Enriching household travel survey data: Experiences from the Microcensus 2000

VS Chalasani, IVT, ETH Zürich

Conference paper STRC 2005

STRC | 5th Swiss Transport Research Conference
Monte Verità / Ascona, March 9-11, 2005

Enriching household travel survey data: Experiences from the Microcensus 2000

VS Chalasani
IVT,
ETH Zürich

Phone:+41 1 633 33 40
Fax: +41 1 633 10 57
email: chalasani@ivt.baug.ethz.ch

March 2005

Abstract

The data collected from any travel survey is not complete for travel demand modelling. One of the potential and prominent ways of procuring reliable and most complete data is collect a porous structured data and later enrich it accordingly. It is highly difficult to design the survey instrument to collect the most important data in a porous structure. It is also important to ensure that the deliberately excluded data can easily be imputed either form the collected information or using the information from other sources. The 2000 travel Microcensus data was enriched in this study for stage imputation, adjustment of arrival and departure times, location geo-coding, main mode of transportation, travel distances, home based trips, location type, accessibility, travel costs and nearest public transport stops

Results obtained are: Around 48% of the missing walk stages were imputed. The two criteria employed in the trip main mode calculations gave a similar modal split. Thus the assumed modal hierarchy is consistent. Comparison of travel distances showed that the reported distance fall between the shortest distance path distances and user equilibrium distances. User equilibrium distances and shortest time path distances follow closely. This study proves that the household travel survey data can be enriched in many aspects using both the collected information and external resources.

Keywords

2000 travel Microcensus, Stage imputation, travel time, travel distance, travel costs, accessibility, geo-coding, enrichments

1. Household travel surveys

Traditionally, transport models have been largely used to predict or assess transport demand and the characteristics of the transport system. Exogenous developments in society have caused increased travel needs and more complex travel behaviour. New and innovative policies were developed to cater society's growing travel needs. Continuous changes in policy, theory development and modelling in general have led to the development of new classes of transport demand models that have in common their increased complexity. From the conventional four-stage methodology involving the prediction of transport demand generation, destination choice, mode choice and route choice, the developments in transport modelling have paved the way for formulation of discrete choice models and activity based models. Reliable data representing the ground realities is essential to evaluate the theoretical concepts of travel demand models. Travel data is mainly obtained from travel surveys, including a wide range of instruments from relatively simple roadside surveys and onboard public transport surveys to complex surveys such as travel and activity dairies. But, developments in the increasingly complex travel demand models seek new kinds of data. For instance, data on activities (purpose, location, time frame, etc.) is essential for activity-based demand models. Innovative data collection methods were implemented to collect reliable and supporting data. In addition to the general purpose travel surveys, valid specialized surveys at different levels of measurement to deal with specific problems are becoming more common. Travel surveys should be selective because substantial amounts of resources are involved in conducting such surveys. It would be interesting to study the usability and reliability of the general purpose travel survey data for different types of travel demand models.

The observable stream of acts must be divided into a sequence of distinct elements which can be counted and characterised in analysis. This study adapts the following division of movements (Axhausen, 2000):

- "A **stage** is a movement with one vehicle (as driver/rider or passenger), or on foot. It includes pure waiting (idle) times immediately before or during that movement.
- A **trip** is a continuous sequence of stages between two activities."

1.1 Data quality and survey instrument optimization

Generally, data collected from surveys is deficient in many aspects due to various errors in survey methodology. The sources of errors can be classified into four different types (Groves, 1989): sampling error, coverage error, non-response error, and measurement error. The errors sampling, coverage and item non-response occur due to improper survey design. Where as, the item non-response and measurement errors result either the theoretically correct survey does not suits to the universe or the survey theme was improperly communicated to the respondents. This study primarily deals with measurement error and ways to improve the affected data.

Survey instrument plays a vital role in colleting the reliable data. In travel surveys, the survey instrument mostly comprises of a set of well defined questionnaires such as household questionnaire, personal questionnaire, daily mobility questionnaire, long journeys, etc. Aimed to collect the comprehensive information, these questionnaires follow in sequence and can also be applied in multiple phases of the survey. Length of the questionnaire depends on the amount of information seeking from the respondents in that particular context. Another important issue that adds complexity is defining the terms (both travel and non-travel). Aimed mobility is usually classified either as pre-defined terms such as trip, stage, etc. (see Axhausen, 2000) or specifically defined for the survey such as excursions, long-distance journeys, etc. Lengthy questionnaire or measuring various movements at different levels simultaneously are the two potential causes of respondent burden. Travel data deficiencies that occur due to respondent burden can be classified as:

1. Missing a part or a full movement (e.g.: missing of one or more stages of a movement which may break the continuity)
2. No answer to a particular question (e.g.: unanswered independent parameters)
3. Vague answers (e.g.: unrealistic and highly variable reported values)

These deficiencies influence the data quality and question its applicability for a wide range of travel demand models. One way to minimize the measurement errors thus improve the data quality is collection of the partial and most critical information. Respondents' burden can substantially be reduced by omitting less important and easily imputable movements i.e. collection of porous information that can be filled with post-analysis. Efficiency of this technique merely depends on derivation of the rule to separate critical movements from the rest and its effective implementation in data collection.

Assessment of data quality in travel surveys is difficult, simply because the number and the characteristics of the trips actually made ("true mobility") cannot be known. There is a growing body of literature that makes data quality as one of the key influencing factors in the development of travel demand models (Kalfs, Meurs, and Saris, 1997; Arentze, Timmermans, Hofman and Kalfs, 1997). Mostly, a part of the data collected from a survey is used in travel demand models. In such circumstances, the quality of selected data is of interest, which might be different from overall data quality. But, the inter-dependency among different variables is difficult to assess. In general it can be concluded that assessment of data quality is purely case specific. However, certain primary indicators can be derived using the rules or assumptions prior to the survey. For instance, number of stages per vehicle based trip should be a minimum of three and a round-trip should have at least two trips. These primary indicators can be used for a preliminary assessment of data quality and the identification of the records in need of correction.

Taking the number of stages per trip as a primary indicator, this study enriches the Swiss household travel survey data. The Microcensus 2000 travel survey is considered in this study. Each stage with a non-walk mode of transport precedes and follows from a walk stage. Accepting this fact and relying on the reported information, this study identifies missing walk stages and imputes the non-reported information. The study also analyses the advantages in enriching the Swiss household travel survey data.

This report combines various enrichments that were performed to the 2000 travel Microcensus at IVT. Most of the documentation has been heavily adopted from different IVT working papers. The structure of this report is: A brief note on the 2000 travel Microcensus will be described in the Chapter 2. Chapter 3 will discuss the deficiencies in and the possible enrichments to the targeted data. All the enrichment process will be discussed in detail in Chapter 4. Conclusions are given in the Chapter 5.

2. Travel Microcensus 2000

The Microcensus is a national household travel survey series that has been conducted every five years by the Swiss Federal Office for Spatial Development (ARE) and the Swiss Federal Statistical Office (BFS) since 1974. The 2000 travel Microcensus is the latest and sixth study in this series that started in 1974, and was conducted in 1979, 1984, 1989, and 1994 (ARE and BFS, 2001; 2002). The 2000 Microcensus incorporated state-of-the-practice survey methods, including pre-notification letters, telephone recruiting, and computer aided telephone interviewing (CATI) to aid in the data collection. The 2000 Microcensus was conducted seven days a week, including all holidays, from January 2000 to December 2000. Survey data were collected from a sample of 29,407 persons from 27,917 households making over 100,000 daily trips. Both urban and rural areas were included. All trips made during a pre-assigned 24-hour travel day by each target person aged six years and above in the sampled households were collected in the survey. In addition to the socio-economic characteristics of the households and the personal characteristics of the targeted persons, following information on travel was collected in the 2000 travel Microcensus:

- Within-day travel: Short stages by all modes below 100m were omitted from reporting. Origin and destination, time of departure and arrival time, and travel duration were collected both at trip level and stage level. Both the trip purpose and in-vehicle travel time were collected at trip level. Mode of transport for each stage and the stage purpose were collected at stage level.
- Multi-day journeys: Respondents were asked to report on the multi-day journeys or journeys with overnight stays that were made in the last three months before the day of reporting.
- Air travel: Considering their rarity of occurrence, brief information on such trips was collected from the target persons. Frequency, origin and destination, and the purpose of these that were made in last one year, 5 years and 10 years were simultaneously recorded from the target persons

Except for the sample weighting, no imputations were performed on the 2000 travel Microcensus data. This study only considers same-day trips for enrichment.

3. Need for household travel survey data enrichments

Household travel survey data enrichments are performed for two primary purposes. First, to mitigate the deficiencies in the collected data caused by unit non-response or intentionally made information pores. This enrichment process uses only the collected data. Second, to impute information collected from other sources. These sources can be transport surveys or other related studies. Using data from external sources and collected data these enrichments not only improve the data quality but also enhance its applications. This study applied both approaches to enrich the 2000 travel Microcensus data.

3.1 Deficiencies in travel Microcensus 2000 data

Although travel Microcensus 2000 efficiently covered structured information, it suffers from the following deficiencies:

- Erroneous stage-sequences: For the trips with at least one non-walk mode stage, the intermediate walk stages were not reported. This results in a series of erroneous stage-sequences
- Inconsistent address information of origin and destination locations: Reported origin and destination location’s address was not verified online. Irrespective of the travel purpose and mode of transport used, reported location addresses were found inconsistent.
- Discrepancy in the reported travel distances: Reported travel distances (both at stage as well as at trip level) reflect the respondent’s ability to estimate the distance travelled. It is hard to distinguish the correctness of these distances.

3.2 Enrichments to the 2000 travel Microcensus

Following enrichments were performed at IVT:

- Imputation of missing walk stages
- Adjustment of arrival and departure times of the modified stage-sequences
- Geo-coding of all the household, origin and destination locations.
- Computation of network distances and crow-fly distances and compare with the reported distances.
- Update the origin and destination location addresses with the household address information.
- Import the accessibility index for each municipality and travel costs for all car trips.

All these enrichment processes are discussed in the next section.

4. Enrichments to the 2000 travel Microcensus

4.1 Stage imputation

The decision to capture independent travel exceeding 100m caused one of the primary deficiencies in the 2000 travel Microcensus. An attempt was made to identify and impute the missing stages. One important assumption in stage imputation process is that the respondents reported all non-walk stages and only walk stages were ignored while reporting.

4.1.1 Stage sequence correction

The stage sequence is the sequence of consecutive stages travelled in a particular trip. Among the various characteristics of a stage, this study selected the mode of transport to describe a stage. For instance, stage-sequence of a three stage trip can be specified as "walk – car – walk", where car is the trip's main mode of transport. Logically each non-walk stage is preceded and succeeded by a walk stage: This can be mathematically represented in the following way:

Number of stages in a trip = (2 x Number of non-walk stages made in the trip) + 1

Generally, walk stages are very short both in duration and length. Their reporting is burdensome for the interviewer and the respondent, which leads potentially to non-response and skipped details. In the Microcensus 2000 short stages below 100m length were omitted by design.

Stage-sequences of all the reported trips in the travel Microcensus 2000 were checked against the theoretically necessary number of stages. Based on the reported stage-sequence, the reported trips were classified as the following:

- Single stage walk trips (e.g. Walk)
- Trips with a correctly reported stage-sequence (e.g. walk – car – walk)
- Trips with an incorrectly reported stage-sequence (e.g. bus – train – bus)
- Trips with illogical stage-sequence (e.g. walk – walk)

Single stage walk trips and trips with a correct stage-sequence were excluded from stage-sequence corrections. Trips with illogical stage-sequence were ignored from stage-sequence corrections and further analysis. After correcting the erroneous stage-sequence trips, both the correctly reported and corrected trips were considered for analysis. Table 1 summarizes these results:

Table 1 Number of stages by category of trips reported in the Microcensus 2000

Description	Number of trips	Percent of trips	Number of stages	Percent of stages	Number of stages/trip
Single stage walk	28,010	26.43	28,010	19.05	1.00
Correctly reported stage sequence	5,146	4.86	16,936	11.52	3.30
Incorrectly reported stage sequence	71,406	67.37	97,957	66.61	1.37
Illogically reported stage sequence	1,417	1.34	4,155	2.83	2.93
All	105,979	100.00	147,058	100.00	1.39

Source: Chalasani and Axhausen, 2004

Two-thirds of the reported trips report an incorrect stage-sequence. Excluding the single stage walk trips, only 5% of the reported multi-stage trips have a correct stage-sequence. The average stage frequency of the trips reported with a correct stage-sequence was larger than that of the incorrect stage-sequence trips, which is the motivating factor for this study. The twenty most frequent reported stage sequences are shown in Table 2.

Table 2 Reported stage-sequence of the same-day trips before correction

Stage sequence	Number of stages	Frequency	Share [%]
Car as driver	1	30281	29.29
Walk*	1	27967	27.05
Car as passenger	1	9090	8.79
Moped	1	6176	5.97
Car as driver – Walk	2	2939	2.84
Walk – Car as driver	2	2726	2.64
Walk – Bus – Walk*	3	1867	1.81
Car as driver – Car as driver	2	1653	1.60
Bus	1	1197	1.16
Motorcycle as driver	1	1067	1.03
Walk – Train – Walk*	3	925	0.89
Car as passenger – Walk	2	830	0.80
Walk – Tram – Walk*	3	818	0.79
Walk – Car as passenger	2	794	0.77
Walk – Walk*	2	758	0.73
Other	1	747	0.72
Moped	1	658	0.64
Walk – Car as driver – Walk*	3	639	0.62
Walk – Bus	2	545	0.53
Bus – Walk	2	529	0.51
Rest	-	11170	10.81

*: Excluded from stages imputation

Source: Chalasani and Axhausen, 2004

Table 2 reveals that nearly half of the reported trips do not contain any walk stages. Excluding the single stage walk trips, only one-fifth of the reported trips contain one or more walk stages. In total 1631 stage-sequence types were reported, but the share of the twenty most frequent stage-sequences is nearly 90%. By adding possible missed walk stages, all the incorrect stage-sequences were corrected. Table 3 shows the most frequent corrected stage-sequences and their distribution by trip purpose.

Table 3 Distributions of trip purposes for the most frequent stage-sequence trips

Corrected stage sequence+	Wo-rk	Educ-ation	Shop-ping	Bus-iness	Leisu-re	Servi-ce trip	Escor-ting	No answer	Total
1_6_1	33.9	1.0	17.9	6.1	32.0	6.2	2.5	0.4	29.29
1**	13.6	14.4	21.6	1.7	44.6	1.5	1.3	1.2	27.05
1_7_1	10.1	4.8	18.8	2.2	57.8	3.4	2.1	0.6	8.79
1_2_1	23.9	18.3	17.7	1.5	37.1	0.9	0.2	0.4	5.97
1_6_1	25.5	1.2	21.5	5.8	41.7	3.3	0.6	0.3	2.84
1_6_1	27.5	1.3	21.7	4.3	36.4	7.5	1.0	0.3	2.64
1_10_1**	27.2	15.9	21.0	1.4	32.4	0.8	0.4	0.2	1.81
1_6_1_6_1	19.3	0	26.3	9.6	30.9	8.8	2.7	2.5	1.60
1_4_1	24.7	27	18.7	1.5	26.6	0.5	0.2	0.3	1.16
1_10_1	45.9	4.9	8.2	4.1	33.8	1.2	1.4	0.3	1.03
1_7_1	38.3	12.9	16.6	2.7	27.8	0.2	0.2	0.4	0.89
1_7_1	6.6	3.5	19.7	0.8	67.0	1.3	0.8	0.3	0.80
1_8_1	26.1	7.7	27.3	1.0	36.4	0.4	0.7	0.4	0.79
1_11_1**	6.6	2.9	20.8	3.5	60.4	3.9	1.1	0.7	0.77
1_90_1	5.1	3.1	38.5	2.1	46.8	2.1	0.4	1.8	0.73
1_3_1	17.2	15.7	8.5	18.6	37.1	0.3	0.5	2.0	0.72
1_1*	31.4	14.5	13.9	1.9	37.3	0.7	0.3	0.0	0.64
1_6_1	29.2	1.1	20.9	4.6	40.0	3.8	0.3	0.0	0.62
1_23_1	22	26.7	17.8	1.4	30.9	0.4	0.2	0.1	0.53
1_10_1	21.3	20.1	22	0.7	34	0.4	1.1	0.3	0.51
Rest	26.3	12.1	14.1	4.5	36.7	1.2	0.6	4.2	10.81

1	Walk	6	Car as driver	11	Tram
2	Bike	7	Car as passenger	23	Small moped
3	Moped	8	Train	90	Others
4	Motorcycle as driver	10	Bus		

*: Illogically reported, excluded from stage-sequence correction

**: Correctly reported stage sequence, excluded from stage-sequence correction

Missing walk stages were more frequent in trips for purposes leisure, shopping and work. Paying more attention on these trips in forthcoming surveys would be a worthwhile.

4.1.2 Imputation of stage attributes

Among the various stage characteristics, only stage number and mode used can be transported directly from the corrected stage-sequences. Other important and essential stage characteristics are origin and destination, arrival and departure time, and distance travelled. This section deals with determining origin and destination, and imputation of missing walk stage distances. Various factors such as, land-use patterns, road network and public transport characteristics, influence the connecting walk distances between the non-walk modes of transport. Based on the position in the stage-sequence, three categories of walk stages were defined:

- First walk stage positioned before first non-walk stage reported
- Intermediate walk stage between two reported non-walk stages, and
- Last walk stage immediately after the last non-walk stage reported.

As stated earlier, walk stages generally are short both in duration and length. In the 2000 Microcensus, respondents were asked to report the postal address (zip code, street name and house number) of each stage origin and destination. Though stage origin and destination zip codes were fully reported, for only 25% stages valid origin and destination addresses were collected. Corresponding municipality numbers are assigned to all stage origin and destination zip codes in the official datasets. Missing walk stage origins and destinations were imputed employing the following assumptions:

- Imputed intermediate walk stages originate at the previous non-walk stage end location and end at immediate next non-walk stage start location.
- An imputed first walk stage ends at the first non-walk stage start location. Origin address of the reported first non-walk stage, which is eventually the corresponding reported trip origin address, was compared with the home address. If the reported first non-walk stage starts at home, then the imputed walk stage originates from home and ends within the same municipality. Only the municipality number was assigned as the first non-walk stage origin address. In case of a non-home based start location reported for the first non-walk stage, imputed walk stage originates in the same municipality and ends at the start location of the reported first non-walk stage start location.
- An imputed last walk stage originates at the previous non-walk stage end location. Similar to the imputed first walk stage, if the reported last non-walk stage's end location is home then the imputed last walk stage end at home and originates in the same municipality. Only the municipality number is retained for the reported last non-walk stage's end location. If the reported last non-walk stage end location differs from home, then the imputed last walk stage ends within the same municipality, but without any specific location.

Most of the imputed walk stages originate and terminate in the same municipality and at unknown origin and destination addresses. Partially reported stage locations and intra-municipal imputed walk stages made the imputed walk stage distance calculations more complex. Geo-codes at different levels (postal address, street name, and post office) were calculated for all the reported stage locations (Jermann, 2003). Crow-fly distance is a distance that can be calculated for stages with valid origin and destination geo-codes. An estimated of the actual distance travelled can be calculated using a pre-defined detour factor[1]. A recent study found an average detour factor of 1.47 for domestic private travel on Swiss roads (Chalasani and Axhausen, 2004). Imputed walk stages with valid home based geo-codes were considered in crow-fly distance calculation and an average detour factor of 1.50 was assumed for distance calculations. Thus, the estimated walk distance of the imputed walk stage with valid geo-codes (at postal address level) is 1.5 times the crow-fly distance. Walk distances for the imputed stages without geo-codes were assumed. Mode of transport used in connected non-walk stages and absolute position of the walk stage were considered in assuming walk distances. The walk distances assumed for the imputed walk stages are shown in Table 4.

4.1.3 Comparison of before and after stage sequence correction

A comparison between the trips with imputed walk stages and with reported stages to assess the imputation significance was made. Results of this comparison are briefly shown in Table 4.

Table 4 Comparison of trip attributes before and after stage sequence correction

Description	Before correction	After correction
Number of trips	105,979	105,579
Number of stages	147,058	283,933
Percent added		48.21
Avg. stage frequency of the trips with a correct stage-sequence	3.30	-
Avg. stage frequency of the trips with an incorrect stage-sequence	1.37	3.29
Average Trip length (kilometres)	13.50	13.58
Percent increased		0.60

Source: Chalasani and Axhuasen, 2004

Comparison results reveal that stage imputation is significant in correcting the stage frequency. But, the average trip length increased is negligible, mainly due to the smaller imputed walk distances.

4.2 Adjustment of stage arrival and departure times

With the assumed distances for the imputed walk stages from the stage imputation and an assumed average walk speed of 75 meters per minute, travel times for the imputed walk stages have been calculated. Model walk times were calculated based on the assumed walk distances which vary as per the mode of transport used in the immediate next stage. The computed travel times for the imputed walk stages range from 0.40 to 20 minutes. Both the departure and arrival times were observed for all reported stages. Though 80% of the reported departure times are multiples of 5 minutes, it is difficult to identify the rounded cases. This point is considered in the travel time adjustments with the following assumption: Travel times are adjusted assuming that the departure and arrival times of each trip are correctly reported and that the travel times of the missing walk stages before the initial and after the final non-walk stages were ignored. In addition, following assumptions have been made for the adjustment of departure and arrival times:

- No waiting time between the imputed first walk stage and the reported first non-walk stage: The first reported stage with non-walk mode resumes immediately after the first imputed walk stage. In other way, people are expected to plan their departures in such a way that they do not have to wait for the first non-walk mode. This assumption is applicable for all trips with at least one non-walk mode and a missing first walk stage.
- No waiting time between a non-walk mode and the imputed walk stage: Intermediate walk stage between two non-walk stages commences immediately after the end of first non-walk stage and ends at the start location of the next non-walk stage. Elapsed time between the intermediate walk stage arrival time and the non-walk stage departure time is considered to be waiting time.
- No waiting time between the reported last non-walk stage and the imputed last walk stage: Similar to the first walk stage, the last walk stage immediately follows from the moment the last reported non-walk stage ends.
- Walk duration for the imputed stages was calculated using an assumed walking speed of 75 meters per minute. Similar to the stage imputation, all the trips with consecutive reported walk stages were ignored in the departure and arrival times' adjustment. At the same time, trips with a correctly reported stage-sequence were excluded from the adjustment.

Departure time and arrival time for all trips with imputed walk stages were calculated based on the position of the imputed walk stage in the stage-sequence. Table 5 summarized the procedure.

Table 5 Adjusted departure time and arrival time calculations

Stage description	Adjusted time calculation
Before the first reported non-walk stage	Arrival time = Departure time of the reported first non-walk stage Departure time = Arrival time – model walk time
Intermediate walk stage: between two consecutive non-walk stages reported	Departure time = Arrival time of the first non-walk stage reported Arrival time = Departure time + model walk time
After the reported last non-walk stage	Departure time = arrival time of the non-walk stage Arrival time = Departure time + model walk time

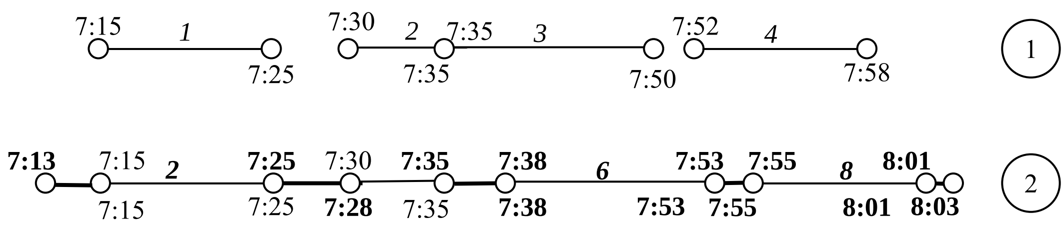
Source: Chalasani and Axhuasen, 2004

Both the imputed initial and final walk stages are simple extensions before the first and after the last reported non-walk stages respectively. Departure time of the first and arrival time of the last imputed walk stages were calculated using the departure time of the first and the arrival time of last reported non-walk stages. In addition to the Table 5, the set of rules in computing the departure and arrival times of the imputed intermediated walk stages followed are:

- When the imputed walk stage ends before the succeeding non-walk stage resumes, the departure time of the non-walk stage would be set as the reported departure time
- When the imputed walk stage ends after the succeeding non-walk stage resumes, the departure time of the non-walk stage will be shifted to the arrival time of the imputed walk stage. From this point onwards, the departure and arrival times for all the succeeding stages are accordingly adjusted.

Figure 1 shows the stage-sequences before and after the departure time and arrival time adjustments. The two stage-sequences represent before (1) and after (2) the stage imputation, that includes without and with imputed walk stages in addition to the reported stages, respectively. Total number of stages was increased from four to nine. Departure times are drawn above the stage nodes and arrival times are below them. Stage numbers are featured in italics. Adjusted departure time, arrival time and stage numbers are shown in bold letters.

Figure 1 Stage-sequence situations without and with imputed walk stages

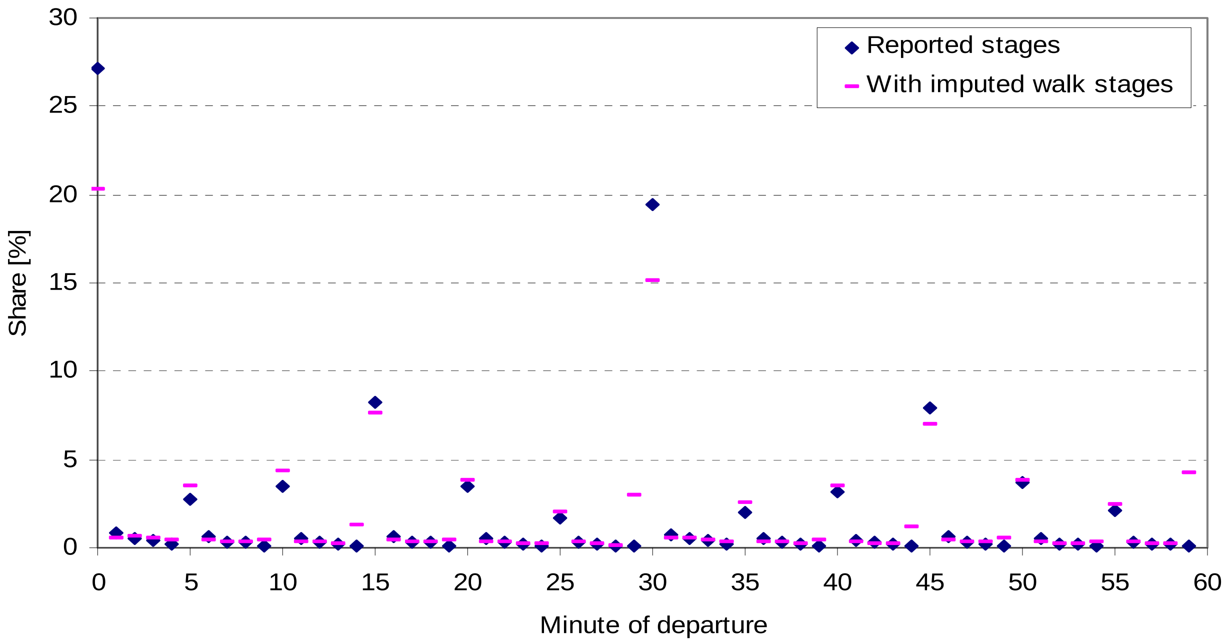


(1) : Reported and without imputed walk stages (2) : Adjusted with imputed walk stages

: Imputed walk stages Source: Chalasani and Axhausen, 2004

Most of the departure times of the imputed walk stages are not multiples of 5 minutes because these are built around the reported departure and arrival times. After rounding the departure and arrival times of the imputed walk stages to the nearest minute, the distribution of the departure all the trips was computed and shown in Figure 2.

Figure 2 Distributions of new minute of departures before and after walk stages imputation

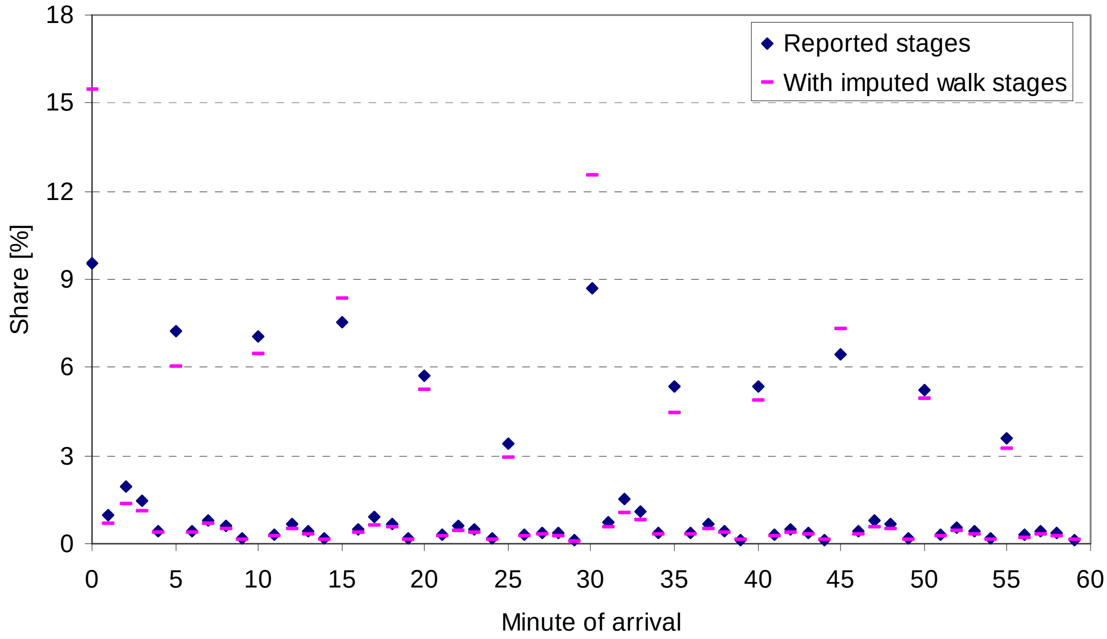


Source: Chalasani and Axhausen, 2004

Compared with the previous distribution, the distribution has smaller peaks and is more spread. A clear reduction in the share of 0th, 15th, 30th, and 45th minute of departures was observed. In contrast, a mixed pattern was observed for the multiples of 5 minutes shares (5th, 10th, 20th, 25th, 35th, 40th, 50th, and 55th). A relative increase in the shares of other minute of departures was observed. An overall reduction of 10% in the share of departure times with multiples of 5 minutes was the main outcome of the departure time adjustments.

Distribution of arrival minutes before and after is shown in Figure 3. Similar to the minute of departure distribution, the minute of arrival distribution has smaller shares except at the quarter hour intervals (0th, 15th, 30th, and 45th), where the shares are larger than that of reported stages. Because of the assumptions behind adjustments, only arrival times of the imputed intermediate and final walk stages differ from the rounded values. The large share of short walks increases the share of the rounded minutes.

Figure 3 Distributions of minute of arrivals before and after walk stages imputation



Source: Chalasani and Axhausen, 2004

4.3 Main mode of transportation

As mentioned earlier, respondents in the 2000 travel Microcensus were asked to report the mode of transport used for all the stages. This decision generated trips with one or more modes of transport, simply multi-modal trips. As most of the travel behaviour analysis is trip based, mode of transport at trip level has been the prime concern. This study has analysed for the trip main mode of transport, the aggregated mode of transport from stages to a trip. Two criteria were considered to calculate the trip main mode:

- Trip main mode by modal hierarchy
- Trip main mode based on distance travelled

In the first criterion, trip main mode was determined using an assumed modal hierarchy. The main mode of transport is the mode with highest priority. In the second criterion, the main mode of transport was ascertained based on the reported travel distance. Table 6 compares the share of main modes of transport for the two criteria. Zero to negligible difference between the two criteria reveals that the assumed modal hierarchy is an alternative to determine trip main mode.

Table 6 Comparison of main mode of transport by hierarchy and travel distance

Mode of transport	Rank*	Percent share by the main mode of transport as per the		
		Modal hierarchy	Distance travelled	Difference
Plane	1	0.07	0.07	0.0
Train	2	3.70	3.60	0.1
Long-haul public transport	3	0.80	0.90	0.1
Short-haul public transport	4	6.20	6.10	0.1
Car	5	50.90	50.50	0.4
Motorcycle	6	1.30	1.30	0.0
Moped	7	0.80	0.80	0.0
Bicycle	8	7.30	7.30	0.0
Other	9	1.00	1.10	0.1
Walk	10	27.20	27.70	0.5

*: Rank of the transport mode as per assumed modal hierarchy

Source: Chalasani and Axhausen, 2005

4.4 Home-based trips

Although all origin and destination addresses were collected in the 2000 travel Microcensus, quality of address information was not a prime concern for the study. A study was conducted to classify the trips based on the origin and destination locations (Reiser, 2004). Trip origin and destination addresses were compared with that of households. Trip origin and destinations were rated based on the extent of matching and interpreted as:

- Home originated trips: If the origin address(Municipality number, street address, and household number) exactly matched with the home address
- Home destined trips: If the destination address (Municipality number, street address, and household number) exactly matches with the home address

In addition to the above matching, when it was tested for the hypothesis: all end trips (1st and last trip of the day) whose origin or destination municipality number matches with household municipality number are home based (either originated or destined), a positive result was obtained for more than 90% of the trips.

4.5 Location geo-coding

Location geo-coding is the most common and important enrichment for any household travel survey data. The 2000 travel Microcensus data was geo-coded in a separate study (Jermann, 2003). The following five geo-databases were used to calculate the geo-codes:

- Gebkoord BfS: All building entrances of Switzerland
- Gebkoord ZH: All building entrances of Canton Zurich
- Microspot: Swiss post office geo-codes
- Haltekoord: Public transport stops in Canton Zurich (Excluding Postbus stops)
- Bahnhoefe ARE: Swiss railway station geo-codes

Using the above geo-coded address data bases, a semi-automatic matching process was implemented after normalising and correcting the location addresses (spelling, punctuation, removal of diacritical marks, et.). The remaining addresses were matched by hand, as far as possible, using maps, telephone books, and information on the internet, especially for place names and leisure facilities. Table 7 details the quality of matching at stage ends.

Table 7 Mikrozensus 2000: Matching quality by stage end

	From														
To	A1	A2	A3	B2	B3	C1	C2	C3	D1	D2	D3	E1	E3	F	Sum
A1	4.0	0.4	0.0	2.6	0.1	3.4	0.1	0.0	1.5	0.3	0.1	6.0	0.0	0.7	19.3
A2	0.4	0.2	0.0	0.2	0.0	0.8	0.0	0.0	0.1	0.0	0.0	1.0	0.0	0.1	3.0
A3	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0	-	0.0	-	0.0	0.1
B2	2.4	0.2	0.0	1.9	0.0	1.0	0.0	0.0	0.5	0.1	0.0	1.3	0.0	0.2	7.9
B3	0.0	0.0	0.0	0.0	0.0	0.1	0.0	-	0.0	0.0	0.0	0.0	-	0.0	0.2
C1	3.2	0.7	0.0	0.9	0.0	12.2	0.3	0.0	0.2	0.1	0.0	4.0	0.0	0.4	22.1
C2	0.1	0.0	0.0	0.1	0.0	0.3	0.1	0.0	0.0	0.0	0.0	0.4	0.0	0.0	1.0
C3	0.0	0.0	-	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-	0.0	0.1
D1	1.5	0.1	0.0	0.5	0.0	0.2	0.0	0.0	1.7	0.2	0.1	0.8	0.0	0.1	5.3
D2	0.3	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.2	0.1	0.0	0.2	0.0	0.0	1.1
D3	0.1	0.0	-	0.1	0.0	0.0	0.0	-	0.1	0.0	0.0	0.1	0.0	0.0	0.5
E1	5.7	1.0	0.0	1.2	0.0	4.1	0.4	0.0	0.9	0.2	0.1	21.4	0.1	0.4	35.5
E3	0.0	0.0	0.0	0.0	-	0.0	0.0	-	0.0	0.0	0.0	0.1	0.1	0.0	0.3
F	0.7	0.0	0.0	0.2	0.0	0.4	0.0	0.0	0.1	0.0	0.0	0.6	0.0	1.5	3.7
Sum	18.5	2.8	0.1	7.9	0.2	22.6	0.9	0.1	5.3	1.1	0.5	36.0	0.3	3.6	100.0

Source: Chalasani et al., 2004

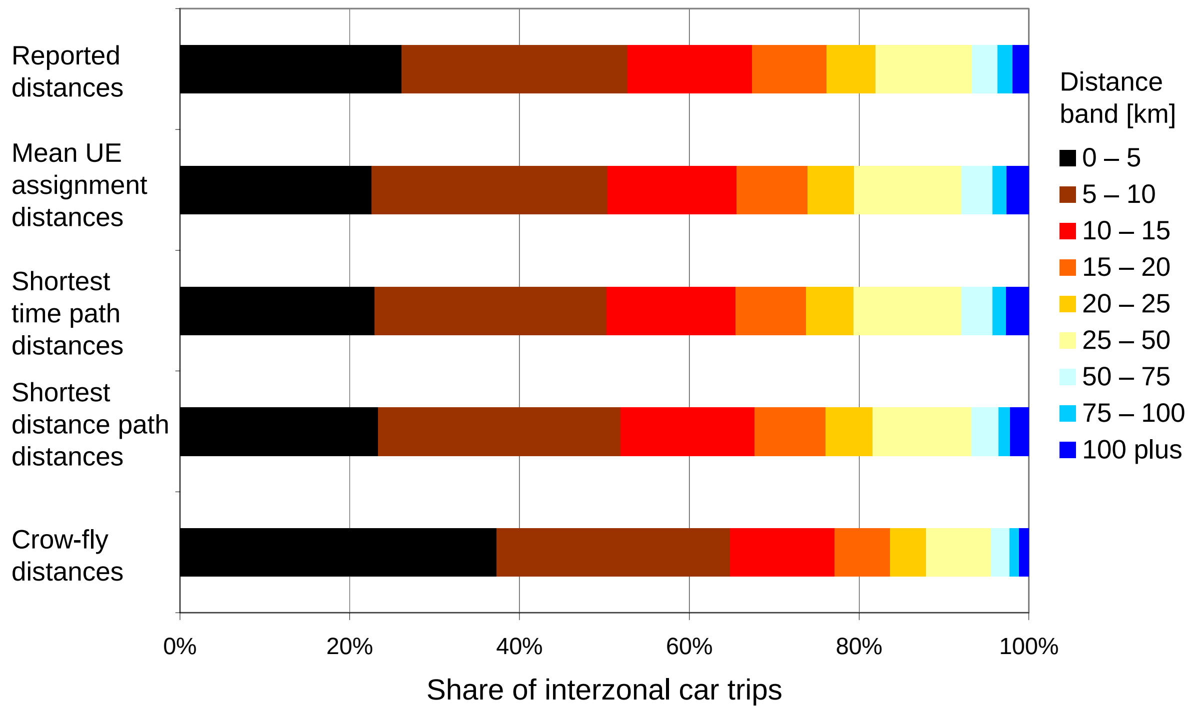
4.6 Travel distance calculations

Because of the difficulties involved in collecting travel distance information from travel surveys, most of the travel surveys inclined to calculate travel distance. The distance travelled is calculated using the geographic information collected from travel surveys. Traditionally, to understand the spatial aspects of the transportation system, transport researchers followed a standard set of distances. This set includes:

- Crow-fly distance
- Network based distances
  - Shortest distance path distance
  - Shortest time path distance
  - Loaded network distance
- Reported distance

A study was conducted (Chalasani and Axhausen, 2005) to compare the above stated distances for all trips with cars as main mode. Except the reported distances, distances were calculated for all stages and aggregated for trip distances. The cumulative trip length distributions of car trips for different distance measures are shown in the Figure 4.

Figure 4 Mikrozensus 2000: Comparison of the distance distributions (34'195 car passenger and driver interzonal trips)



Source: Chalasani et al., 2004

In many cases, it is useful to convert one distance estimate to another. Such conversion or detour factors have been previously reported, but only for certain pairs of distance estimates. Table 8 provides six comparisons for 2000 travel Microcensus based on the estimates described above. A clear difference can be observed in detour factors change patterns. Calculations are based on all observations in the sample, even if crow-fly distances were longer than model based estimates. This can happen, especially for shorter trips, when the distance between zonal centroids is smaller than actual distance travelled (see above). Detour factors fall as crow fly distances become longer.

One of the prime goals of the travel distance calculation is to assess the reported distances stand among the set of distances, which comprises of three network distances (shortest distance path distance, shortest time path distance, and loaded network distance) and crow-fly

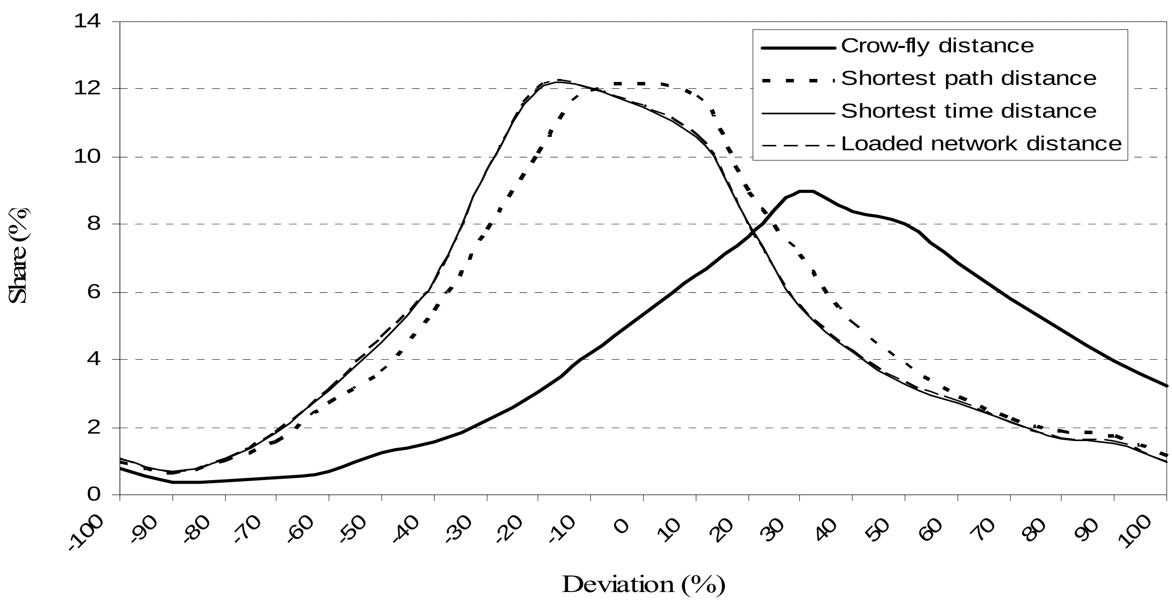
Table 8 Mikrozensus 2000: Detour factors between different distance estimates (34'195 car passenger and driver interzonal trips)

Average detour factor with	Crow fly distance			Shortest distance paths		Shortest time distance
Distance band	Shortest distance path distance	Shortest time path distance	Mean user equilibrium distance	Shortest time path distance	Mean user equilibrium distance	Mean user equilibrium distance
0 to 5 km	1.83	1.87	1.88	1.01	1.02	1.01
5 to 10 km	1.39	1.46	1.46	1.04	1.05	1.00
10 to 25 km	1.35	1.47	1.47	1.09	1.09	1.00
25 to 50 km	1.31	1.46	1.46	1.11	1.11	1.00
50 to 75 km	1.31	1.47	1.47	1.12	1.12	1.00
75 to 100 km	1.32	1.49	1.49	1.13	1.13	1.00
100km and more	1.26	1.48	1.48	1.16	1.16	1.00
Total	1.54	1.62	1.62	1.05	1.05	1.00

Source: Chalasani et al., 2004

calculated. Figure 5 shows the distributions of deviations of the reported distances against the four distinct computed distances.

Figure 5 Distributions of reported distance deviations from calculated distances



4.7 Travel costs

Travel cost is one of the factors that substantially influence travel. An exclusive study was conducted (Koenig, 2004) to compute the costs incurred in travel by car in Switzerland. Travel costs for all car trips were calculated using the following equation:

$$(V*1.35/100 + R/40000 + W/JFL + (1+2)*S/JFL + P/JFL)*A$$

Where

V	Consumption per 100km
R	Cost of a set of tyres
W	Maintenance costs
JFL	Distance travelled (km/year)
S	Annual taxes (Canton Zürich)
P	Annual parking costs
A	Factor based on the vehicle age

With the information about distance travelled, year of production, location and approximate cost of tyres and parking, travel costs for all car trips in the 2000 travel Microcensus were calculated.

4.8 Location type and accessibility index

Based on the land-use type and accessibility to the transport facilities, Federal office for spatial development (ARE) has categorized all the Swiss municipalities into 5 categories and are tabulated in Table 9.

Table 9 Spatial pattern classification of Swiss municipalities by transport

	Group	Spatial pattern type by transport	Description
Urban areas	1	V1	Grosszentren
	2		Nebenzentren der Grosszentren
	3	V2	Innere Gürtel der Agglomerationen der Grosszentren
	4		Äussere Gürtel der Agglomerationen der Grosszentren
	5	V3	Mittelzentren
	6	V4	Innere Gürtel der weiteren Agglomerationen
	7		Äussere Gürtel der weiteren Agglomerationen
Rural areas	8	V5	Kleinzentren
	9		Wegpendlergemeinden
	10		Industrielle und tertiäre Gemeinden
	11		Semiagrarische Gemeinden
	12		Agrarische Gemeinden
	13		Touristische Gemeinden

Source: ARE, 2002

A special study on accessibility of Swiss municipalities has been carried out (Frohlich, 2004). As per this study, accessibility can be understood as the number of opportunities available for social and economic life that can be reached within a travel time appropriate to the relevant purpose. One way to measure accessibility is to weight attractiveness of the activity points with the necessary travel time to these points by means of a negative potential function. The main challenge here is to find the appropriate exponent, β, that determines the destination choice of the people, which can vary over time. In the past, the β factor should have been larger than today, but these values would have to be derived from old traffic surveys, which are not available yet. In this study, the β factor is taken as 0.1 and kept constant across periods.

$$A_i = \log \sum\nolimits_{\forall ij} X_j e^{-\beta c_{ij}}$$

Where

A_i	accessibility of location *i*
X_j	Number of opportunities of interest at location *j*
c_{ij}	Generalised costs or network impedance between locations *i* and *j*
β	Weighting parameter

Accessibilities fort he year 2000 obtained from this study were assigned to all the municipalities involved in the 2000 travel Microcensus.

4.9 Nearest public transport access points

The latest enrichment to the 2000 travel Microcensus is addition of a set of nearest public transport access points or stops. In an exclusive study for Canton Zurich a choice set of alternative routes was developed for all the trips within Canton Zurich in the 2000 travel Microcensus. The 2000 travel Microcensus Canton Zurich trips geo-database is matched with the public transport network geo-database to identify all the public transport stops within 2km radius. Three nearest (by Crow-fly distance) public transport stops were selected and a choice set for each stage origin and destination was prepared.

5. Conclusions

Experiences from the twofold enrichment process implemented on the 2000 travel Microcensus can be concluded as follows:

- Deficiencies in household travel survey data are common. These deficiencies largely influence the data quality. Eradication of household travel survey data deficiencies is extremely difficult. Instead, they can be reduced in size and effect.
- After identifying the deficiencies and with the knowledge about the external resources, a set of possible enrichments should be determined. The enrichment sequence can be decided based on the information interaction among different elements of the household travel survey data.
- Internal enrichments with the collected household travel survey data results in many advantages such as, improves the data quality, reduce the systematic errors, support better for secondary use, etc.
- Addition information from external sources not only maximises the data value, but also enhance the application area database.

The following enhancements were implemented on the 2000 travel Microcensus:

- Stages imputation: Missing walk stages can easily be identified and imputed. However, it is not possible to depict the missing short length (less than 100m) non-walk stages.
- Arrival and departure time adjustments: It was hard to differentiate the waiting time and walk stage time in the adjustment process of arrival and departure times for the modified stage-sequences.
- Main mode of transport: Modes of transport reported for all the stages in a trip were aggregated to identify the trip main mode. Assumed modal hierarchy equals the distance based aggregation.
- Home based trips: Origin and destination address information updated to with respect to the reported household address information found consistent with the correctly reported trips.
- Location geo-coding: A very large amount of resources were required for location geo-coding some time after the survey. Many errors such as spell checks, etc., could have been avoided with online geo-coding.
- Travel distance calculations: From the comparison of reported distances with four calculated distances, it is concluded that the reported distances are close to the loaded network distances.
- Travel costs: As the parking cost sand taxes vary widely, huge amounts of information are required to calculate the travel costs.

From the travel behaviour analysis's view, the enriched 2000 travel Microcensus has acquired the potential to support mode-choice and route-choice modelling.

6. References

Arentze, T., H. Timmermans, F. Hofman and N. Kalfs (1997) Data needs, data collection, and data quality requirements of activity-based transport demand models, *Transportation Research Circular*, **E-C008**, II-J/1 – II-J/35, Transportation Research Board, Washington.

Axhausen, K. W., S. Beige, M. Bernard und J. Jermann (2003) Grundlagenbericht für die Perspektiven des Schweizer Personenverkehrs bis 2030, *Schlussbericht für das Bundesamt für Raumentwicklung*, Bern.

Axhausen, K.W. (2000) Definition of movement and activity for transport modeling. In D. Hensher and K. Button (eds.) *Handbook in Transport: Transport Modelling*, 271-283, Elsevier, Oxford.

Battelle Transportation Division (1997) *Lexington Area Travel Data Collection Test: Global Positioning Systems for Personal Travel Surveys*, U.S. Department of Transportation, Federal Highway Administration.

Bundesamt für Raumentwicklung (2002) Verkehrliche Raumgliederung (V+-V%) ausgehend von der „Raumgliederung 2002“ (1-13), Bern

Chalasani, V.S. and K.W. Axhausen (2004) Enriching household travel survey data: The case of the Microcensus 2000, *Arbeitsbericht Verkehrs- und Raumplanung*, **223**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Chalasani, V.S., J.M. Denstali, Ø. Engebretsen und K.W. Axhausen (2004) Precision of geocoded locations and network distance estimates, *Arbeitsbericht Verkehrs- und Raumplanung*, **256**, IVT. ETH Zürich, Zürich.

Chalasani V.S. and K.W. Axhausen (2005) Travel distance computation from household travel survey data: The case of the Microcensus 2000, *Arbeitsbericht Verkehrs- und Raumplanung*, **252**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Fröhlich, Ph. (2004) Sensitivity of the accessibility measurements to the underlying transport network model, *Arbeitsbericht Verkehrs- und Raumplanung*, **245**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Groves, R.M. (1989) *Survey Errors and Survey Costs*, John Wiley and Sons, New York.

Jermann, J. (2003) Geokodierung Mikrozensus 2000, *Arbeitsbericht Verkehrs- und Raumplanung*, **177**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Kalfs, N., H. Meurs and W. Saris (1997) Quality indicators, *Transportation Research Circular*, **E-C008**, II-E/1 – II-E/14, Transportation Research Board, Washington.

König, A., K.W. Axhausen, G. Abay, J.J. Bates und M. Bierlaire (2004) Zeitkostenansätze im Personenverkehr Eine aktuelle Schweizer Studie, *Arbeitsberichte Verkehr- und Raumplanung*, **195**, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Rieser, M. (2004) *Berechnung von Nachfragematrizen mit VISEM*, Semesterarbeit, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Rietveld, P. (2002) Rounding of arrival and departure times in travel surveys: An interpretation in terms of scheduled activities, *Journal of Transportation Statistics*, **5** (1) 71-82.

Swiss Federal Office for Spatial Development and Swiss Federal Statistical Office (2001) Travel behaviour results from the 2000 Microcensus travel, *Mobility in Switzerland*, SFSO and ARE, Bern.