ASSESSING THE EFFECTS OF A MIXED-MODE DESIGN IN A LONGITUDINAL HOUSEHOLD TRAVEL SURVEY

Christine Weiss
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Tel: +49 721 6084 7737
Email: Christine.Weiss@kit.edu

Bastian Chlond
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Tel: +49 721 6084 2257
Email: Bastian.Chlond@kit.edu

Clotilde Minster
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Tel: +49 721 6084 2257
Email: Clodilde.Minster@kit.edu

Christian Jödden
KANTAR TNS Germany
Landsberger Strasse 284, 80687 München, Germany
Tel: +49 89 5600 1558
Email: Christian.Joedden@tns-infratest.com

Peter Vortisch
Institute for Transport Studies, Karlsruhe Institute of Technology (KIT)
Kaiserstrasse 12, 76131 Karlsruhe, Germany
Tel: +49 721 6084 2255
Email: Peter.Vortisch@kit.edu

Paper prepared for presentation and publication at the
96th Annual Meeting of the Transportation Research Board
5,970 words (abstract, text, references)
2 figures, 4 tables (6 * 250 words = 1,500 words)
Total: 7,470 words (including abstract, text, tables, figures and references)
Submission date: November 15, 2016 (revised from original submittal)

ABSTRACT

The German Mobility Panel (MOP) is a national household travel survey that has been collecting data on the travel behavior of the German population since 1994. One of the MOP's central assets is the ability to provide time series data on travel behavior. Thus, the comparability of survey results from different years is a mayor objective on the survey method used. Within the last decade declining survey participation rates amongst different socio-economic groups resulted in the need to implement a mixed-mode design of the MOP in 2013, both for the sampling stage (landline and mobile phone recruitment) and in the data collection stage (paper and web). We analyze whether the adaptations in survey modes do affect the survey results and if so why. Ideally, we have increased the representativeness of the MOP. However, measurement biases because of the mixed-mode design are also conceivable. For the decomposition of survey mode effects we applied the propensity score weighting method. This method aims to impute the hypothetic responses participants would have given under different survey modes; disparities between actual responses and potential outcomes under another mode are traced back to the mixed-mode design. Our analyses indicate that trip rate biases on shopping and leisure trips and on short trips are partly caused by the mixed-mode design; quantities in time spent in the transportation system and trips made by car and public transportation and commuting trips are barely biased.

INTRODUCTION

Within the last decade many repeated and longitudinal travel surveys have been facing declining participation rates (*2; 1*). Declining participation rates diminish the coverage and representativeness of a survey, because rates decrease not equally amongst all socio-demographic groups, but certain populations groups are particularly affected. Furthermore, they result in higher survey costs, because it is necessary to contact more potential survey participants than before to ensure comparable sample sizes. As a consequence, different approaches to improve coverage are usually introduced to compensate these problems, e.g. alternative and additional recruitment procedures and other channels for participants to report (*3*).

The German Mobility Panel (MOP) is a longitudinal national household travel survey (NHTS) that has been collecting data on the travel behavior of the German population since 1994. Every year about 2,000-2,400 individuals fill in a trip diary for one week. One of the MOP's central assets is the ability to provide time series data on travel behavior. To ensure the comparability of survey results from different survey years, survey design adaptations should be avoided. Within the last decade we have observed declining participation rates amongst different socio-economic groups, particularly young adults. Despite concerns about survey comparability, we adapted the survey design of the MOP in 2013 in the sampling design stage and in the data collection stage in order to improve the representativeness and coverage of the survey. Hence, it can be regarded as crucial that any adaptations to the survey modes may affect the outcome in terms of the reported quantities of travel. This paper deals with the question whether the survey mode adaptations made in the MOP do affect the survey results and if so, why the adaptations do have an impact on the survey results. Ideally, we increased the representativeness of the MOP by reducing coverage errors and non-response errors. This outcome is named selection effect (*SE*). Biases in survey results due to measurement effects (*ME*) are also conceivable. *ME*s occur when respondents answer a survey question differently or fill in the trip diary in a different way solely because of the mode in which the question is administered.

Within this study we aim to better understand of how survey mode adaptations and mixed-mode designs in travel behavior surveys affect the results. Based on analyses of the outcomes of different survey modes in the recruitment and the data collection stage we apply propensity score estimation methods to identify and quantify the effects of the survey modes to the results.

LITERATURE REVIEW

Mixing modes in a survey gives researchers the possibility to compensate the weaknesses and disadvantages of each mode, but also to reduce the survey costs (*4*). Different mixes of data collection modes are discussed in the literature, including the recruitment phase, during the selection, or during the data collection phase (*4*). However, mixed-mode surveys have some

consequences and thus disadvantages in terms measurement (*5*). It must be assumed that all these changes will affect the composition of respondents and beyond this, also the way and completeness of reported travel.

Mixed-mode surveys have been applied in several disciplines, for example, health surveys. Studies related to health highlight that different population groups are reached with different modes (*6; 5*). This issue is for example addressed in the "Behavioral Risk Factor Surveillance System", a telephone based survey on health conditions and risks behavior in the United States (*8; 7*). The issue of weighing is thus discussed, underlying the need to understand differences between subsamples (*7*). In social science literature, analyses conducted with mixed-mode surveys, with a paper-and-pencil (PAPI) and a web survey (CAWI) data collection mode show that *MEs* are very small (*10; 9*).

Similar issues may be found for travel surveys. However, the outcome might be different in travel behavior surveys, because the participants are asked to fill in trip diaries. That means that the participants need to fill in the survey not at one point of time only but at certain points of time, e.g. once a day. There, the survey mode might also affect the frequency to fill in the diary and thus the survey quality. Several travel surveys are using mixed-mode approaches. The Regional Household Travel Survey conducted in the New York City megaregion seems to be the survey offering the most survey modes: they combine face-to-face (F2F) interviews with telephone interviews (CATI), with PAPI, CAWI and GPS-based data-collection methods (*11*). Research conducted on this survey highlighted that households with different socio-economic characteristics have been reached by different survey modes (*11*). Other examples of mixed-mode surveys in transportation include for example a travel survey conducted in Quebec City once in 2002-2003 as a PAPI and in 2009 in a CAWI (*12*). The authors have highlighted higher dropout rates with CAWI, especially among the participants aged 50 years and older (*12*). The authors highlight that the sociodemographic characteristics of CAWI and CATI respondents differ. However, travel quantities were similar in both groups (*12*). A household travel survey which was conducted in 2006 in Lyon, France showed different results. In this survey, households who either have not been willing to respond F2F or who were not reachable had the option to answer by CAWI. The authors found that, compared to F2F respondents, CAWI respondents live in bigger households, own more cars, and are better equipped in communication tools (internet access and mobile phone) (*13*). These socio-characteristics differences underlines, as assumed by the authors, differences in travel behavior: CAWI respondents are less mobile than F2F respondents (*13*). This is explained by the authors by, at the one hand, a higher share of days without trip making, but on the other hand, a lower number of trips reported due to the omission (*13*). Christensen analyzed the effects of introducing CAWI in the Danish NHTS (*14*). Denmark was indeed the first country to introduce CAWI in a NHTS in 2006 (*14*). Differences in socio-economic characteristics of respondents in different survey modes were also found. Christensen highlighted that the number of reported trips by CAWI respondents depend of the length of the trip: short trips (< 2km) are less reported (*14*). However, trips between 2 and 10 km seems to be better reported by CAWI than by CATI (*14*).

As different studies implement mode alterations in a different way and as various survey mode adaptations are introduced concurrently, it is sometimes difficult to identify and quantify the effects of the respective modes. In the literature the need for further research is addressed, i.e. to include the identification and quantification of mode effects (*13*) in travel surveys.

DATA

This section gives an overview of the MOP, summarizes the survey mode adaptations made since 2013 and introduces the sample of investigation of this study.

The German Mobility Panel (MOP)

We used the data of the MOP for our analyses. The MOP is a German NHTS which is conducted every year since 1994. The survey is carried out on behalf of and funded by the German Federal Ministry of Transport and Digital Infrastructure. The market research firm KANTAR TNS is responsible for the field work (i.e. recruitment and data collection) and the Institute for Transport Studies of the Karlsruhe Institute of Technology is in charge for the design and scientific supervision of the survey (*15; 16*).

The general sample is controlled by spatial categories, household type and car ownership. Thus it is guaranteed that the whole spectrum of rural areas up to inner cities is covered as well as different types of households. The sample size is 1,000-1,500 households with 2,000-2,400 persons (aged ten years and older). Participants are asked to take part in the MOP during three consecutive years. Every year a part of the households is dropped from the subsequent wave and replaced with new households.

The MOP survey takes place in autumn every year and the survey weeks are meant not to contain any school holidays or bank holidays (“everyday travel”). The participants are asked to provide a complete trip diary containing information about all their trips during a whole week, i.e. distances, means of transportation used, purposes and start respectively arrival times. Moreover, socio-demographic information about the participants and the availability of cars, bicycles and transit passes are questioned.

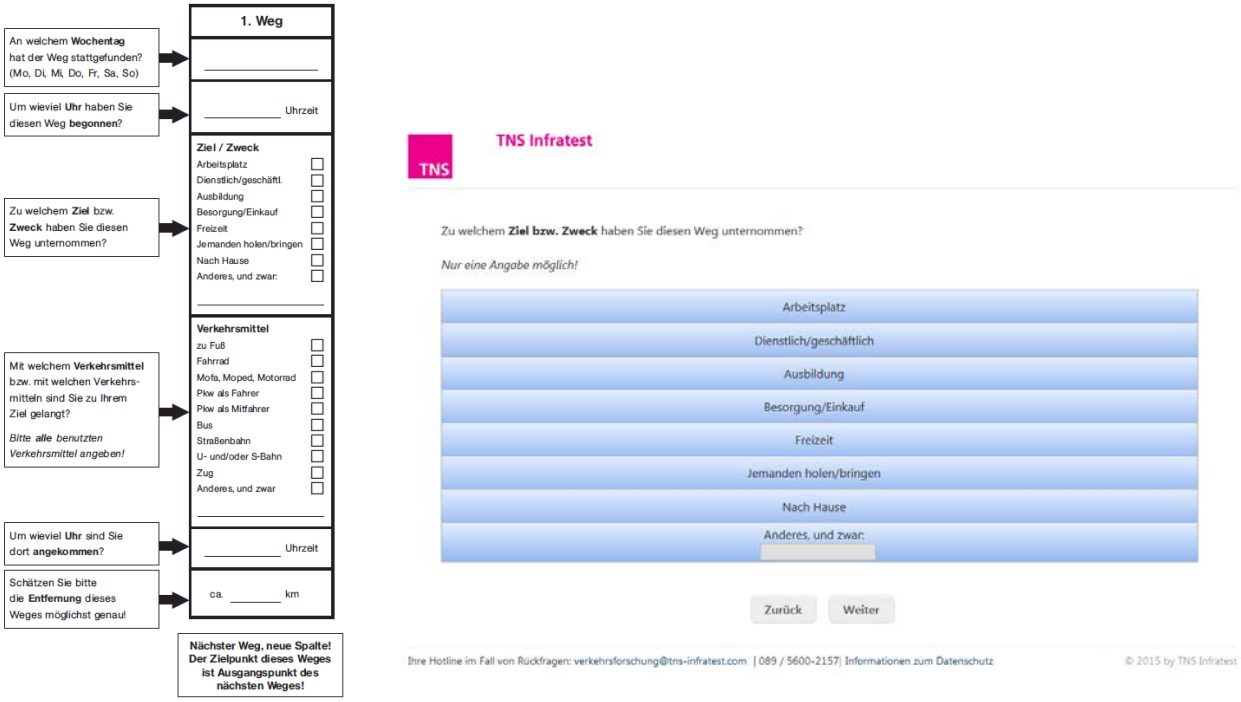
Survey mode adaptations made

In the prior survey design, households were contacted only by landline phone. In 2013 we introduced a dual-frame sampling approach and draw two samples now, a landline sample, and a mobile sample. Here, we aim to reduce the coverage error since households without landline connection – especially small households with young household members – are covered in the sampling frame now. Also households that do have a landline connection but are in fact only mostly reachable by their mobiles are better covered in the sample now.

Working with two independent sampling frames requires a design weighting that adjusts the different chances of being selected. This design weighting considers the overall probability of

a household to be included in the sample by combining the probability of being interviewed by landline phone or mobile phone (*17*).

Previously, the household questionnaire and the trip diary were PAPI surveys. Since 2013, participants can decide whether they complete a PAPI or a CAWI survey (Figure 1, concurrent mixed-mode design). The link for CAWI is imprinted on the PAPI questionnaire in combination with individual log-in data. CAWI is accessible via IP address and via QR code. The CAWI form is device agnostic and can be completed on PCs, laptops, tablets and smartphones. Therefore, we chose a survey design of only one question at once in order to enable an easy use also of mobile devices (*18*). However, CAWI might require more time to fill in the questionnaire than PAPI. To reduce *MEs*, we have not included plausibility checks when filling in the CAWI survey. Respondents have thus the same degree of freedom when filling in the diary, both in PAPI and in CAWI. The aim of CAWI is a reduction non response errors and an improvement of representativeness, because certain population groups might be rather willing to use the CAWI instead of the PAPI diary.



a) PAPI diary b) CAWI diary

Figure 1 Illustration of the PAPI (a) and the CAWI (b) trip diary in the MOP.

Sample for investigation

For our study on survey mode effects we analyzed the survey waves 2013, 2014 and 2015. Wirtz et al. (*19*) show that the survey report process of participants changes with repeating participation in the MOP; this influences the measured travel quantities. In order to exclude that effect from our analysis, our sample consists of first-year reporters only, i.e. respondents that

take part in the MOP for the first time. Weighting factors are not used for this analysis since this study focusses on the changes in travel behavior due to survey mode adaptations and not on quantities in travel behavior.

The size of the sample is 3,566 persons. The households of 2,341 participants were recruited via landline phone and the households of 1,225 participants were recruited via mobile phone. For the following analyses we divide the sample into four survey mode groups by their survey modes within the recruitment stage (landline, mobile) and the data collection stage (PAPI, CAWI). From the landline sample 2,054 persons filled in the PAPI trip diary (58% of the total sample) and 287 persons used CAWI (8%). 1,054 persons of mobile sample (30% of the total sample) completed the PAPI survey and 171 persons filled in the CAWI trip diary (5%). For the following analyses we distinguish the sample by these four survey mode groups.

METHODS

For the analysis of a mixed-mode survey design on survey results in the MOP we utilize a method from social science literature. Therefore, we define different survey effects, conduct descriptive analyses on the socio-demographics of the survey participants and on the reported travel behavior and introduce the propensity score weighting method used to decompose survey mode effects.

Definition of survey mode effects

The definition of survey mode effects utilized in (*9; 23; 22; 21; 20*) are in the following paragraphs summarized and applied to our case.

Let μ_t^m denote a continuous random variable for the outcome on a given question posed under survey mode *m* of the sample *t*. For our case μ might be a variable conducted in the trip diary, such as trips made or mileage driven per week. Because the MOP participants choose between PAPI and CAWI in the data collection stage, we divide the sample into these two groups. This means that $\mu_{paper}^{paper}$is the outcome on a given question of the PAPI subsample within the survey mode PAPI. Equally, μ_{web}^{web} is the outcome on a given question of the CAWI subsample within the survey mode CAWI.

Once, the survey participants have participated either under mode PAPI or under mode CAWI, we can calculate a pure total effect (*TE*) as

$$TE\ (\mu) = \mu_{web}^{web} - \mu_{paper}^{paper}$$

The *TE* value on a target variable only indicates whether and how the mean outcomes of PAPI sample and the CAWI sample differ. This analysis does not explain, why the average values of the two samples vary. Thus, the *TE* can be broken down to a *SE* and to a *ME* as

$$TE(\mu) = SE(\mu) + ME(\mu).$$

We need to define a benchmark mode. The benchmark mode is supposed to be the prevalent mode of the survey. We choose PAPI as reverence mode since it was used in the MOP before the mode adaptions were implemented.

The *SE* is defined as

$$SE = \mu_{web}^{paper} - \mu_{paper}^{paper}$$

with μ_{web}^{paper} as the potential outcome members of the CAWI subsample would have given if they had answered the survey under PAPI.

The *SE* compares the outcomes of the two different subsamples under the same hypothetic mode. *SE ≠0* indicates that the CAWI subsample answers differently to the given question than the PAPI subsample. This means for our case that the travel behavior of the PAPI subsample and the CAWI subsample differs. The main reason of a different travel behavior are different socio-demographic characteristics of the two subsamples.

The *ME* is defined as

$$ME = \mu_{web}^{web} - \mu_{web}^{paper}.$$

The *ME* compares the average values of the same subsample, assuming that they would have answered the same questions in different modes. The *ME* suggests how the measurement bias of the focal mode differs from the measurement bias of the benchmark mode. Survey collectors of mixed-mode surveys want that the same participants to provide the same answers to given questions in different modes, i.e. they aim as a *ME* as small as possible. *ME=0* indicates measurement equivalence.

Descriptive analyses on socio-demographics and on travel behavior

To gain a better understanding of the socio-demographic characteristics and the travel behavior of the sample, we conduct descriptive analyses for the four survey mode groups. Descriptive analyses on travel behavior enable to examine *TEs* only.

Propensity score weighting to decompose mode effects

The potential survey outcomes μ_{web}^{paper} are unknown since we have not asked the participants of the CAWI subsample to fill in a PAPI diary additionally to CAWI. For the following analyses we use propensity score (*ps*) weighting as estimation technique for the potential outcomes of μ_{web}^{paper}. The goal of this approach is to decompose the effects of a mixed-mode design on the results. The potential outcome framework was primary shaped by Rubin and Rosenbaum (e. g. *24; 25*). This concept was not initially introduced to decompose mixed-mode effects but to analyze causal effects in nonrandomized studies. They define the *ps* as the conditional probability of assignment to a particular treatment given a vector of unobserved covariates (*25*).

The *ps* π_{web} is the probability of a person to be selected into the focal mode CAWI. A binary logistic regression is used to estimate π_{web}. Auxiliary variables for the model are provided by socio-demographic information known from the survey. Hence, one needs to ensure with the selection of variables that the responses to those variables are not influenced by the

survey modes. We thus assume that the responds to the socio-demographic variables used do are not affected by survey modes since PAPI and CAWI are both self-administered modes.

It is assumed that the *ps* depend on auxiliary variables X only:

$$\hat{\pi}_{web} = P(M = web|X)$$

The hat on μ indicates that the *ps* is an estimate. Next, the logistic regression results are applied to the sample and the response sample of the PAPI mode is weighted to the CAWI mode, as

$$\mu_{web}^{paper} = \mu_{paper}^{paper} * w_{web}.$$

The weighting factor w_{web} is defined as

$$w_{web} = \frac{\hat{\pi}_{web}}{1 - \hat{\pi}_{web}}$$

Once we have calculated μ_{web}^{paper} we can determine *ME* and *SE*.

We conduct the survey mode effect analyses separately for the sample of landline recruits and of mobile recruits. This is because participants could not choose their recruitment mode but they chose the data collection mode.

RESULTS

The methods used enable us to present descriptive results for the four survey mode groups, as defined earlier. We are furthermore able to explain by the use of a *ps* weighting model and an underlying logistic regression model why the reported travel behavior of the four survey mode groups differs.

Descriptive analyses – socio-demographic characteristics

Descriptive analyses of socio-demographic characteristics for the four survey mode groups are shown in Table 1. These findings indicate that the different mode groups vary in terms of their gender, age, employment status, household income and place of residence.

Table 1 Socio-demographic characteristics of the sample, grouped by survey mode groups

	Recruitment: landline				Recruitment: mobile			
	PAPI diary		CAWI diary		PAPI diary		CAWI diary	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
Gender: male	47%	0.499	57%	0.496	51%	0.500	60%	0.492
Age group: 10-25 years	11%	0.318	13%	0.340	14%	0.343	12%	0.329
Age group: 26-35 years	6%	0.232	13%	0.332	14%	0.347	24%	0.428
Age group: 36-50 years	20%	0.402	32%	0.466	26%	0.439	36%	0.482
Age group: 51-60 years	24%	0.428	21%	0.407	22%	0.413	13%	0.342
Age group: 61-70 years	19%	0.396	15%	0.354	15%	0.359	13%	0.336
Age group: 71 years and older	19%	0.392	7%	0.255	9%	0.293	1%	0.108
Employment status: employed	51%	0.500	57%	0.496	59%	0.493	70%	0.459
Employment status: in education	11%	0.313	14%	0.347	13%	0.341	10%	0.300
Level of education: secondary school (Hauptschule)	19%	0.396	13%	0.332	22%	0.413	12%	0.322
Level of education: university-entrance diploma (Abitur)	13%	0.333	12%	0.328	14%	0.349	15%	0.360
Monthly household income: 1.500€ and less	12%	0.326	8%	0.267	15%	0.362	4%	0.199
Monthly household income: 3.000€ and more	47%	0.499	62%	0.486	43%	0.495	54%	0.500
Place of residence: newly-formed Germany states	22%	0.412	19%	0.394	28%	0.450	20%	0.400
Other survey participants in the household: no	34%	0.474	32%	0.469	37%	0.483	41%	0.493
Mobile phone: yes	81%	0.393	94%	0.230	89%	0.310	96%	0.185
Public transit pass: yes	18%	0.386	21%	0.410	20%	0.401	26%	0.442
Carsharing member: yes	0%	0.070	1%	0.102	1%	0.087	4%	0.199
Mobility restrictions: yes	11%	0.317	7%	0.261	9%	0.285	9%	0.284
Sample size	*2,054 (58%)*		*287 (8%)*		*1,054 (30%)*		*171 (5%)*	

CAWI respondents are more often male, younger than 50 years, went rarely to secondary school, are more often part time or full time employed and their household income is rather high, no matter whether their household was recruited via landline or via mobile phone. Interestingly, the share of participants without other household members participating in the survey as well is with 34% versus 32% higher for PAPI respondents amongst landline recruits but lower amongst mobile recruits (37% vs. 41%). The mobile phone coverage is in all groups with more than 80% quite high. However, we only know whether participants have a mobile phone or not but it is not surveyed whether they have a smartphone, tablet or rater a simple mobile phone without internet connection. The mobility tools transit pass and car ownership are more spread amongst CAWI respondents within both recruitment groups. This might indicate that CAWI participants combine different modes of transportation.

These results indicate furthermore that – unlike our expectations before we implemented the survey mode adaptations – most participants in both recruitment samples prefer PAPI instead of CAWI. Limited time available to fill in an CAWI survey for busy people, or limited access to a computer for population groups with limited resources are reasons evoked in the literature explaining a low response rate by CAWI (*14*).

Descriptive analyses – travel behavior

When describing travel behavior, the quantities trips made, distance travelled, time spent in the transportation system and the number of days in the survey period with any trip making are commonly used.

Table 2 Quantities on travel behavior and data quality (Mean, (StDev)), grouped by survey modes groups

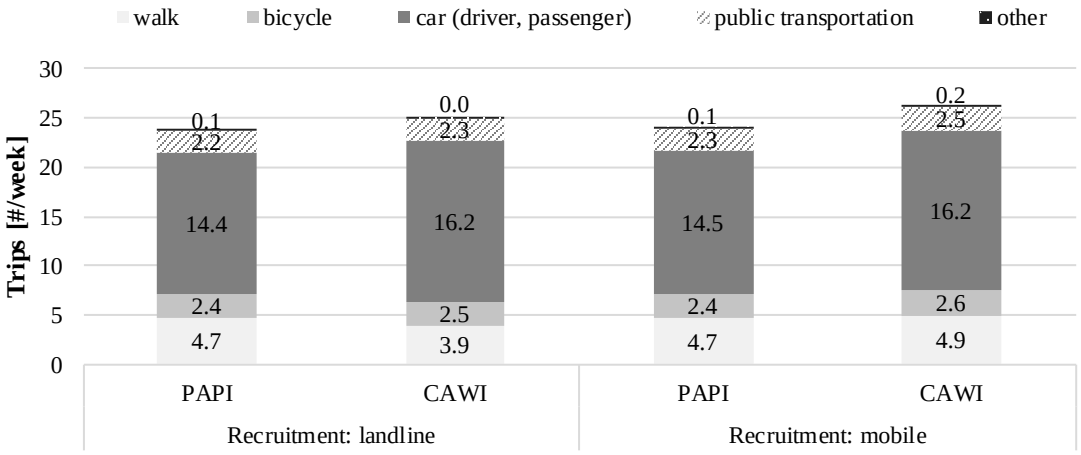
Quantities per week	Recruitment: landline		Recruitment: mobile	
	PAPI diary	CAWI diary	PAPI diary	CAWI diary
Trips made [#]	23.7 (9.4)	24.9 (10.5)	24.0 (9-1)	26.4 (11.1)
Distance travelled [km]	281.8 (280.4)	368.0 (395.4)	332.1 (329.1)	398.5 (369.7)
Time spent in the transportation system [min]	581.3 (329.6)	598.7 (338.0)	617.4 (352.1)	654.6 (333.4)
Days with any trip-making	6.4 (1.0)	6.3 (1.0)	6.5 (0.8)	6.6 (0.9)

We assume that mode effects are in particular linked to the number of trips reported. Survey participants might rather not report certain trips in their trip diaries within different survey modes, e.g. because the trips were to short and they omitted it when filling in the diary at a later time. Another explanation might be that the reporting procedure was felt too time-consuming for the participants and they thus decided to summarize several short trips to one long trip or they did not report the (short) trip at all. This is why we analyze the number of trips reported in a greater detail. This kind of reporting behavior impacts further quantities indirectly.

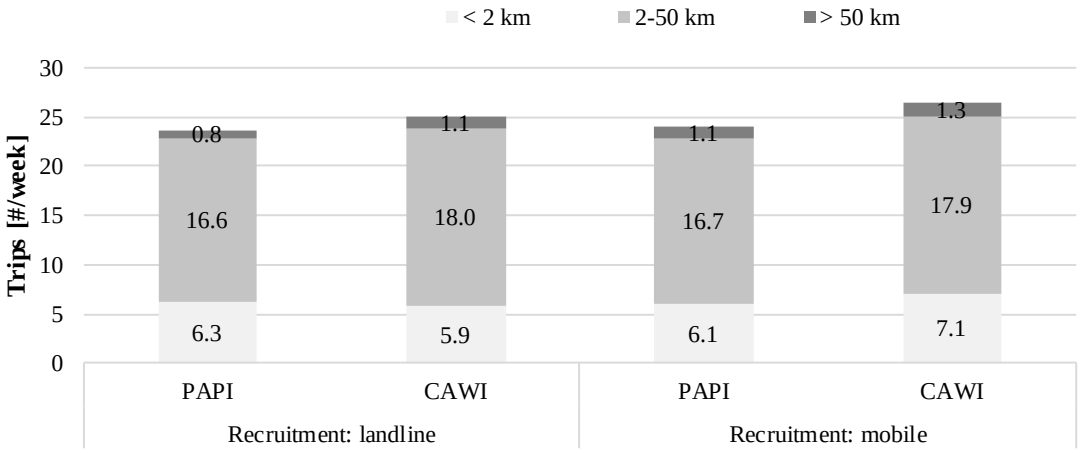
Another reason for mode effects could be that participants fill in the trip diaries at a later point of time than they would have done it using another survey mode; they can maybe not remember the detailed trip characteristics, such as time of departure, time of arrival and distance travelled but they report approximate information only (e.g. 11:00 a.m. instead of 11:04 a.m., 50 km instead of 46 km). Different levels of approximation of trip characteristics amongst different survey mode groups leads to *ME* for the quantities distance travelled and time spent in the transportation system.

Our in analysis on the characteristics of the trips reported (Figure 2) shows that CAWI participants report more trips, both in the landline and in the mobile recruitment sample. The trip level by bicycle, public transportation and other modes is similar within the four survey mode groups. However, CAWI reporters among the landline recruits report about one trip less per week by walk than the other survey mode groups. Furthermore, CAWI reporters make about 1.5 car trips more per week than PAPI reporters. Regarding trips made in different distance classes we observe that CAWI reporters amongst mobile recruits have most trips below 2 km and trips above 50 km. Analyses on trip quantities for different purposes show that mobile recruits report more commuting trips while the number of shopping trips is higher amongst landline recruits.

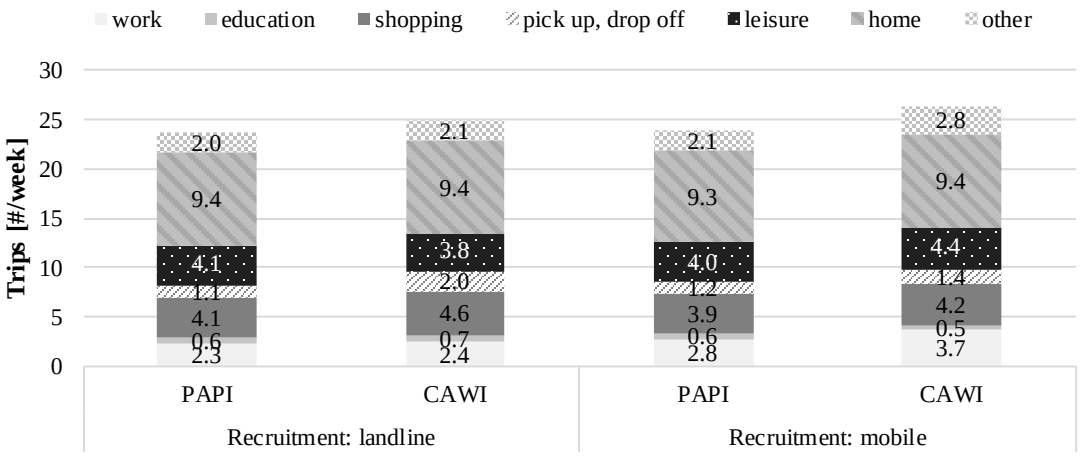
Our analyses indicate that the travel behavior amongst the four survey mode groups varies: obviously there are differences in travel behavior to observe between the different form of recruitment on the one side (mobile recruited persons seem to be more active than landline recruits) whereas CAWI participants report higher quantities of travel than the PAPI respondents. As shown above there is obviously the need to split up the *TE* into the *SE* and into a *ME*.



(a) Modes of transportation



(b) Trip distances



(c) Trip purposes

Figure 2 Number of reported trips for the four survey mode groups, distinguished by (a) modes of transportation, (b) trip distances, and (c) trip purposes.

Logistic regression – estimation results

For the *ps* weighting approach, it is necessary to understand in which way different socio-economic and socio-demographic characteristics show influence on the probability that a person will choose a certain mode for its trip diary. Therefore, binary logistic regression models have been introduced in order to estimate the relative influence of the relevant characteristics that a participant uses CAWI, both for the landline and for the mobile recruitment sample. Table 3 shows the estimates and the significance levels for the parameters included in the model. All socio-economic and socio-demographic characteristics have been defined as binary variables.

Table 3 Estimation Results of the Binary Logistic Regression Models. Probability that a Survey Participant Uses CAWI

	Recruitment: landline		Recruitment: mobile	
Parameter	Estimate	P > |t|	Estimate	P > |t|
Intercept	-3.374	.000	-2.287	.000
Gender: male	0.462	.001	0.352	.044
Age group: 10-25 years	-	-	0.764	.078
Age group: 26-35 years	1.251	.000	0.833	.001
Age group: 36-50 years	0.954	.000	0.603	.006
Age group: 51-60 years	0.414	.067	-	-
Age group: 71 years and older	-0.822	.002	-1.797	.015
Employment status: employed	-0.703	.000	-	-
Employment status: in education	-	-	-0.913	.037
Level of education: secondary school (Hauptschule)	-0.340	.083	-0.508	.056
Level of education: university-entrance diploma (Abitur)	-0.278	.163	-	-
Monthly household income: 1.500€ and less	-	-	-1.547	.000
Monthly household income: 3.000€ and more	0.473	.001	-	-
Place of residence: newly-formed Germany states	-	-	-0.398	.061
Further survey participants in the household: no	0.252	.094	0.389	.034
Mobile phone: available	1.054	.000	-	-
Public transit pass ownership: yes	-	-	0.365	.097
Carsharing member	-	-	1.280	.023
Health related restrictions of mobility: yes	-	-	0.642	.047
Number of observations	*2,341*		*1,225*	
Log likelihood at convergence	*1,623*		*902*	
McFadden index	*0.07*		*0.09*	

The regression models indicate that in both samples, men, young age groups and households with higher incomes prefer the CAWI diary. In the landline sample, mobile phone availability affects CAWI positively. Amongst the mobile recruits carsharing membership, transit pass availability and the place of living in the so called “old” German federal states (Germany in the borders until 1990) prefer CAWI. Other socio-demographic variables such as car ownership, household size, the existence of infants in the household and the spatial type were tested but occurred not to be significant in the regression models.

The McFadden index suggests that, although most of the variables used are significant at least on the 10% level, the accuracy of both models is reasonably good only. However, the person-related variables conducted in the MOP are no sufficient to improve the model further. Therefore, the inclusion of questions regarding technology affinity, e.g. the availability of smartphones and tablets, usage of mobility apps, etc. might be advantageous.

Propensity score weighting

The results of the *ps* weighting analysis are shown in Table 4. Some explanations concerning the interpretability of the results: *TEs* greater than zero occur when the average values of the CAWI sample are higher than the average values of the PAPI sample. The results of the variable numbers of trips made per week (landline sample) show that the *TE* is almost completely explained by the *SE*, the *ME* is negligible. This means that different survey modes in the data collection stage do not impact the outcome of the travel quantities number of trips made. For the quantity of distance travelled per week in the landline sample the *TE* is caused by the *ME* and the *SE* in approximately the same magnitude. This means that about 50% of the effect can be explained by different travel behaviors amongst the subsamples and the other 50% by different modes of reporting. It may happen that the *SE* and the *ME* compensate each other partly. E.g. the *TE* of the quantity number of days with any trip-making is -0.103 (landline sample). The decline would be even greater with the *ME* only; the fact that the CAWI participants amongst the landline recruits have a higher general level of travel quantities leads to a positive *SE* and thus to a partly compensation of *ME*.

Table 4 Resulting TE, SE and ME on travel quantities, grouped by recruitment mode

	Recruitment: landline					Recruitment: mobile				
	TE	*SE*	*ME*	PAPI Mean	CAWI Mean	*TE*	*SE*	*ME*	PAPI Mean	CAWI Mean
Trips made per week [#]	1.249	1.256	-0.007	23.685	24.934	2.396	1.542	0.854	23.954	26.351
Distance traveled per week [km]	86.177	42.425	43.752	281.833	368.010	66.428	56.875	9.552	332.086	398.513
Time spent in the transportation system per week [min]	17.370	19.579	-2.209	581.285	598.655	37.166	42.227	-5.061	617.448	654.614
Days with trip making [#]	-0.103	0.067	-0.171	6.431	6.328	0.017	0.076	-0.060	6.545	6.561
Trips made per week grouped by modes of transportation [#]										
Walk	-0.831	-0.111	-0.720	4.726	3.895	0.250	0.315	-0.065	4.686	4.936
Bicycle	0.123	0.011	0.112	2.369	2.491	0.114	0.168	-0.055	2.448	2.561
Car (driver / passenger)	1.866	1.357	0.509	14.378	16.244	1.711	0.823	0.888	14.453	16.164
Public transportation	0.098	-0.007	0.105	2.156	2.254	0.192	0.215	-0.023	2.299	2.491
Trips made per week grouped by trip distance [#]										
< 2 km	-0.424	0.203	-0.627	6.292	5.868	1.053	0.408	0.645	6.082	7.135
2-50 km	1.404	0.861	0.543	16.572	17.976	1.135	0.919	0.216	16.749	17.883
>50 km	0.269	0.192	0.077	0.821	1.091	0.209	0.216	-0.007	1.124	1.333
Trips made per week grouped by trip purposes [#]										
Work	0.091	0.371	-0.279	2.313	2.404	0.904	0.822	0.082	2.792	3.696
Education	0.126	0.124	0.002	0.557	0.683	-0.141	-0.121	-0.021	0.603	0.462
Shopping, running errands	0.452	-0.022	0.474	4.109	4.561	0.288	-0.110	0.397	3.894	4.181
Picking someone up, dropping someone off	0.847	0.293	0.555	1.129	1.976	0.169	0.263	-0.094	1.229	1.398
Leisure	-0.661	-0.034	-0.627	5.124	4.463	0.229	0.110	0.119	4.835	5.064

The results from the landline recruitment sample show that differences in the travel quantities trips, time spent in the transportation system and days with trip making can be mainly explained by *SEs*. However, the *ME* for the kilometers travelled is higher than the *SE*. This indicates the CAWI diary leads to an overestimation of trip distances compared to the PAPI mode. This is maybe the case when participants fill in the trip diary only one time for the full week and they cannot remember the exact distance but overestimate. We found negative *ME*s for both, trips by foot and trips shorter than 2 km (these results are highly correlated, since most short trips are walking trips), indicating that CAWI participants tend to report fewer short walking trips as if they would have filed in the PAPI survey. The higher number of car trips amongst the CAWI sample is mainly explained by the socio-demographics of that sample. Different trip numbers for mandatory trip purposes, such as trips to work or education and trips to pick someone up or drop someone off, are mainly explained by *SEs* whereas discretionary trips, such as shopping and leisure trips, face higher *MEs*. A potential reason is that mandatory trips happen one a more regular basis, e.g. once a day, the destination, departure time and mode of transportation stays the same and it is thus easier to remember, even when the trip diary id filled out only once for the whole survey week.

For the mobile recruitment sample negative *MEs* amongst all transportation modes but the car indicate that CAWI respondents reported fewer trips per walk, car and public transportation as they would have reported when using PAPI. However, *TEs* are still positive due to the different sample in terms of socio-demographics. As a matter of fact, we find no explanation why the *ME* is positive for trips shorter than 2 km. Again, *SEs* are higher than *MEs* for mandatory trip purposes and the other way around for discretionary trip purposes.

A comparison of the outcomes of the mobile and the landline sample indicates, that disparities in travel quantities are more often traced back to *SEs* in the mobile sample than in the landline sample. This leads to the questions whether the CAWI participants amongst the landline recruits are rather bad risks since they are not that much used to online tools than the CAWI sample amongst mobile recruits.

CONCLUSION

Within this work, we analyzed the effects of a mixed-mode design in the MOP to the survey results. Therefore, we utilized the method of *ps* weighting in order to impute hypothetic responses participants would have given under other survey modes. Disparities between actual responses and hypothetical responses under another mode are caused by *MEs*. Our analyses indicate that differences in quantities of kilometres travelled, time spent in the transportation system and trips made by car and public transportation are mainly explained by a *SE*. Different trip rates on shopping and leisure trips and on short trips below 2 km are partly be traced back to *MEs*. Furthermore, disparities in reporting behaviour due to different survey modes occur more often amongst landline recruits.

Our analysis helps to identify whether differences in reported travel behavior are caused by different survey modes. These findings can be utilized to adapt focal modes of a survey further. For the MOP, it might be advantageous to provide an app for CAWI with automatic regular reminders to fill in the diary. In this way, CAWI respondents might forget fewer short trips.

A mayor shortcoming of our analysis is that the fits of our logistic regression models underperform. Though, socio-demographic information only might not be sufficient to estimate the probability of a survey participant to choose CAWI. To overcome these shortcomings and thus to improve the accuracy of the *ps* weighting, we suggest to include questions regarding technology affinity in the survey (e.g. the availability of smartphones and tablets, usage of mobility apps).

Not only the survey mode itself might lead to *MEs* but also deviations in the point in time and the frequency of trip diary completion. Thus, they might differ for the survey modes. Recording the timestamps of survey completion in the CAWI survey might help to gain a better understanding.

There is a broad discussion in social science literature, which statistical methods are most suitable to decompose survey mode effects. We used *ps* weighing for our analyses, since it is a straight and less complex method. Other approaches for statistically adjusting mode effects are *ps* matching (distance measures between all pairs of individuals are estimated), double robust regressions (combination of *ps* based and regression based estimations) and multiple imputation (mode effects are conceptualized as missing data problems) (*26; 21; 20*). Further in depth research may use various statistical measures to give additional insight.

Overall, our analyses indicate that survey mode effects need to be taken into consideration when analyzing and comparing travel survey outcomes, since the mode in which a travel survey is conducted might influence the outcome.

ACKNOWLEDGEMENT

This paper presents analyses of the German Mobility Panel funded by the German Federal Ministry of Transport and Digital Infrastructure.

REFERENCES

1. Weiss, C., B. Chlond, T. Hilgert, and P. Vortisch. *Deutsches Mobilitätspanel (MOP) - Wissenschaftliche Begleitung und Auswertungen Bericht 2014/2015: Alltagsmobilität und Fahrleistung,* 2016.
2. Christensen, L. Possible Explanations for an increasing share of no-trip respondents. In *Travel Survey Methods - Quality and Future Directions,* Stopher, P., Stecher, Ch., ed. Elsevier, Oxford, Amsterdam, 2006, pp. 303–336.

3. De Leeuw, E. D. To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics*, Vol. 21, No. 2, 2005, pp. 233–255.
4. Chlond, B., T. Streit, G. Abler, and P. Vortisch. Balancing Innovation and Continuity – Experiences with Survey Design Adaptations of the German Mobility Panel. *Transportation Research Procedia*, 2015, pp. 43–59.
5. Dillman, D. A., G. Phelps, R. Tortora, K. Swift, J. Kohrell, J. Berck, and B. L. Messer. Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. *Social Science Research*, Vol. 38, 2009, pp. 1–18.
6. Brambilla, D. J., and S. M. McKinlay. A comparison of responses to mailed questionnaires and telephone interviews in a mixed mode health survey. *American Journal of Epidemiology*, Vol. 126, No. 5, 1987, pp. 962–971.
7. Link, M. W., M. P. Battaglia, M. R. Frankel, L. Osborn, and A. H. Mokdad. Reaching the U.S. Cell Phone Generation - Comparison of Cell Phone Survey Results with an Ongoing Landline Telephone Survey. *Public Opinion Quarterly*, Vol. 71, No. 5, 2007, pp. 814–839.
8. Hu, S. S., L. Balluz, M. P. Battaglia, and M. R. Frankel. Improving public health surveillance using a dual-frame survey of landline and cell phone numbers. *American Journal of Epidemiology*, Vol. 173, No. 6, 2011, pp. 703–711.
9. Klausch, T., J. J. Hox, and B. Schouten. Measurement Effects of Survey Mode on the Equivalence of Attitudinal Rating Scale Questions. *Sociological Methods & Research*, Vol. 42, No. 3, 2013, pp. 227–263.
10. Beuckelaer, A. de, and F. Lievens. Measurement Equivalence of Paper-and-Pencil and Internet Organisational Surveys. A Large Scale Examination in 16 Countries. *Applied Psychology*, Vol. 58, No. 2, 2009, pp. 336–361.
11. Wolf, J., J. Wilhelm, J. Casa, and S. Sen. Case Study: Multiple Data Collection Methods and the NY/NJ/CT Regional Travel Survey. In *Transport Survey Methods - Best Practice for Decision Making*, J. Zmud, M. Lee-Gosselin, M. Munizaga and J.A. Carrasco, eds. Emerald Group Publishing, 2013, pp. 71–90.
12. Thériault, M., M. Lee-Gosselin, L. Alexandre, F. Théberge, and L. Dieumegarde. Web versus Pencil-and-Paper Surveys of Weekly Mobility: Conviviality, Technical and Privacy Issues. In *Transport Survey Methods - Best Practice for Decision Making*, J. Zmud, M. Lee-Gosselin, M. Munizaga and J.A. Carrasco, eds. Emerald Group Publishing, 2013, pp. 225–246.
13. Bayart, C., and Bonnel Patrick. The mixing of survey modes: application to Lyon web and face-to-face household travel survey. In *12th World Conference in Transportation Research*, Lisbon, Portugal, 2010.

14. Christensen Linda. The Role of Web Interviews as Part of a National Travel Survey. In *Transport Survey Methods - Best Practice for Decision Making*, J. Zmud, M. Lee-Gosselin, M. Munizaga and J.A. Carrasco, eds. Emerald Group Publishing, 2013, pp. 115–153.
15. Streit, T., B. Chlond, C. Weiss, and P. Vortisch. *Deutsches Mobilitätspanel (MOP) - Wissenschaftliche Begleitung und Auswertungen Bericht 2013/2014: Alltagsmobilität und Fahrleistung*, 2015.
16. Zumkeller, D., and B. Chlond. Dynamics of Change: Fifteen-Year German Mobility Panel. In *TRB 88th Annual Meeting Compendium of Papers*, Washington, D.C., 2009.
17. Gabler, S., and Ö. Ayhan. Gewichtungen bei Erhebungen im Festnetz und über Mobilfunkt: Ein Dual Frame Ansatz. In *Mobilfunktelefonie. Eine Herausforderung für die Umfrageforschung*, S. Gabler, ed. ZUMA, Mannheim, 2007, pp. 39–46.
18. Arn, B., S. Klug, and J. Kołodziejski. *Evaluation of an Adapted Design in a Multi-device Online Panel*, 2015.
19. Wirtz, M., T. Streit, B. Chlond, and P. Vortisch. On New Measures for Detection of Data Quality Risks in Mobility Panel Surveys. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2354*, 2013, pp. 19–28.
20. Vannieuwenhuyze, J. T. A., and G. Loosveldt. Evaluating Relative Mode Effects in Mixed-Mode Surveys. Three Methods to Disentangle Selection and Measurement Effects. *Sociological Methods & Research*, Vol. 42, No. 1, 2013, pp. 82–104.
21. Kolenikov, S., and C. Kennedy. Evaluating Three Approaches to Statistically Adjust for Mode Effects. *Journal of Survey Statistics and Methodology*, Vol. 2, No. 2, 2014, pp. 126–158.
22. Klausch, T., J. Hox, and B. Schouten. Selection error in single- and mixed mode surveys of the Dutch general population. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 178, No. 4, 2015, pp. 945–961.
23. Klausch, T. *Lecture 1: Decomposing total mode effects in mixed mode data. Lecture notes of the GESIS summer course "mixed mode and mixed device surveys"*, 2015.
24. Rosenbaum, P. R. Model-Based Direct Adjustment. *Journal of the American Statistical Association*, Vol. 82, No. 398, 1987, p. 387.
25. Rosenbaum, P. R., and D. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, Vol. 70, No. 1, 1983, pp. 41–55.
26. Kemmick Pintor, J. B., D. McAlpine, T. J. Beebe, and P. J. Johnson. Propensity score matching to measure the effect of survey mode on reports of racial and ethnic discrimination in health care. *Medical care*, Vol. 53, No. 5, 2015, pp. 471–476.