**	3,314	1,950	2,834	7,433	**393,816**
Rail	64	56	85	1,748	7,298	**60,270**	691	258	298	**70,768**
Tram	-	49	2	128	396	60	**20,174**	149	16	**20,974**
Walk	80	3,807	456	1,224	9,960	868	868	**514,944**	638	**532,845**
	2,634	**33,043**	**1,422**	**47,279**	**400,909**	**64,829**	**24,112**	**521,845**	**10,750**	**1,106,823**

CONCLUSION

This work makes multiple contributions to the literature on conducting tracking-based mobility studies, and demonstrates the feasibility of running an incentive-based field experiment using a tracking app. We analysed the effect of the mobile device operating system on GPS tracking studies, and identified certain areas where the difference in OS needs to be considered when undertaking such studies. The impact on participant retention is significant. While this effect is strongest at the start of the study, it persists throughout. The on-boarding of Android users into the study took significant resources, and we suggest this be accounted for when planning and budgeting such studies. Correspondence by email was effective, and participant engagement did not decline over the 8 weeks. The mode detection algorithm was also sufficiently accurate to support the calculation of external costs in the field experiment. Finally, concerns that participants would manipulate the study by ‘correcting’ their trips in the app were unfounded, with participants adhering to the spirit of the study. Socio-demographic differences in the correction rate do, however, indicate that some participants were more engaged than others.

REFERENCES

1. Cecil Arthur Pigou. The economics of welfare. *New Brunswick & Londres: Transaction Publishers*, 4, 1920.
2. Kenneth A Small et al. Road pricing and public transport. *Road pricing: Theory and evidence*, 9:133–58, 2004.
3. Erik Verhoef, Peter Nijkamp, and Piet Rietveld. Second-best congestion pricing: the case of an untolled alternative. *Journal of Urban Economics*, 40(3):279–302, 1996.
4. Lucas Meyer de Freitas, Oliver Schuemperlin, Milos Balac, and Francesco Ciari. Equity effects of congestion charges: An exploratory analysis with matsim. *Transportation Research Record*, 2670(1):75–82, 2017.
5. Artem Chakirov. *Urban mobility pricing with heterogeneous users*. PhD thesis, ETH Zurich, 2016.
6. Ihab Kaddoura. Marginal congestion cost pricing in a multi-agent simulation investigation

of the greater Berlin area. *Journal of Transport Economics and Policy (JTEP)*, 49(4): 560–578, 2015.

7. Georgina Santos and Blake Shaffer. Preliminary results of the London congestion charging scheme. *Public Works Management & Policy*, 9(2):164–181, 2004.
8. Jonathan Leape. The London congestion charge. *Journal of Economic Perspectives*, 20 (4):157–176, 2006.
9. Jonas Eliasson, Lars Hultkrantz, Lena Nerhagen, and Lena Smidfelt Rosqvist. The Stockholm congestion–charging trial 2006: Overview of effects. *Transportation Research Part A: Policy and Practice*, 43(3):240–250, 2009.
10. Milenko Vrtic, Nadine Schuessler, Alexander Erath, and Kay W Axhausen. The impacts of road pricing on route and mode choice behaviour. *Journal of Choice Modelling*, 3(1): 109–126, 2010.
11. Kevin Washbrook, Wolfgang Haider, and Mark Jaccard. Estimating commuter mode choice: A discrete choice analysis of the impact of road pricing and parking charges. *Transportation*, 33(6):621–639, 2006.
12. Zheng Li and David A Hensher. Congestion charging and car use: A review of stated preference and opinion studies and market monitoring evidence. *Transport Policy*, 20: 47–61, 2012.
13. Milenko Vrtic, Nadine Schüssler, Alexander Erath, and Kay W Axhausen. Design elements of road pricing schemes and their acceptability. In *TRB 86th Annual Meeting Compendium of Papers*, pages 07–1548. Transportation Research Board, 2007.
14. Cecilia Jakobsson, Satoshi Fujii, and Tommy Gärling. Determinants of private car users' acceptance of road pricing. *Transport Policy*, 7(2):153–158, 2000.
15. Otto Anker Nielsen. Behavioral responses to road pricing schemes: Description of the danish akta experiment. In *Intelligent Transportation Systems*, volume 8, pages 233–251. Taylor & Francis, 2004.
16. Transurban. Changed conditions ahead. The transport revolution and what it means for Australians, Melbourne Road Usage Study Report. Report, Melbourne, 2016.
17. L. A. Martin and Sam Thornton. To drive or not to drive? a field experiment in road pricing. Working paper, Parkville, Australia, 2017.
18. Maxim Janzen, Maarten Vanhoof, Zbigniew Smoreda, and Kay W Axhausen. Closer to the total? Long-distance travel of French mobile phone users. *Travel Behaviour and Society*, 11:31–42, 2018.
19. Jean Wolf, Michael Loechl, Miriam Thompson, and Carlos Arce. Trip rate analysis in GPS-enhanced personal travel surveys. *Transport survey quality and innovation*, 28:483–498, 2003.
20. Peter Stopher, Camden FitzGerald, and Min Xu. Assessing the accuracy of the Sydney Household Travel Survey with GPS. *Transportation*, 34(6):723–741, 2007.
21. Nadine Schuessler and Kay W Axhausen. Processing raw data from global positioning systems without additional information. *Transportation Research Record*, 2105(1):28–36, 2009.
22. Alessio D Marra, Henrik Becker, Kay W Axhausen, and Francesco Corman. Developing a passive gps tracking system to study long-term travel behavior. *Transportation research part C: emerging technologies*, 104:348–368, 2019.

23. Li Shen and Peter R Stopher. Review of GPS travel survey and GPS data-processing methods. *Transport Reviews*, 34(3):316–334, 2014.
24. Marija Nikolic and Michel Bierlaire. Review of transportation mode detection approaches based on smartphone data. In *17th Swiss Transport Research Conference*, number CONF, 2017.
25. Lara Montini, Sebastian Prost, Johann Schrammel, Nadine Rieser-Schüssler, and Kay W Axhausen. Comparison of travel diaries generated from smartphone data and dedicated GPS devices. *Transportation Research Procedia*, 11:227–241, 2015.
26. Chris Harding. *From Smartphone Apps to In-Person Data Collection: Modern and Cost-Effective Multimodal Travel Data Collection for Evidence-Based Planning*. PhD thesis, 2019.
27. Oliver Burkhard, Henrik Becker, Robert Weibel, and Kay W Axhausen. On the requirements on spatial accuracy and sampling rate for transport mode detection in view of a shift to passive signalling data. *Transportation Research Part C: Emerging Technologies*, 114: 99–117, 2020.
28. Christopher Tchervenkov, Joseph Molloy, and Kay W Axhausen. Estimating externalities from GPS traces using MATSim. In *18th Swiss Transport Research Conference (STRC 2018)*. STRC, 2018.
29. Sebastian Hörl, Milos Balać, and Kay W Axhausen. Pairing discrete mode choice models and agent-based transport simulation with MATSim. In *2019 TRB Annual Meeting Online*, pages 19–02409. Transportation Research Board, 2019.
30. BFS and ARE. Verkehrsverhalten der Bevölkerung – Ergebnisse des Mikrozensus Mobilität und Verkehr 2015, 2017.
31. Comparis. Beliebtheit von iPhones in der Schweiz nimmt ab (German). online, [`https://www.comparis.ch/telecom/handy-gadgets/analyse/smartphone-markt-schweiz`], accessed 13/07/2020, 2019.
32. Christopher Tchervenkov, Joseph Molloy, Alberto Castro Fernández, and Kay W Axhausen. Mobis study - a review of common reported issues. In *20th Swiss Transport Research Conference (STRC 2020)*. STRC, 2020.
33. Andreas Allström, Ida Kristoffersson, and Yusak Susilo. Smartphone based travel diary collection: Experiences from a field trial in stockholm. *Transportation research procedia*, 26:32–38, 2017.
34. Elizabeth Greene, Leah Flake, Kevin Hathaway, and Michae Geilich. A Seven-Day Smartphone-Based GPS Household Travel Survey in Indiana. In *TRB 95th Annual Meeting Compendium of Papers*. Transportation Research Board, 2012.
35. Hamid Safi, Behrang Assemi, Mahmoud Mesbah, Luis Ferreira, and Mark Hickman. Design and implementation of a smartphone-based travel survey. *Transportation Research Record*, 2526(1):99–107, 2015.
36. Laurie Wargelin, Peter Stopher, et al. Gps-based household interview survey for the cincinnati, ohio region: executive summary report. Technical report, Ohio. Dept. of Transportation. Office of Research and Development, 2012.
37. J Livingston. Atlanta regional commission-regional travel survey-final report. *PTV NuStata, Austin, TX*, 2011.
38. Peter R. Stopher, Vivian Daigler, and Sarah Griffith. Smartphone app versus gps logger:

A comparative study. *Transportation Research Procedia*, 32:135 – 145, 2018. Transport Survey Methods in the era of big data:facing the challenges.

39. Bat-hen Nahmias-Biran, Yafei Han, Shlomo Bekhor, Fang Zhao, Christopher Zegras, and Moshe Ben-Akiva. Enriching activity-based models using smartphone-based travel surveys. *Transportation Research Record*, 2672(42):280–291, 2018.

40. Linlin Wu, Biao Yang, and Peng Jing. Travel mode detection based on GPS raw data collected by smartphones: A systematic review of the existing methodologies. *Information*, 7(4):67, 2016.

41. Statistica. Skiing in Europe - Statistics & Facts. online, [`https://www.statista.com/topics/3922/skiing-in-europe/`], accessed 13/07/2020, 2018.