

12<sup>th</sup> International Conference on Transport Survey Methods

# Identifying and understanding long-distance travel demand by combining official transport statistics and survey data

Angelika Schulz<sup>a\*</sup>, Miriam Magdolen<sup>b</sup>, Bastian Chlond<sup>b</sup>, Tobias Kuhnimhof<sup>c</sup>, Katja Köhler<sup>a</sup>

<sup>a</sup>German Aerospace Center (DLR), Institute of Transport Research, Rutherfordstr. 2, 12489 Berlin, Germany

<sup>b</sup>Karlsruhe Institute of Technology (KIT), Institute for Transport Studies, Otto-Ammann-Platz 9, 76131, Germany

<sup>c</sup>RWTH Aachen University, Institute of Urban and Transport Planning, Mies-van-der-Rohe-Straße 1, 52074 Aachen, Germany

---

## Abstract

While much is known about everyday travel of the German population, long-distance travel is still underreported. The main data source, the national travel survey “*Mobility in Germany (MiD)*”, cannot simply be used to describe the demand: complex extrapolations and complementary data are necessary to obtain a consistent picture. The presented approach of ‘data fusion’ integrates different data sources to provide the overall long-distance travel demand.

The result reveals that almost half of the total transport performance of the residential population in Germany (46 % of passenger kilometers) is accounted for by trips of at least 100 km (one-way distance).

© 2023 The Authors. Published by ELSEVIER B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the International Steering Committee for Transport Survey Conferences (ISCTSC)

*Keywords:* long-distance travel; data fusion; data consistency; data merging

---

## 1. Introduction

In contrast to everyday travel, long-distance passenger travel continues to increase for the German population resulting in a share of roughly 40 to 50 % of total passenger kilometers. Although data are available for some types of travel (e.g., long-distance rail, air), there is no up-to-date and consistent overall picture of this specific demand, neither in terms of total transport volumes, nor in terms of socio-demographic characteristics of the corresponding population and the driving forces. A dedicated survey – already carried out in the early 2000s (*INVERMO 2002/2003*) – revealed that long-distance travel was extremely unevenly distributed among the population: only a small proportion was

---

\* Corresponding author. Tel.: +49-30-67055-294.

E-mail address: [angelika.schulz@dlr.de](mailto:angelika.schulz@dlr.de)

responsible for the majority of travel volumes. This small proportion was characterized by high education levels and higher incomes (Zumkeller et al. 2005). The question arises whether this fact has already changed in view of current societal processes, such as an increasing level of education, the conversion to a knowledge-based society, long-term socialization in terms of traveling to distant destinations, the emergence of multi-local lifestyles or wide-ranging social networks, all of them likely to result in changing travel volumes and shifting proportions between different groups of travelers.

Analyses in this respect require a meaningful data framework. The challenge is to gather and prepare available data according to an appropriate definition of 'long-distance travel' and to close data gaps based on complementary data collection or other empirically founded assumptions.

Despite their focus on everyday mobility, conventional National Household Travel Surveys (NHTS) continue to be the main data source for describing and quantifying long-distance travel. However, in the case of the German NHTS (*MiD – Mobility in Germany*) there are some aspects to be considered when extracting long-distance trips. On the one hand, such trips are subject to underreporting: While respondents are only asked to report actual trips made on a designated reporting day, long-distance trips not only tend to occur less frequently, but are also very unevenly distributed across the population. Furthermore, unit nonresponse is a problem: highly active individuals, who are responsible for a large share of long-distance trips, are very likely to be absent during the respective survey period. To address these issues, the *MiD* survey comprises a dedicated 'journey module' for respondents aged 14 and over, aiming to cover overnight journeys during the last three months regardless of their distance. Due to the criterion 'overnight stay', however, day trips, which may also encompass longer distances, are not included. The same applies to journeys made by younger children. On the other hand, a number of multi-day trips may be partially recorded twice: First, when either the outward or the return trip is reported on the reporting day, and a second time, if the entire overnight journey is also reported in the 'journey module'. To eliminate the effects of unit nonresponse and potential double counting (or even over-reporting for example of air travel due to its subjectively perceived importance), original microdata requires several steps of post-processing, imputation and reweighting. To this end, a 'data fusion model' was developed to generate a consistent trip dataset containing trips and journeys<sup>†</sup> of all distances for the entire German population, but without overlaps, double counting and missing trips. Furthermore, the 'data fusion model' was calibrated using further socio-economic and transport statistics.

The resulting dataset provides the quantitative basis to determine key figures for long-distance travel (e.g., travel demand and volume, trip purposes, modal split), also in relation to everyday mobility. It can be broken down by different modes, socio-demographic characteristics of travelers, travel purposes, etc.

Further sources are used to complement the picture, in particular data that are not provided by conventional NHTS (e.g., public transport passenger statistics). The compilation of data from different sources usually is a major challenge: Each data source is collected for a particular purpose, uses different conceptual definitions (e.g., single trip vs. complete journey including outbound and return trip) or threshold values to identify long-distance trips (e.g., a minimum of 50, 100 or 300 km one-way distance vs. a minimum number of overnight stays). Last but not least, they follow different methodological approaches (e.g., passenger counting vs. questionnaire-based trip interview). All this together may cause additional methodological biases, as the different surveys are usually not coordinated with each other.

The structure of this paper is as follows: First, the challenge of merging data is described against the background of diverging definitions of long-distance travel. Second, the adopted approach of 'data fusion' is presented. The current picture of long-distance travel in Germany is then outlined. Finally, methodological issues and challenges of data collection and data merging focusing on long-distance travel are discussed.

---

<sup>†</sup> In the context of this paper, a distinction is made between 'trips' and 'journeys': While trips do not include any overnight stay, journeys are considered as travel events consisting of two trips (outward and return) with at least one overnight stay in between.

## 2. Data on long-distance travel

In many countries, data on long-distance travel are fragmentary and not harmonized. This also applies to Germany, where there is no comprehensive data source to identify and describe long-distance travel. However, there are many independent data sources that cover at least parts of long-distance travel. Information from different data sources can thus be assembled like a puzzle. Careful attention must be paid to what is included in the data, e.g., the number of trips or journeys (outbound and return), and which travelers are included. In addition, when reviewing the available data, it is important to consider that travel events may appear in multiple data sources. Particularly important is the distinction between the territory-based principle on the one hand and the resident-based principle on the other. The first encompasses all travel within the country's borders, both travel by the domestic population and by tourists from abroad or transit passengers; travel beyond national borders, however, is not included. The resident-based principle, in contrast, addresses travel of the resident population only, regardless of whether it takes place within or outside national borders. As a consequence, there is no information on the travel of inbound tourists and transit passengers. Figure 1 illustrates the two principles for Germany.

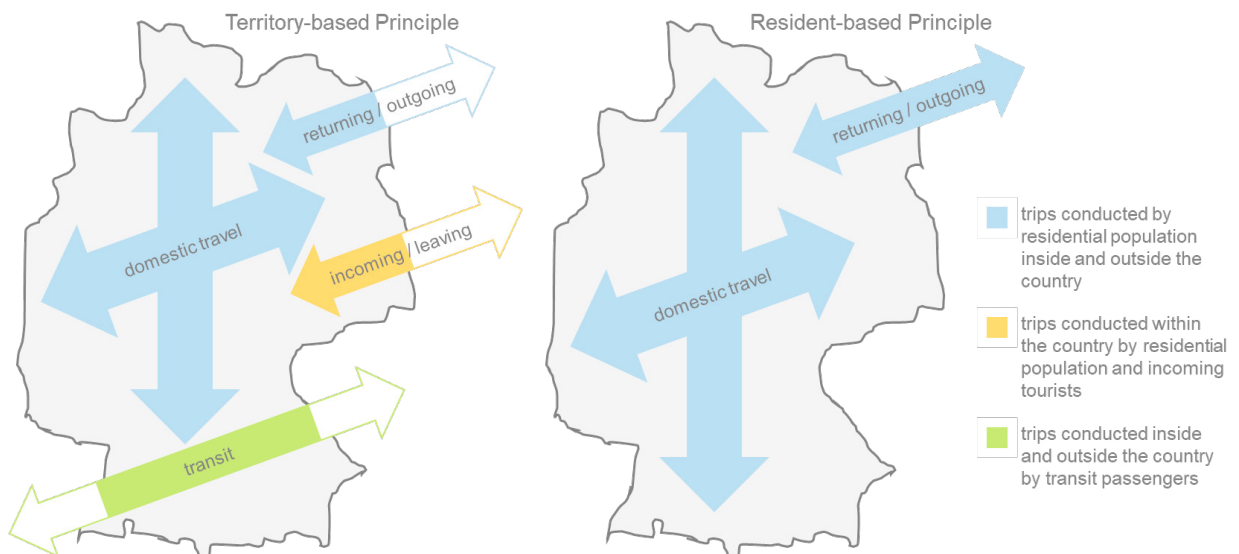


Figure 1. Distinction between the territory-based and the resident-based principle for Germany

As mentioned above, there are many different data sources on mobility, some of which also cover long-distance travel. They either use the resident-based or the territory-based principle and can be roughly divided into three groups. A brief overview is given below.

### *Data from official statistics*

Transport companies (e.g., long-distance rail and bus operators) and the tourism industry must report selected services to the Federal Statistical Office. Resulting figures published by official statistics are usually aggregated, with little differentiation according to the travelers' characteristics. Following the territory-based principle, statistically speaking all travel events within the country are reported. Thus, data also include trips of non-residents within the country. If no clear distinction between local and long-distance travel is applied, not only the transport volume (number of passengers and passenger kilometers) of long-distance services is recorded, but also journeys of short distances may be included. Conversely, long-distance trips conducted with local or regional transport services are not transferred to long-distance transport statistics. Another problem when using and interpreting figures from official statistics arises from double counting. For example, in the case of an intermodal trip by rail and air, each trip or flight is reported separately. This also applies to air travel statistics, when passengers with transfer connections are recorded twice or even four times (for each departure and each arrival).

*Data from companies and service operators*

Transport and travel service operators collect their own statistics and data, particularly for corporate or market research purposes. Examples are surveys of touristic multi-day trips (short- and long-term vacations), data from the cruise industry, or surveys conducted by the business travel association. These surveys are typically selective: only specific groups of travelers (e.g., commuters) or companies are addressed (e.g., associated members only, minimal turnover), or mainly economically relevant aspects are captured (e.g., number of overnight stays, expenses at destination). These data collections are mostly carried out following the resident-based principle: samples include either the residential population or companies with a registered office in Germany, thus capturing travel within and outside of Germany.

*Data from travel surveys*

Travel surveys are the main data source to describe long-distance travel. Usually, such surveys are conducted following the resident-based principle, primarily focusing on everyday travel. Long-distance travel may also be included, but only incompletely due to frequent absence of respondents on the reporting date. Additional modules (e.g., the *MiD* journey module) are therefore used to record trips with overnight stays. Even surveys designed to be representative are subject to certain socio-demographic and socio-economic selectivities as well as over- and underrecording of specific travel types. One example is the participation of especially those people who are interested in the topic of the survey, another is the recall effect due to the limited ability to remember journeys made in the past.

Figure 2 provides an overview on available data sources containing information on long-distance travel or at least some parts thereof. Each survey or data source (all represented by colored rectangles, respectively) covers specific parts of travel.

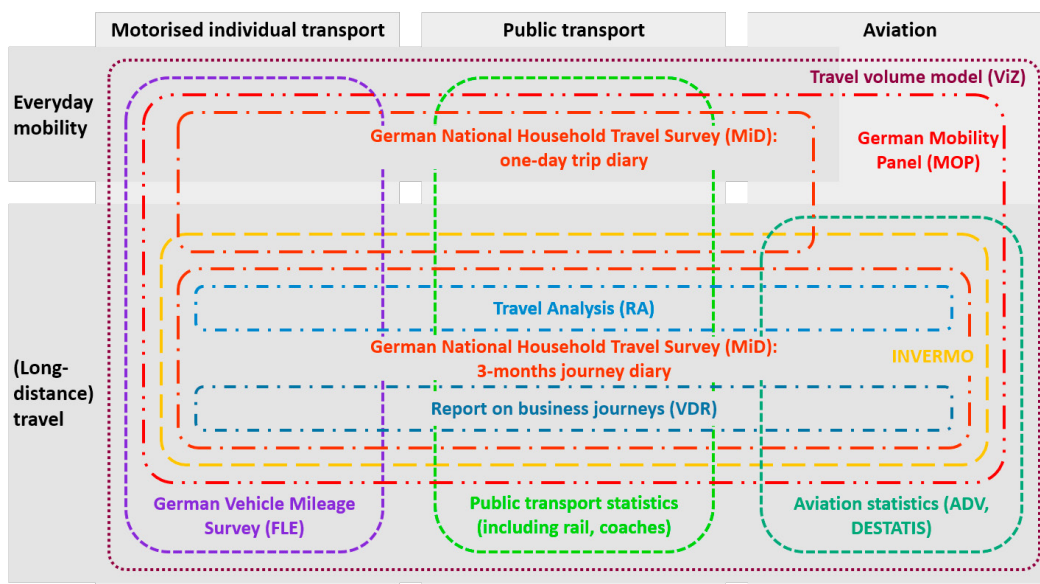


Figure 2. Overview on selected data sources containing information on long-distance travel in Germany (Schulz et al. 2020, p. 70, modified and translated). Since long-distance travel is not typically undertaken on foot or by bike, the coverage of these non-motorized modes is not shown. Notes on abbreviations and references: ADV = Arbeitsgemeinschaft Deutscher Verkehrsflughäfen (ADV 2018), Destatis (2019a); FLE = Fahrleistungserhebung (Bäumer et al. 2014a, 2014b); INVERMO = Zumkeller et al. (2005); MiD = German NHTS “Mobilität in Deutschland” (BMVI 2019); MOP = German Mobility Panel (Ecke et al. 2020); Public transport statistics = Destatis (2019b); RA = Reiseanalyse (FUR 2018); VDR = Verband Deutsches Reisemanagement (VDR 2021); ViZ = Verkehr in Zahlen (BMVI 2021)

The one-day trip diary of the German National Household Travel Survey (*MiD*) for instance (see the upper light red rectangle with dash-dot line) covers everyday mobility by all modes, but only a small part of long-distance travel. In contrast, the *INVERMO* survey (in yellow) equally covers all modes, but only long-distance travel. The admittedly somewhat confusing visualization nicely illustrates the complex task of selecting appropriate data, taking into account

potential overlaps or missing information. In addition to the discussed challenge of applying either the resident-based or the territory-based principle, the definition of long-distance travel varies between data sources. While some surveys explicitly focus on overnight trips only, e.g., the *Travel Analysis* (German: *Reiseanalyse*, FUR (2018)), other data sources include all types of trips regardless of trip duration, e.g., aviation statistics.

The *MiD* dataset is the most comprehensive source of data on the transport demand of the domestic population. Besides the one-day trip diary covering typical everyday trips, the additional journey module addresses multi-day journeys, which frequently involve longer distances. The *MiD* data files consist of more than 150,000 households, 300,000 individuals, 960,000 trips, and 38,000 journeys. The mere sample size in combination with the broad spectrum of questions allows for an equally broad spectrum of in-depth analyses.

This brief review of the available data sources shows that none of them, neither those from the transport sector nor those from a tourism perspective, provide a comprehensive overall picture of long-distance travel. The generation of a consistent picture of long-distance travel demand therefore requires a dedicated methodological approach that brings together central, complementary data bases in a single framework.

### 3. Model-based data fusion

As outlined in the preceding section, *MiD* data are principally well suited to determine long-distance travel. However, given its methodological characteristics and the particular focus on everyday mobility within Germany, the original data require careful interpretation and therefore adequate post-processing in order to select relevant trips and journeys. The data package consists of several individual files (BMVI 2019). Apart from files providing information on private households, household members and household-owned cars, two other files are dedicated to individual travel of household members. Since both files differ in terms of coverage (Table 1), they cannot be easily combined with each other for several reasons. First, they refer to different samples (total population on the one hand, persons aged 14 and over on the other). Second, due to its parallel questionnaire modules (one-day trip-diary vs. three-month journey module), some trips may be included twice, i.e., in both the trip file and the journey file. This is the case when a trip is an outward or return trip of a multi-day journey that began or ended on the individual reporting day. The opposite applies for one-day trips without overnight stay: The associated outward and return trips will not be reported at all if they were not conducted on the individual reporting day. Furthermore, overnight journeys by children below the age of 14 are not covered.

Table 1. Methodological characteristics of *MiD* questionnaire modules

	Trip module	Journey module
<b>Sample</b>	<ul style="list-style-type: none"> <li>all age groups (proxy for children below the age of 14)</li> </ul>	<ul style="list-style-type: none"> <li>adolescents and adults (14 years and above)</li> </ul>
<b>Reporting period</b>	<ul style="list-style-type: none"> <li>one-day trip-diary for individual reporting day</li> </ul>	<ul style="list-style-type: none"> <li>individual retrospective three-month reporting period</li> </ul>
<b>Type of travel</b>	<ul style="list-style-type: none"> <li>everyday trips</li> <li>trips as part of journeys with overnight stay (either outward or return trips)</li> </ul>	<ul style="list-style-type: none"> <li>only journeys with at least one overnight stay</li> </ul>
<b>Travel distance</b>	<ul style="list-style-type: none"> <li>up to 1,000 km (corresponding to the maximum distance to be assumed within Germany)</li> </ul>	<ul style="list-style-type: none"> <li>one-way distance without any limitation</li> </ul>

Separate analysis and subsequent totaling would inevitably result in skewed and – when compared to other data sources – partially implausible key figures for the total travel volume (i.e., the number of trips/journeys) and the associated travel performance. Therefore, a fusion model was developed to obtain a consistent picture of the total travel (including both short- and long-distance travel) of the German population. The basic idea of this approach was to post-process and harmonize original *MiD* data and to reweight and calibrate them using additional external data.

To address the methodological characteristics of the German NHTS, important steps within the post-processing procedure include the exclusion of potentially double-counted outward or return trips, the imputation of journeys conducted by children, and a doubling of the number of journeys (i.e., one journey is converted into a distinct outward and return trip, assuming the two being similar to each other in terms of transport mode and distance).

Overall, the *MiD* survey with its dedicated modules covers the total travel demand of Germans both within and outside Germany and can therefore be regarded as appropriate data basis. However, these survey data have some methodological shortcomings (e.g., overestimation of certain types of long-distance travel such as vacation travel by air, which is still considered as status symbol) and are therefore not directly suitable for accurately deriving information on total travel volumes. This was taken as motivation to reweight and/or recalculate the *MiD* survey outcome. It should be emphasized that, despite its disadvantages, *MiD* data build the fundament of our approach. It only requires additional (external) reference data and key figures to be used for calibration purposes within the fusion model (see below). The modeling process as such is described in more detail by Kuhnimhof et al. (2022). However, in order to determine the long-distance travel of the German population more precisely, other external data were selected for calibration. Which other data sources were used and how they were integrated is described below.

### *Air travel statistics*

German air travel statistics (Destatis 2019a) count every traveler either departing from or arriving at a German airport. As mentioned above, an air-traveler within Germany is counted twice (at the departing airport as well as at the destination). The same holds true for travelers who change planes on an international flight connection (e.g., from Delhi to Lisbon with a stopover in Munich). Although probably not conducted by German citizens, these flights are counted by national air travel statistics. These statistics provide some details about flights (such as origin and destination), but no further information about passengers. To draw conclusions about travelers, we used the outcomes of the regularly conducted *Airport Travel Survey*, which asks air travelers about their origins and final destinations (ADV 2018). In addition, socio-economic characteristics and travel purposes are captured. These data allow a breakdown of the airport statistics figures (which are likely to be valid) by traveler, travel purpose, and type of travel.

Table 2 illustrates the outcome of this data coupling. The total of 117.5 million passengers departing from German airports in 2017 (Destatis 2019) can be broken down into different groups of travelers and – according to the requirements of the data captured in the *MiD* journey module – be assigned to the number of journeys by plane of German residents for calibrating the fusion model.

Table 2. Derived calibration figures based on German *Air Travel Statistics* (Destatis 2019) and the *Airport Travel Survey* (ADV 2018)

Spatial relation of journeys	Type of journey	Number of journeys 2017 (origin in Germany)
within Germany	private journey	1.76 million
	business journey	5.38 million
from Germany to abroad	private journey within Europe	30.51 million
	intercontinental private journey	6.48 million
	business journey within Europe	12.40 million
	intercontinental business journey	3.07 million
Total		59.60 million

### *Vehicle Mileage Survey*

The German *Vehicle Mileage Survey* consists of two parts:

- A survey of car owners in Germany on the use and mileage of their cars and other vehicles, covering both the mileage driven in Germany and abroad. It allows a distinction between vehicle types, but not in terms of mileage by trip length or in geographical terms (Bäumer et al. 2017b).
- A survey of the vehicle mileage performed on German territory. Data are obtained from a sample of traffic counts on the German road network, which are then transferred to the entire German road network. It includes all mileage within Germany, both that of vehicles registered in Germany and that of foreign vehicles (Bäumer et al. 2017a).

A careful analysis of these two complementary surveys provides the percentage of mileage of German cars that is driven abroad; this can then be extrapolated to the total German car fleet and car use.

### Public transport demand data

German operators of rail and urban transit services are required to provide statistical data to the Federal Statistical Office, including both the total number of trips per year and the total number of kilometers traveled by different modes of transport (Destatis 2019b). However, as described above, these data are not free from overlaps, either due to double counts (for instance mode change from bus to rail within one trip, which is typical for public transport) or to an unclear distinction between long-distance rail services (which may include trips below 100 km) and regional rail services (i.e., medium and short-distance services which also may be used for long-distance trips). Only the reported total kilometers can be used here – for urban as well as for long-distance and regional modes. However, these data include both trips and mileage of non-residents. They must be reduced to estimate the volume of the residential demand for urban transit, rail services, and interurban bus services to match the demand represented by the *MiD* survey data. Some simplifications, assumptions and conclusions drawn from other data are necessary, to derive the reference information required:

- The share of travel demand by non-residents, i.e., passenger kilometers traveled primarily by long-distance trains and long-distance buses, for which no additional information about the customers is available, is assumed to be equal to the car use, where respective information could be deduced from the *Vehicle Mileage Survey*.
- The occupancy rate of long-distance car trips is assumed to be higher than for short-distance car trips (Schulz et al. 2020).

Now back to the *MiD* data fusion: In order to obtain a combined person-day dataset, both the trip file and the journey file must be referred to one and the same reporting period first (i.e. 90 days, since the journey module covers 3 months) and then be expanded to an entire year. After merging both files, an iterative weighting procedure takes place based on an iterative proportional fitting approach using several external data sources for calibration as described above. External data include the number of journeys by air, the total mileage of trips either made by car or by collective modes such as public transit, intercity bus/coach services or railway (all of them referring to the residential population). In addition, the new dataset is weighted based on the socio-demographic distribution. The final outcome is a trip file containing ALL trips and ALL journeys of the total residential population in Germany including adapted weighting and extrapolation factors (Figure 3).

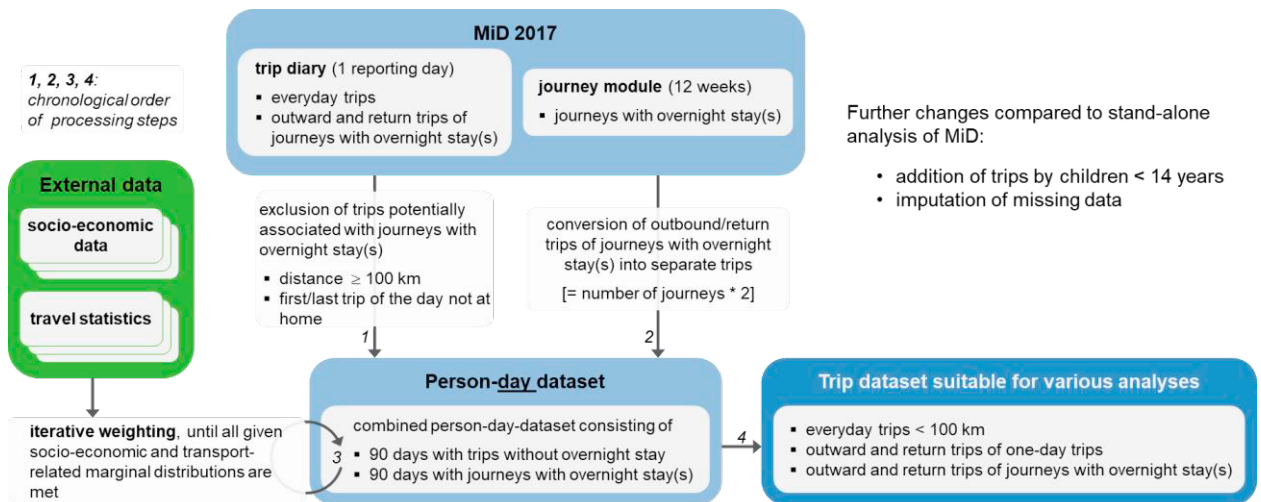


Figure 3. Procedure for the generation of an enhanced *MiD* trip dataset (Schulz et al. 2020, p. 99, modified and translated). Weighting variables: (a) socio-economics: age combined with sex, household size combined with car ownership, spatial typology of place of residence ‚RegioStaR7‘ (BMVI 2020, BMDV 2021), weekday and month of reporting day, (b) transport key figures: passenger car mileage as calculated for the ‘vehicle kilometer traveled model’ used for *Verkehr in Zahlen* (BMVI (2021)), volume and performance of public transport, volume of air travel by type of journey (private/business combined with destination Germany/Europe/Overseas).

The weighted trips are consistent with the statistical volumes and figures of travel demand broken down by modes and corrected by the demand generated by the non-residential population (e.g., foreign tourists traveling in Germany). The final weighting and extrapolation factors reflect socio-demographic characteristics as well as the total travel demand.

Since the trip file is still based on *MiD* data, the entire set of *MiD* variables can be matched using the individual person ID. Hence, the dataset can be used for a broad range of analyses, for example, focusing on different socio-demographic characteristics, different types of travel (e.g., long-distance travel versus everyday mobility), or – based on travel performance in combination with transport modes – potentially resulting emissions. This allows contrasting different population groups to illustrate which socio-demographic characteristics are related to which emissions. At this point it would be possible to go beyond the resident-based principle and also to look at emissions produced abroad. Furthermore, the resulting trip file facilitates the assessment of the effectiveness of measures in terms of which part of the travel market (travel segments such as types of journey, types of modes or travelers with certain characteristics etc.) might be influenced.

#### 4. Results

Both travel volume and travel performance of the German residential population was determined using the fusion model based on *MiD* data, supplemented by other official statistics (see section 3). The model was calibrated using key figures from transport statistics and thus is consistent with these data sources. The resulting travel volumes of everyday and long-distance travel are given in Table 3. It should be noted that long-distance travel is distinguished from everyday travel on the basis of a minimum distance of 100 km. However, the resulting dataset could also be analyzed with other delimitations.

Table 3 shows that the overall travel demand of the residential population in Germany is about 95 billion trips and more than 1,500 billion passenger kilometers traveled. Less than 2 % of the trips have a minimum distance of 100 km. However, in terms of travel performance, such long-distance trips account for 46.3 %. These results underline the importance of long-distance travel, not in terms of number of trips, but in terms of kilometers traveled. Based on the original *MiD* survey data alone, the total mileage per person per year was around 23,000 km, with the proportion of long-distance trips being significantly higher (57 %). This overestimation results – as already discussed above – from the presumable overreporting of long-distance travel, especially air travel. One possible explanation for this over-coverage may be that air travel tends to be well remembered due to its exceptional nature on the one hand, and is still regarded as a status symbol in larger parts of the population on the other. This underlines the relevance of our approach to identify and address such typical errors in conventional surveys.

Table 3. Travel volume and travel performance of the residential population based on the fusion dataset

	Trip distance $\geq$ 100 km (share)	Trip distance < 100 km (share)	Total (share)
<b>Travel volume</b>	1.589 billion trips (1.7%)	93.113 billion trips (98.3%)	94.702 billion trips (100%)
<b>Travel performance</b>	702.386 billion pkt (46.3%)	815.421 billion pkt (53.7%)	1,517.806 billion pkt (100%)

\* pkt = passenger kilometers traveled

Source: own calculation based on the fusion dataset

The breakdown of the overall travel performance into everyday and long-distance travel including the kilometers traveled by each mode is shown in Table 4. It becomes clear that some modes are primarily used for long-distance travel, namely airplane and ship (ferries and cruise ships). However, also motorized individual transport is used for more than a third of all long-distance kilometers. For public transport, about 40 % of the travel performance result from long-distance travel events with a minimum distance of 100 km (one-way).



Table 4. Travel performance of long-distance and everyday mobility differentiated by mode

Travel performance of the residential population	Walking	Cycling	Motorized individual transport	Public transport	Airplane	Ship	Other/ not specified
[billion passenger kilometers]							
				[billion passenger kilometers]			
				[%]			
Long-distance mobility as of 100 km	0.0 0%	1.2 3%	315.3 34%	70.1 40%	290.7 100%	13.8 95%	11.2 43%
Total: 702.4 46.3%							
Everyday mobility up to less than 100 km	35.8 100%	41.6 97%	618.4 66%	104.1 60%	0.0 0%	0.7 5%	14.9 57%
Total: 815.4 53.7%							
<b>Total mobility: 1,517.8</b> <b>100%</b>	<b>35.8</b> <b>100%</b>	<b>42.8</b> <b>100%</b>	<b>933.7</b> <b>100%</b>	<b>174.2</b> <b>100%</b>	<b>290.7</b> <b>100%</b>	<b>14.5</b> <b>100%</b>	<b>26.1</b> <b>100%</b>

Reference year: 2017.

Source: own calculation based on the fusion dataset

The fusion dataset also allows us to distinguish between socio-demographic characteristics of travelers, so that travel volume and travel performance can be analyzed for different groups of the population. One example is given in Figure 4.

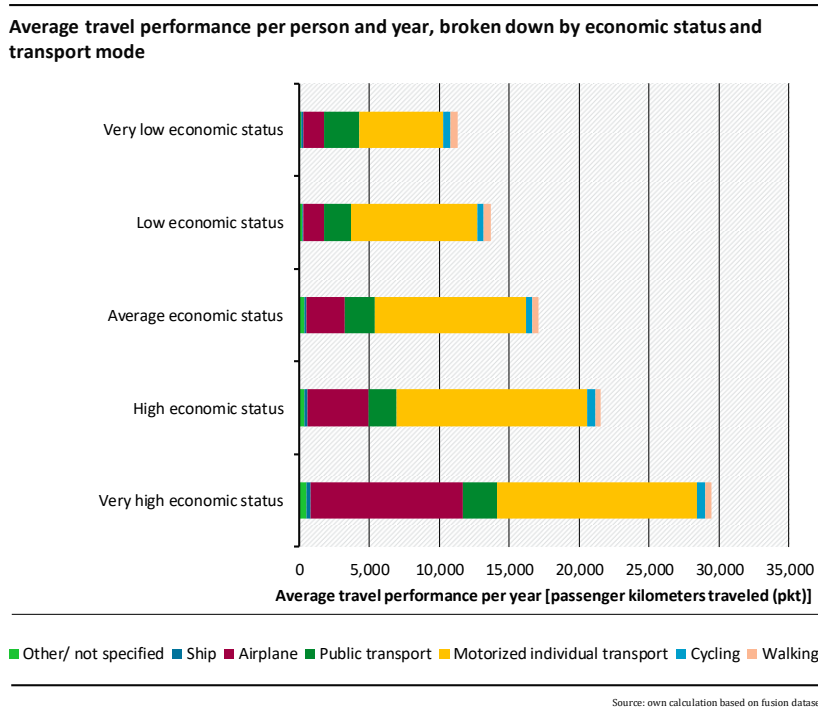


Figure 4. Travel performance per person and year, broken down by economic status and transport mode (analysis of the fusion dataset) (Magdolen et al. (2022b), modified and translated). The *MiD* variable “economic status of a household” was determined using the equivalized income that reflects the differences in a household’s size and composition. Using a matrix of household net income and weighted household size, each household was assigned an economic status ranging from ‘very low’ to ‘very high’ (for threshold values see infas & DLR (2019), part ‘Variablenuaufbereitung’, p. 11).

The overall travel performance per person and year strongly increases with the economic status. Moreover, this increase is mainly due to passenger kilometers traveled by air. These results confirm the heterogenous distribution of long-distance travel in the population already found in the *INVERMO* project (Zumkeller et al. 2005): A rather small group of the population (very high economic status; about 6 million out of 82 million Germans) is disproportionately active in long-distance travel.

With the fusion dataset it is possible to identify characteristics associated with high long-distance travel demand. It facilitates not only the quantification of the total demand, but also the differentiation by mode and trip purpose.

## 5. Conclusion

With respect to the data fusion approach, the following conclusions are drawn:

The added value of the analysis based on the fusion dataset is evident in two respects. First, the resulting figures plausibly describe the mobility of the German population. Deviations from official statistics result from the application of the resident-based principle when collecting data for the NHTS. Official statistics, in contrast, usually follow the territory-based principle and consequently only include travel conducted within Germany's borders. They therefore do not provide any information on the residential population's travel abroad. Second, the fusion dataset is structured as a trip dataset, so that – in addition to other information such as trip purpose – in particular information on the trip distance is available for each individual trip. Thus, it can be analyzed in a flexible and differentiated manner. In our case, for instance, we used the criterion of distance to distinguish between everyday travel and long-distance travel, but any other criterion would have been equally applicable. To the authors' best knowledge, this is the first model capable of dividing the passenger kilometers traveled by the German population for each mode into everyday and long-distance travel in a convenient and flexible way.

The fusion dataset offers yet another important advantage compared to official transport statistics: It allows to distinguish travelers according to their socio-demographic characteristics. This way, both travel volume and travel performance can be analyzed for different groups of the population. Combined with information on which mode of transport is used on each individual trip, resulting emissions can be calculated. Thus, the flexible dataset allows to highlight the unequal distribution of travel-related emissions among the population. Our results show that especially people with a high economic status are responsible for an extremely high travel demand.

However, mere quantification reveals hardly any idea of how travel demand can be influenced, since no background information on the decision-making processes related to long-distance travel is included. Therefore, further research is needed, especially targeted empirical work, to better understand the interrelation between socio-economic characteristics and long-distance travel and, for example, to derive recommendations for influencing travel behavior towards more sustainable travel.

Yet, some limitations and unresolved issues remain: Because the vast majority of data sources continue to use different definitions and delimitations or refer to different populations, some simplifying assumptions are still necessary, e.g., to estimate the travel performance of incoming tourists using public transport, which is also recorded in the German transport statistics. Overall, there is still a huge data gap with regard to the travel of foreign tourists in Germany, but this should also apply to other countries.

In principle, the approach presented is transferable to other countries, provided that some conditions are met. A conventional NHTS should be in place as a baseline, ideally covering not only everyday travel, but also long-distance travel and multi-day journeys with overnight stays. Since NHTS are likely to overrepresent certain types of travel (e.g., long-distance holiday trips by plane) or there is some other type of inevitable measurement error, external statistics are needed that allow for addressing such typical shortcomings of travel surveys. Hence, thorough and complete transport statistics are required to determine the demand for all modes of transport, such as passenger vehicle mileage (either recorded by means of a survey or calculated based on fuel sales), passenger mileage of different collective modes (public transit, intercity bus/coach services, railways) as well as comprehensive air travel statistics. Needless to say, that the external data used must also be thoroughly examined and understood in terms of the definitions, delimitations and assumptions on which they are based, before they are used for calibration purposes.

Among the most difficult may be the quantification of the non-domestic share of international travel or the travel volume of transit passengers. This task is likely to be much easier, especially for large-area countries with a comparatively low share of cross-border travel and/or low transit travel, than for Germany with its rather small spatial

size and with a high level of transit travel. The reliability of the calculated key figures is directly dependent on the quality and complementarity of the input data used. The challenge is a careful selection of data sources, followed by an appropriate application of definitions and threshold values, depending on the underlying research question.

Altogether, we conclude that the data fusion approach produces more reliable results than relying solely on the original survey data, which are in some respects and unavoidably biased by reporting and measurement errors.

## Acknowledgements

This research was supported by funding of two related projects provided by the German Environment Agency: Magdolen et al. (2022a), Schulz et al. (2020). The idea of ‚data fusion‘ was originally developed in the context of modeling passenger transport volume and performance for the statistical pocketbook „Verkehr in Zahlen“ on behalf of the German Federal Ministry of Transport (BMVI 2021).

## References

- ADV (2018): Airport Travel Survey 2018 – Zahlen, Fakten, Trends. Flughafenverband ADV (Arbeitsgemeinschaft Deutscher Verkehrsflughäfen), Berlin.
- Bäumer, M.; Hautzinger, H.; Pfeiffer, M.; Stock, W.; Lenz, B.; Kuhnimhof, T.; Köhler, K. (2017a): Fahrleistungserhebung 2014 "Inlandsfahrleistung und Unfallrisiko". Heft V291. Bericht zum Forschungsprojekt FE 82.0584/2013: Fahrleistungserhebung 2014: Begleitung und Auswertung.
- Bäumer, M.; Hautzinger, H.; Pfeiffer, M.; Stock, W.; Lenz, B.; Kuhnimhof, T.; Köhler, K. (2017b): Fahrleistungserhebung 2014 – „Inländerfahrleistung“. Heft V290. Bericht zum Forschungsprojekt FE 82.0584/2013: Fahrleistungserhebung 2014: Begleitung und Auswertung.
- BMVDV (2021): Regionalstatistische Raumtypologie (RegioStaR). Last update December 2021. German Federal Ministry of Transport and Digital Infrastructure (BMVI). <https://www.bmvi.de/SharedDocs/DE/Artikel/G/regionalstatistische-raumtypologie.html>. For english terminology see BMVI (2020).
- BMVI (2019): MiD 2017 – Mobilität in Deutschland. Dataset B1. Survey on behalf of the German Federal Ministry of Transport and Digital Infrastructure (BMVI). Project homepage: <http://www.mobilitaet-in-deutschland.de>. Access to micro data: <https://daten.clearingstelle-verkehr.de/279/>.
- BMVI (2020): RegioStaR – Regional Statistical Spatial Typology for Mobility and Transport Research. German Federal Ministry of Transport and Digital Infrastructure (BMVI). <https://www.bmvi.de/SharedDocs/DE/Anlage/G/regiostar-raumtypologie-englisch.pdf>
- BMVI (2021) (Ed.): Verkehr in Zahlen 2021/2022. Editor: German Federal Ministry for Digital and Transport (BMDV). Compilation: Deutsches Zentrum für Luft- und Raumfahrt (DLR) and Deutsches Institut für Wirtschaftsforschung (DIW). <https://www.bmvi.de/SharedDocs/DE/Artikel/G/verkehr-in-zahlen.html>.
- Destatis (2019a): Verkehr – Luftverkehr. Statistisches Bundesamt (Destatis; German Federal Statistical Office)). Fachserie 8, Reihe 6.
- Destatis (2019b): Personenverkehr mit Bussen und Bahnen 2017. Statistisches Bundesamt (Destatis; German Federal Statistical Office)). Fachserie 8, Reihe 3.1.
- Ecke, L.; Chlond, B.; Magdolen, M.; Vortisch, P. (2020): Deutsches Mobilitätspanel (MOP)—Wissenschaftliche Begleitung und Auswertungen. Bericht 2019/2020: Alltagsmobilität und Fahrleistung. Project No. (FE) 70.952/19 on behalf of the German Federal Ministry of Transport and Digital Infrastructure (BMVI).
- FUR. (2018): Erste ausgewählte Ergebnisse der 48. Reiseanalyse zur ITB 2018. Forschungsgemeinschaft Urlaub und Reisen e.V. (FUR.), Kiel.
- infas & DLR (2019): MiD 2017 – Mobilität in Deutschland. Nutzerhandbuch. Ausgabe Januar 2019. Anhang 4: Variablenaufbereitung. Survey conducted by infas, DLR, IVT and infas360 on behalf of the German Federal Ministry of Transport and Digital Infrastructure (BMVI). Project No. 70.904/15. [http://www.mobilitaet-in-deutschland.de/pdf/4\\_MiD2017\\_Variablenaufbereitung.zip](http://www.mobilitaet-in-deutschland.de/pdf/4_MiD2017_Variablenaufbereitung.zip)
- Kuhnimhof, T.; Köhler, K.; Eisenmann, Chr.; Kunert, U.; Radke, S. (2022): Data Fusion to Compensate for Insufficient Survey Coverage – Example of the German national PKT-Model. 12<sup>th</sup> ISCTSC Conference on Transport Survey Methods, March 20-25, 2022, Portugal. Further publication in Transportation Research Procedia (forthcoming).

- Magdolen, M.; Chlond, B.; Schulz, A.; Nobis, C.; Jödden, Chr.; Sauer, A.; Führer, M.; Frick, R. (2022a): Handlungsoptionen für eine ökologische Gestaltung der Langstreckenmobilität im Personenverkehr. Abschlussbericht. Project No. (FKZ) 3717 58 102 0 on behalf of the German Environment Agency.
- Magdolen, M.; Chlond, B.; Schulz, A.; Nobis, C.; Jödden, Chr.; Sauer, A.; Führer, M.; Frick, R. (2022b): Options for action for an ecological design of long-distance mobility in passenger and freight transport. Executive summary. Part A – Passenger Transport. Pp. 11-23. Project No. (FKZ) 3717 58 102 0 on behalf of the German Environment Agency.
- Nobis, C. Kuhnimhof, T. (2019): MiD 2017 – Mobilität in Deutschland. Ergebnisbericht. Survey conducted by infas, DLR, IVT and infas360 on behalf of the German Federal Ministry of Transport and Digital Infrastructure (BMVI). Project No. 70.904/15. [http://www.mobilitaet-in-deutschland.de/pdf/MiD2017\\_Ergebnisbericht.pdf](http://www.mobilitaet-in-deutschland.de/pdf/MiD2017_Ergebnisbericht.pdf)
- Schulz, A.; Nobis, C.; Chlond, B.; Magdolen, M.; Chlond, B.; Bergk, F.; Kämper, C.; Knörr, W.; Kräck, J.; Jödden, Chr.; Sauer, A.; Führer, M.; Frick, R. (2020): Klimawirksame Emissionen des deutschen Reiseverkehrs. UBA Texte 141/2020. Project No. (FKZ) 3717 16 108 0 on behalf of the German Environment Agency. [https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2020-07-20\\_texte\\_141-2020\\_emissionen-reiseverkehr\\_0.pdf](https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2020-07-20_texte_141-2020_emissionen-reiseverkehr_0.pdf)
- Schulz, A.; Kuhnimhof, T.; Köhler, K.; Magdolen, M.; Chlond, B. (2019): Current long-distance travel in Germany: How to merge distinct data sources into a meaningful and consistent picture? RGS-IBG Annual International Conference 2019, August 28-30, 2019, London, UK.
- VDR (2021): VDR Business Travel Report 2021 – Management Summary. [https://www.vdr-service.de/fileadmin/services-leistungen/fachmedien/geschaeftsreiseanalyse/VDR-Business-Travel-Report-2021\\_Management-Summary.pdf](https://www.vdr-service.de/fileadmin/services-leistungen/fachmedien/geschaeftsreiseanalyse/VDR-Business-Travel-Report-2021_Management-Summary.pdf)
- Zumkeller, D.; Manz, W.; Last, J.; Chlond, B. (2005): Die intermodale Vernetzung von Personenverkehrsmitteln unter Berücksichtigung der Nutzerbedürfnisse (INVERMO) – Schlussbericht. Projekt im Auftrag des Bundesministeriums für Bildung und Forschung (BMBF).