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

Angelika Schulz, Miriam Magdolen, Bastian Chlond, Tobias Kuhnimhof, Katja Köhler

Workshop B15
Innovations in Travel Survey Collection to Support Long-distance and Tourism Travel Analyses
Long-distance travel: Why should we care?

- Everyday mobility
  - Shopping
  - Leisure
  - Commuting

- Long-distance travel
  - Worldwide living
  - Tourism
  - Multi-local
  - Internationalisation
  - Business travel

- GHG emissions
  - Monitoring task
  - Kyoto Protocol
  - Paris Agreement
  - Emission inventories
  - Influential measures
Long-distance travel: What do we need to know?

**TASK**
Quantitative & qualitative description of
- travel volumes & resulting emissions

**APPROACH**
Comprehensive data base
- transparent & up-to-date
- consistent & reliable
- accessible for research

**CHALLENGE**
- long-distance travel is a very heterogeneous phenomenon
- no “all-in-one” source of data that explains everything

**SOLUTION**
Combined data base of several complementary sources (survey data, official statistics)
Long-distance travel: How to define and delimit?

- **purposes** range from activities carried out almost every day or only once a year (if at all)
- **distances** vary as well, and respective threshold values are set according to the research question
- in terms of **duration**, there are both one-day trips as well as journeys with overnight stays

Source: KUHNIMHOF & SCHULZ, in NOBIS & SCHULZ (2017)
Long-distance travel: How to define and delimit?

- **purposes** range from activities carried out almost every day or only once a year (if at all)

- **distances** vary as well, and respective threshold values are set according to the research question

- in terms of **duration** there are one-day trips as well as journeys with overnight stays

Source: KUHNIMHOF & SCHULZ, in NOBIS & SCHULZ (2017)
Available data sources: Example of Germany

- data landscape resembles a puzzle or a toolbox:
  - for each area of travel there are dedicated data sources

- BUT: hardly no information about individual background of long-distance travelers:
  - motives
  - preferences
  - attitudes
  - travel-related decision making

Source: SCHULZ et al. (2020), modified and translated
Coverage of long-distance travel: Methodological ‘shortcomings’ of data sources

**Surveys on everyday mobility**
- under- or over-coverage due to survey design
- one-day trip diaries vs. retrospective, multi-month journey module
- missing trips (only outward/return trip of a multi-day journey)
- potential double counting (both in trip diary as well as in journey module)
- incomplete mileage due to distance thresholds (focus on domestic travel)
- selectivity due to sampling (highly active people are hard-to-reach)

**Surveys on long-distance travel**
- under-coverage due to recall effect (retrospective reporting)
- maybe biased reporting (only ‘interesting’ or extraordinary trips are reported)
- selectivity due to sampling (highly active people are hard-to-reach)

**Official public transport statistics**
- no distinction between residential population, incoming tourists & transit passengers
- potential double counting in case of intermodal trips/journeys

**Data from companies and service providers**
- particular coverage and/or selectivity due to commercial purpose of data collection
- limited and/or costly access to data (confidential business data)
Different coverage according to territorial or national principle

- **Territorial principle**
  - trips conducted within the country by residential population and incoming tourists

- **National principle**
  - trips conducted by residential population inside and outside the country

→ German NHTS “Mobility in Germany (MiD)
Central data source for Germany: National Household Travel Survey (MiD)

- coverage of residential population only (national principle)
- focus on domestic travel (distance threshold 1,000 km)
- dedicated survey instruments for …
  - everyday trips
    (one-day trip diary, inside Germany)
  - journeys with overnight stay(s)
    (up to 3 journeys, 3 months retrospectively, no spatial limitation)
- few information on trips abroad (outgoing travel)
- no information on
  - incoming tourists
  - transit passengers
Combined MiD file consisting of trips & journeys ('fusion dataset')

**MiD 2017**

- **trip diary** (1 reporting day)
  - everyday trips
  - outward and return trips of journeys with overnight stay(s)

- **journey module** (12 weeks)
  - journeys with overnight stay(s)

**Further changes compared to stand-alone analysis of MiD:**

- addition of trips by children < 14 years
- imputation of missing data

**External data**

- **socio-economic data**
- **travel statistics**

**Person-day dataset**

- combined person-day-dataset consisting of
  - 90 days with trips without overnight stay
  - 90 days with journeys with overnight stay(s)

**Trip dataset suitable for various analyses**

- everyday trips < 100 km
- outward and return trips of one-day trips
- outward and return trips of journeys with overnight stay(s)

**Source:** SCHULZ et al. (2020), modified and translated
Fusion model: Use of various additional data sources for calibration

**Statistical pocketbook**
“Verkehr in Zahlen 2018/2019” (ViZ))
- Passenger travel volume model

**Vehicle mileage survey 2014 (FLE)**
- MIV travel performance in Germany*
- MIV travel performance of residents**

**Official public transport statistics**
- Transport demand
- Transport performance

**Official aviation statistics**
- Number of passengers

MiD 2017

* Territorial principle
** National principle

* Travel of residents and non-residents; ** Travel in Germany and abroad

Source: own illustration based on SCHULZ et al. (2020)
Travel volumes ‘before & after’: Improved key figures after data fusion

**Analysis based on original MiD data**
- extrapolation of trips/journeys reported in …
  - one-day travel diary and
  - 3-month journey diary

> 960,000 trips and > 38,000 journeys

- overall annual travel volume

23,000 km (~ 14,300 miles)

per person per year

- share of long-distance travel (≥ 100 km)

57 %

**… based on enhanced MiD fusion data**
- correction of overlaps
- imputation of missing data
- iterative re-weighting procedures

> 995,000 trips

- overall annual travel volume

18,400 km (~ 11,400 miles)

per person per year

- share of long-distance travel (≥ 100 km)

46 %
Touristic and long-distance travel of the German residential population: 2017 key figures based on different definitions

**Travel volume**
Total: 94.7 billion journeys/ trips
- outside usual environment/ ≥ 100 km
- inside usual environment/ < 100 km

**Travel performance**
Total: 1,517.8 billion passenger kilometers
- outside usual environment/ ≥ 100 km
- inside usual environment/ < 100 km

---

(Touristic) travel events
Journey = destination „outside the usual environment“ (UNWTO definition)
- 51% inside usual environment
- 49% outside usual environment
- 8% total travel events

Long-distance travel
Long distance = trip distance ≥ 100 km
- 46% inside usual environment
- 54% outside usual environment
- 2% total long-distance travel

Source: own calculation and illustration based on SCHULZ et al. (2020) and MAGDOLEN et al. (in print)
Average travel performance of different groups of people

**Verkehrsleistung pro Person und Jahr**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gesamt</td>
<td>1.517.8</td>
<td>1.517.8</td>
</tr>
<tr>
<td>Kreisfreie Großstadt</td>
<td>9.621</td>
<td>9.621</td>
</tr>
<tr>
<td>Städtischer Kreis</td>
<td>8.341</td>
<td>8.341</td>
</tr>
<tr>
<td>Ländl. Kreis mit Verdichtungsansätzen</td>
<td>7.561</td>
<td>7.561</td>
</tr>
<tr>
<td>Dünn besiedelter ländlicher Kreis</td>
<td>8.011</td>
<td>8.011</td>
</tr>
<tr>
<td>Erwerbstätiger HH mit Kindern*</td>
<td>7.046</td>
<td>7.046</td>
</tr>
<tr>
<td>Erwerbstätiger HH ohne Kinder*</td>
<td>11.988</td>
<td>11.988</td>
</tr>
</tbody>
</table>

**Weiblich**

- Männlich: 8.498
- (noch) kein Abschluss: 8.498
- Volks- oder Hauptschule: 8.777
- Realschulabschluss: 8.267
- Abitur, Fachhochschulreife: 9.356
- Fachhochschul- oder Universitätsabschluss: 10.385
- Sonstiges und keine Angabe: 9.356

**Verkehrsleistung pro Person und Jahr**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gesamt</td>
<td>1.517.8</td>
<td>1.517.8</td>
</tr>
<tr>
<td>0-13-Jährige</td>
<td>5.647</td>
<td>9.021</td>
</tr>
<tr>
<td>14-18-Jährige</td>
<td>4.266</td>
<td>7.986</td>
</tr>
<tr>
<td>19-24-Jährige</td>
<td>9.319</td>
<td>10.801</td>
</tr>
<tr>
<td>25-34-Jährige</td>
<td>12.425</td>
<td>12.335</td>
</tr>
<tr>
<td>35-44-Jährige</td>
<td>11.131</td>
<td>12.725</td>
</tr>
<tr>
<td>45-54-Jährige</td>
<td>11.380</td>
<td>12.003</td>
</tr>
<tr>
<td>55-64-Jährige</td>
<td>10.926</td>
<td>10.971</td>
</tr>
<tr>
<td>65-74-Jährige</td>
<td>6.335</td>
<td>6.773</td>
</tr>
<tr>
<td>75+ Jährige</td>
<td>5.108</td>
<td>5.058</td>
</tr>
</tbody>
</table>

**Verkehrsleistung in km pro Person und Jahr**

- Wege ≥ 100 km
- Wege < 100 km

**Verkehrsleistung in km pro Person und Jahr**

- Wege ≥ 100 km
- Wege < 100 km

**Quelle:** Eigene Berechnung auf Basis des Fusionsdatensatzes

---

**Average travel performance of different groups of people**

<table>
<thead>
<tr>
<th>Segment</th>
<th>Verkehrsleistung in km pro Person und Jahr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gesamt</td>
<td>5.000 10.000 15.000 20.000 25.000 30.000 35.000</td>
</tr>
<tr>
<td>Kreisfreie Großstadt</td>
<td>9.621 8.630 10.000</td>
</tr>
<tr>
<td>Städtischer Kreis</td>
<td>8.341 10.000</td>
</tr>
<tr>
<td>Dünn besiedelter ländlicher Kreis</td>
<td>8.011 11.229</td>
</tr>
<tr>
<td>Erwerbstätiger HH mit Kindern*</td>
<td>7.046 2.344</td>
</tr>
<tr>
<td>Erwerbstätiger HH ohne Kinder*</td>
<td>11.988 12.996</td>
</tr>
<tr>
<td>Nicht-Erwerbstätiger HH*</td>
<td>8.498 6.127</td>
</tr>
<tr>
<td>Sehr niedriger ökonomischer Status</td>
<td>8.320 7.857</td>
</tr>
<tr>
<td>Niedriger ökonomischer Status</td>
<td>8.414 8.814</td>
</tr>
<tr>
<td>Mittlerer ökonomischer Status</td>
<td>7.986 9.674</td>
</tr>
<tr>
<td>Hoher ökonomischer Status</td>
<td>10.497 11.105</td>
</tr>
<tr>
<td>Sehr hoher ökonomischer Status</td>
<td>18.498 10.928</td>
</tr>
<tr>
<td>Gesamt</td>
<td>1.517.8</td>
</tr>
<tr>
<td>0-13-Jährige</td>
<td>5.647 9.021</td>
</tr>
<tr>
<td>14-18-Jährige</td>
<td>4.266 7.986</td>
</tr>
<tr>
<td>19-24-Jährige</td>
<td>9.319 10.801</td>
</tr>
<tr>
<td>25-34-Jährige</td>
<td>12.425 12.335</td>
</tr>
<tr>
<td>35-44-Jährige</td>
<td>11.131 12.725</td>
</tr>
<tr>
<td>45-54-Jährige</td>
<td>11.380 12.003</td>
</tr>
<tr>
<td>55-64-Jährige</td>
<td>10.926 10.971</td>
</tr>
<tr>
<td>65-74-Jährige</td>
<td>6.335 6.773</td>
</tr>
<tr>
<td>75+ Jährige</td>
<td>5.108 5.058</td>
</tr>
</tbody>
</table>

**Quelle:** Eigene Berechnung auf Basis des Fusionsdatensatzes

---

**Schulz, Magdolen, Chlond, Kuhnimhof, Köhler: Identifying and understanding long-distance travel demand by combining official transport statistics and survey data**

**ISCTSC 2022 – Workshop B15**

**Chart 14**

- employed HH without children
- very high economic status
- high level of education
- intermediate age groups
- male
Complementary empirical work ‘in brief’

- **Net sample:** 1,002 individuals aged 18+
- **Recruiting:** online panel with screening (journeys ≥ 100 km within last 2 months, total n° of journeys during last year)
- **Composition of sample:** disproportionate inclusion of frequent travelers (50%) with at least 10 trips ≥ 100 km
  - allows for targeted analyses
- **Data collection:** Spring (5/23-3/6/2019) & Fall (9/3-15/9/2019)
  - taking seasonality into account
- **Length of interview:** 21.9 minutes (median)
- **Survey instrument:** CAWI

### Chart 15

- **Socio-demographics (Part 1)**
- **One-year summary of travel events**
- **Details for up to 2 travel events**
- **Overall travel behavior and attitudes towards travel**
- **Socio-demographics (Part 2)**
Complementary empirical work:
Focus on motivation, decision making and behavioral dimensions

- **Parameters of decision**
  - Destination(s) (including accessibility)
  - Trip purpose
  - Travel date
  - Availability of transport modes
  - Special requirements (e.g. luggage, ...)

- **Parameters of the assessment**
  - Duration (travel time)
  - Expenses
  - Subjective convenience (organization of the trip, trip itself, ...)

- **Context of decision**
  - Who plans and decides?
  - Flexibility in terms of destination, time and mode choice

- **Socio-demographics**
  - Household and individual characteristics (e.g., household structure, vehicle ownership, income, driver’s license, occupational status, etc.)
Private journeys: Reason for traveling ‘just like this’

- Almost 80% of the journeys are linked to the destination regions
  - personal contacts
  - fixed appointments
  - explicit touristic interest

- Monetary considerations such as bargains play a subordinate role.

► limited options to influence travel behavior by any measures
Methodological challenges

Modeling approach …

• requires detailed (methodological) knowledge about genuine input data as well as data used for calibration
• underlying definitions (e.g. distance threshold or overnight stay)
• population addressed (e.g. residential population only vs. all travelers, age groups etc.)
• spatial coverage (e.g. domestic vs. international travel)

• if different data sources are used in an appropriate way, they will complement each other to a more comprehensive picture
• relying on one single data source might lead to ‘wrong’ results and misleading interpretation

• informed assumptions based on ‘external’ sources will continue to be necessary
Unresolved issues

• depending on research question and applied definition, the actual volume of long-distance travel may be incomplete

• travel volumes of incoming tourists and transit passengers remain unknown
  ▶ subject to assumptions

• ‘shortcomings’ of traditional household travel surveys with respect to journeys

• no detailed information on trip stages and respective transport modes (only main mode)

• in particular with respect to overnight journeys:
  no information on trips conducted at the destination (local mobility)
Conclusion

• combination of different data sources based on a ‘fusion model’ …
  • reduces the ‘shortcomings’ of individual data sources
  • helps to get a consistent picture, provided that initially heterogeneous input and calibration data are adequately post-processed and harmonized

• resulting data set …
  • covers the travel volume of the German residential population
  • allows for flexible and multi-purpose data analyses, e.g. to identify and analyze …
    • different socio-economic groups of travelers
    • different types of trips and journeys
    • modal share
References


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

Angelika Schulz, Miriam Magdolen, Bastian Chlond, Tobias Kuhnimhof, Katja Köhler

Workshop B15
Innovations in Travel Survey Collection to Support Long-distance and Tourism Travel Analyses

Corresponding author: angelika.schulz@dlr.de