Transportation Research Part A 153 (2021) 66–82

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Comprehensive data validation of a combined weekly time use and travel survey

Florian Aschauer [a,*], Reinhard Hössinger [a], Sergio Jara-Diaz [b,c], Basil Schmid [d], Kay Axhausen [d], Regine Gerike [e]

[a] Institute for Transport Studies, University of Natural Resources and Life Sciences, Vienna, Austria
[b] Department of Civil Engineering, University of Chile, Santiago, Chile
[c] Instituto de Sistemas Complejos de Ingeniería (ISCI), Santiago, Chile
[d] ETH Zurich, Institute for Transport Planning and Systems (IVT), Zurich, Switzerland
[e] Technische Universität Dresden, Chair of Integrated Transport Planning and Traffic Engineering, Dresden, Germany

ARTICLE INFO

Keywords:
Data quality
Data validation
Item-nonresponse
Mobility behaviour
Survey quality
Travel survey methods

ABSTRACT

A well-known problem in Household Travel Surveys (HTS) is item-nonresponse, which occurs when complete trips are not reported or only partial information for trips is given. This article presents a comprehensive data validation effort (data completion and correction through immediate call-backs of the survey participants) on travel and non-travel information as part of a one-week Mobility-Activity-Expenditure Diary Survey. Validation recovered 2.5 percent reporting days of mobile persons with complete information on trips and increased the number of trips of pre-existing mobile reporting days by 3.5 percent. The characteristics of these trips confirm and extend the findings from item-nonresponse studies without validation: The majority of the new trips begin in the afternoon; they are mainly short and irregular; the most important trip purposes are home, work, shopping, and leisure. 37 percent of the new trips generated completely new tours; females seem to under-report less than men do. The analysis of the information on the time use (activities) between trips showed that the data was recorded with high accuracy and completeness even before validation. This study confirms typical item-nonresponse patterns and provides a solution for their mitigation through direct follow-up validation. The validation effort increased the time and cost of fieldwork by 12 percent.

1. Introduction

Household Travel Surveys (HTS) have been an important data source for transport planning and research since the 1970s. Survey methods have been improved continuously, and current HTS are mainly mixed-mode surveys which are combinations of interview types such as computer-assisted personal (CAPI), telephone (CATI), web (CAWI), and, to some extent, also paper-and-pencil (PAPI) (Armoogum et al. 2014; Eisenmann et al. 2018). Innovative tracking methods with dedicated GPS-devices or smartphones are not yet fully established in HTS practice but are widely used in research (Gershuny et al. 2017; Safi et al. 2017). As in all types of surveys, HTS are subject to a variety of errors. For example, the standard mixed-mode HTS is subject to several error types (Richardson et al. 1995;

* Corresponding author.
E-mail addresses: florian.aschauer@boku.ac.at (F. Aschauer), r.hoessinger@boku.ac.at (R. Hössinger), jaradiaz@ing.uchile.cl (S. Jara-Diaz), basil.schmid@ivt.baug.ethz.ch (B. Schmid), axhausen@ivt.baug.ethz.ch (K. Axhausen), regine.gerike@tu-dresden.de (R. Gerike).

https://doi.org/10.1016/j.tra.2021.08.011
Received 17 September 2020; Received in revised form 16 July 2021; Accepted 23 August 2021
Available online 14 September 2021

Groves and Lyberg 2010; Hubrich et al. 2018). Sampling errors occur when the sample is biased e.g. in its proportions of specific person groups. These can be mitigated by data weighting. Non-sampling errors include all issues that are not directly related to the sampling process (Bonnel et al. 2015; Hubrich et al. 2018). This study focuses on one important non-sampling error; the item-nonresponse problem either when complete trips are missing in respondent answers or when the information provided on trips is incomplete. For example, certain items such as addresses or transport mode might not be reported correctly or information about persons/household such as income or available mobility tools might be missing.

Item-nonresponse errors can be sorted into the following three types (Brög et al. 1982; Richardson 2003): (i) errors which can be detected during data processing without any additional observed information; (ii) errors which can only be detected by means of follow-up assessments; and (iii) errors which are non-detectable because the respondent is unwilling to disclose the information. While type (i) errors can be corrected without re-contacting the respondents and type (iii) errors cannot be corrected at all, errors of type (ii) can be corrected by immediate follow-up data completion and validation procedures after survey participation. Incoming data from respondents is checked promptly, and the fieldwork team gets back to respondents if type (ii) errors are suspected. The fieldwork team completes or corrects the data based on the telephone conversation with the respondent with the final goal of getting as close as possible to the actual behaviour on the reporting day(s). These activities are referred to as "(data) validation" in the following text.

Item-nonresponse issues in HTS have been widely investigated with a focus on the number and characteristics of missing trips. These studies have been based on comparisons of different HTS methods (Ampt & Stopher 2006; Bayart and Bonnel 2012; Wolf et al. 2013; Armoogum et al. 2014; Chlond et al. 2015), HTS with time use surveys (TUS) (Gerike et al. 2015), HTS with GPS-based innovative survey approaches (Bricka and Bhat 2006; Taniar et al. 2014; Safi et al. 2017), HTS with further data sources such as wearable cameras (Kelly et al. 2014; Gershuny et al. 2017), traffic counts (Ashley et al. 2009), and bankcard transaction data (Yoshimura et al. 2018), and comparison of HTS data series over time (Eisenmann et al. 2018; Wittwer et al. 2018).

Only a few studies could be identified that address the possibilities for mitigating type (ii) item-nonresponse errors through follow-up validation directly within the HTS as described above (Brög and Meyburg 1980; Brög et al. 1982; Brög and Winter 1990; Richardson 2003; Brög 2015). A possible reason for this might be that ex-post validation is rarely done; the costs are often perceived as outweighing the benefits. A detailed analysis of trips identified in addition to the ones originally reported by the respondent during validation as well as an item-by-item comparison of pre- and post-validation data has only been found in the studies done by the group led by Werner Brög (Brög and Meyburg 1980; Brög et al. 1982; Brög and Winter 1990; Brög 2015). These publications by Werner Brög and associates (1) only originate from one research group; (2) are quite old; (3) are not sufficiently detailed as they do not consider the insights gained in recent research on item-nonresponse based on studies without validation; (4) do not include any assessment on resources used (money or work-hours) to validate information and deal with incomplete or missing items; (5) are devoid of a context, i.e. refer to travel alone. Point 5 is particularly relevant in our research as our validation procedure applies to individuals that were surveyed under a new format, combining three survey disciplines in an integrated diary format: a time use survey, a consumer expenditure survey, and a travel survey. This allowed for validating information on trips and on activities including the address and location of activities as two highly relevant variables for time use analysis.

This study contributes to this strand of research by means of the following innovations: (i) We add a new case study to the few existing ones done by the research group led by Werner Brög (Brög and Meyburg 1980; Brög et al. 1982; Brög and Winter 1990; Brög 2015). (ii) The analysis of the validation results is up-to-date and much more comprehensive than former research efforts; it builds on the insights gained by the item-nonresponse studies carried out since the earlier data validation studies have been published (including also the ones without validation). (iii) We analyse and quantify for the first time the effort (time of work hours respectively money) of validation. (iv) By using a combined sample of travel behaviour and time use, we analyse and compare the effects of validation for these two types of data – a main novelty especially for the latter type.

The insights gained in this study based on the analyses on both types of information - e.g., the number of trips to determine the trip rate as well as specific trip information (or on activities) support particularly two groups of practitioners, (1) those who collect data on travel behaviour, (2) the users of data on travel behaviour (mainly transport planners and modellers). Both groups benefit from the empirical analysis of under-reporting in terms of travel and activities. The first group of practitioners gets up-to-date insights on efforts and effects of validation and might optimise their techniques for data collection based on this information. The second group gets a sound data basis for assessing possible biases in travel data to be used for planning and modelling.

This study is based on a PAPI survey but is also relevant for CAWI surveys. CAWI provides various opportunities for prompting and reminding respondents during the questionnaire process. However, trips (e.g., mid-day tours for lunch) can still be forgotten since they do not cause obvious inconsistencies and thus cannot be added in automated or manual post-processing without re-contacting the respondents. The completion of trip characteristics (e.g., addresses) is only partly possible in CAWI when respondents are not involved. Validation is also relevant for smartphone-based surveys since they are highly ambitious in terms of data quality yet lack the capacity for detecting trips or their respective characteristics with complete accuracy. Immediate validation together with the respondents might be more effective and might yield more accurate results compared to solely checking incoming data in the phase of post-processing without re-contacting respondents. Seeing that most HTS currently are mixed-mode surveys and will continue to be mixed-mode surveys in the coming years, these issues are of highest relevance for enhancing data quality and also the cost-effectiveness of HTS.

2. Literature

The reporting of results from validation efforts as described in this article could only be identified in a limited amount of literature. Therefore, the following summary of relevant literature mainly includes findings from item-nonresponse studies that do not focus on

validation but instead are based on alternative approaches as described in Section 1. Main findings from the few identified studies reporting the results of validation are summarized in the last paragraph of this section.

The literature on item-nonresponse consistently shows under-reporting effects in HTS for the number of trips (Brög and Meyburg 1980; Wolf et al. 2013; Armoogum et al. 2014; Taniar et al. 2014; Aschauer et al. 2018b; Aschauer et al. 2018a). Mainly short and irregular trips are under-reported in HTS, which results in diminished trip rates particularly for discretionary ('leisure') trips compared to mandatory ('work', 'education') and non-discretionary (e.g. 'shopping', 'errands') trips (Richardson 2007; Stopher et al. 2007; Stopher and Shen 2011; Gerike et al. 2015). Resulting from high under-reporting effects for short trips, the average trip durations are higher in HTS compared e.g. to GPS surveys (Stopher et al. 2007; Stopher and Shen 2011; Aschauer et al. 2018a).

Stopher et al. (2007) quantify the overall amount of trip under-reporting with 7 percent based on a comparison of data from GPS devices that were carried by a subsample of the Sydney Household Travel Survey (face-to-face interviews). (Wolf et al. 2003a; Wolf et al. 2003b) report a similar effort of GPS-subsamples for the 2001 HTS in Atlanta, California and Ohio (CATI) and find substantially higher under-reporting up to 32 percent. The duration of the subsequent activity following the trip is an important determinant of trip under-reporting. Stopher et al. (2007) find that almost half of missed trips arrive at a subsequent activity with a short duration of $\leq$10 min. Aschauer et al. (2018a) find trip under-reporting of almost 20 percent for trips with a duration of the subsequent activity of $\leq$10 min and values of maximum 5 percent for longer durations of the subsequent activity.

Findings for the temporal pattern of trip under-reporting consistently show highest under-reporting effects for the afternoon peak (Wolf et al. 2003b; Stopher et al. 2007; Aschauer et al. 2018a). Findings for the morning peak and off-peak hours are mixed but under-reporting in these time periods is always lower compared to the afternoon peak.

Elias et al. (2011) stress the importance of interviewers. They compare a standard HTS (face-to-face and telephone interviews) with face-to-face interviews conducted by especially skilled and trained local interviewers in a relaxed environment for three Arab towns in Israel and find substantial under-reporting in the standard HTS, particularly for non-home based trips.

Aschauer et al. (2018a) study differences in the number of tours (sequence of trips beginning and ending at home) in addition to trip analysis, based on a comparison of the MAED survey (used also in this study, see Section 3) with a standard mixed-mode HTS. They show that under-reporting is mainly an issue of trips, whereas the number of tours was found to be stable; under-reporting mainly occurs within tours also in HTS.

Under-reporting is higher for heavy travellers and for persons with complex (Bricka and Bhat 2006; Stopher et al. 2007) travel pattern. These respondents might have less time for filling out the diary compared to others due to their many activities and the task of filling out the diary is more burdensome and challenging as more trips in longer tours need to be recalled and reported. Using a similar argument, Schmid et al. (2018) show that car owners – exhibiting a significantly higher number of trips on average – tend to show a decreasing response behaviour over the reporting period. Furthermore, trip under-reporting is higher among young adults, men, persons with less than high school education and unemployed persons (Bricka and Bhat 2006).

Studies reporting the results of validation efforts within HTS confirm the main lines of the above findings (Brög and Meyburg 1980; Brög et al. 1982; Brög and Winter 1990; Brög 2015). Validation increased the overall trip rates by 9 percent (Brög and Winter 1990), 10 percent (Brög et al. 1982) and 13 percent (Brög 2015). Walking trip rates increase most as the result of validation. Brög et al. (1982) report that the number of walking trips increased by 23 percent due to validation. Findings for the other modes are mixed. In terms of trip purposes, under-reporting is consistently high for shopping trips followed by leisure. Brög et al. (1982) find nonreporting rates of 17 percent for both discretionary and non-discretionary trips and 6 percent for mandatory trips. Brög et al. (1982) show in addition that validation mainly generates short trips. They find nonreporting rates of 26 percent for trips with distances $\leq$0.4 km, 23 percent for trips 0.5–0.9 km and values below 10 percent for longer trip distances.

As for activities, which are mainly collected through time use surveys, no validation studies on item-nonresponse could be identified. Stinson (1999) stresses the importance of so-called follow-up probes by the interviewer team to fill in gaps in the daily schedule and to retrieve missing information, but no analyses on the received data completeness have been found.

3. Data sources and methodology

3.1. Mobility activity expenditure diary (MAED)

The data used for the analysis were obtained from a Mobility-Activity-Expenditure Diary (MAED), a novel survey format, which

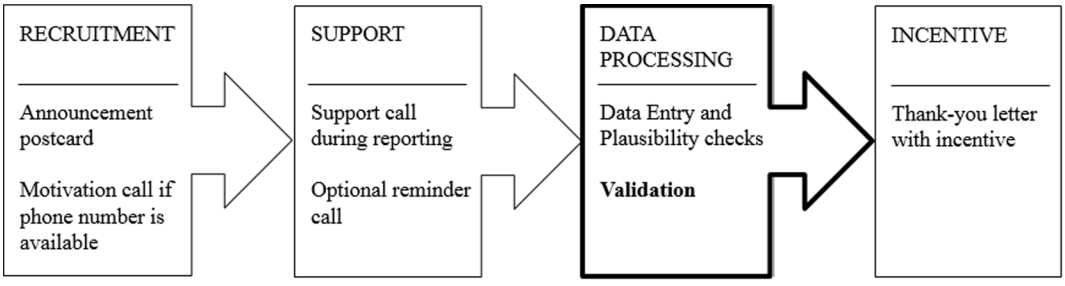

Fig. 1. Survey procedure of the MAED-survey.

combines three survey disciplines in an integrated diary format: a travel survey, time use survey, and consumer expenditure survey. A detailed description of the diary and the survey approach can be found in Aschauer et al. (2019). The MAED-survey is a self-administered mail-back survey with a one-week reporting period and detailed questions about all trips as well as all activities for each diary day. It was conducted in spring and autumn 2015. The sample is based on a random selection of Austrian households for 18 pre-defined strata arranged by region and level of urbanisation (urban, intermediate, rural). Telephone numbers were added to the sampled addresses if available. Only employed persons were selected for the MAED-survey as a wage rate was needed for modelling the trade-off-processes between time and money using the transport economic models described in Jara-Díaz et al. 2008.

The survey procedure is shown in Fig. 1. It follows the traditional New KONTIV Design (NKD) (Socialdata 2009; Fellendorf et al. 2011) with slight modifications due to the high respondent burden and the necessity for a screening phase in order to recruit only employed persons. An announcement postcard was sent to all sampled households as a first step, followed by a first phone call to households with telephone number. Households without a telephone number were asked via an announcement postcard to provide their contact details. Households who provided their contact details received the survey material and were treated similarly to the households with telephone number from then on. In the first phone call to households with available phone number, employed persons were identified by a screening question about the employment status of all household members. Selected persons were asked whether they would like to receive the survey material. All participants received at least one call for technical support at the beginning of the reporting week. Respondents sent the survey material back to the survey team after having completed the survey for their reporting week.

Data processing consisted of two steps:

- Data entry and plausibility checks: The survey team completed plausibility checks immediately after return of the survey material similar to the NKD procedure (Socialdata 2009) based on an automatically generated, detailed error protocol describing missing and implausible information from the questionnaire, after it has been entered in the database. The design of the 168-hour MAED diary covering a whole week required a continuous recording of trip-activity sequences from the participants and allowed missing information (e.g. unreported trips or trip characteristics) to be detected by time gaps within the diary. Objectively imputable items of time gaps and missing information of trips and activities that were not filled out by the participants were added manually. Such items can be assumed with certainty (e.g. missing trip distance of a trip back home, with same address and same transport mode as the first trip) or with high probability (missing transport mode within a trip chain, missing activity at the address of the usual workplace or shopping destination). Since the sample consisted of employed persons only, it was also possible to impute missing information within the daily routine. This procedure aimed at the retrieval of type (i) errors as described in Section 1: All errors that could be corrected in the course of data preparation can be assigned to this error type, according to Brög et al. (1982, see also Richardson 2003).
- Validation (data completion and correction): In case of further missing (irretrievable) information, respondents were called back immediately by the survey staff (usually 2–3 days after the end of the reporting week) in order to retrieve missing information (i.e. completely unreported trips and following activities at the destination), to correct or validate implausible answers and to complete missing items such as missing attributes of trips and activities (incompletely / incorrectly reported trips). The missing or incomplete information was asked directly from the participants themselves, as generally no proxy participation in the survey or the validation process was accepted. This method of data retrieval aimed at correcting errors of type (ii). A €40 incentive was sent to the respondents by mail after their questionnaire has been validated over the phone. A payment of the incentive prior to the start of the survey would have resulted in reduced data quality, as the willingness of the participants to cooperate diminishes with increasing frequency of telephone contacts by the survey staff (Aschauer et al., 2019).

The survey design included a series of measures for quality management designed to minimise item-nonresponse of any kind in order to reduce the subsequent validation effort as much as possible: The questionnaire contained numerous completion aids; During the support phase of the survey, the survey staff provided detailed assistance within the support calls to prevent “typical” diary completion errors based on the experience of the first survey wave; Additionally, the need for address information at the destination of each trip was pointed out in each call. The participants were reminded that the payment of the incentive was linked to the submission of a completely filled out MAED diary.

During the fieldwork of the survey, two precautions were taken in view of the imminent validation analysis: (i) all recorded data of a person directly after having completed data entry and plausibility checks were ’frozen’ prior to validation[1]; and (ii) the members of the survey staff recorded all hours worked for the ex-post validation separately.

3.2. Survey response and validation procedure

The overall response rate of the survey was 12%, according to the AAPOR response rate calculator (Response Rate 1). 424

[1] The information from the returned questionnaires was entered into the database completely (including incorrect and incomplete information). The data from each respondent was saved before the validation call and a backup file was created. This prevented the data from being overwritten with the new/modified information resulting from the validation process.

respondents took part in the second wave of the MAED survey[2]. After data entry and plausibility checks by the survey staff, each questionnaire received a final usability status: (1) Three questionnaires were classified as a priori not usable, a validation by phone was not possible for these cases. (2) Five questionnaires would have been eligible for validation but contacting the respective participants was either not possible or they refused to cooperate for validation. (3) 314 respondents (three quarters) received a validation call and agreed to cooperate, their questionnaires needed to be validated; (4) For the remaining 102 questionnaires validation was not necessary, as their survey completion quality did not leave any information open. In total, 416 questionnaires could be further used but 'freezing' of data before validation was only successful for 393 respondents. These 393 persons serve as data basis in this article.

Respondents showed a high willingness to cooperate in the validation calls. Possible reasons for this are: the conditional incentive (respondents knew that they would only receive the €40 voucher after having completed all steps of the survey – including a possible validation); the personal relationship established during the whole survey process so far; the announcement that this is the last and final call in the survey process; the appreciation by the survey team of the initial completion of the questionnaires (the validation calls started with a big thank you to the respondents for their work done so far); and also personal interest in the topic and in what the survey staff uses the information for provided by the respondents.

Table 1 gives an overview of the time spent by the survey staff for different parts of the fieldwork. The validation took on average 15.6 min per respondent. This amounts to 12 percent of total time for all activities of the survey staff within the whole fieldwork. It should be noted that the validation time investment also includes the validation of MAED expenditure data, which is not part of this article.

3.3. Sample description

The MAED survey for this article was aimed at obtaining a representative sample of 393 Austrian gainfully employed persons; it includes all trips (travel decisions), activities (non-travel related) and expenditures for each respondent over a period of one week. Table 2 lists the sample characteristics compared to the data of the Austrian working population (respectively of comparable households with an employed reference person[3]) from the Austrian National Census ('Registerzählung 2011'; Statistics Austria 2014). Compared to the official statistics, females and the groups of 40+ year old and well-educated persons would be overrepresented in the survey, a well-known phenomenon in household surveys (Eisenmann et al. 2018). Fewer than average single person households and households from urban areas took part in the survey. The high response from households of rural areas causes a higher share of car ownership and private parking space availability in MAED compared to the Census data. Data was not weighted for the analysis because the research focus of this paper is on comparing samples before and after validation as well as newly generated trips in the validation process.

3.4. Data preparation

The 'frozen' before-validation data are referred to as *Ante-Val* datasets, the validated data as *Post-Val* datasets. The different analyses carried out in this study (see Section 4) required several steps for data processing resulting in six sub-datasets as shown in Fig. 2.

The original Ante dataset (*Ante-Val Original*) contains 160 trips from 6 persons with trip-IDs that do not match with any trip in the Post dataset (*Post-Val Raw*). These are persons with artificial trips that were created by the survey staff during the phase of data entry and plausibility checks, but all entries of these persons were removed during the validation calls (drop-outs). The dataset cleared for these trips (*Ante-Val Matching Trips*) contains only trips that exist with the Matching Trips before and after the validation calls.

The dataset resulting from the validation calls (*Post-Val Raw*) increased, compared to *Ante-Val Matching Trips*, by 349 additional trips. These trips were completely omitted (deliberately or not) by the survey participants during their initial completion of the MAED survey. 173 of these trips, however, belong to six persons, who had been recorded as immobile persons without any trips and activities before validation. They reported their entire trip and activity data in the validation calls. These six participants (that are not the same as the ones removed because of unmatchable trips earlier) and their respective trips (and activities) are part of the validation success and are therefore part of the *Post-Val Raw* sample, but they constitute completely new-recorded travel diaries. Hence, they do not represent the typical characteristics of missing trips with regard on the overall's sample survey and validation process[4]. Therefore, these 173 trips (and respective persons) were removed to form *Post-Val Complete*. The remaining 176 new trips were stored in *Post-Val New Trips*. Finally, *Post-Val Matching Trips* contains the equivalent trips of *Ante-Val Matching Trips*, and hence represents those trips that already existed in the *Ante-Val* dataset but their items such as trip attributes could have been affected by the completion and correction process of validation. *Ante-Val* and *Post-Val* datasets for the analysis of activities were created in the same way as described above for the trip data.

The datasets divided into subgroups serve as different reference levels, depending on the item non-response types to be examined:

[2] The complete MAED sample consists of 748 persons, but the 'frozen datasets' with the unvalidated information were only generated in the 2nd half of the survey (wave 2); only these datasets provide a basis for comparison and can be used to analyse the effect of validation.

[3] The household's reference person according to Statistics Austria is defined as the oldest person of the nuclear family (for single-family households) or as the oldest person of that family, which represents the middle generation (for two- or multi-family households).

[4] The six completely new participants received do not have substantially different characteristics than the rest of the sample. They have slightly more mobility options on average, and come from households with more persons. They reported more trips (28.8) in their reporting week than the sample average (25.1).

Table 1
Shares of different activities of the survey team.

Survey staff activity	Time spent per participant [min]	Share [%]
Motivation	25.4	19.1
Support & reminders	12.0	9.0
Data entry	80.1	60.2
Validation	**15.6**	**11.7**
Others	0.1	0.1
Total	**133.1**	**100.0**

Table 2
Sample characteristics of MAED data and Austrian National Census.

	MAED 2015	Stat. Austria 2011
n households	256	–
n persons	393	
n person reporting days	2,552	
Gender		
Female	51.1	46.7
Male	48.9	53.3
Age		
15–19	3.1	5.0
20–29	7.9	19.5
30–39	19.6	22.6
40–49	31.6	29.1
50–59	33.3	20.0
60+	4.6	3.8
Highest level of education[4]		
Compulsory school	3.1	17.8
Apprenticeship, college	40.0	50.9
Matura	23.3	15.9
University, FH	34.6	15.4
Household size		
1 person	14.1	30.2
2 persons	28.9	23.1
3 persons	25.0	19.0
4 or more persons	32.0	27.7
Level of Urbanisation[5]		
Urban	20.7	33.5
Intermediate	28.5	29.9
Thin	50.8	36.7
Personal mobility tools available		
Car parking	88.0	86*
Car availability	91.1	76*
PT season ticket	23.7	22*

* not collected by Statistik Austria, but by the Austrian Ministry for Transport, Innovation and Technology. The shares include the whole population (and not only employed persons).
[4]Matura refers to the Austrian school leaving examination for high school education that qualifies for university (usually at the age of 18 years).
[5]Definition for the MAED-survey and Statistics Austria: According to the Degree of Urbanisation (DEGURBA) - classification by the European commission (Eurostat 2011).

Ante Original and *Post-Val Raw* datasets are used for analysing the number and characteristics of newly added trips or activities in the process of validation. *Ante/Post-Val Matching* datasets of trips and activities are compared for investigating missing attributes of trips or activities that were completed in the process of validation.

The *Post-Val* datasets introduced in Fig. 2 comprise the *Ante-Val* datasets and the results from the validation effort. All percentage values given for comparing *Ante-Val* and *Post-Val* therefore use the values of *Post-Val* as 100 percent reference.

3.5. Methodology for data analysis

The effects of the validation procedure on item-nonresponse - the correction of type (ii) errors - with respect to the two types of missing information introduced in Section 1 (completely missing trips/activities respectively incomplete information on trips/

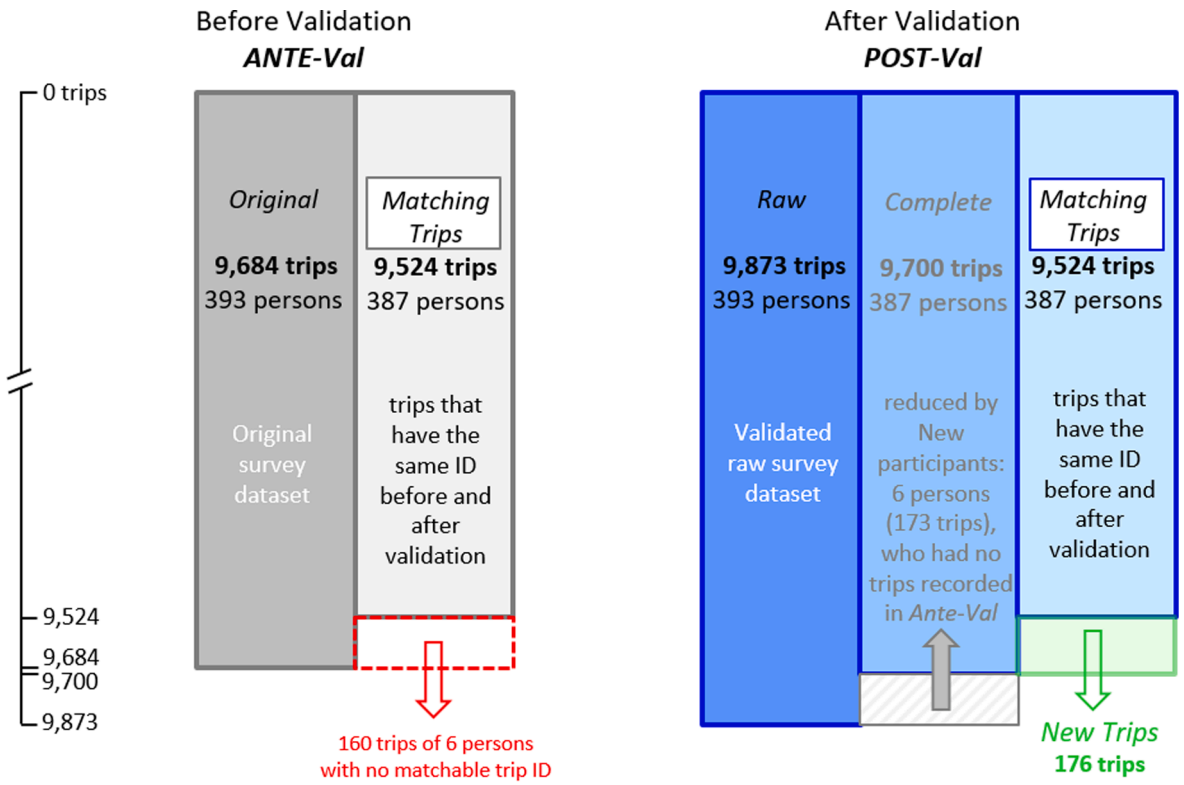

Fig. 2. Sub-datasets (names in italic) from before (Ante) and after (Post) validation used in the trip-validation analysis.

activities) are analysed along the following dimensions:

a) Characteristics of newly added trips during validation: we examine item-nonresponse (missing information due to nonreported trips) by analysing 176 newly added trips stored in *Post-Val New Trips*, and by comparing *Post-Val Complete* and *Ante-Val Matching Trips* for examining the overall change of mobility indicators caused by validation.
b) Change in the attributes of trips that existed before and after validation: we examine 'Item incorrect/incomplete-response' through the analysis of trips that existed before validation but may have been completed (missing items) or corrected during validation (comparison of *Ante/Post-Val Matching Trips*).
c) Characteristics of (non-travel) activities before and after validation: the activity data is analysed for the number of overall activities as well as activities with missing or wrong item information such as activity type and duration.

The six new persons (including their 42 new reporting days and 173 trips) were surveyed completely with CATI during the validation. These persons are part of the *Post-Val Raw* sample, but – as the sample – not part of the in-depth analyses because their newly recorded travel diaries are unlikely to represent the typical characteristics of missing trips and therefore do not contribute to the increase in quality of the data through validation.

3.6. Sample sizes and characteristics

Table 3 shows the characteristics of samples before and after validation. The *Post-Val Raw* dataset contains 1.9 percent additional trips compared to *Ante-Val Original*. The share of trips added in the validation process increases when compared to *Ante-Val Matching Trips* (leaving only matchable trips in the *Ante-Val* data). Compared to these 9524 trips with matchable trip IDs, the validated dataset increased by 349 (3.5%) additional trips. The number of reporting days and tours increased by 2.5 percent (days) and 2.6 percent (tours) in *Post-Val Raw*. These values gained by validation are lower than reported in the literature (Brög and Meyburg 1980), which may be caused by several reasons: due to their regular daily routine, we assume that the employed persons of our sample generally report more accurately than persons without employment. In addition, Brög's studies may have included participants in the validation that in our study were not contacted for validation due to their a-priori satisfactory completion quality. Therefore, no new trips/activities could be found for this group, which represents about 25% of our sample. Another possible reason is that MAED in general has more reported trips than a conventional travel diary (see Aschauer et al. 2018a). We hypothesise that respondents in the MEAD sample inserted more trips into the diary because reporting of trips is embedded in reporting of all daily (travel and non-travel) activities; this makes it less likely that trips are forgotten or deliberately omitted. In addition, the support calls made by the survey staff during the

Table 3
Dataset sizes with regard to different indicators.

	Total Number			share [%]			difference [%]	
Sample indicators	Ante-Val Original	Ante-Val Matching Trips	Post-Val Raw	Ante-Val Original	Ante-Val Matching Trips	Post-Val Raw	Ante-Val Original	Ante-Val Matching Trips
Persons	393	387	393	99.0	98.5	100	*1.0*	*1.5*
Reporting days	2,510	2,489	2,552	98.4	97.5	100	*1.6*	*2.5*
Tours	3,743	3,709	3,809	98.3	97.4	100	*1.7*	*2.6*
Trips	9,684	9,524	9,873	98.1	96.5	100	*1.9*	*3.5*
Activities	36,464	35,693	37,763	96.6	94.5	100	*3.4*	*5.5*

Definition for the MAED-survey and Statistics Austria: According to the Degree of Urbanisation (DEGURBA) - classification by the European commission (Eurostat 2011).

reference week probably had a positive effect on the completeness of reported information. These effects result in a relatively accurate data basis prior to validation, contrary to Brög's studies. The high baseline quality of the MAED data already before validation leads to the relatively small number of 176 new trips considered for the analysis in *Post-Val New Trips*. However, we hypothesise that their attributes represent typical characteristics of trip under-reporting and item-nonresponse well. Compared to *Ante-Val*, validation generated 5.5 percent additional activities in the *Post-Val* activity dataset. These consist of (i) time not assigned to any activity (time gaps in the diary) and (ii) activities that were divided into several individual activities in the process of validation resulting in more detailed/disaggregated reporting of activities.

4. Validation results

4.1. Key mobility indicators

Table 4 shows key mobility indicators before and after validation. *Post-Val Complete* has an increased trip rate / share of mobile persons of 0.8/ 0.9 percent compared to *Ante-Val Matching Trips*. The average trip duration decreases by 2.6 percent indicating that mainly short trips have been under-reported. The information on modes in the validated dataset increased by 2.7 percent trips. Mode under-reporting was highest for car-driver trips (2.6%-increase) and walking trips (1.2%).

4.2. Characteristics of newly added trips during validation

The following Fig. 3 to Fig. 5 compare the characteristics of the 176 trips that were newly added in the validation process, with the data before validation. Under-reporting mainly occurs in the afternoon peak hour, as 33 percent of the new trips were added in the time between 4 pm and 6 pm. These might be trips that are forgotten on the way back home from longer tours e.g. for the main tour purpose of work. The period between lunch (12 pm) and 3 pm comprises around 6 percent of newly added trips per hour (start time);

Table 4
Mobility indicators for *Ante-Val* and *Post-Val* datasets.

Sample indicators	Ante-Val Matching Trips	Post-Val Complete	
Persons	387	387	
Reporting_days	2,489	2,512	
Tours	3,709	3,753	
Trips	9,524	9,700	
Mobility indicators			Ante-Val Matching Trips Difference in share [%]
Share of mobile persons [%]	91.9	92.7	*0.9*
Trip rate [trips per day]	3.8	3.9	*0.8*
Tour rate [tours per day]	1.5	1.5	*−0.7*
Length of tours [trips per tour]	2.6	2.6	*0.7*
Trip duration per trip [min]	24.5	23.9	*−2.6*
Mode choice [%]			
Missing	2.9	0.3	*−2.7*
Walk	11.8	13.0	*1.2*
Bike	4.2	4.3	*0.0*
Motorbike	0.6	0.6	*0.0*
Car Driver	60.9	63.5	*2.6*
Car Passenger	8.5	8.7	*0.2*
Public transport	9.3	9.7	*0.4*
Total	98.2	100.0	*1.8*

these might be under-reported trips as part of lunch breaks or early back home tours for part-time employed persons. The morning peak hours have low proportions of newly added trips. This might be either an indication for less "side-activities" in the morning e.g. on the way to work - or for less under-reporting in the morning when respondents are still attentive to their behaviour. Another reason might be more routine trips in the morning compared to the afternoon, when higher numbers of irregular trips occur such as errands, shopping or leisure trips.

Fig. 4 shows that mainly short trips are under-reported. 30 percent of the newly added trips in *Post-Val New trips* have a duration of $\leq$5 min. The percentages of trips with longer durations are consistently below their shares in *Ante-Val Matching Trips*; only trips with a duration of $\leq$20 min and of $\leq$30 min deviate, this might be an interaction with other variables such as the trip purpose.

The distribution of new trips across trip purposes[5] is shown in Fig. 5. 15 percent of the newly added trips have the purpose 'work', these are often part of newly added mid-day tours e.g. for having lunch. These trips were rarely not reported (compared to 22 % of 'work' trips in the Ante-Val dataset). 32 percent of the newly added trips are trips 'back home'. This high share of the trip purpose *home* corresponds with the literature that consistently shows that trips back home tend to be forgotten more often than trips to other destinations (Stopher and Shen 2011; Gerike et al. 2015). However, the share of trips 'back home' is still lower than in *Ante-Val* dataset because a major proportion of these trips has a start time in the afternoon peak hours and they are often split into a trip from work to a 'stopover' and a trip from the 'stopover' back home. The 'stopover' trips are often 'shopping' or 'errand' trips, and are heavily affected by underreporting.

With regard to the distribution of trips along their main transport modes, two main overlapping effects generating the differences between *Ante-Val Matching Trips* and *Post-Val New trips*. On the one hand, the main mode of under-reported trips is dominated by the mode used for the whole tour. This leads to the high proportion of 65 percent of car driver trips in *Post-Val New trips (comparing to* 63 percent in *Ante-Val Matching Trips*, as more than half of the respondents live in rural areas. On the other hand, mainly short trips are under-reported and these are mostly walking trips. The share of walking trips in *Post-Val New trips* is with 24 percent is twice as high as the proportion of walking trips in *Ante-Val Matching Trips* which is 12 percent of all trips.

4.3. *Effects of validation with respect to sample units and population segments*

Table 5 gives an overview about how the 176 newly added trips are distributed among persons, reporting days and tours: 107 tours within 95 reporting days of 61 persons received at least one new trip. At the person level (column 1) almost half of respondents received more than one new trip (47.5 %). Multiple new trips at person level are caused by regular trips, which were forgotten more than once in the reporting week. This also explains the high share (18 %) of respondents with more than 4 newly added trips by validation. Column 2 shows that 61 percent of those reporting days where new trips were added received a single new trip, followed by 21 percent of reporting days with two new trips. The tour level analysis (column 3) shows that 56 percent of new trips were added as single trips to an existing tour and 34 percent to tours with 2 new trips, whereas only very few tours got more than two new trips. This indicates that the basic mobility pattern obtained from the non-validated MAED data was rather complete and mostly 'particular gaps' were filled by the validation. Column 4 shows that 35 percent of the new trips generated completely new tours, while 65 percent were added to existing tours containing n trips before validation.

Column 5 indicates the sequential number of the new trip within the tour to which it was added. 40 percent of the new trips were added as the second trip compared to 30 percent added as the first trip. This is in line with the finding that omitted trips are often those from a 'major destination' (work place) to a 'minor destination' (shopping) before people go back home.

Table 6 shows how trip under-reporting is distributed across different population segments. Different from the typical travel behaviour pattern of complete populations (Tomschy et al. 2016), in our sample of employed persons, women have almost the same trip rate as men, resulting in a share of 49 percent of the overall number of trips. Part-time employed persons have substantially lower underreporting levels than full-time employed persons do. The increased under-reporting of full-time employees may be related to two effects: (i) full-time employees have less time to fill in the questionnaire carefully and/or (ii) they can only fill in the questionnaire at longer intervals compared to part-time employees and therefore their response burden is greater. Both lead to increased item-nonresponse. Gender, trip rates and mean length of tours might be additional correlates as described below. No differences in underreporting have been identified across education levels.

Table 7 shows the relationship between travel pattern and trip under-reporting.

- Group 1 'Validated reporting days without any validated trips' (*Post-Val Matching Trips* - which were already existing before validation) are compared with the other two types of reporting days that were affected by validation:
- Group 2 'Validated reporting days with at least one validated trip': Reporting days which were previously 'immobile' (had no reported trips) but became 'mobile' during validation (received at least one trip).
- Group 3 'Newly created reporting days by validation': Reporting days, which had reported trips before validation but received at least one trip during validation.

In order to identify relationships between the three groups—using the indicators trips per day, tours per day and trips per tour—an

[5] The distribution of trip purposes in the MAED data slightly differ from the standard Austrian HTS trip purposes because the MAED is a combined travel-time use survey with coding not only of travel but of all in-home and out of-home activities. For further information on activity coding in MAED see Aschauer et al. 2018a.

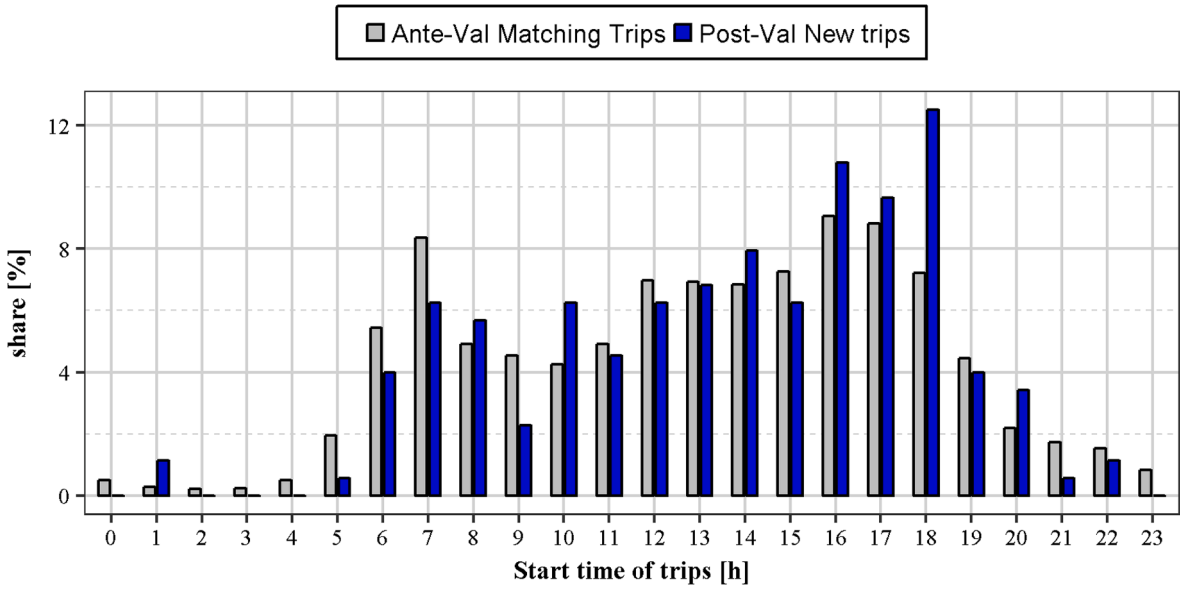

Fig. 3. Start time of new trips (n = 176) compared to Ante-Val Matching Trips (n = 9524).

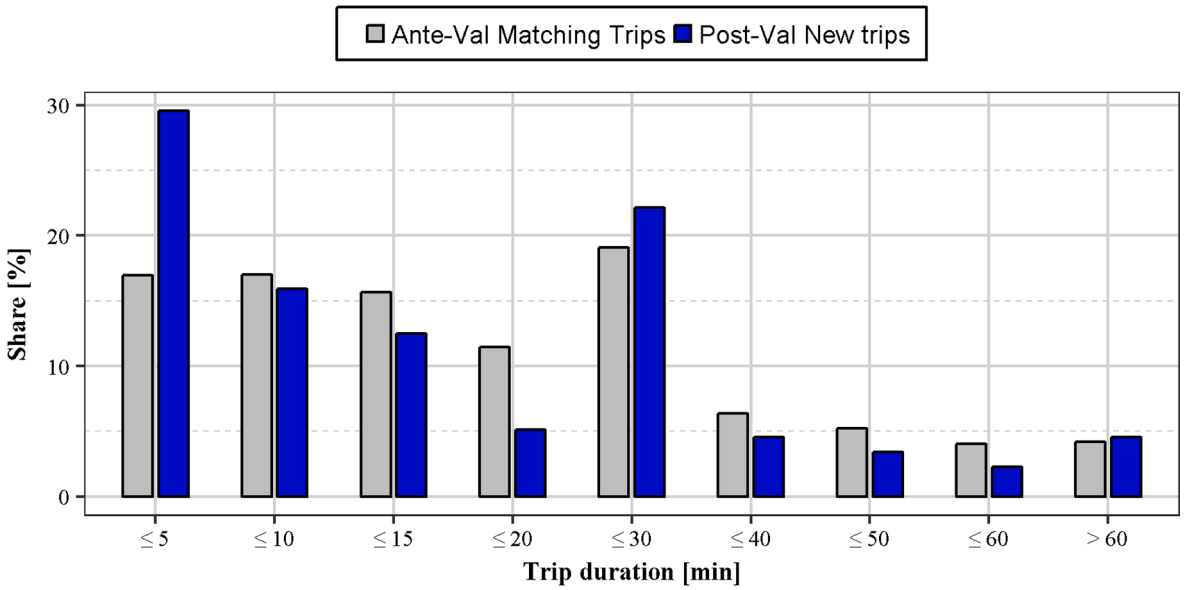

Fig. 4. Distribution of new trips according to their trip duration (n = 176) compared to Ante-Val Matching Trips (n = 9365).

Anova with post hoc tests was applied. Reporting days of Group 3 do not significantly differ from reporting days without validated trips. However, reporting days of group 2 have significantly more trips per day (5.31 vs. 3.82) and also more trips per tour (3.86 vs. 2.56) than reporting days without validated trips.

From this, we conclude that validation serves on the one hand to increase the rate of mobile days by adding trips to those days that have previously been reported as immobile; these days seem to have similar trip characteristics like days without validated trips. On the other hand, it corrects the problem that single trips tend to be forgotten on reporting days with a large number of trips caused by long and complicated tours. It must be noted that respondents in group 3 were completely surveyed with CATI interviews. We hypothesise that biases compared to the PAPI diaries provided by all the other respondents are limited because of the extensive validation efforts for all respondents and diaries.

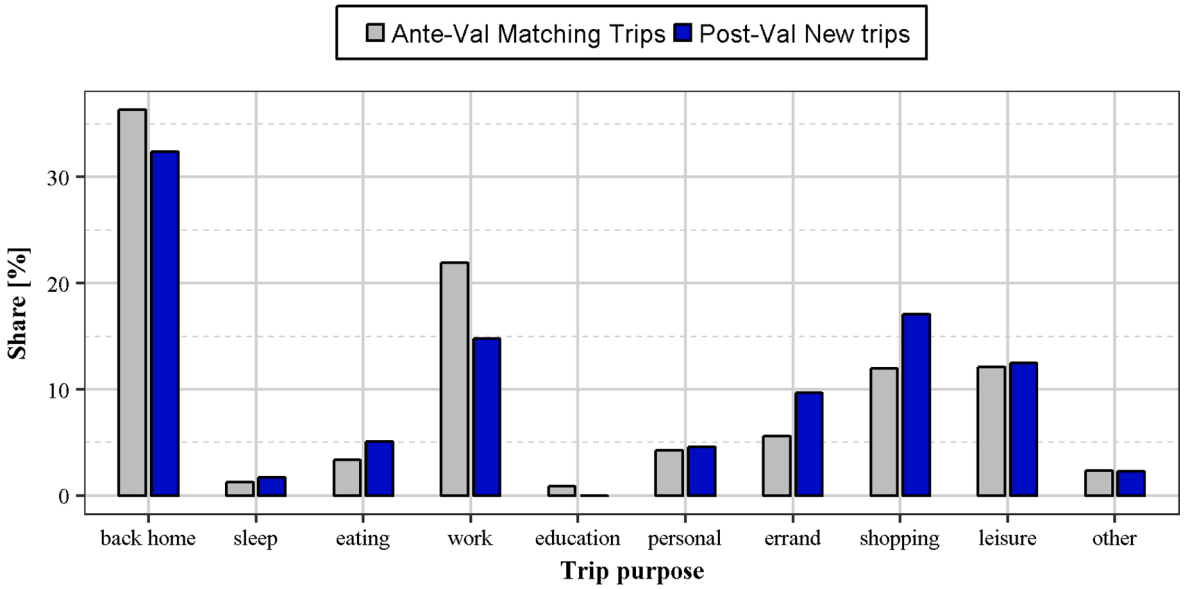

Fig. 5. Distribution of new trips according to their trip purpose (n = 176) compared to Ante-Val Matching Trips (n = 9499).

Table 5
Distribution of new trips among persons, reporting days and tours.

Col. Nr.	**1**	**2**	**3**	**4**	**5**
Share of [%]	**n trips added per**			**trips added to tours of n trips prior to validation**	**n position of the added trip in the tour**
n	person	reporting day	tour		
0	–	–	–	34.6	
1	52.5	61.1	56.1	17.8	29.6
2	21.3	21.1	34.6	22.4	39.8
3	4.9	4.2	5.6	14.0	14.2
4	3.3	7.4	0.9	4.7	6.3
≥5	18.0	6.3	2.8	6.5	10.2
n affected by new trips	61	95	107	107	–

Table 6
Differences in validation among person characteristics.

	Matching Trips		**Nr. of trips**		**Share of trips [%]**	
	Nr. of reporting days (∑=2494)	Trip rate	*Matching Trips* (∑=9524)	*New trips* (∑=176)	*Matching Trips*	*New trips*
Gender						
female	1,213	3.81	4,625	73	48.6	41.5
male	1,281	3.82	4,699	103	51.4	58.5
Level of employment						
part-time (<35 h/week)	788	4.13	3,258	40	34.20	22.7
full-time (≥35 h/week)	1,706	3.67	6,266	136	65.80	77.3
Education						
Low (without Matura)	1,059	3.59	3,806	72	40.0	40.9
High (with Matura)	1,435	3.84	5,718	104	60.0	59.1

4.4. Validation of trips with matching trips

Those 9,524 consistent trips, which exist in both datasets *Ante-Val Matching Trips* and *Post-Val Matching Trips* are compared in order to quantify the enhancements in data richness of trip attributes achieved by validation. The analysis focusses on those attributes that are important for mode-choice models. Table 8 shows how the validation changed the data with respect to (i) address information, (ii) means of transport information and (iii) personal information required for the generation of valid choice sets in such models. The largest improvement concerns missing or wrong addresses. Moreover, person characteristics were also often missing; these are

Table 7
Characteristics of mobility indicators for validated reporting days with and without newly added trips.

Indicator i		trips per day	tours per day	trips per tour
Anova (F value)[5]		12.14***	2.0	39.88***
1	Validated reporting days without any validated trip			
	n	2,417	2,417	3,646
	mean	**3.82**	**1.49**	**2.56**
	post-hoc (p adj)[8] i to:	2*** / 3	2 / 3	2*** / 3
2	Validated reporting days with at least one validated trip			
	N	77	77	70
	Mean	**5.31**	**1.64**	**3.86**
	post-hoc (p adj)[7] i to:	1*** / 3***	1 / 3	1*** / 3***
3	Newly created reporting days by validation			
	N	18	18	37
	Mean	**3.11**	**1.33**	**2.22**
	post-hoc (p adj)[7] i to:	1 / 2***	1 / 2	1 / 2***

[8]Signif. codes: '***' <0.001; '**' <0.01; '*'< 0.05

required for annotating alternative-specific attributes in mode choice models (driving license, availability of public transport season ticket, etc.). Altogether, the usability of the data for mode-choice models could be increased by 25 percent thanks to the validation.

Fig. 6 shows the relationship between correctness of address information and main activity type at the destination. Incorrectly reported addresses include trips for which either no address was reported (missing address) or information on the reported address before validation was insufficient (incomplete address). The highest share of incorrect addresses (39 %) was found for shopping trips. This is plausible as the detailed postal address of shops is often unknown. The large shares of incorrect home and work addresses (14 and 25%, respectively) might have two main reasons: (i) these addresses are visited many times per week, one single missing address causes many missing items in this case; (ii) respondents deliberately might not complete the address information especially for their home and work address for privacy issues. The proportion of missing home addresses is still with 13.5 percent the lowest compared to all other activity types.

Fig. 7 shows the relationship between correctness of address information and frequency of visit of the destination; a frequency of seven visits means that the destination was visited once a day during the reporting week. The share of incorrect addresses is highest for those destinations which were visited only once or twice during the reporting week (31% each), which indicates that non-routine addresses are most affected by incorrect reporting. However, the improvement to frequent visits is not very large; there is substantial rate of incorrectness between 15 and 30 percent across all numbers of visits. Correctly reported addresses were visited on average 5.8 times by the respondents while incorrectly reported addresses were visited only 1.6 times. The result illustrates the sensitivity of collecting correct addresses in paper (and presumably also in online) questionnaires.

Further important trip attributes such as trip duration, departure time and arrival time were reported more correctly, as at least 97% of those attributes remained unchanged by validation.

4.5. Validation of activity data

Underreporting of activities was analysed similarly to the trip analysis reported above. Table 9 shows that 3.8 percent of the total time (one reporting day resp. 1440 min) was not reported and another 2.2 percent were reported with missing activity type in the *Ante-Val* dataset. The activity types most affected by under-reporting are sleeping (3.2 %), leisure (0.8 %) as well as personal and domestic activities (0.4 and 0.5 %). Table 10 shows a comparison of consistent activities which exist in both datasets of *Ante-Val* and *Post-Val Matching Activities*. 96 percent of these activities were complete and correct before validation such that no changes were necessary. For 1.4 percent of activities in *Ante-Val* the activity types were completed or added, for another 1.8 percent the duration was modified. The result indicates that the reporting quality of activities was generally more satisfactory than that of trips for two reasons: (i) trips are

Table 8
Model-relevant attributes of *Matching Trips* data and respective changes by validation.

	Nr. of trips		Share [%]		Difference
	Ante-Val Matching Trips	Post-Val Matching Trips	Ante-Val Matching Trips	Post-Val Matching Trips	
Trips without any missing or different info	7,087	9,472	74.4	99.5	*−25.1*
Trips with any missing or different info	2,437	52	25.6	0.5	*25.1*
with missing or different address	1,664	1	17.5	0	*17.5*
with missing or different mode	315	23	3.3	0.2	*3.1*
with missing person variables	458	28	4.8	0.3	*4.3*
Total	**9,524**	**9,524**	**100.0**	**100.0**	

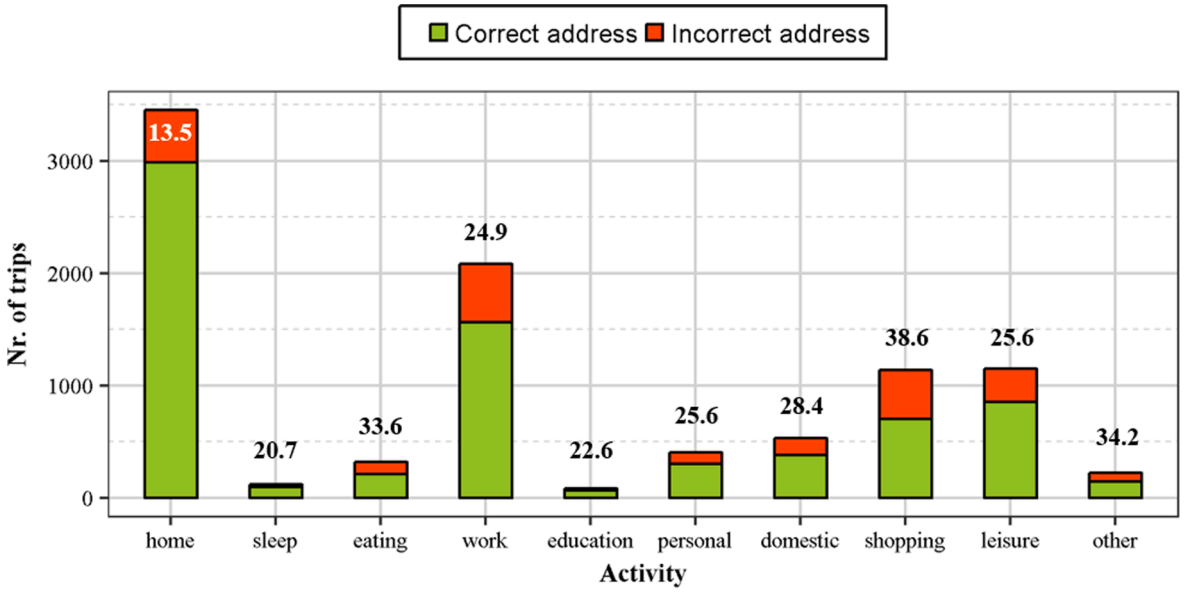

Fig. 6. Number of trips by the main activity type at the destination and by correctness of address information (n = 9499) (numbers above bars indicate the share of incorrectly – missing or incompletely – reported addresses).

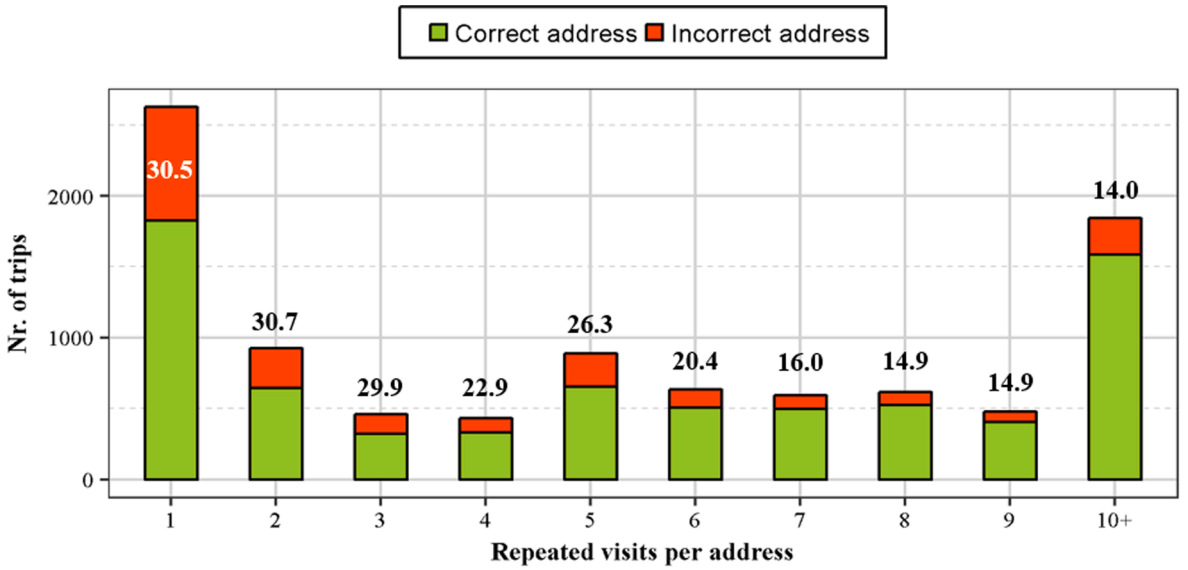

Fig. 7. Number of trips by the frequency of visits of the destination and by correctness of address information (n = 1843) (numbers above bars indicate the share of incorrectly reported addresses).

affected by the cautious address issue and (ii) the amount of required information is higher for trips if the data are to be used in mode choice models.

5. Quantified effort of validation

Taking into account the total time (work hours) spent on validation and an assumed hourly wage of 22 € for the fieldwork staff, this results in a total amount of 2243 € in staff cost for conducting the validation of 393 participants. The staff cost for the remaining activities of the fieldwork (motivation, support and reminder calls) and data entry was 16,939 €. The cost of incentives for this survey was 15,720 €. The total costs of survey staff costs and incentives thus amount to 34,902 €.

Table 11 shows how the benefits arising from the validation process might be quantified and monetised (based on staff costs and incentives).

Table 9
Activities by activity type in *Ante-Val* and *Post-Val* datasets.

	Minutes per day [min]		**Share [%]**		**Difference [%]**
Activity type	Ante-Val Original	Post-Val Raw	Ante-Val Original	Post-Val Raw	Ante-Val Original
travel	85	86	5.9	6	*0.1*
sleep	437	482	30.3	33.5	*3.2*
eating	75	81	5.2	5.6	*0.4*
work	289	295	20.0	20.5	*0.4*
education	10	11	0.7	0.7	*0.0*
personal	83	89	5.7	6.2	*0.4*
domestic	103	111	7.2	7.7	*0.5*
shopping	16	18	1.1	1.2	*0.1*
leisure	252	263	17.5	18.3	*0.8*
other	4	3	0.3	0.2	*0.0*
unspecific	0	0	0.0	0	*0.0*
missing	32	0	2.2	0	*−2.2*
Total	**1,386**	**1,440**	**96.2**	**100**	***3.8***

Table 10
Attributes of activity data and respective changes by validation in *Ante-Val* and *Post-Val Matching Activities* datasets.

Indicator	Ante-Val Nr. of Act.	Post-Val Nr. of Act.	Ante-Val [%]
Matching activities	35,693	35,693	100
Activities without any missing or different info	34,374	35,693	96.3
Activities with missing or different type	515	0	1.4
Activities with missing or different duration	652	0	1.8
Activities with missing or different start time	152	0	0.4

1. Sample indicators: Validation costs per item might be compared with the average costs per newly collected item without validation. It is shown that the costs of a new item gained by validation (column "validation cost per item") clearly exceed those of collecting a completely new item (column "av. costs per new item"). For example, each new person added by validation is 5.5 times as expensive as if a completely new person had been recruited for the survey (including staff costs and incentives).
2. Model-relevant attributes: Average costs might be also computed per validated trip when missing attributes such as addresses have been added in the validation. This is shown in the lower part of Table 11.

The costs in Table 11 always refer to total costs with/without validation. This table gives an indication of the achievements from validation compared to the efforts needed for generating new items. It does not show the effects on sample characteristics. The analyses in this article demonstrate biases in validated trips and reporting days. For example, more missing trips have been found in the afternoon compared to other times of day. These biases are reduced by validation but not by increasing sample sizes without validation. It needs to be considered that the costs and cost factors of validation presented in the table represent rough estimates and need to be seen in the context of the specific survey design given. Although they generally point towards a positive assessment of the validation

Table 11
Quantified and monetised benefits through validation.

	Without validation	**Including validation**	**Diff.**			
overall cost [€]	32,160	34,902	2742			
1 Sample indicators	Ante-Val Matching Trips	Post-Val Raw		validation cost per item [€]	av. cost per new item [€]	additional cost factor
Persons	387	393	6	457	83	5.5
Reporting days	2,489	2,552	63	44	13	3.4
Tours	3,709	3,809	100	27	9	3.2
Trips	9,524	9,873	349	8	3	2.3
Activities	35,693	37,763	2,070	1	1	1.5
2 Model-relevant attributes	Ante-Val Matching Trips	Post-Val Matching Trips				
Nr. of trips without any missing or different information	7,087	9,472	2,385			
Total nr. of trips (100%)	9,524	9,524				
Estimated cost to receive the total nr. of trips without missing or different info [€]	43,219	35,093	−8,126			

method applied within this study, the respective cost framework and survey design must be taken into account for the transferability of the estimates to larger travel surveys.

6. Conclusions and outlook

This article presents an item-nonresponse study to investigate data quality enhancement achieved by immediate follow-up data completion and validation as part of a combined mobility-activity survey. Incoming data from respondents were checked promptly, and the fieldwork team re-contacted respondents if missing or incomplete information was observed during data processing that could not be imputed without additional information. Data on persons, trips, and activities were completed in telephone calls with the goal of getting as close as possible to the actual behaviour. As a result of this validation process, the number of trips increased by 3.5 percent. This share is substantially lower than the values of around 10 percent found by Werner Brög and associates (Brög and Meyburg 1980; Brög et al. 1982; Brög and Winter 1990; Brög 2015). The main reasons for this might be the combined travel and activity format of the MAED questionnaire used in this study. It appears to yield more complete responses compared to traditional travel surveys, even without validation.

The characteristics of newly added trips (not reported prior to validation) are in line with the findings from item-nonresponse studies done without validation and include some interesting further insights thanks to the analyses that are added to previous work: The new trips begin mainly in the afternoon with a second peak during lunch time; the most important trip purpose is home, although relatively speaking, less regular trips such as errands, shopping and leisure were most often underreported and uncovered by validation; the trips added through the validation process were mainly short. 35 percent of the new trips generated new tours; the remaining 65 percent were amended to existing tours, mainly as a second trip on the way back home from the destination for which the main activity of the tour was performed. Females seem to under-report less than men do; no difference was found with respect to the level of education. Underreporting is higher on reporting days with many trips and for tours with many trips. 32 percent of the new trips added through the validation process changed an immobile reporting day to a mobile one, i.e. they increased the share of mobile days. The new mobile days reveal similar trip rates, tour rates and lengths of tours as the mobile days in the original dataset before validation.

Underreporting of activities was analysed similarly to trips. We found a high consistency and completeness of activity information already before validation. The shares of missing, incorrect, or incomplete activities are comparable to those found for trips.

The findings of this study confirm and extend the literature based on surveys without validation and also the few studies with validation done by Werner Brög and associates (Brög and Meyburg 1980; Brög et al. 1982; Brög and Winter 1990; Brög 2015). The presented validation procedure proved suitable for mitigating item-nonresponse issues. The effort for validation accounted for 12 percent of the overall fieldwork effort (time and cost, respectively). This seems reasonable given that item-nonresponse is not just a random error but systematically affects the travel indicators: It decreases both the share of mobile persons and trip rates, and influences the average trip characteristics in the sample. This study shows that almost all respondents are willing to cooperate with the validation process. The main reason for this is the appreciation for the information provided by the respondents demonstrated through the validation calls.

The validation procedure analysed in this study might be applied not only for Paper-and-Pencil Surveys (PAPI) but also for online questionnaires (CAWI) and smartphone-based travel surveys. While this PAPI survey included a fill-in aid at the beginning if the questionnaire, CAWI surveys provide various opportunities for supporting and prompting respondents directly in the process of filling out the online questionnaire including also the options to reduce participant burden e. g. by inserting recurring addresses automatically. However, not all item-nonresponse issues can be addressed by such automated and rule-based plausibility checks. Smartphone-based approaches detect trips automatically but do not yield completely correct information. Both survey methods might benefit from validation efforts similar to the ones developed for PAPI in this study.

Further studies that measure data quality of travel surveys before and after a validation procedure are needed, such as the work of Ipsos based on Australian surveys. Studies based on such surveys that include the total population (not only employed persons) and also using online and other types of questionnaires would help to validate the results found in this study and in the studies done by Werner Brög and associates. We hypothesise that validation would yield similar magnitudes in terms of corrected and newly added trips but differences in the details. For example, other person groups such as retired persons rarely have any working trips. These are regular trips with often longer distances compared to other trip purposes. These trips tend to be reported with higher accuracy compared e. g. to irregular shopping trips but at the same time, they are often combined with short activities on the way back home from work which show substantial underreporting. Literature shows that non-employed person groups might have less regular trips but instead shorter tours with mainly one trip to the destination and the second trip from the destination back home. Overall, we think that these differences between person groups might level out to comparable overall magnitudes in the share of validated trips.

In addition, the influence of methodological characteristics of surveys on the validation success, such as the incentive paid out after successful completion of participation in this survey, would need to be analysed in more detail. Based on such research, correction factors could be developed for standard HTS, which cannot afford a comprehensive validation process, to maximize extracted information from the original survey. Insights from this study might also feed into the advancement of time-use surveys for which validation could be considered with the final goal of enhancing data quality and of collecting data that is as close as possible to actual respondent behaviour. The question about whether or not survey managers should integrate validation routines into their surveys or whether they should better spend the resources for recruiting new respondents could only be answered on a case-by-case basis. The authors think that validation cost are justified for all surveys with particularly high ambitions in data quality and in longer duration surveys, where each respondent is valuable. Alternative survey methods such as CAWI and smartphone-based tracking may help to

substantially increase the efficiency of validation and to reduce the specific cost per newly added or completed item.

Funding

The authors gratefully thank to the Austrian Science Fund (FWF) for funding the research project Valuing (Travel) Time (Award Number I 1491-G11), from which this article arises. Prof. Jara-Diaz acknowledges funding from Fondecyt, Chile, Grant 1200157, and CONICYT-Basal, Chile, Grant FB0816.

CRediT authorship contribution statement

Florian Aschauer: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Reinhard Hössinger:** Data curation, Writing – review & editing. **Sergio Jara-Diaz:** Writing – review & editing. **Basil Schmid:** Writing – review & editing. **Kay Axhausen:** Writing – review & editing. **Regine Gerike:** Methodology, Writing – original draft, Supervision.

Declaration of Competing Interest

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

References

Ampt, E., Stopher, P.R., 2006, Mixed Method Data Collection in Travel Surveys: Challenges and Opportunities, Sydney, viewed 23 December 2018, from http://hdl.handle.net/2123/19471.

Armoogum, J., Bonsall, P., Browne, M., Christensen, L., Cools, M., Cornelis, E., et al., 2014, Survey Harmonisation with New Technologies Improvement –SHANTI.

Aschauer, F., Hössinger, R., Axhausen, K.W., Schmid, B., Gerike, R., 2018a, 'Implications of survey methods on travel and non-travel activities: A comparison of the Austrian national travel survey and an innovative mobility-activity-expenditure diary (MAED)', European J. Transport Infrastructure Res. 18(1), 4–35.

Aschauer, F., Hössinger, R., Schmid, B., Gerike, R., 2018b. Reporting quality of travel and non-travel activities: A comparison of three different survey formats. Transp. Res. Proc. 32, 309–318.

Aschauer, F., Rösel, I., Hössinger, R., Kreis, H.B., Gerike, R., 2019. Time use, mobility and expenditure: an innovative survey design for understanding individuals' trade-off processes. Transportation 46 (2), 307–339.

Ashley, D., Richardson, T., Young, D., 2009, Recent information on the under-reporting of trips in household travel surveys, Australasian Transport Research Forum (ATRF), 32nd, Auckland, New Zealand, viewed 23 December 2018, from http://atrf.info/papers/2009/2009_Ashley_Richardson_Young.pdf.

Bayart, C., Bonnel, P., 2012. Combining web and face-to-face in travel surveys: comparability challenges? Transportation 39 (6), 1147–1171.

Bonnel, P., Bayart, C., Smith, B., 2015. Workshop Synthesis: Comparing and Combining Survey Modes. Transp. Res. Proc. 11, 108–117.

Bricka, S., Bhat, C.R., 2006. Comparative Analysis of Global Positioning System-Based and Travel Survey-Based Data. Transport. Res. Record: J. Transport. Res. Board 1972 (1), 9–20.

Brög, W., 2015. Surveys on Daily Mobility are not "Surveys to go". Transp. Res. Proc. 11, 98–107.

Brög, W., Erl, E., Meyburg, A.H., Weymuth, M.J., 1982, 'Problems of nonreported trips in surveys of nonhome activity patterns', Transport. Res. Record: J. Transport. Res. Board 891, 1–5, viewed 23 December 2018, from http://onlinepubs.trb.org/Onlinepubs/trr/1982/891/891-001.pdf.

Brög, W., Meyburg, A.H., 1980. Nonresponse Problem in Travel Surveys: An Empirical Investigation. Transport. Res. Record: J. Transport. Res. Board 775, 34–38.

Brög, W., Winter, G., 1990, Investigation of "Non-Reported Trips" in Business-Related Passenger Transport in Household Travel Surveys, Final Report, On Behalf of the German Transport Ministry [in German], Bonn.

Chlond, B., Streit, T., Abler, G., Vortisch, P., 2015. Balancing Innovation and Continuity – Experiences with Survey Design Adaptations of the German Mobility Panel. Transp. Res. Proc. 11, 43–59.

Eisenmann, C., Chlond, B., Minster, C., Jödden, C., Vortisch, P., 2018, 'Assessing the effects of a mixed-mode design in a longitudinal household travel survey', Transportation 2018, from https://doi.org/10.1007/s11116-018-9879-2.

Elias, W., Bekhor, S., Shiftan, Y., 2011. Analysis of travel behavior in Arab communities in Israel: a comparison of household surveys. J. Transp. Geogr. 19 (1), 162–169.

Eurostat, 2011, Eurostat Degree of Urbanisation, Brussels, Statistical Office of the European Communities, viewed 23 December 2018, from http://ec.europa.eu/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=DSP_DEGURBA.

Fellendorf, M., Herry, M., Karmasin, H., Klementschitz, R., Kohla, B., Meschik, M., et al., 2011, Handbook for Travel Surveys. KOMOD – concept study travel data, Vienna, viewed 23 December 2018, from https://www2.ffg.at/verkehr/file.php?id=346.

Gerike, R., Gehlert, T., Leisch, F., 2015. Time use in travel surveys and time use surveys – Two sides of the same coin? Transport. Res. Part A: Policy Pract. 76, 4–24.

Gershuny, J., Harms, T., Doherty, A., Thomas, E., Milton, K., Kelly, P., et al., 2017. CAPTURE24: Testing self-report time-use diaries against objective instruments in real time. University of Oxford, Oxford viewed 23 December 2018, from https://www.timeuse.org/sites/default/files/2017-10/CTUR%20WP%2010%202017.pdf.

Groves, R.M., Lyberg, L., 2010. Total Survey Error: Past, Present, and Future. Public Opinion Quarterly 74 (5), 849–879.

Hubrich, S., Wittwer, R., Gerike, R., 2018. Quality Indicator Set for Household Travel Surveys. Transp. Res. Proc. 33, 219–226.

Jara-Díaz, S.R., Munizaga, M.A., Greeven, P., Guerra, R., Axhausen, K., 2008. Estimating the value of leisure from a time allocation model. Transport. Res. Part B: Methodol. 42 (10), 946–957.

Kelly, P., Doherty, A., Mizdrak, A., Marshall, S., Kerr, J., Legge, A., Godbole, S., Badland, H., Oliver, M., Foster, C., 2014. High group level validity but high random error of a self-report travel diary, as assessed by wearable cameras. J. Transport Health 1 (3), 190–201.

Richardson, A.J., 2003. Behavioral Mechanisms of Nonresponse in Mail-Back Travel Surveys. Transport. Res. Record: J. Transpor. Res. Board 1855 (1), 191–199.

Richardson, A.J., 2007, Immobility in urban travel surveys, 30th Australasian Transport Research Forum. Crawley, Australia, viewed 23 December 2018, from https://atrf.info/papers/2007/2007_Richardson.pdf.

Richardson, A.J., Ampt, E., Meyburg, A.H., 1995. Survey methods for transport planning. Eucalyptus Press, Melbourne, Vic., Australia viewed 23 December 2018, from http://www.no2hcf.co.uk/docs/Traffic_survey_form.pdf.

Safi, H., Assemi, B., Mesbah, M., Ferreira, L., 2017, 'An empirical comparison of four technology-mediated travel survey methods', J. Traffic Transport. Eng. (English Edition) 4(1), 80–87.

Schmid, B., Balac, M., Axhausen, K.W., 2018, 'Post-Car World: Survey methods and response behavior in a multi-stage travel survey', Transportation 2018, from https://doi.org/10.1007/s11116-018-9968-2.

Socialdata, 2009, The New KONTIV-Design (NKD), Munich, viewed 23 December 2018, from http://www.socialdata.de/.

Statistics Austria, 2014, Census 2011 Austria. Results of the Register-based Census, Vienna.

Stinson, L.L., 1999, Measuring how people spend their time: a time-use survey design, Monthly Labor Rev. 122, 12–19.
Stopher, P.R., FitzGerald, C., Xu, M., 2007, 'Assessing the accuracy of the Sydney Household Travel Survey with GPS', Transportation 34(6), 723–741, from https://doi.org/10.1007/s11116-007-9126-8.
Stopher, P.R., Shen, L., 2011. In-Depth Comparison of Global Positioning System and Diary Records. Transport. Res. Record: J. Transport. Res. Board 2246 (1), 32–37.
Taniar, D., Rasouli, S., Timmermans, H., 2014, Mobile Technologies for Activity-Travel Data Collection and Analysis, IGI Global.
Tomschy, R., Herry, M., Sammer, G., Klementschitz, R., Riegler, S., Follmer, R., et al., 2016, Austrian Household Travel Survey 2013/2014. Final Report.: On behalf of: BMVIT, ASFINAG, ÖBB Infrastructure AG, Provincial government of Burgenland, Lower Austria, Styria, Tyrol. Vienna. In German.
Wittwer, R., Hubrich, S., Wittig, S., Gerike, R., 2018. Development of a new method for household travel survey data harmonisation. Transp. Res. Proc. 32, 597–606.
Wolf, J., Loechl, M., Moers, J., Arce, C., 2003a. Trip rate analysis in GPS-enhanced personal travel surveys. In: Stopher, P.R., Jones, P. (Eds.), Transport survey quality and innovation. Emerald Group Publishing Limited, Bingley, UK, pp. 483–498.
Wolf, J., Oliveira, M., Thompson, M., 2003b. Impact of Underreporting on Mileage and Travel Time Estimates: Results from Global Positioning System-Enhanced Household Travel Survey. Transport. Res. Record: J. Transport. Res. Board 1854 (1), 189–198.
Wolf, J., Wilhelm, J., Casa, J. & Sen, S., 2013, 'A Case Study: Multiple Data Collection Methods and the NY/NJ/CT Regional Travel Survey', in J.P. Zmud, et al. (eds.), Transport survey methods: Best practice for decision making, pp. 71–90, Emerald, Bingley.
Yoshimura, Y., Sobolevsky, S., Bautista Hobin, J.N., Ratti, C., Blat, J., 2018. Urban association rules: Uncovering linked trips for shopping behavior. Environ. Plann. B: Urban Analy. City Sci. 45 (2), 367–385.