ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Koszowski, Caroline; Hubrich, Stefan; Wittwer, Rico; Gerike, Regine

Article — Published Version

From trips to stages: a methodology for Generating Stage Information in trip-level Household travel surveys

Transportation

Provided in Cooperation with:
Springer Nature

Suggested Citation: Koszowski, Caroline; Hubrich, Stefan; Wittwer, Rico; Gerike, Regine (2024) : From trips to stages: a methodology for Generating Stage Information in trip-level Household travel surveys, Transportation, ISSN 1572-9435, Springer US, New York, Vol. 53, Iss. 4, pp. 2361-2379, https://doi.org/10.1007/s11116-024-10567-5

This Version is available at:
https://hdl.handle.net/10419/342700

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



http://creativecommons.org/licenses/by/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Transportation (2026) 53:2361–2379
https://doi.org/10.1007/s11116-024-10567-5

From trips to stages: a methodology for Generating Stage Information in trip-level Household travel surveys

Caroline Koszowski[1] · **Stefan Hubrich**[1] · **Rico Wittwer**[1] · **Regine Gerike**[1]

Accepted: 21 November 2024 / Published online: 3 December 2024
© The Author(s) 2024

Abstract
Trip-level household travel surveys (HTS) are an efficient and widely used instrument in transport planning and research and are expected to remain in this role for at least the near future. Mode information is typically assigned to trips in these surveys based on a hierarchy of transport modes that hides important information on the individual stages which is particularly relevant for walking. This study develops a methodology for estimating detailed stage-level information for trip-level HTS that contain some information on stages, this is the sequence of used transport modes and the number of transfers in Public Transport (PT) trips. The methodology is developed based on detailed stage-level data from a sub-sample in the German National HTS MiD 2017 and directly applied to the German city-based HTS SrV 2018 which is a trip-level survey but contains stage-level information on modes and PT transfers. Linear Regression Models for estimating walking stage duration in PT and car trips are combined with simple heuristic estimations for the less frequent types of intermodal trips. Trip purpose, accompaniment and total trip duration are important predictors for walking stage duration. Trip-level and stage-level modal-split figures for the number, duration, and distance of trips and stages in SrV 2018 are computed with the developed methodology. About half of all stages and 30% of trips are done by walking. Walking stage duration is with around 38% considerable, this share drops to around 12% for walking stage distance.

Keywords Household travel survey · Travel diary · Intermodal trips · Walking · Public transport · Modal split

✉ Caroline Koszowski
caroline.koszowski@tu-dresden.de

[1] "Friedrich List" Faculty of Transport and Traffic Sciences, Institute of Transport Planning and Road Traffic, Chair of Mobility System Planning, TUD Dresden University of Technology, 01062 Dresden, Germany

Introduction

Household travel surveys (HTS) have a long tradition and are regularly conducted in many countries on the national level (National Household Travel Surveys, NHTS), and on regional and municipal levels. They are crucial for transport planning and modelling, for preparing decision-making processes, for the evaluation of transport policy measures, and also research on travel behavior heavily relies on this data. HTS can collect trip-level and/or stage-level data where trips are defined as the movement outside home from an origin to a destination for a specific purpose with one or more transport modes. Each segment of a trip covered by one single transport mode is called a stage (Clifton and Muhs 2012; Hubrich et al. 2019; Nobis and Kuhnimhof 2018). The following three main groups of HTS and particularly NHTS can be identified:

1. NHTS with information only on trips, such as NHTS Austria 2013/2014, NTS Canada 2022, NHTS South Africa 2020, and the main survey of the NHTS "Mobility in Germany (MiD) 2017" (Nobis and Kuhnimhof 2018; BMVIT 2016; Republic of South Africa 2021; Statistics Canada 2022)
2. NHTS with information at trip level which is complemented (a) by some information on stages for each trip such as the sequence of used transport modes or the number of transfers within Public Transport (PT) (German HTS SrV 2018 (Hubrich et al. 2019)), or (b) by data on access to and egress from PT stations (U.S. NHTS (Federal Highway Administration 2018, 2019, 2020)); HTS from this group are called “enhanced trip-level HTS” from now on
3. NHTS with detailed information on each stage of a trip including the transport mode, duration and distance; these data are collected for example in the Swiss NHTS (Schweizerische Eidgenossenschaft et al. 2021), the Dutch NHTS 2020 (Statistics Netherlands 2022, 2023) and in the non-representative stage-level sub-sample of the German NHTS MiD 2017 (Follmer 2019)

The majority of trips in HTS is monomodal, meaning that only one transport mode is used for the entire trip. Exemplarily for the city of Berlin, around 60% are monomodal trips (Gerike et al. 2019). These trips are well covered by the trip-level approaches in group 1. PT trips are by definition intermodal trips, meaning that at least two transport modes are combined within a single trip, but also trips by car or bicycle might have access and/or egress stages to/from car respectively bicycle (Jarass and Oostendorp 2017; Kagerbauer et al. 2015; Krygsman and Dijst 2001). In trip-level HTS, these intermodal trips are assigned to a main transport mode based on a predefined hierarchy of modes. This hierarchy is typically derived from stage-level investigations and usually assigns modes with longer distances in their stages a higher priority. PT is typically ranked highest, followed by car, bicycle, and walking (Hubrich et al. 2019; Nobis and Kuhnimhof 2018; Follmer 2019; Burke and Brown 2007). This means that in the commonly used trip-level statistics, walking as a transport mode only becomes visible for monomodal trips, when respondents completely walk from the origin to the destination without using any other mode because this other mode would automatically become the main transport mode for the entire trip. This limited visibility of walking as the core basis of our daily mobility in trip-level statistics is problematic because

researchers cannot understand, planners cannot design, and politicians cannot prioritize what they do not see (Clifton and Muhs 2012; Burke and Brown 2007).

Stage-level HTS are a promising option for a better visibility of walking as the basis for a broader understanding of intermodal travel behavior and walking stages with their various positive effects also for urban planning and public health, these HTS can provide more detailed input data for transport modelling and better support integrated approaches in infrastructure planning. However, most existing HTS belong to group 1. They contain only trip-level information and are with this an efficient way to collect the data that cover most of the needs in research and practice.

Collecting detailed information on stage level as this is the case in group 3 increases respondent burden even when innovative survey approaches (i.e., smartphone-based) are applied. This leads to lower response rates, increased risks of selectivity and nonresponse bias, and finally to higher survey costs. For this study, we therefore assume that even in times of digitization and smartphone-based tracking technologies, "lean" trip-level HTS will continue to be a relevant survey method in the future, possibly in combination with innovative smartphone-based approaches, and that it is feasible and worthwhile to advance them towards group 2-designs and thus to uncover stage-level information.

Literature on stage-level travel surveys (Lu et al. 2023; van Soest et al. 2020) shows that PT trips have the highest proportion of intermodal trips followed by car. Walking is the most relevant mode for accessing the main mode vehicles and for getting to the final destination. Distance and duration of these walking stages are influenced by characteristics of the person, the built environment, PT supply, time of the day, and trip purposes (Jarass and Oostendorp 2017; van der Waerden et al. 2017; van Soest et al. 2020; Lu et al. 2023). Women and persons with higher income tend to have shorter walking stages to/from PT stages but findings are inconclusive and socio-demographic variables such as age, vehicle availability, driver license, household size, education, or employment hardly get significant or are even not considered in studies investigating first/last-mile stages in PT trips. People walk longer distances to rail-based PT vehicles than to busses and to PT services with higher frequency. In terms of trip characteristics, walking stages in work trips are longer than for other purposes in some studies but shorter in others. The reason might be opposite effects. People might prefer rail-based fast and reliable PT services for their trips to work with longer access/egress stages and at the same time, leisure trips might have less time pressure as well as a higher willingness to walk further. Walking stages are shorter in the morning than in the afternoon and for trips with less transfers between PT vehicles within one trip; they get longer with increasing distance of the entire trip. Typical durations of walking stages in PT trips vary between six and nine minutes per stage and 300 m to 800 m. Hardly any literature could be identified on typical characteristics of walking stages in intermodal trips with car or bicycle as the main mode (van der Waerden et al. 2017).

The literature review provides insights on the characteristics of walking stages in PT trips but at the same time, the following three main research gaps are identified: First, previous studies solely rely on descriptive statistics, no model-based analyses could be found. Second, only intermodal PT trips have been covered to far, empirical evidence on the characteristics of walking stages in intermodal car trips is missing. Third, no method could be identified for computing stage-level travel estimates based on enhanced trip-level HTS.

Against this background, three main goals are formulated for this study: (1) to investigate the characteristics of walking stages in intermodal PT and car trips with the help of descrip-

tive and model-based statistics, (2) to develop, on this basis, a methodology for estimating the number, duration, and distance of stages for enhanced trip-level HTS from group 2 as introduced above and (3) to directly apply this methodology for German inner-city trips in cities with ≥ 100,000 inhabitants.

HTS in group 2 are trip-based and thus "lean" in their core approach but ask for the sequence of used transport modes in intermodal trips and also for the number of transfers within the same types of PT vehicles. The methodology in this study is developed based on detailed stage-level data from a non-representative sub-sample in the German NHTS MiD 2017 (Follmer 2019) for which detailed information for individual stages was collected in addition to the trip-level data that is available for the entire sample. The methodology is then directly applied for the German city-based HTS SrV 2018 (Hubrich et al. 2019) which belongs to group 2 in order to demonstrate the applicability of the methodology and also the scope of information that can be generated while still keeping the advantages of the mainly trip-level survey design with only slight modifications. The methodology to be developed in this study has a great potential to enrich existing trip-level HTS of group 2 and to generate key indicators of travel behavior at stage-level such as the modal split by number, distance, and duration as well as to provide adequate visibility particularly for walking.

The remainder of this paper is organized as follows: The subsequent section describes the two German HTS SrV 2018 and MiD 2017 and introduces the approach for computing stage durations for PT and car trips with walking stages based on stage-level data from MiD 2017. Linear Regression Models are estimated and integrated into a comprehensive methodology to compute duration and distance of all stages in SrV 2018. The developed methodology is finally applied to estimate trip-level and stage-level information for SrV 2018. The remaining section discusses and summarizes the results and gives an outlook for further research.

Data and overview of methodology

The German NHTS "Mobility in Germany (MiD) 2017" (Follmer 2019) is used in this study to estimate stage-level travel characterics. MiD 2017 is a nation-wide trip-level HTS which collects representative travel data for Germany as a whole. The MiD 2017 surveyed more than 316,000 people in 156,000 households, and it was the first German NHTS that collected stage-level data for a disproportional and non-representative sub-sample of respondents (aged 14 years and more). These respondents recorded stage characteristics for selected intermodal trips involving more than one mode to provide methodological insights and detailed contextual evidence on intermodal trips (Follmer 2019). In total, 4,465 trips with 11,000 stages were recorded in this sub-sample of MiD 2017 (Federal Ministry for Digital and Transport 2017).

SrV 2018 (Hubrich et al. 2019) is the enhanced trip-level HTS (from group 2) for which the methodology for estimating stage characteristics is developed in this study, based on the detailed information from the sub-sample in MiD 2017. SrV was developed at TUD Dresden University of Technology as a trip-diary based household survey (Wittwer et al. 2024). The SrV-sample is drawn from the population registers of the participating municipalities. All members of the randomly selected households who are willing to participate record their travel behavior at trip level on a randomly selected reporting day and provide in addition information about their individual socio-demographic and socio-economic setting as well

as on the household structure. The reporting days are distributed over the entire field time of one year and are mainly average working days (Tuesday to Thursday, except vacations and bank holidays) (Hubrich et al. 2019; Wittwer et al. 2024).

In SrV 2018, 649,103 trips from 197,566 individuals out of 87,299 German households were recorded (TU Dresden 2018). In addition to the trip data, information on the sequence of used transport modes (for all trips) and on the number of transfers between different types of PT vehicles (for all PT trips) was collected in SrV 2018 which is a unique characteristic of this mainly trip-level survey. 916,489 stages were identified for the 602,691 valid trips, for which trip information was complete. Information on the duration and distance of individual stages was not collected in SrV 2018, this is added in this study based on data from MiD 2017 as introduced above. Out of the whole SrV 2018 dataset, only data on inner-city trips in cities with ≥ 100,000 inhabitants are considered for this study for the following reasons. First, it can be assumed that larger cities with ≥ 100,000 inhabitants have denser and more walkable building structures as well as a better PT supply with possibly different stage characteristics than smaller cities (Gerike et al. 2020). Second, limiting data to inner-city trips avoids biases from rare long-distance trips and ensures a consistent focus on intermodal trips as part of the typical daily travel behavior in urban environments.

The methodology to calculate stage characteristics for the enhanced trip-level HTS SrV 2018 consists of four main steps which are introduced below (see also Fig. 1). Detailed information on data processing will follow in the subsequent section.

1. **MiD 2017 data processing:**

   Consistency checks were carried out as the first step of data processing for the stage-level sub-sample in MiD 2017. Only stages with complete information on the transport mode, duration, and distance were retained for further analysis. Stages with ≤ 100 m are systematically not included in the MID 2017 stage module (Follmer 2019). Missing access and egress stages for PT trips were therefore imputed with a standard duration of one minute, in order to account for the fact that access and egress stages can be assumed to exist for all PT trips. After having completed these two steps of data processing, the data frame included 3,581 trips with 10,166 stages. In the next step, data were filtered by city size and trip distance, only respondents from cities with ≥ 100,000 inhabitants, and inner-city trips with ≤ 30 km were retained leading to a final sample of 1,674 trips with detailed information on 5,237 stages. The threshold of ≤ 30 km in MiD 2017 was set to ensure comparability between MiD 2017 and SrV 2018. 99.8% of all inner-city trips in SrV 2018 (n = 289,156 trips; n = 452,898 stages, city size group of ≥ 100,000 inhabitants) have a distance ≤ 30 km. Therefore, this value was used to define inner-city trips in MiD 2017 which does not distinguish between inner- and intercity trips in the own survey variables.

2. **Estimation of stage duration for MiD 2017 and application to SrV 2018:**

   The majority of intermodal trips (around 90%) in the MiD 2017 stage-level sub-sample are PT trips with walking stages followed by car trips with walking stages. Hence, for these two most relevant types of intermodal trips, Multiple Linear Regression Models were set up to estimate the duration of walking stages as the dependent variable. In the

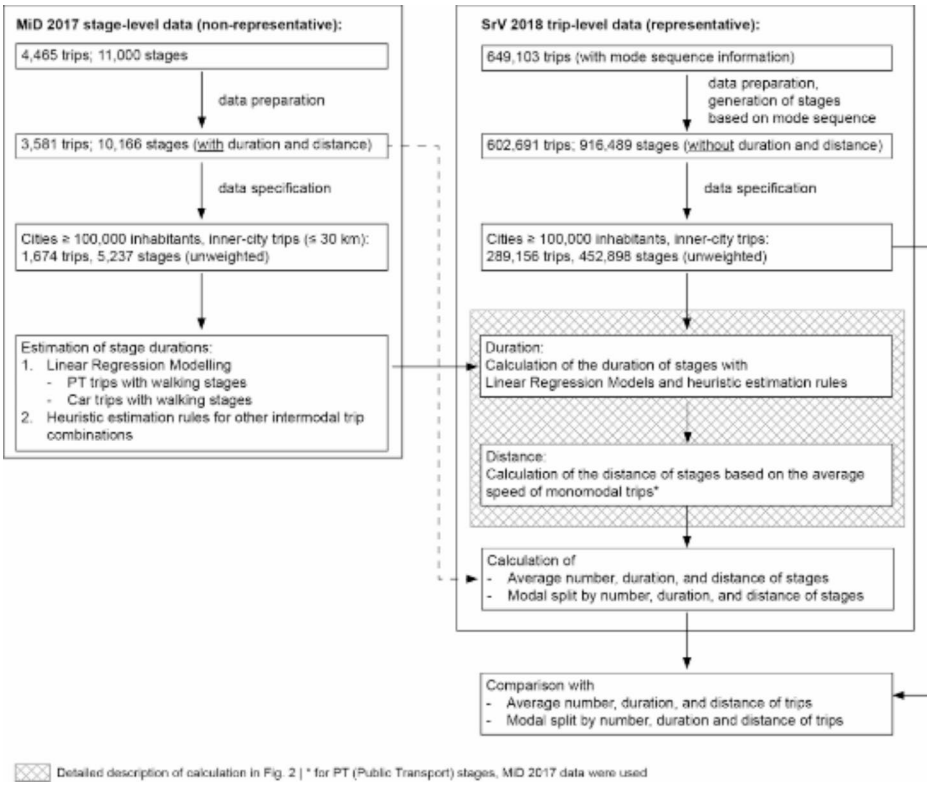


Fig. 1 General overview of the methodology for estimating stage-level information from the enhanced trip-level HTS SrV 2018

SrV 2018 sample, 98% of the trips are monomodal or covered by the Multiple Linear Regression Models. For the remaining intermodal 2% of all trips, heuristic estimations based of other combinations of transport modes in MiD 2017 were developed.

3. **Calculation of stage distance for SrV 2018 using the average speed from monomodal trips:**

   For calculating the distances of walking, cycling, and car stages, the average speed of monomodal trips in SrV 2018 was used. The average speed of PT stages in MiD 2017 was used for calculating distances of PT stages because there are no monomodal PT trips by definition and thus, the trip-level data from SrV 2018 could not be used.

4. **Calculation of stage-level modal-split values for SrV 2018:**

   Finally, the number, duration and distance of stages can be computed for SrV 2018. Based on this, stage-level modal-split proportions and travel characteristics can be calculated.

Development of a methodology for estimating stage-level information from enhanced trip-level HTS

The methodology to be developed in this study foresees the estimation of walking-stage durations for *PT trips with walking stages* as well as for *car trips with walking stages* in MiD 2017 as the two most relevant types of intermodal trips. The selection of independent variables was done based on the initial literature review which was complemented by own assumptions particularly for intermodal car trips for which literature rarely could be found. Two traditional OLS Linear Regression Models were set up (Bortz and Schuster 2010) with duration as the metric dependent variable and dummy-coded (0/1) independent variables. For statistical analysis, IBM SPSS Statistics Version 28 was used.

Initial statistical tests with MiD 2017 data

Multicollinearity of dummy-coded explanatory variables was checked in the first step to ensure the suitability of the MiD 2017 data for computing the Multiple Linear Regression Models. Pearson correlation coefficients show values of $r < 0.5$ for all independent variables (Bortz and Schuster 2010) which means low to moderate correlation. U-tests have less strong assumptions than their parametric counterpart (t-test) (Chang et al. 2016). They are thus preferable for comparing differences of two subsamples (0/1-coded factors) in MiD 2017 and were applied to test for the statistical significance of differences between specific person groups and types of trips which are desirable for subsequent regression modelling.

Table 1 shows the results of these comparisons. For *PT trips with walking stages*, the share of walking stages in the total trip duration differs significantly for the variables *trip purpose* and *trip with accompaniment*. These results are in line with other studies that show longer durations of walking stages in PT trips for leisure (van Soest et al. 2020). It seems like people have more flexibility in these trips and accept or even appreciate higher shares of the slow walking stages. For example, Wang et al. (2021) show in their study for couples with and without children that accompanied leisure trips have a positive effect on daily walking trip time measured in minutes per person (Wang et al. 2021). There are no statistically significant differences between the shares of walking stages for the variables gender, occupation, or educational degree which is also in line with the literature (van Soest et al. 2020). The correlation between *total trip duration [min]* and the *duration of walking stages [min]* was tested with the Pearson correlation coefficient and is highly significant ($r = 0.601$) which is again in line with the literature (van Soest et al. 2020) and a core input for setting up the Multiple Linear Regression Models.

For *car trips with walking stages*, the variables *trip purpose, trip with accompaniment* and *occupation* show significant differences in this comparison. For trip purpose and accompaniment, we assume similar mechanisms as for PT trips. Longer distances to parking garages or on-street parking locations for people with no occupation might be possible reasons behind the significant differences for this variable. Similar to PT trips, a strong correlation between the *total duration of a trip [min]* and the *duration of walking stages [min]* was found for car trips with walking stages ($r = 0.638$).

Table 1 Relationship between independent variables and the share of the duration of walking stages of the total trip duration in MiD 2017 [%]

PT trips with walking stages					
		Share of the duration of walking stages of the total trip duration [%]			
		Mean (%)	Median (%)	p-value	n
Gender	Man	31.9	28.6	0.210	1,141
	Woman	33.5	30.0		
Drivers license	No	32.8	28.6	0.849	1,101
	Yes	32.7	28.6		
Occupation	No	32.6	28.6	0.462	1,141
	Yes	32.7	30.0		
Trip purpose	Leisure	34.6	30.0	0.004**	1,140
	Utilitarian	29.5	28.2		
Trip with accompaniment	Yes	36.5	33.3	0.003**	1,138
	No	31.5	28.6		
Education degree	Other degrees	33.3	30.0	0.058	1,141
	High school diploma	30.2	25.9		
Pearson Correlation r: Total duration of the trip [min] * duration of walking stages [min]		0.601		<0.001***	1,141
Car trips with walking stages					
		Share of the duration of walking stages of the total trip duration [%]			
		Mean (%)	Median (%)	p-value	n
Gender	Man	25.1	16.7	0.138	424
	Woman	27.5	19.4		
Drivers license	No	22.6	15.5	0.448	423
	Yes	26.3	18.2		
Occupation	No	32.9	23.1	<0.001***	424
	Yes	23.2	16.7		
Trip purpose	Leisure	28.6	20.0	<0.001***	424
	Utilitarian	19.9	14.3		
Trip with accompaniment	Yes	29.6	23.1	<0.001***	424
	No	23.5	16.7		
Education degree	Other degrees	26.6	20.0	0.174	424
	High school diploma	24.5	16.7		
Pearson Correlation r: Total duration of the trip [min] * duration of walking stages [min]		0.638		<0.001***	424

p-values: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Multiple linear regression models for estimating stage-level information from MiD 2017 data

Multiple Linear Regression Models are applied to estimate durations of walking stages in PT trips (Model 1) and in car trips (Model 2) (see Table 2). This applies to 14% of all trips as intermodal PT trips with walking stages and to 12% of all trips as intermodal car trips with walking stages in the SrV 2018-database.

Before running the statistical models, the assumptions of OLS regression models were carefully checked. Linearity in parameters is only relevant for non-categorical variables. Linearity between walking duration and the total duration of the trip is sufficiently met within the range of our data points (≤30 km). A normal distribution of residuals in our outcome variables (duration of walking stages in PT and car trips) can be assumed based on visual checks of the residual histograms and Q-Q-Plots. Deviations in the histogram of residuals are reasonably small, deviations in the Q-Q-Plots were mainly found at the lower and upper ends which is acceptable. Observations are not fully independent, as trips are nested within individuals. However, since we treated intermodal PT and car trips separately, the intra-cluster correlation is negligible. Multicollinearity was checked by calculating the correlation matrix of potential predictors but is not present. The assumption of homoscedasticity was tested which is a typical problem in cross-sectional data (Long and Ervin 2000). Heteroscedasticity may lead to wrong standard errors and p-values in OLS Regression Models. For dealing with this problem, the approach of "Heteroscedasticity Consistent Standard Errors" is recommended for estimating robust values. In MiD 2017, the assumption of homoscedasticity is violated. Therefore, HC3 was used for calculating standard errors in the Multiple Linear Regression Models in this study (Long and Ervin 2000).

Two HC3 Linear Regression Models were computed (see Table 2). Model 1 for *PT trips with walking stages* and Model 2 for *car trips with walking stages*. For both models, only

Table 2 Linear Regression Models (HC3) on the impact of individual and trip characteristics on the duration of walking stages in PT trips (Model 1) and in car trips (Model 2)

Dependent variable: Duration of walking stages [min]	Model 1: PT trips with walking stages		Model 2: Car trips with walking stages	
	Modell 1.1 Reduced Model (HC3)	Modell 1.2 Full Model (HC3)	Modell 2.1 Reduced Model (HC3)	Modell 2.2 Full Model (HC3)
	Beta (95% CI)	Beta (95% CI)	Beta (95% CI)	Beta (95% CI)
Intercept	8.8373*** (8.2–9.5)	-3.5053 (-8.2–1.2)	4.4900*** (3.5–5.5)	-8.4358 (-18.7–1.8)
Occupation: No (Ref. Yes)	–	–	5.6123* (0.5–10.7)	2.6923 (-0.9–6.3)
Trip purpose: Leisure (Ref. Utilitarian)	0.0586 (-0.9–1.0)	2.1934*** (1.0–3.4)	1.2150 (-1.3–3.7)	6.5322*** (2.9–10.1)
Trip with accompaniment: Yes (Ref. No)	3.0235** (1.0–5.0)	2.3616** (1.0–3.7)	4.4473* (0.4–8.5)	1.8413 (-1.0–4.7)
Total duration of trips [min]		0.3757*** (0.2–0.5)		0.3555* (0.0–0.7)
Adjusted R^2	0.013	0.382	0.034	0.444
n	1,137	1,137	424	424

p-values: * p<0.05; ** p<0.01; *** p<0.001

significant variables were introduced in Table 1. The *total duration of trips* is found to be the main predictor for the duration of walking stages for both PT trips and car trips. The significant influence of this variable on walking stage duration in intermodal PT trips has been shown in the literature before (van Soest et al. 2020). We assume two possible reasons for this: Persons in the less dense outer parts of a city might have longer distances to their next PT stop and longer total PT trip distances. Persons in any part of the city might be willing to walk longer to PT stops for longer trips. We calculated a full model with and a reduced model without variable *total duration of trips* to facilitate, based on the comparison of these two models, the assessment of the explanatory power particularly for this variable.

As a result, controlling for occupation, trip purpose, and accompaniment, the total duration of trips [min] substantially contributes to the explanatory power in both models. The goodness of fit criterion "Adjusted R^2" increases in Model 1.2 up to the value of 0.382. In Model 2.2, 44.4% of the variance can be explained when adding this variable. Besides the adjusted R^2-statistics, the stability of signs of regression coefficient as well as the logical interpretability of the Beta-coefficients show a sufficient explanatory power of the presented models.

In Model 1.2, *trip purpose* as well as *accompaniment on trips* contribute as reasonably significant predictors to the model. The duration of walking stages extends by about 2.2 min for leisure PT trips with walking stages in contrast to utilitarian trips, for accompanied trips by around 2.4 min compared to trips without accompaniment. The duration of walking stages proportionally increases by around 0.4 min with an increase of one minute in the *total duration of PT trips* which confirms the high relevance of this variable in the model.

In Model 2.2, *trip purpose* emerged as a particular strong and significant predictor with an extension of 6.5 min of walking stages for leisure car-trips compared to utilitarian car-trips. In contrast to utilitarian trips, these trips are probably less time constrained. The duration of walking stages increases by around 0.4 min for each additional minute in the *total duration of car trips* which is very similar to the results for PT trips. *Occupation* and *trip accompaniment* remain statistically insignificant within the full model 2.2. Those factors seem to have only a marginal effect on walking stage duration in intermodal car-trips.

Integration of the models into a comprehensive methodology to estimate duration and distances of all stages in SrV 2018

The developed Multiple Linear Regressions were integrated into a comprehensive methodology to estimate duration and distances of all stages in the SrV 2018 database. Multiple Linear Regression Models are valid for most of the trips in this database but do not produce meaningful results for very short stages and trips and in addition, they only cover intermodal PT trips and car trips with walking stages. Heuristic simple rules were therefore defined to complement the Multiple Linear Regression Models and to estimate characteristics also of stages that were not covered so far. This approach is illustrated in Fig. 2 and described in detail below:

I. **Calculation of stage duration based on the Linear Regression Models and heuristic estimations:**

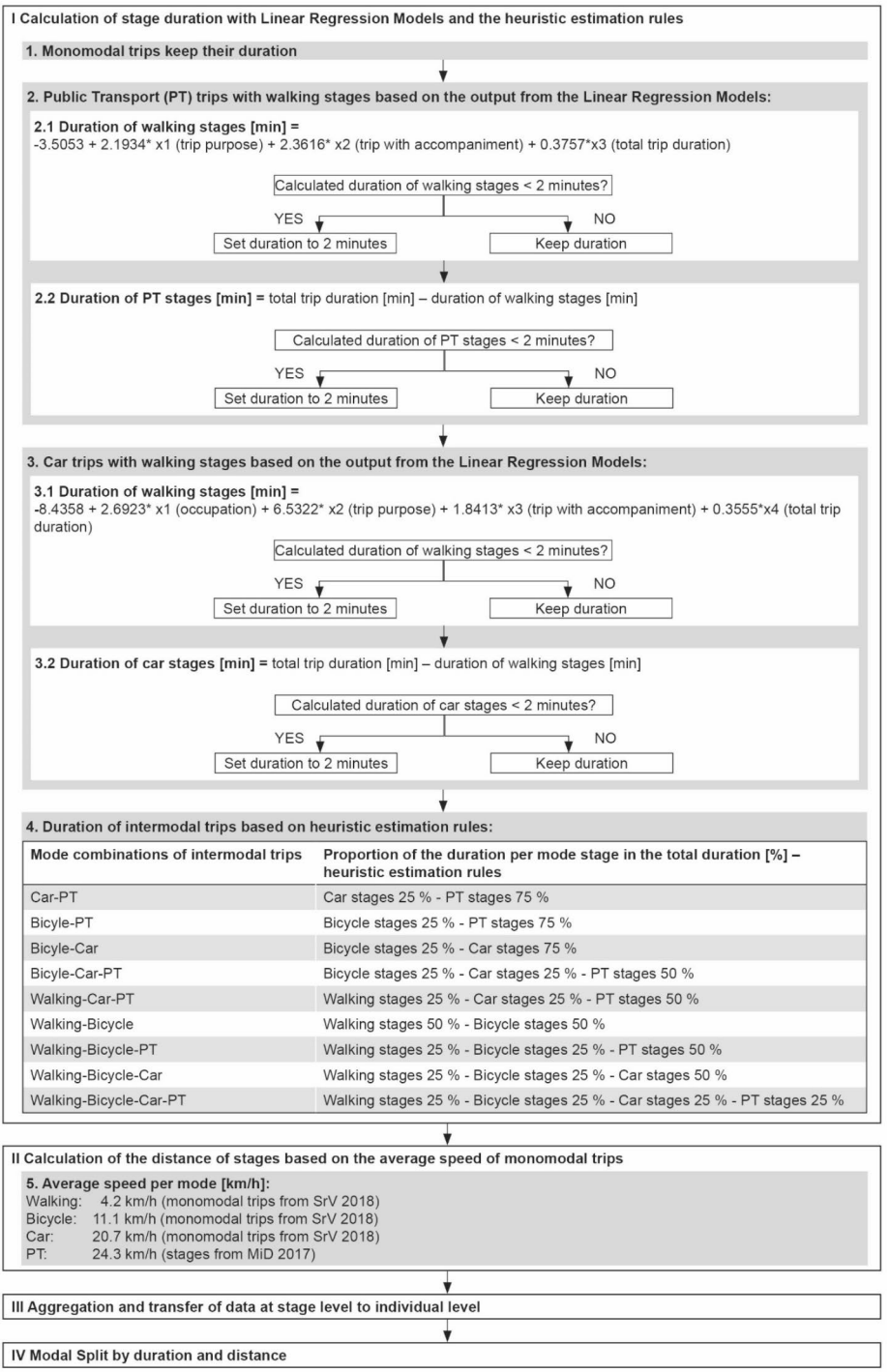


Fig. 2 Detailed overview of the methodology for computing stage durations and distances

For estimating the durations of stages in SrV 2018, all different combinations of transport modes within one trip need to be considered and rules needed to be defined for how to deal with them.

1. In cities with ≥ 100,000 inhabitants, about 70% (n = 205,288) of the trips are monomodal trips with only one stage (walking, bicycling, car trips). The original duration of these trips in SrV 2018 can be retained directly.
2. PT trips are, by definition, intermodal trips as they always include access and egress stages. The Multiple Linear Regression Models developed above were used for PT trips with walking stages as shown in general terms in Equation 1 (see e.g. Bortz and Schuster 2010):

$$\hat{y} = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \cdots + \beta_k * x_k \quad (1)$$

Using the intercept β_0 and the regression coefficients $\beta_{1...k}$ allowed to easily apply the regression coefficients from the full models from Table 2 to SrV 2018 and to estimate the duration of walking stages. For trips with short durations, the models generated very low values for walking stage duration which are not plausible. Two minutes were considered as a minimum value for walking stage duration, based on the assumption, that even short PT trips have walking stages and that one minute for each access to and egress from PT might be a suitable lower threshold. This assumption corresponds to a minimum distance for walking stages of 70 m per stage with an average walking speed of 4.2 km/h. This minimum value of two minutes was also applied for PT stages if these fall below this value when being computed as the difference between total trip duration and the duration of the walking stages within these trips.

3. The procedure described in step 2 was similarly applied to car trips with walking stages.
4. Only 2% (n = 5,344) of all inner-city trips in SrV 2018 (cities with ≥ 100,000 inhabitants) are "other" intermodal trips, which are not covered by the model estimations yet. For these trips, a simple heuristic estimation with 25 % - increments was applied with faster modes (and possibly longer distances) being assigned higher shares of the total duration.

II. **Calculation of the stage distance based on the average speed of monomodal trips:**

Stage distances were calculated by multiplying stage duration by the average speed by mode. For walking, bicycling and car stages, the distances were computed based on the average speed of the respective monomodal trips in SrV 2018. Distances of PT stages were determined based on the average speed of PT stages in MiD 2017 since there are no monomodal PT trips by definition.

III. **Aggregation from stage level to individual level:**

Stage-level information on duration and distance were aggregated to the individual level in the next step.

IV. **Modal split by duration and distance:**

Finally, the weighted shares of the modal split at the stage and trip level by number, duration, and distance, as well as the absolute values of these indicators, could be calculated.

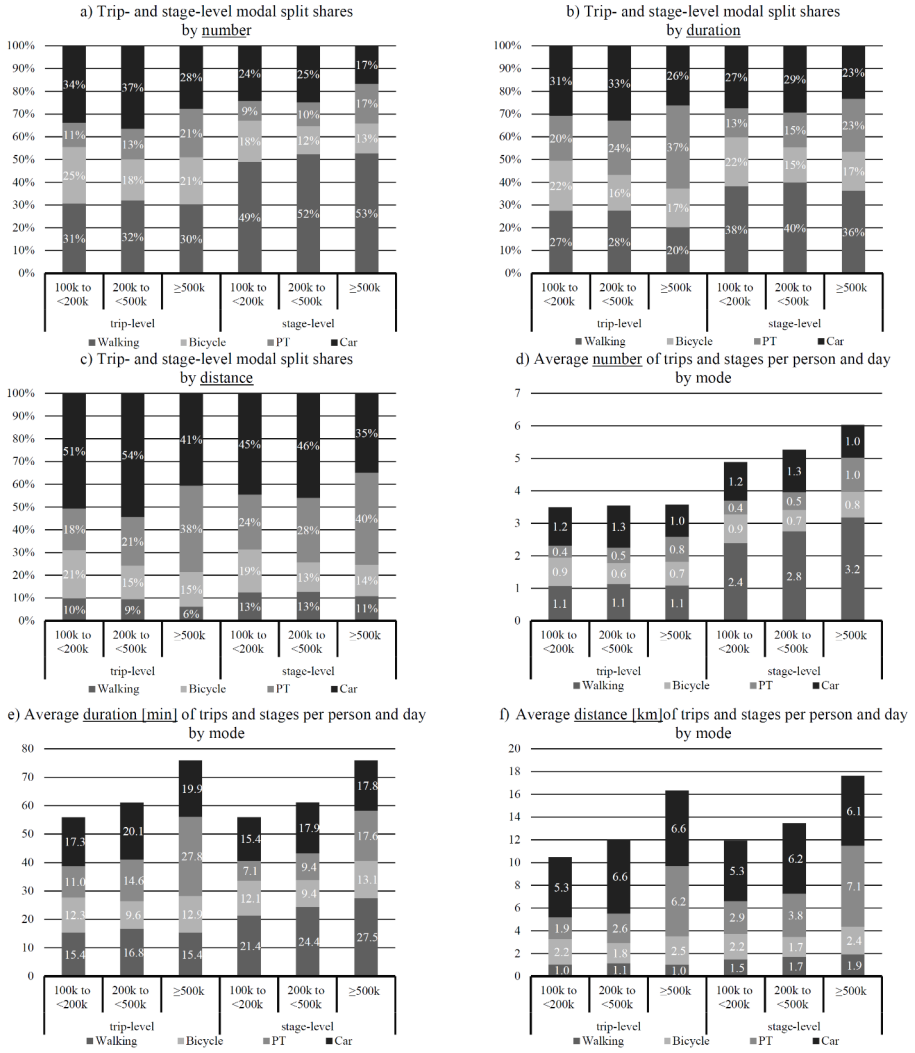
Application of the developed methodology to estimate trip- and stage-level information for SrV 2018

Figure 3a–f show trip- and stage-level estimates by number, duration, and distance for three different city size groups (100,000 to < 200,000 inhabitants; 200,000 to < 500,000 inhabitants and ≥ 500,000 inhabitants) for inner-city travel as it can be now computed with the help of the newly developed methodology.

The comparison of trip-level and stage-level modal split proportions by number (Fig. 3a) confirms the high relevance of walking in stage-level analysis. Around 30% of trips and 50% of all stages are done by walking. The increased shares of walking in stage-level analysis compared to trips are compensated by decreased shares particularly for bicycle and car which are the typical monomodal trips. PT trips have transfers also within the PT mode and thus show less distinct differences between trip-level and stage-level analysis. The comparison of city size groups in Fig. 3a shows similar tendencies for trips and stages. The shares of walking are surprisingly similar across city size. The shares of car and PT are quite similar for cities with < 500,000 inhabitants and decrease only for the largest cities with ≥ 500,000 inhabitants. The shares of bicycle are higher for the smallest cities < 200,000 inhabitants but quite similar for the two higher city size groups. Better PT supply seems to have a significant effect particularly in the biggest city size group. It reduces trips and stages both with bicycle and car whereas walking remains similar in its shares.

Duration and distance of trips and stages in Fig. 3b and c show significantly different magnitudes but similar tendencies in the comparison between trips and stages and also between city size groups. The proportion of walking drops from around 50% for trip numbers to around 38% for the duration and 12% for the distance. Walking as the slowest mode with the shortest distances is thus most sensitive to the different indicators but still covers a substantial share of the overall trip duration. Interestingly, the share of walking drops for duration and distance between the two smaller and the biggest city size group. The reason for this might be the better PT supply in cities with ≥ 500,000 inhabitants; distances to PT stops and walking stages are possibly shorter on average so that the share of walking drops even though the PT share increases. The drop from trips to stages across all city size groups is most substantial for PT, this shows the substantial share of walking in these trips. Distance-based modal split values in Fig. 3c show the least differences between trips and stages. Walking matters for the number and duration of stages but hardly makes a difference in terms of distance.

Figure 3d–f with the average values of number, duration and distance for trips and stages per mode, person and day are interesting in the comparison of trips and stages but particularly also in the comparison between the city size groups. The overall number of trips is very similar for all three city size groups whereas the number of stages increases substantially in the largest group with ≥ 500,000 inhabitants. These differences are mainly caused by walk-



Abbreviation: 100k = 100,000 inhabitants; PT = Public Transport
n (100 to < 200k) = 12,017 trips//58,657 stages | n (100 to < 200k) = 22,151 trips//116,592 stages | n (≥ 500k) = 46,069 trips//277,769 stages

Fig. 3 **a**–**f** Comparisons of trip- and stage-level estimates by number, duration, and distance for cities with ≥ 100,000 inhabitants (weighted, inner-city travel)

ing stages and PT stages. The number of PT stages doubles from the second group (200,000 to < 500,000 inhabitants) to the third group (≥ 500,000 inhabitants) with higher PT usage and more transfers within PT trips in this group being the main reasons. The number of trips and stages per person are almost the same for all city size groups for car and bicycle as the two typical monomodal transport modes.

For duration and distance in Fig. 3e and f, differences between city size groups are apparent for trips and stages. People in cities with ≥ 500,000 inhabitants travel longest both in terms of duration and distance. Differences mainly come from PT. Trip durations for PT are substantially longer for cities with ≥ 500,000 inhabitants than for smaller cities, the differ-

ence is less distinct for stages because parts of PT trip duration are now included in walking stage duration. Differences between city size groups are less distinct for distance in Fig. 3f, but still the increased overall distances of trips and stages for cities with ≥ 500,000 inhabitants are mainly caused by PT.

Overall, the analyses reveal interesting pattern in the comparison of trips and stages in terms of number, duration and distance and also of the city group sizes. Sensitivity analysis shows the stability of the results. Important parameters from Fig. 2 were varied, such as the threshold of two minutes for very short stages of walking for PT and car (step I.2 and I.3), the simplified heuristic estimates for other intermodal trips (step I.4), as well as the average speed per mode (step II). Deviations in all indicators are marginal in these sensitivity analyses, they reach from − 0.8 to + 0.8 percentage points.

Discussion and conclusion

This study has generated methodological insights and in addition, it contributes to a better understanding of travel behavior in cities with ≥ 100,000 inhabitants. The Linear Regression Models confirm the high relevance of the overall trip distances for the distances of walking stages in intermodal trips. This is in line with the literature (van Soest et al. 2020) but has not been demonstrated based on model-based analysis before. We controlled for socio-demographic attributes and trip characteristics but still, the overall trip distances are most relevant for understanding walking at the stage level. These mechanisms hold similarly for intermodal PT trips and car trips which is an important contribution to the literature where hardly any study on intermodal car trips could be identified. The Linear Regression Models are complemented by simple heuristic estimates for the less frequent combinations of transport modes in intermodal trips. This approach is reasonable and robust because the proportion of these trips is small.

Overall, the developed methodology for estimating stage information for enhanced trip-level HTS proves to be viable and insightful. It increases the level of detail in the travel data substantially with only slight increases in respondent burden and survey costs. The methodology can be directly applied to any specific German city or region because it was developed based on the stage module of the German NHTS MiD 2017 which should be valid across the entire country. Future studies could estimate Linear Regression Models for stage-level surveys in other countries and thus investigate the transferability of the mechanisms that have been identified in this work for the German context. The average speed of each transport mode as the basis for deriving stage distances from durations needs also to be verified but apart from these possible differences in travel characteristics between countries, the overall scheme for computing stage characteristics should be transferable to any other context.

It would be also interesting to adapt the developed methodology to rural areas and to include also trips between municipalities. However, the relevance of stage-level analysis is expected to be lower here, mainly because of lower PT use compared to cities with ≥ 100,000 inhabitants which are in the focus of this study. The combination of enhanced (but still lean) trip-level HTS with smartphone-based tracking studies is promising. Regression models (Linear Models or more complex models such as Generalized Linear Models or Generalized Linear Mixed Models) on the duration of walking stages in intermodal trips similar to this study might be computed based on the detailed stage-level smartphone-based tracking

data, they can be used then in the enhanced trip-level HTS which at least to date are less costly than their smartphone-based counterparts. They could be run with far larger sample sizes but still provide stage-level information. Larger stage-level samples would also allow to distinguish between road- and rail-based PT vehicles in the statistical models and might add valuable information. The use of Large Language Models (LLMs) to study walking in intermodal trips is unchartered territory and might open up new avenues for future research.

In terms of content, this study provides insights on the number, duration and distance of trips and stages in absolute values and in proportions by transport mode. Walking trips and PT trips are particularly sensitive to the two analysis-levels of trips and stages. About half of all stages and 30% of trips are done by walking. Walking stage duration is with around 38% still significant, this share drops to around 12% for walking stage distance. In absolute numbers, this means that on average 3.2 out of 6.0 stages are being walked per mobile person and day with an average duration of 27.5 min and an average distance of two kilometres. The high proportion of walking stage duration was not expected before. Travellers spend substantial shares of their time for walking which is good because walking is space-efficient, supports lively streets and cities and is beneficial for public health.

Differences between city size groups are also interesting. The shares of walking for trips and stages in terms of number is almost the same for all groups even though PT usage is substantially higher in the largest cities with ≥ 500,000 inhabitants. This means that people have a similar number of walking stages but that these stages belong more to monomodal walking trips in smaller cities and to intermodal trips mainly in combination with PT or car in the larger cities. The interactions between the modes in the different city size groups show that better PT supply reduces the use of the car but also of walking (and bicycling) as monomodal trips. It is therefore essential to plan for the environmentally friendly modes walking, cycling and PT in an integrated way and in combination with restrictive policy measures for the car.

Besides these contributions to the scientific literature, this study provides insights for practical survey designers and transport planners outside academia: First, the developed methodology for computing stage-level information based on enhanced HTS can be directly applied to any German region and also to regions outside Germany if the statistical models are checked before for their applicability. Second, this study directly provides new empirical evidence on walking at stage level in German cities, it demonstrates the high relevance of this mode which is less visible in the trip-level statistics that are commonly used in transport planning so far. This evidence can be used to support initiatives to promote walking as a mode of transport, as a physical activity and also as an effective way to populate streets and public spaces with people which is essential for improving the quality of life and the economic success of cities.

The results of this study are encouraging for survey designers to go for enhanced trip-level HTS, possibly in combination with stage-level smartphone-based surveys. The additional insights on detailed stage-level travel characteristics that can be gained compared to traditional trip-level HTS approaches clearly outweigh the additional costs and provide valuable information for research and for planning practice.

Acknowledgements This study is part of the PhD thesis by C. Koszowski and has received no funding. We thank the German Federal Ministry for Digital and Transport for providing the MiD 2017 data. SrV 2018 data is used for research purposes with kind permission of SrV funding cities and organizations.

Author Contribution The authors confirm contribution to the paper as follows: study conception and design: C. Koszowski; S. Hubrich; analysis and interpretation of results: C. Koszowski, S. Hubrich, R. Wittwer and R. Gerike; draft manuscript preparation: C. Koszowski, S. Hubrich, R. Wittwer and R. Gerike. All authors reviewed the results and approved the final version of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability Data is available upon request for SrV and can be requested from the German Transport Ministry for MiD.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

BMVIT: Österreich unterwegs 2013/2014". Edited by Bundesministerium für Verkehr, Innovation und Technologie (2016). Available online at https://www.bmk.gv.at/themen/verkehrsplanung/statistik/oesterreich_unterwegs/berichte.html, checked on 5/13/2023

Bortz, J.; Schuster, C.: Statistik für Human- und Sozialwissenschaftler. 7., vollständig überarbeitete und erweiterte Auflage. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg (Springer-Lehrbuch) (2010). Available online at http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10448295

Burke, M., Brown, A.L.: Distances people walk for transport. Road Transp. Res. **16**(3), 16–29 (2007)

Chang, J., Shao, Q.-M., Zhou, W.-X.: Cramér-type moderate deviations for Studentized two-sample *U*-statistics with applications. Ann. Statist. (2016). https://doi.org/10.1214/15-AOS1375

Clifton, K., Muhs, C.D.: Capturing and representing multimodal trips in travel surveys. Transp. Res. Record **2285**(1), 74–83 (2012). https://doi.org/10.3141/2285-09

Federal Highway Administration: NHTS—Main Study Retrieval Questionnaire (2018). Available online at https://nhts.ornl.gov/assets/2016/NHTS_Retrieval_Instrument_20180228.pdf, checked on 5/13/2023

Federal Highway Administration: National Household Travel Survey (NHTS), 2017 NHTS Data User Guide (2019). Available online at https://nhts.ornl.gov/assets/NHTS2017_UsersGuide_04232019_1.pdf, checked on 5/13/2023

Federal Highway Administration: NHTS Household File—Public Use Codebook (2020). Available online at https://nhts.ornl.gov/assets/codebook_v1.2.pdf, checked on 5/13/2023

Federal Ministry for Digital and Transport: Database Mobilität in Deutschland (MiD) 2017. Own analysis based on MiD 2017

Follmer, R.: Mobilität in Deutschland - MiD Wegeerfassung im Etappenkonzept. Studie von infas, DLR und IVT im Auftrag des Bundesministers für Verkehr und digitale Infrastruktur (FE-Nr. 70.904/15). Bonn, Berlin (2019). Available online at www.mobilitaet-in-deutschland.de

Gerike, R., Hubrich, S., Ließke, F., Wittig, S., Wittwer, R.: Tabellenbericht zum Forschungsprojekt Tabellenbericht zum Forschungsprojekt, Mobilität in Städten—SrV 2018" in Berlin, updated on 2020 (2019)

Gerike, R., Koszowski, C., Hubrich, S., Wittwer, R., Wittig, S., Pohle, M. et al.: Aktive Mobilität: Mehr Lebensqualität in Ballungsräumen. Abschlussbericht. Forschungskennzahl 3716 58 104 0. Edited by Umweltbundesamt (2020). Available online at https://www.umweltbundesamt.de/publikationen/aktive-mobilitaet-mehr-lebensqualitaet-in, checked on 7/20/2023

Hubrich, S., Ließke, F., Wittwer, R., Wittig, S., Gerike, R.: Methodenbericht Methodenbericht zum Forschungsprojekt „Mobilität in Städten—SrV 2018", Technische Universität Dresden (2019). Available online at https://tu-dresden.de/srv

Jarass, J., Oostendorp, R.: Intermodal, urban, mobil—Charakterisierung intermodaler Wege und Nutzer am Beispiel Berlin. RuR (2017). https://doi.org/10.1007/s13147-017-0478-z
Kagerbauer, M., Hilgert, T., Schroeder, O., Vortisch, P.: Household travel survey of intermodal trips—approach, challenges and comparison. Transp. Res. Procedia **11**, 330–339 (2015). https://doi.org/10.1016/j.trpro.2015.12.028
Krygsman, S., Dijst, M.: Multimodal trips in the netherlands: microlevel individual attributes and residential context. Transp. Res. Record **1753**(1), 11–19 (2001). https://doi.org/10.3141/1753-02
Long, J.S., Ervin, L.H.: Using heteroscedasticity consistent standard errors in the linear regression model. Am. Stat. **54**(3), 217 (2000). https://doi.org/10.2307/2685594
Lu, Y., Kimpton, A., Prato, C.G., Sipe, N., Corcoran, J.: First and last mile travel mode choice: a systematic review of the empirical literature. Int. J. Sustain. Transp. **1**(1), 1–14 (2023). https://doi.org/10.1080/15568318.2023.2218285
Nobis, C., Kuhnimhof, T.: Mobilität in Deutschland—MiD Ergebnisbericht. Studie von infas, DLR, IVT und infas 360 im Auftrag des Bundesministers für Verkehr und digitale Infrastruktur (FE-Nr. 70.904/15). Bonn, Berlin (2018). Available online at https://www.mobilitaet-in-deutschland.de/archive/pdf/MiD2017_Ergebnisbericht.pdf, checked on 5/13/2023
Republic of South Africa: Department: statistics South Africa; national household travel survey 2020. Statistical release (2021). Available online at https://www.statssa.gov.za/publications/P0320/P03202020.pdf
Schweizerische Eidgenossenschaft; Bundesamt für Statistik; Bundesamt für Raumentwicklung: Mobilitätsverhalten der Bevölkerung. Ergebnisse des Mikrozensus Mobilität und Verkehr 2021 (2021). Available online at https://www.bfs.admin.ch/bfs/de/home/statistiken/mobilitaet-verkehr/erhebungen/mzmv.assetdetail.24165261.html, checked on 5/13/2023
Statistics Canada: National Travel Survey Canada 2022 (2022). Available online at https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=1517074
Statistics Netherlands: ODiN, the Dutch National Travel Survey (2022). Available online at https://unece.org/sites/default/files/2022-06/1%20Netherlands%20WP%206%20presentation%20June%202022%20Dutch%20National%20Dutch%20Travel%20Survey.pdf, checked on 5/13/2023
Statistics Netherlands: Dutch National Travel survey (Website) (2023). Available online at https://www.cbs.nl/en-gb/our-services/methods/surveys/brief-survey-description/dutch-national-travel-survey, checked on 5/13/2023
TU Dresden: Database SrV 2018. Own analysis based on SrV 2018. Chair of Mobility System Planning (2018)
van der Waerden, P., Timmermans, H., de Bruin-Verhoeven, M.: Car drivers' characteristics and the maximum walking distance between parking facility and final destination. JTLU (2017). https://doi.org/10.5198/jtlu.2017.568
van Soest, D., Tight, M.R., Rogers, C.D.F.: Exploring the distances people walk to access public transport. Transp. Rev. **40**(2), 160–182 (2020). https://doi.org/10.1080/01441647.2019.1575491
Wang, Z., Ettema, D., Helbich, M.: Objective environmental exposures correlate differently with recreational and transportation walking: a cross-sectional national study in the Netherlands. Environ. Res. **194**, 110591 (2021). https://doi.org/10.1016/j.envres.2020.110591
Wittwer, R., Hubrich, S., Gerike, R.: New evidence on nonresponse in household travel surveys. Transp. Res. Procedia **76**, 233–245 (2024). https://doi.org/10.1016/j.trpro.2023.12.051

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Caroline Koszowski has been a research associate and lecturer at TU Dresden since 2015. She studied spatial planning at the Technical University of Kaiserslautern (2008–2012) and completed a master's degree in urban and regional planning at the University of Kassel (2012–2015). Her research focuses on walking, active mobility, place activities, travel behavior, integrated transportation planning, and urban street design. A central aspect of her research is understanding the needs and behavior of users in public spaces (https://tu-dresden.de/vkw/msp/die-professur/koszowski).

Stefan Hubrich is a senior researcher and research group leader at TU Dresden. He earned his diploma in transportation engineering with a specialization in transport planning and traffic engineering in 2008 and completed his PhD in 2017. His research is dedicated to survey methodology and statistical data analysis in the field of travel surveys. He is responsible for the repeated large-scale German household travel survey "Mobility in Cities – SrV", with a focus on survey design and quality assurance. Additionally, he is involved in research on the application of smartphone tracking apps in travel surveys (https://tu-dresden.de/vkw/msp/die-professur/hubrich).

Rico Wittwer has been a senior researcher and lecturer at TU Dresden since 2003. He studied civil engineering, specializing in transport planning as well as city and regional planning. He holds a PhD in transport planning from TU Dresden (2008) and a postdoctoral qualification (habilitation) in the same field from TU Dresden (2014). He is actively involved in the key project Mobility in Cities (SrV), contributing to sampling theory, weighting, data analysis, and visualization. With extensive experience in data science, he specializes in travel and attitude surveys, as well as travel behavior analysis using multivariate statistics (https://tu-dresden.de/vkw/msp/die-professur/wittwer).

Regine Gerike is the head of the Chair of Mobility System Planning at TU Dresden. Before joining TU Dresden, she chaired the Institute for Transport Studies at the University of Natural Resources and Life Sciences (BOKU) in Vienna, Austria. Regine Gerike did her PhD at TU Dresden in the field of modelling transport and its environmental effects. From 2008 to 2012 she was assistant professor at Technische Universität München and head of the research centre mobility and transport (mobil.TUM). Her research interests focus on transport planning and empirical research on travel behavior (https://tu-dresden.de/vkw/msp/die-professur/inhaberin).