

Investigating the Transferability of National Household Travel Survey Data

Abolfazl Mohammadian, Ph.D.
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor St.
Chicago, IL 60607
Phone: 312-861-0259
Fax: 312-996-2426
Email: kouros@uic.edu

Yongping Zhang
Ph.D. Candidate
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor St.
Chicago, IL 60607
Phone: 312-413-7447
Fax: 312-996-2426
Email: yzhang31@uic.edu

Paper to appear in the
Transportation Research Record: Journal of the Transportation Research Board, 2007

Final Revised Paper submitted on April 1, 2007

Total Words: 5,806 + 1750 (2 figures + 5 tables) = 7,556 words

ABSTRACT

Traditionally, Metropolitan Planning Organizations are required to have their models calibrated on a continuing basis using new data. However, new survey data required to support these models do not exist in most urban areas. This makes it very difficult to calibrate existing models or develop new travel demand models using emerging modeling techniques. As a result, the need to assess potential approaches and develop knowledge on how to transfer information collected in one context and use it in another context is becoming critical. This paper presents the process of developing a model that can facilitate household travel data transferability, which can reduce or eliminate the need for a large data collection in the application context. Household records from the 2001 National Household Travel Survey are clustered into several homogeneous groups representing various types of households and their lifestyles. The clustering model is based on several variables that are easily obtainable from available sources. In addition to household characteristics, the variables of land use, built environment, and transportation system characteristics are included in the analysis. Various statistical measures of travel behavior are estimated for each cluster. Using an artificial neural network model, cluster membership rules and travel statistics are then transferred to other geographic areas for the purpose of comparison and validation of the transferability method.

INTRODUCTION

Traditionally, Metropolitan Planning Organizations (MPOs) of urbanized areas throughout the United States are required to conduct transportation planning activities using federal and state funds for their regions. Conformity regulations require these MPOs to have their models calibrated on a continuing basis using new data. On the other hand, emerging modeling techniques (e.g., microsimulation models) that require much richer data sets are becoming available to planners. However, new survey data required to support these models do not exist in most metropolitan areas, so many questions are emerging with respect to the transferability of household travel surveys data. The main idea behind the current study is to develop a household travel data transferability model that can facilitate the use of travel data collected in one context in a new context. This can reduce or eliminate the need for a large data collection in the application context (1).

Transferability can be defined as the effectiveness of a transferred model, information, or theory in a new context and the extent to which findings can be applied in the new environment. It is important to distinguish between transferability and generalizability. Generalizable findings must apply across all environments, while transferability allows knowledge to be applied to other environments. Transferability can be achieved through in-depth description and sufficiently detailed analysis of the information that brings the subject explicitly into the environment under investigation (2).

While data transferability is an emerging research topic, travel demand model transferability has been the subject of numerous studies. There are several empirical studies in the literature to assess the effectiveness of model transferability (1). Many of these studies have focused on examining model transferability from one spatial context to another (3-5). Several others studies have attempted to examine the effects of temporal transferability of these models (6-8). More recently, a few but growing number of researchers have attempted to investigate the transferability of transportation planning data. Trip generation tables for various land use applications are probably the most widely used form of transferred transportation planning data (9). The National Highway Cooperative Highway Research Program (NCHRP) Report #365 provides a review of the four-step travel demand process and transferable parameters that can be used in simple planning analyses. This can be useful to planners in smaller urban areas that cannot afford to develop area-specific parameters (10). Wilmot and Stopher (2001) extracted aggregate data such as trip rates, mode shares, and trip-length frequency distributions from the 1995 NPTS Survey and the North Central Texas Council of Governments Survey of 1996 and compared them with data from the Baton Rouge Personal Transportation survey conducted in 1997 (11). Reuscher et al. used aggregate data from the NPTS to estimate regional or local travel behavior which can be used by state or municipal transportation planners (12). They compared four household data classification levels of census tract, MSA, census division, and census region using 1995 NPTS data against the add-on data sets from New York, Massachusetts, Oklahoma, and Baton Rouge separately. They

concluded that, in most cases, the census tract clustering method predicts travel better than the other three methods, except in the case of small sample sizes. Greaves and Stopher combined local socio-demographic data for individuals and households from the census with probability distributions of activity/travel patterns from other travel surveys, such as NPTS, to simulate local travel survey data (13). Following similar approach, Stopher et al. simulated household travel survey data for Dallas-Fort Worth and Salt Lake City metropolitan regions (14) and in Australia (15). They concluded that a classification scheme is critical in simulating household travel survey data. The more the population and the dependent variables are subdivided into homogeneous groups, the better the chance the simulation will ultimately have of reflecting these differences.

In this study, records from the 2001 National Household Travel Survey (NHTS) are clustered into several homogeneous groups representing various types of households. Next, statistical measures of travel behavior (e.g., trip rates by purpose, travel mode, commute distance) are estimated within each cluster for various household life cycles. Using an artificial neural network model, cluster membership rules and values of travel-characteristic measures are then transferred to other geographic areas for the purpose of comparison and validation of the transferability method.

DATA SOURCES

The National Household Travel Survey (NHTS) is the nation's inventory of daily and long-distance travel (16). The survey includes demographic characteristics of households, people, vehicles, and detailed information on daily and longer-distance travel for all purposes by all modes. NHTS survey data are collected from a sample of U.S. households and expanded to provide national estimates of trips and miles by travel mode, trip purpose, and household attributes. The data set includes information on 69,817 households, 160,758 people, 139,382 vehicles, and 642,292 trips. The 2001 NHTS data reflects average travel behavior in the nation and has been designed to provide statistically valid estimates of trip rates and travel statistics at the national level, at the census division level, and at the seven area-type levels (divided by population size and the availability of transit services). The NHTS data is not recommended for areas smaller than a census division because the sample is generally too small to be used at a local level. Thus, the problem is how to transfer the data from the census division level to smaller areas.

Other sources of information used in this study include Census Transportation planning Package (CTPP 2000), Census 2000 TIGER/Line GIS data files, and Texas Transportation Institute's (TTI) Urban Mobility Study report (17).

Data and Variables

In order to establish a relationship between travel behavior, household characteristics, and land-use and built environment attributes, exploratory regression analysis was performed

that resulted in the selection of explanatory variables in this study. Demographics and socio-economics and built-environment related variables were extracted from various sources. These variables are in individual, household, and census tract levels.

Individual-Level Variables:

Dummy variables were generated from those categorical variables in Person file of NHTS2001. These include *age group* (0-15, 16-24, 25-34, 35-44, 45-54, 55-64, 65+ years), *gender*, *race/ethnicity* (White, Black, Hispanic, Asian), *educational level* (less than high school, high school diploma, college degree, or graduate degree), *occupation* (blue collar [manufacturing, construction, maintenance, or farming]; service [sales or service]; or professional [professional, managerial, or technical]).

Household-Level Variables:

Household-level aggregate variables were generated using individual-level variables. These include number of household members in various *age groups*, *race/ethnicity*, *education levels*, and *occupations*. Other household-level variables are *household size*, *income*, *number of adults*, *number of vehicles*, *number of drivers*, and *number of workers* that were extracted directly from the Household file of the NHTS 2001.

Census Tract-Level Variables:

Housing density (number of housing units per square mile), *employment density* (number of employment opportunities per square mile), and *population density* (number of people per square mile) were directly extracted from the NHTS 2001 Household file. Several new variables were also generated that are used as proxies to reflect the pedestrian environment, transit usage, and congestion factor.

Pedestrian Environment: Several measures were estimated to reflect whether each census tract is pedestrian friendly. These measures include *intersection density* (number of intersections/area size), *road density* (road length/area size), and *block size* (road length/number of intersections). Dense urban areas with smaller block size and higher road and intersection numbers may offer a superior pedestrian-friendly environment. Road lengths and intersection counts are measured for all census tracts in the nation. These are estimated by overlapping street network and census tract shape-files in a GIS environment. As an example, the intersection density for a small area is illustrated in Figure 1.

Transit Usage: In order to determine whether each census tract is transit friendly, a transit usage measure has been introduced by matching the proportion of transit users to the total number of workers in the census tract. The CTPP 2000 data set has been used to estimate this measure. Areas with higher transit ridership are assumed to be more transit friendly.

Congestion Factor: Several different measures were estimated and tested to determine whether living in a congested area has an impact on travel behavior. The first factor was

estimated by dividing the percentage of workers driving to work by the road density in that census tract. Presumably, if two census tracts have the same level of road density, the one with the higher share of auto-drive mode choice can be more congested. With the same logic, the second factor was estimated by dividing total number of auto users by road density. Finally, the third congestion factor was extracted from the Travel-Time Index Tables of the Urban Mobility study (17). Travel time indices and five levels of congestion are reported for 85 cities across the nation. These are matched with 74 MSA areas of the NHTS 2001 data set to form travel time indices that are used as proxy to present the congestion level that an average household experiences at the MSA level.

Dependent Variables:

The dependent variables that are the subject of the transferability study were all derived from the trip file of the national-level NHTS 2001 (excluding add-on observations) that present household-level travel behavior. They include VMT, number of trips, number of trips by purpose (e.g., mandatory, maintenance, discretionary), number and percentage of transit users, number of transit trips, number of private vehicle trips, number of non-motorized trips, number of tours, average trips per tour, average trip distance in miles for all members, number and percentage of carpool users, percentage of public transit usage, and total and average commute distance in the household.

Data Preparation

There are nine add-on areas accounting for 43,779 of the total 69,817 observed households in the NHTS 2001 data set. In this study, 26,038 households in the national-level data set were used for modeling, leaving add-on observations to examine the validity of the transferred data. Previous studies have also suggested excluding Manhattan cases from the analysis (12). It is assumed that Manhattan residents behave so differently from the rest of the nation that national travel data should not be used as a basis for estimates about Manhattan. There are totally 473 households surveyed in Manhattan as part of both national and New York add-on samples. In order to test the suitability of the Manhattan elimination assumption, all 69,817 households in the NHTS data set were clustered into four homogeneous groups according to their travel characteristics using a two-step clustering method. The cluster membership shares for each group were 24.7%, 23.8%, 14.5%, and 33.8%, respectively; 3.2% were outliers and did not fall into any of the four groups. Analyzing those 473 Manhattan cases, it became clear that their cluster membership rate among four homogeneous groups were 25%, 26%, 5%, and 3%, respectively, leaving 41% of the Manhattan cases as outliers. Manhattan residents have very different travel characteristics compared to the rest of the nation, so a decision was made to exclude those cases.

ANALYSIS AND FINDINGS

The analyses include factor analysis of several explanatory variables of the national

sample and households clustered into homogeneous groups based on extracted factors. Travel characteristics of the national data set are then transferred to the add-on areas according to their cluster membership. Finally, data from add-on data sets were used to validate the transferability of the travel data.

Principal Component Analysis

The household demographics, socio-economic, and built-environment related variables discussed earlier that are selected for analysis present a special difficulty when calibrating a model. Many of these variables are highly correlated. For example, population and residential densities or number of vehicles and number of workers in a household are highly correlated. This high multi-collinearity between variables might create problems of identification of the influence of land-use and household characteristics on travel behavior. Factor analysis, and more particularly, principal components analysis provides a means to solve this problem. In very broad terms, factor analysis is a method for reformulating a set of natural or observed independent variables into a new set (usually fewer in number) of independent variables, such that the latter set has certain desired properties specified by the analyst. Principal component analysis is the search through data to try to find the factors or components that may reduce the dimensions of variations and may be given a possible meaning (18).

Thirty-three explanatory variables of demographics, socio-economic status, and land use that were discussed earlier were chosen for principal component analysis. Numerous attempts were made to extract the best and most reasonable factors from the data set. Models with five to seventeen factors were tested with different combinations of explanatory variables. Finally, eleven factors were extracted that explain 71% of variance in the data set. Each of the eigenvalues for the eleven factors is greater than 1. The results of the principal component analysis using the Varimax rotation method (with Kaiser normalization) are presented in Table 1. The factor loadings of each of the explanatory variables onto each of the factors provide an initial understanding of the interdependencies between each of the variables. These factors can be identified as follows:

- a) **High Density and High Transit Usage:** This factor consists of four dominant variables of population, residential, employment densities and transit usage that are highly correlated. They define dimensions relating to land use.
- b) **Vehicles, Drivers, and Workers:** Five dominant variables forming this factor are number of drivers, adults, workers, vehicles, and individuals between 16 and 24 years old. This factor is the representative of driving dimension of the households.
- c) **Middle-Aged Family:** These are young white households with children. Dominant variables include number of people between ages 35 and 44, number of children younger than 15, and household size.
- d) **Highly Developed:** This factor presents urban form and accessibility measure. Dominant variables include block size, road density, intersection density, and travel

time index.

- e) Higher Income and Educated: This factor is heavily influenced by professionals with advanced degrees and a high income.
- f) Working Class: This factor describes working-class families who are not retired and have clerical and manufacturing jobs.
- g) College Degree: This factor mainly describes households with college degrees.
- h) Hispanic: This factor refers to households with Hispanic background who also seem to have less than a high school education.
- i) Young Singles or Couples: This factor describes households that are formed by young people between 25 and 35 years old.
- j) Black or Asian: Minorities seem to be dominant variables in this factor.
- k) Seniors: This final factor is heavily influenced by households with members more than 55 years old.

Cluster Analysis

With the preceding factors obtained through the principal component analysis as the keystone, the main goal of the study was to understand how each factor can be used to define the household and explain the observed travel behavior. In order to assign households to homogeneous lifestyles, cluster analysis was performed. Cluster analysis groups observations into relatively homogenous collections by essentially minimizing the variance or spread across defining variables of interest within the clusters and maximizing that between clusters.

Clustering schema is a critical issue. Theoretically, the further the population and the dependent variables are subdivided into homogeneous groups, the better. However, large numbers of clusters are difficult to work with. Different clustering methods were examined in this study, including K-means, hierarchical, C&RT, two-step, and artificial neural networks. Each of these methods offers advantages and disadvantages. K-means and hierarchical methods are not as fast as a two-step approach. Additionally, while using these approaches, it is not easy to draw conclusions on the optimal number of clusters. C&RT method has the disadvantage of providing only binary choices at each node of the tree, which do not necessarily reflect reality. The black-box nature of the artificial neural networks may also result in an ambiguous clustering schema. After numerous experiments with various clustering methods, a two-step cluster analysis routine was ultimately employed in this study. The two-step clustering schema is recognized to be efficient when dealing with very large data sets that contain both continuous and categorical variables (19). Using log-likelihood distances, households are pre-clustered forming a cluster feature tree. The resulting sub-clusters are then further categorized by comparing their distances to a specified threshold. In case the distance is larger than the threshold, the two clusters can be merged. The distance between two clusters j and s is defined as

$$d(j, s) = \xi_j + \xi_s - \xi_{\langle j, s \rangle}, \quad (1)$$

where

$$\xi_v = -N \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{vk}^2) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right), \quad \text{and} \quad (2)$$

$$\hat{E}_{vk} = - \sum_{l=1}^{L_k} \frac{N_{vkl}}{N_v} \log \frac{N_{vkl}}{N_v}. \quad (3)$$

In the above equations, K^A is the total number of continuous variables used in the procedure, K^B is the total number of categorical variables used, L_k is the number of categories for the k^{th} categorical variable, N_v is the number of observations in cluster v ,

$\hat{\sigma}_k^2$ is the estimated variance of the k^{th} continuous variable in all data, $\hat{\sigma}_{jk}^2$ is the variance of the k^{th} continuous variable in cluster j , N_{jkl} is the number of data records in cluster j whose k^{th} categorical variable takes the l^{th} category, and $\langle j, s \rangle$ is the index that represents the cluster formed by combining clusters j and s . The $\hat{\sigma}_k^2$ term solves the

problem of the undefined natural logarithm. The distance between clusters j and s would represent the decrease in log-likelihood when the two clusters are combined. The distance between two clusters can be defined by the Euclidean distance between the two cluster centers that can only be applied if all variables are continuous. A cluster center is defined as the vector of cluster means of each variable. In calculating the log-likelihood, continuous variables are assumed to be normally distributed and categorical variables are assumed to follow multinomial distributions (19).

After numerous trials, it became clear that compared to other clustering techniques, the two-step cluster analysis approach can converge faster and save a great amount of time. Additionally, the optimal number of clusters can be determined using a Bayesian information criterion (19). Using 21,472 cleaned household records for which all required variables were available from the national level NHTS data set, a two-step clustering method with 10% noise allowance was employed to group these households on the basis of the preceding extracted factors that reflect their socio-economics and land-use characteristics. These clusters should be helpful in defining various lifestyles that are assumed to shape household behavior, including travel activities. Cluster memberships, descriptive statistics, and cluster centers are all presented in Table 2.

Each cluster is named by exploring cross-tabulations of cluster memberships against socio-economics and demographics, life cycles, land use, and urban form characteristics of the household members in each group. The eleven clusters found in this study include:

1. **Rich and Smart:** The cluster represents middle-aged families with professional or managerial white collar jobs. They usually have graduate degrees and earn high

- incomes. Majority live in suburbs or towns. Greater part are White but also some Asian ethnicity.
2. **Young Achievers:** The group comprises young couples without children or mainly with pre-school children. They tend to have college degrees and primarily have white collar jobs in sales, service, technical, and professional with mid-range income. Higher percentages live in suburb or rural areas.
 3. **Kids-centered Families:** These are middle-aged and working class families with pre-school and school-age children. They usually have a college education and earn mid-range to high level income. They are primarily White and live in a suburb or town.
 4. **Rural Blues:** The cluster includes working class, middle-aged families with pre-school and school-age children. They are mainly high school graduates in blue collar jobs (farming, manufacturing, etc) and earn low- to mid-range income. Mostly White and live in rural areas or small towns.
 5. **Working Mixing Pot:** They tend to be working-class White, Black, Asian, or Hispanic single adults or couples with college or high school education and low- to mid-range income. Majority live in suburb or rural area, but some in urban areas.
 6. **Mainstream Families:** The cluster comprises mid-scale, upper middle age, White, working-class couples or families with older children. Usually have college or high school education, earn mid- to high-level income and live in suburb or rural areas.
 7. **Senior Couples:** These are senior couples, majority working and some are retired. Greater part is White but include some Black, Asian, or Native American. They live primarily in suburbs or rural areas.
 8. **Sustaining Minority Families:** The cluster represents low-income, middle-aged, working-class families. They are mainly Hispanic or Black but also some Asian and White. Majority have not finished high school. They tend to have service, sales, manufacturing, farming, or construction jobs.
 9. **Forever Youngs:** These are White senior couples. Most of them are retired and empty-nested but some have sales, service, or managerial jobs and earn low- to mid-range income.
 10. **Traditional Seniors:** They mainly comprise retired single individuals. A number of them are retired couples with low income. Majority are White but some Black, Asian, or Native American.
 11. **Neo Urbans:** These are small families/couples or single individuals living in dense urban areas. They typically have college education and low- to mid-range income from sales, service, or professional jobs. Their dominant race is White but a significant number are Black, Asian, and Hispanic.

Weighted means and standard deviations of household- and person-level travel characteristics for each of the eleven clusters are presented in Table 3. This table offers detailed information on lifestyles of households and their travel characteristics. Travel

behavior within each cluster varies significantly and variations in household-level variables are larger than those of the person-level variables. This can be attributed to differences in household size within each cluster. Further investigation of household travel behavior became possible by assigning household members of each cluster to smaller groups representing ten different life cycles. Life cycles are the typical sequence of household evolution and different stages a household passes through from early formation to the retirement years. The household life cycle as defined in NHTS data codebook, consists of ten developmental stages:

1. one adult, non-retired, no children
2. 2+ adults, non-retired, no children
3. one adult, youngest child 0-5
4. 2+ adults, youngest child 0-5
5. one adult, youngest child 6-15
6. 2+ adults, youngest child 6-15
7. one adult, youngest child 16-21
8. 2+ adults, youngest child 16-21
9. one adult, retired, no children
10. 2+ adults, retired, no children

Weighted statistics of travel estimates for various life cycles in each cluster are presented in Table 4. While the statistics of various life cycles within each cluster in Table 4 present significant variations, it should be noted that due to the size of the NHTS survey sample, some of these life cycles do not have more than few observations. Therefore, extra caution should be exercised when interpreting the results presented in Table 4.

It should be noted that while previous studies suggest using different combinations of characteristics of people, households, and trips for the different travel characteristics (13- 15), additional attempts in this study using *Classification and Regression Trees* (C&RT) failed to further partition travel characteristics in most clusters as function of different combinations of household characteristics. Therefore, current results were considered as the best achievable results and a decision was made to simulate all travel characteristics using the single set of characteristics that define cluster membership.

ASSESSMENT OF RESULTS

The performance of the household-level cluster estimates presented in Tables 3 and 4 was assessed by comparing them to the statistics from specific geographic areas. The sources of the extra data used for the comparison were the add-on samples of the NHTS. Add-on program allows states and MPOs to purchase additional samples of the NHTS in their planning area. This will help them to receive data for use in trip generation and other travel-demand forecasting model inputs and transportation planning applications. From this study's point of view, the add-on data sets present an excellent opportunity for

comparison with the national-level data. These data sets are consistent and have the exact same variables as the national-level data. Add-on participants in the 2001 NHTS included five states, including New York, Wisconsin, Texas, Kentucky, and Hawaii and four MPOs, including Baltimore, MD, Des Moines, IA, Lancaster County, PA, and Oahu, HI, resulting in 43,779 total sampled households.

Using the same 33 independent variables and factor score coefficients presented in Table 1, factors are generated for household members of add-on data sets. The next step was to transfer the national level cluster membership schema to the add-on data sets. This was not an easy task, as there are several factors that influence cluster membership, and interaction between these factors is difficult to record and transfer. An artificial neural network model was developed to simulate cluster membership assignment as a function of eleven factors that are estimated for each household.

Transferability

An artificial neural networks (ANN) model (20) was developed to simulate cluster membership of each household in the national NHTS data set. The model also utilizes a genetic algorithm (GA) to optimize network parameters tuning to select influential factors and to improve the explanatory power of the model. Inputs to the model are factors obtained from principal component analysis and outputs are cluster membership. To avoid overtraining, data sets of 21,461 national household records were partitioned to provide a randomly chosen test set of 6,438 (30%). The remaining records were further partitioned into a 12,877 (60%) record training set and a 2,146 (10%) record validation set to be used to test for over-fitting in the training process. In order to evaluate the predictive potential of trained neural network, the test data set was introduced to the trained network as input and compared to the actual desired response. Utilizing GA, extensive testing efforts were undertaken to obtain the best networks in terms of prediction potential. The final network that generated the best model fit offered 92.4% accuracy on the test set.

The ANN model calibrated on the national data set was then used to simulate cluster membership for each household in various add-on data sets. Travel estimates are then transferred from national clusters to each household record in the add-on areas according to their cluster membership. Weighted observed and predicted travel statistics are compared to validate the transferability model and evaluate its performance.

Assessment of Estimates

In order to assess the performance of the transferability approach and the quality of the travel estimates, weighted average values of transferred estimates from national-level data are compared to the observed statistics in each add-on sample. It worth noting that sample sizes of add-on surveys vary significantly, from 1,030 for Lancaster, PA, to 17,012 for the Wisconsin data set. Each add-on data set should be further partitioned

into eleven lifestyles (e.g., clusters). Unfortunately, smaller add-on samples do not offer enough records in each cluster to draw a conclusion on average values of travel statistics. In order to test the quality of add-on samples for the transferability test in this study, minimum sample sizes required for each cluster are estimated based on values of means and standard deviations of number of trips per household from the national-level NHTS data set using the following equation:

$$\frac{t_{\alpha/2}}{n^{1/2}} \leq \frac{d}{S.D.}, \quad (4)$$

where n is the minimum required sample size, $t_{\alpha/2}$ is t -stat at 95% confidence level (1.96), $S.D.$ is the standard deviation from national-level data, and d is the tolerance level. Minimum required sample sizes were estimated for each cluster. It became clear that none of the add-on areas have enough observation in all eleven clusters to be used for comparison. Therefore, the two largest add-on data sets (New York and Wisconsin) that have enough observations in most of the clusters were selected to validate the performance of the transferability model. A third sample (Baltimore) was selected to show the performance of the transferred data compared to a smaller data set. Due to the limited number of observations in add-on areas for the basis of comparison, it was not possible to further partition each cluster into life cycles (as shown in Table 4); therefore, assessments were performed only on cluster (lifestyle)-level data.

While comparison tests have been performed for various travel statistics (except for VMT, as best miles estimates are not available for add-ons), due to page limitation, this paper presents only few selected travel estimates. Figure 2 depicts observed and transferred estimates in each cluster for selected travel characteristics. Average statistics from two large add-on data sets (Wisconsin and New York) demonstrate better fit while compared to the transferred estimates. The gaps between observed and transferred estimates get wider for Baltimore, where the sample size is smaller. Table 5 presents a more detailed comparison between observed and transferred estimates and their percentage differences for selected travel statistics. While the results are highly encouraging, it appears that statistics associated with trips and tours by purpose and automobile trips are more transferable than other travel attributes such as commuting distance, transit use, and carpool use. This was expected, as transit usage or commute distance are highly influenced by urban form, built environment, and supply of transit services. Exploring these factors and improving the current model using a small local sample and Bayesian updating is subject of another study by authors.

SUMMARY AND CONCLUSION

Household records from the 2001 National Household Travel Survey (NHTS) are clustered into several homogeneous groups representing various household types and their lifestyles. The clustering model is based on several quantifiable variables that can be easily predicted or are available from other sources (e.g., PUMS, CTPP, NHTS, Tiger

files). In addition to household characteristics, variables representing land use, built environment, and transportation system information are included in the analysis. Various statistical measures of travel behavior (e.g., trip rates by purpose, commute distance, travel mode) are estimated for each cluster and different household life cycles. Cluster membership rules and values of travel estimates are then transferred to other geographical areas for the purpose of comparison and validation of the transferability method. Using an artificial neural network model, households from add-on areas of the NHTS were assigned to the same cluster schema developed for the national data set. Travel estimates from national data are transferred to the add-on areas based on household cluster membership and are then compared with the observed travel statistics for the same households. The results of the comparisons are very promising. It appears that statistics associated with trips and tours by purpose and automobile trips are easily transferable while other travel attributes such as commuting distance, transit use, and carpool use need to be further processed using a small local sample and Bayesian updating to be transferable. Further research is underway combining current travel data transferability model with a population-synthesizer attempting to simulate household travel data.

ACKNOWLEDGEMENT

This study was supported by Federal Highway Administration (FHWA). The authors would like to thank Ed Christopher, Elaine Murakami, Susan Liss, and Nancy McGuckin for their time and valuable comments and for sharing their various data sets for use in this study.

REFERENCES

1. Karasmaa, Nina. The spatial transferability of the Helsinki metropolitan area mode choice models, Paper presented at the 5th Workshop of the Nordic Research Network on Modelling Transport, Land-Use and the Environment, September 28-30, 2001, Nynäshamn.
2. Twining, Joanne. A Naturalistic Journey into the Collaboratory: In Search of Understanding for Prospective Participants. PhD dissertation, School of Library and Information Studies, Texas Woman's University, 1999.
3. Atherton, T.J., and M.E. Ben-Akiva. Transferability and Updating of Disaggregate Travel Demand Models, *Transportation Research Record*, 610, 1976, pp. 12-18.
4. Galbraith, R.A. and D.A. Hensher. Intra-Metropolitan Transferability of Mode Choice Models, *Journal of Transport Economics and Policy*, XVI(1), 1982, pp. 7-29.
5. Koppelman, F.S., G.K. Kuah, and C.G. Wilmot. Transfer Model Updating with Disaggregate Data, *Transportation Research Record*, 1037, 1985, pp.102-107.
6. Train, K.E., A Comparison of the Predictive Ability of Mode Choice Models with Various Levels of Complexity, *Transportation Research*, Vol. 13A, No. 1, 1979, pp. 11-16.
7. Badoe, D.A., and E.J. Miller. Modelling Mode Choice with Data from two Independent Cross-Sectional Surveys: an Investigation. *Transportation Planning and Technology*, Vol. 21, 1998, pp. 235-261.
8. Elmi, A.M., D.A. Badoe and E.J. Miller. Transferability Analysis of Work-Trip Distribution Models. Presented at the 78th Annual Meeting of Transportation Research Board, Washington D.C., January 1999.
9. Institute of Transportation Engineers. Trip Generation, 7th edition., *ITE*, Washington D.C., 2003, 1700 pp.
10. Martin, W.A., and N.A. McGuckin. NCHRP Report 365: Travel Estimation Techniques for Urban Planning, *TRB*, Washington DC, 1998.
11. Wilmot, C.G. and P.R. Stopher. Transferability of Transportation Planning Data. *Transportation Research Record*, 1768, 2001, pp. 36-43.
12. Reuscher, T.R., R.L. Schmoyer, and P.S. Hu. Transferability of Nationwide Personal Transportation Survey Data to Regional and Local Scales. *Transportation Research Record*, 1817, 2002, pp. 25-32.
13. Greaves, S.P. and P.R. Stopher. Creating a Synthetic Household Travel/ Activity Survey -Rationale and Feasibility Analysis. *Transportation Research Record* 1706, 2000, pp. 82-91.
14. Stopher P.R., P. Bullock, and S. Greaves. Simulating household travel survey data: Application to two urban areas, *82nd Annual Meeting Transportation Research Board*, Washington D.C., January 2003.
15. Stopher, P.R., S. Greaves, and M. Xu. Using Nationwide Household Travel Data for Simulating Metropolitan Area Household Travel Data, Paper presented at *TRB Conference on 2001 National Household Travel Survey*, Washington, D.C.,

November 2004.

16. Federal Highway Administration (FHWA). National Household Travel Survey (NHTS), version 4 datasets, <http://nhts.ornl.gov/2001/index.shtml> , accessed July, 2005.
17. Texas Transportation Institute. Urban Mobility Study - Key Mobility Measures, http://mobility.tamu.edu/ums/congestion_data/tables/national/table_1.pdf , accessed March 2005.
18. Stopher, P.R., and A.H. Meyburg. Survey Sampling and Multivariate Analysis for Social Scientists and Engineers, Lexington Books, USA, 1979.
19. Chiu, T., D. Fang, J. Chen, Y. Wang, and C. Jeris, A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*, 263, 2001.
20. Principe, J. C., N. R. Euliano, and W. C. Lefebvre. *Neural and Adaptive Systems: Fundamentals through Simulations*. John Wiley & Sons, New York, 2000.

LIST OF FIGURES

FIGURE 1 Intersection Density in Chicago Region.

FIGURE 2 Observed and Transferred Travel Statistics in Each Cluster.

LIST OF TABLES

TABLE 1 Results of Factor Analysis

TABLE 2 Cluster Memberships, Statistics, and Cluster Centers

TABLE 3 Means and Standard Deviations of Travel Characteristics in Each Cluster

TABLE 4 Travel Characteristics for Various Lifecycles in Each Cluster

TABLE 5 Comparisons of Observed vs. Transferred Travel Statistics

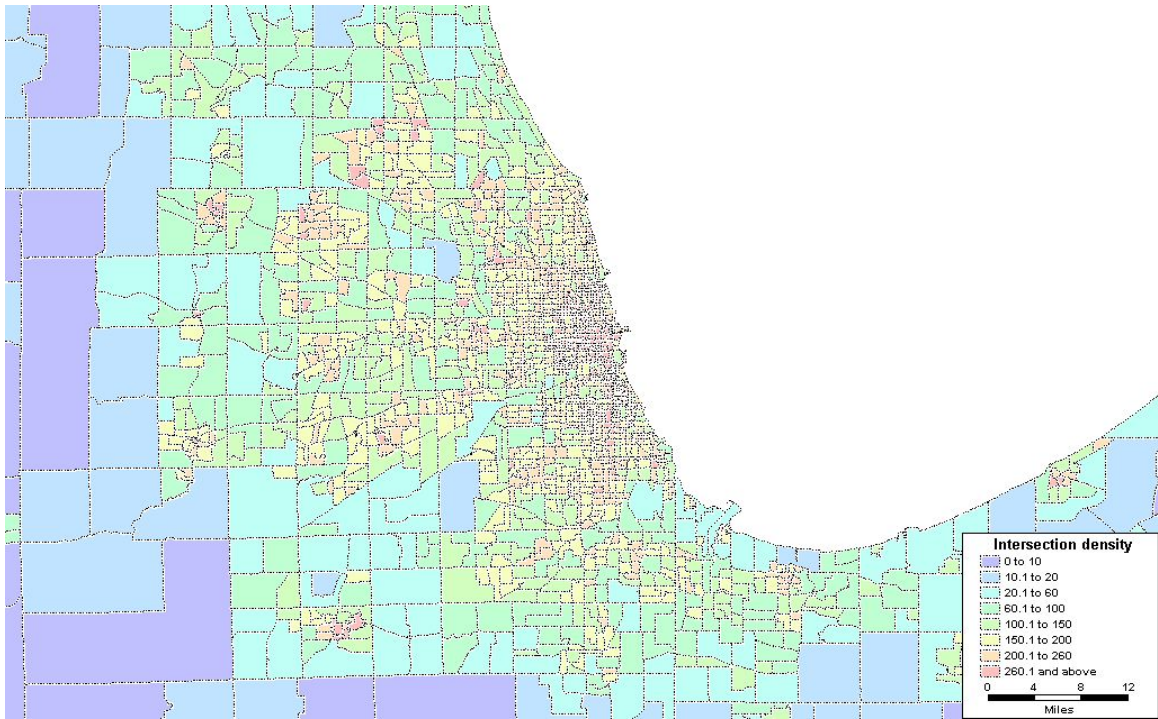


FIGURE 1 Intersection Density in Chicago Region.

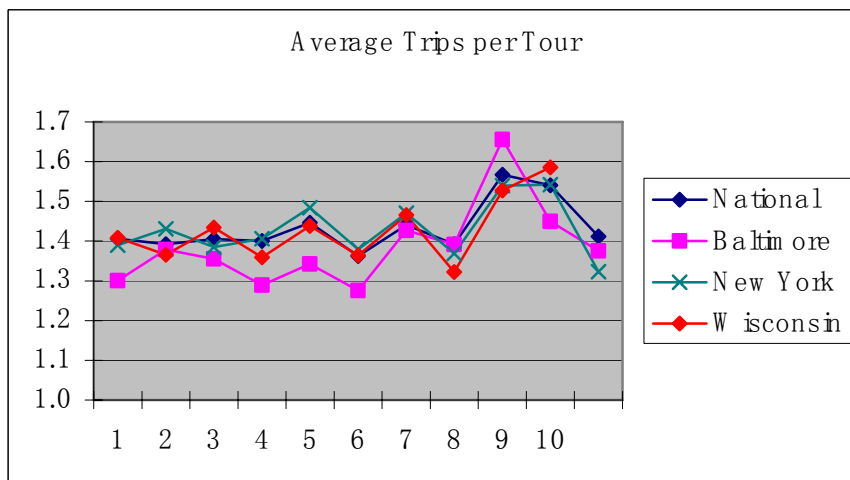
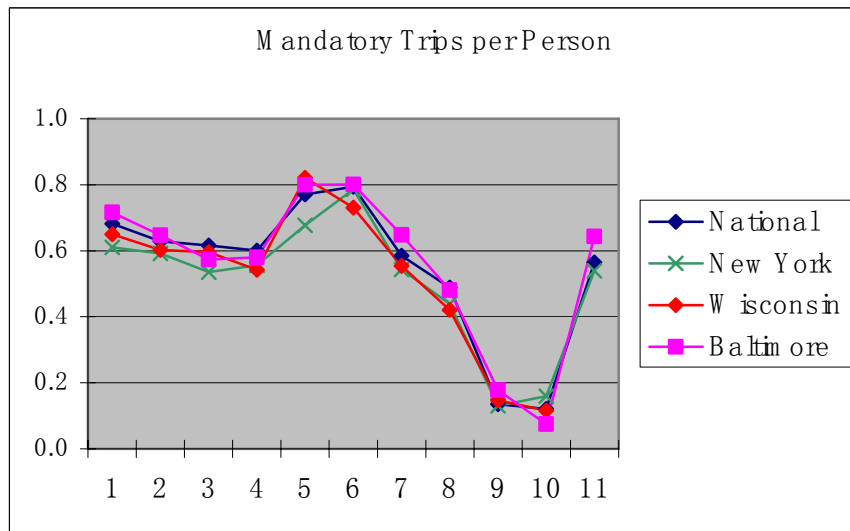
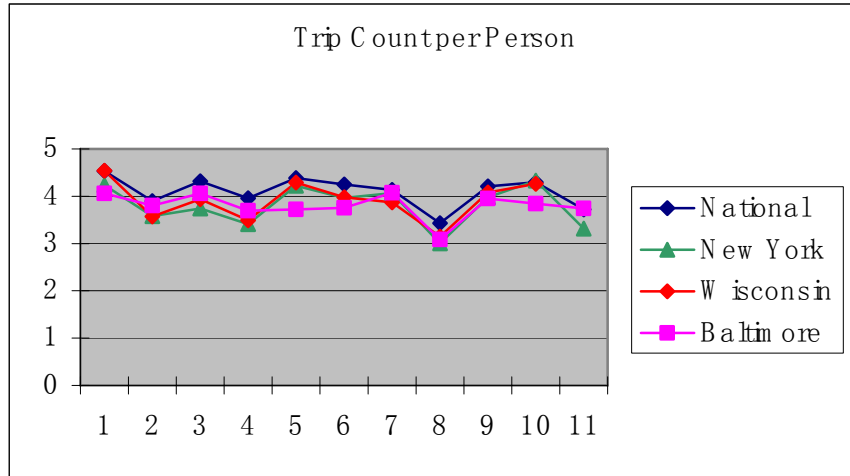


FIGURE 2 Observed and Transferred Travel Statistics in Each Cluster
 (note: National data is the transferred one).

TABLE 1 Results of Factor Analysis

Household Variables	Factors										
	High Density and Transit	Vehicles, Drivers, and workers	Middle Age Family	Highly Developed	Higher Income and Educated	Working Class	College Degree	Hispanic	Young Singles or Couples	Black or Asian Minorities	Seniors
Population Density in CT	0.953	-0.030	-0.009	0.185	-0.018	0.010	0.004	0.039	0.012	0.032	-0.004
Residential Density in CT	0.950	-0.049	-0.023	0.180	-0.010	-0.001	0.006	0.002	0.011	0.010	-0.007
Employment Density in CT	0.944	-0.032	-0.015	0.212	0.008	0.015	0.014	0.018	0.015	0.009	0.003
Transit Usage in CT	0.843	-0.051	-0.015	0.147	0.026	-0.001	0.000	0.011	0.002	0.109	0.000
No. of Drivers	-0.109	0.868	0.135	-0.035	0.160	0.063	0.022	0.017	-0.020	-0.083	0.101
No. of Adults	0.009	0.867	0.116	0.021	0.074	-0.030	-0.031	0.175	0.015	-0.031	0.159
No. of Workers	0.016	0.689	0.120	0.006	0.233	0.465	0.010	0.004	0.169	0.086	-0.112
No. of Persons Aged 16-24	0.036	0.632	-0.008	0.012	-0.108	0.063	-0.005	0.069	-0.086	0.154	-0.285
No. of Vehicles	-0.157	0.625	0.053	-0.171	0.184	0.163	0.027	-0.034	-0.075	-0.154	0.161
No. of Service Jobs	-0.028	0.543	-0.056	0.046	-0.199	-0.090	-0.042	-0.236	0.250	0.267	-0.229
No. of Persons Aged 0-15	-0.025	0.074	0.839	-0.016	0.064	0.036	0.022	0.062	0.275	0.039	-0.116
No. of Persons Aged 35-44	-0.019	0.043	0.760	0.008	0.112	0.169	-0.007	-0.020	-0.289	0.117	-0.009
Household Size	-0.005	0.584	0.678	0.001	0.084	0.029	0.003	0.160	0.204	0.055	-0.033
White Ethnicity	-0.093	0.449	0.546	-0.111	0.094	-0.054	0.013	-0.230	0.126	-0.517	-0.072
Block Size	-0.089	0.033	0.031	-0.828	-0.022	0.023	-0.018	0.043	-0.025	-0.017	0.043
Road Density in CT	0.515	-0.049	-0.030	0.749	-0.033	-0.002	-0.014	0.020	0.030	0.027	-0.040
Intersections Density in CT	0.524	-0.044	-0.025	0.742	-0.037	-0.001	-0.020	0.009	0.030	0.006	-0.040
Travel Time Index in MSA	0.254	0.001	0.045	0.544	0.178	0.045	0.074	0.061	-0.036	0.065	0.080
No. of Graduate Degrees	0.005	0.006	0.049	0.065	0.758	-0.205	-0.203	0.033	-0.040	0.042	-0.018
No. of Professional Jobs	0.023	0.135	0.085	0.007	0.733	0.109	0.211	-0.115	0.120	0.000	-0.063
Income	-0.040	0.249	0.097	0.105	0.626	0.194	0.187	-0.229	-0.043	-0.128	0.099
No. of Persons Aged 65+	-0.046	-0.062	-0.106	0.043	-0.244	-0.712	-0.026	0.066	-0.315	-0.208	0.119
No. of Clerical Jobs	-0.009	0.151	0.061	0.115	-0.189	0.515	0.036	-0.066	-0.222	-0.101	0.173
No. of Manufacturing Jobs	0.008	0.199	0.132	-0.169	-0.175	0.431	-0.239	0.407	0.021	-0.164	-0.004
No. of HS Diplomas	-0.036	0.203	0.225	-0.049	-0.229	0.099	-0.822	-0.125	0.071	-0.029	0.023
No. of College Degrees	-0.016	0.260	0.373	0.023	-0.132	0.127	0.799	-0.174	0.140	-0.027	-0.034
Less than HS Education	0.037	0.070	-0.053	-0.085	-0.110	-0.210	0.064	0.706	-0.055	0.020	-0.004
Hispanic Ethnicity	0.017	0.020	0.077	0.145	-0.039	0.119	-0.043	0.669	0.175	0.088	-0.062
No. of Persons Aged 25-34	0.035	0.042	0.100	0.032	0.029	0.059	0.029	0.110	0.876	-0.075	0.086
Black Ethnicity	0.106	0.032	0.083	-0.017	-0.076	0.009	-0.044	-0.021	-0.014	0.785	-0.020
Asian Ethnicity	0.004	0.048	0.072	0.134	0.133	0.008	0.101	0.122	-0.041	0.317	0.200
No. of Persons Aged 55-64	-0.014	0.103	-0.287	-0.064	0.023	0.103	-0.047	-0.091	0.024	0.057	0.746
No. of Persons Aged 45-54	-0.016	0.354	-0.286	-0.048	0.239	0.307	0.034	-0.012	-0.166	-0.093	-0.544

TABLE 3 Means and Standard Deviations of Travel Characteristics in Each Cluster

Cluster		No. of HH Trips	No. of Auto Trips	No. of Non-Motorized Trips	% of Transit	% of Carpool	Avg Commute Distance	No. of Tours	Avg Trip Per Tour	Avg Annual VMT	Mandatory Trips per Person	Maintenance Trips per Person	Discretionary Trips per Person	Auto Trips per Person	Transit Trips per Person	Non-Motorized Trips per Person	Tours per Person
1	Mean	14.84	12.90	1.47	0.03	0.05	19.43	10.77	1.41	13,008	0.68	1.27	1.10	3.96	0.05	0.44	3.28
	S.D.	8.47	11.74	3.18	0.11	0.19	132.95	9.43	0.40	11,984	0.70	1.10	0.95	2.06	0.21	0.74	1.45
2	Mean	12.95	11.50	1.10	0.02	0.07	15.90	9.47	1.39	13,450	0.63	1.09	0.91	3.47	0.04	0.32	2.87
	S.D.	8.17	11.08	2.95	0.11	0.23	36.80	8.81	0.44	14,713	0.63	1.02	0.91	1.92	0.22	0.68	1.39
3	Mean	17.02	14.88	1.43	0.02	0.04	13.53	12.26	1.41	13,501	0.62	1.20	1.04	3.79	0.03	0.36	3.11
	S.D.	9.69	13.67	3.13	0.09	0.16	40.93	10.30	0.37	12,627	0.64	1.05	0.92	2.07	0.14	0.63	1.44
4	Mean	15.68	13.69	1.24	0.02	0.07	12.54	11.41	1.40	12,705	0.60	1.08	0.93	3.47	0.04	0.31	2.89
	S.D.	8.86	12.78	2.72	0.10	0.23	22.21	9.73	0.44	14,721	0.57	0.96	0.90	1.94	0.24	0.57	1.34
5	Mean	5.62	4.97	0.51	0.04	0.04	15.26	4.01	1.45	11,309	0.77	1.21	1.05	3.87	0.08	0.40	3.12
	S.D.	3.49	4.67	1.46	0.18	0.18	59.75	3.39	0.60	13,337	1.02	1.32	1.24	2.65	0.43	0.99	1.72
6	Mean	13.06	11.94	0.83	0.02	0.07	14.30	9.68	1.36	12,655	0.79	1.09	0.95	3.91	0.03	0.26	3.14
	S.D.	7.68	11.01	2.31	0.09	0.21	47.65	8.67	0.39	12,233	0.74	1.07	0.92	2.09	0.18	0.56	1.41
7	Mean	9.16	8.42	0.63	0.02	0.04	15.77	6.56	1.44	10,995	0.59	1.19	1.01	3.80	0.03	0.29	2.96
	S.D.	5.31	7.65	1.85	0.10	0.18	47.41	5.75	0.53	12,107	0.79	1.16	1.02	2.21	0.19	0.70	1.56
8	Mean	12.38	10.40	1.29	0.05	0.10	12.09	9.14	1.39	10,998	0.49	0.97	0.77	2.92	0.09	0.34	2.51
	S.D.	8.24	11.76	3.43	0.15	0.25	21.55	9.74	0.45	13,200	0.54	0.97	0.83	2.01	0.34	0.68	1.37
9	Mean	8.91	8.11	0.70	0.01	0.04	10.49	6.06	1.57	8,606	0.14	1.41	1.21	3.84	0.01	0.32	2.86
	S.D.	4.95	6.86	1.79	0.09	0.18	20.18	4.95	0.68	8,484	0.40	1.25	1.07	2.27	0.14	0.69	1.57
10	Mean	4.61	3.95	0.51	0.04	0.02	8.60	3.21	1.54	5,972	0.12	1.45	1.25	3.69	0.07	0.47	2.99
	S.D.	2.72	3.53	1.35	0.19	0.15	14.03	2.64	0.77	8,278	0.46	1.50	1.25	2.58	0.42	1.00	1.64
11	Mean	6.58	3.39	2.03	0.27	0.05	21.71	4.90	1.41	7,504	0.56	1.05	0.90	1.88	0.64	1.19	2.72
	S.D.	4.27	4.80	3.35	0.39	0.17	174.94	4.69	0.51	12,759	0.73	1.19	1.03	2.18	1.07	1.63	1.60

TABLE 4 Travel Characteristics for Various Lifecycles in Each Cluster

Cluster	Life Cycle	Count	No. of HH Trips		No. of Mandatory		No. of Maintenance		No. of Discretionary		% of Transit		% of Carpool	
			Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	1	0
	2	340	9.98	4.79	1.83	1.93	1.83	1.93	2.67	2.38	0.09	0.37	0.06	0.28
	3	11	12.03	4.92	2.25	1.14	2.25	1.14	2.88	2.62	0.10	0.30	0.08	0.27
	4	256	17.02	9.47	2.23	2.17	2.23	2.17	4.89	4.49	0.09	0.37	0.11	0.42
	5	43	14.15	7.61	1.91	1.70	1.91	1.70	5.53	4.87	0.03	0.26	0.04	0.19
	6	386	17.73	8.73	2.41	2.11	2.41	2.11	4.82	4.19	0.12	0.39	0.12	0.38
	7	7	7.80	4.65	1.02	1.12	1.02	1.12	2.80	2.50	0.14	0.35	0.31	0.46
	8	91	16.50	7.87	2.80	2.40	2.80	2.40	3.95	3.36	0.32	0.65	0.07	0.25
	9	0
	10	61	11.10	5.61	0.87	1.29	0.87	1.29	3.40	2.37	0.05	0.25	0.04	0.18
2	1	0
	2	569	8.79	4.14	1.91	1.68	1.91	1.68	2.17	2.13	0.18	0.50	0.09	0.34
	3	43	13.38	8.34	1.49	1.39	1.49	1.39	4.75	5.37	0.04	0.19	0.00	0.00
	4	1043	15.00	8.92	1.91	1.96	1.91	1.96	4.53	4.23	0.11	0.35	0.06	0.27
	5	24	12.64	6.74	1.69	1.46	1.69	1.46	4.07	3.98	0.00	0.00	0.19	0.47
	6	249	14.45	8.45	2.30	2.10	2.30	2.10	3.66	3.86	0.09	0.32	0.09	0.41
	7	0
	8	5	13.93	8.47	2.35	2.41	2.35	2.41	3.27	2.92	0.00	0.00	0.00	0.00
	9	0
	10	28	8.22	4.19	0.89	0.83	0.89	0.83	2.54	2.14	0.05	0.22	0.03	0.17
3	1	0
	2	178	9.10	4.41	1.78	1.60	1.78	1.60	2.42	2.44	0.12	0.39	0.02	0.15
	3	29	14.79	7.66	1.75	2.09	1.75	2.09	4.83	3.67	0.00	0.00	0.01	0.17
	4	685	18.53	10.75	2.09	2.34	2.09	2.34	5.59	5.02	0.06	0.25	0.07	0.31
	5	117	13.12	6.66	2.10	2.13	2.10	2.13	3.40	2.73	0.03	0.17	0.05	0.29
	6	920	18.12	9.17	2.62	2.43	2.62	2.43	4.81	4.50	0.08	0.32	0.10	0.39
	7	0
	8	41	14.00	5.39	3.25	2.83	3.25	2.83	3.14	2.52	0.28	0.62	0.02	0.15
	9	0
	10	16	8.26	3.90	1.68	1.53	1.68	1.53	1.77	1.71	0.00	0.00	0.08	0.27
4	1	0
	2	108	8.50	5.17	1.58	1.62	1.58	1.62	2.24	2.29	0.28	0.61	0.08	0.27
	3	17	15.33	8.62	2.89	1.98	2.89	1.98	4.19	4.54	0.00	0.00	0.14	0.52
	4	327	17.53	10.12	2.01	1.98	2.01	1.98	5.20	4.97	0.14	0.45	0.08	0.32
	5	53	12.20	6.50	1.66	1.65	1.66	1.65	2.44	2.57	0.05	0.21	0.10	0.48
	6	597	16.80	8.07	2.71	2.34	2.71	2.34	4.54	3.89	0.11	0.37	0.11	0.41
	7	2	5.09	1.78	3.00	0.00	3.00	0.00	0.82	1.34	0.27	0.45	0.00	0.00
	8	57	12.51	6.58	2.02	1.87	2.02	1.87	3.14	3.31	0.44	0.82	0.03	0.18
	9	0
	10	27	11.30	9.25	1.25	0.93	1.25	0.93	3.99	5.44	0.24	0.57	0.17	0.68
5	1	2059	4.99	2.70	0.99	1.17	0.99	1.17	1.32	1.42	0.03	0.16	0.04	0.19
	2	859	6.24	4.11	1.24	1.53	1.24	1.53	1.72	1.98	0.10	0.35	0.07	0.26
	3	50	6.55	4.06	0.68	0.81	0.68	0.81	2.13	2.15	0.02	0.14	0.08	0.30
	4	70	5.39	3.36	0.65	0.94	0.65	0.94	1.57	1.85	0.08	0.27	0.02	0.15
	5	167	9.30	4.86	1.54	1.35	1.54	1.35	2.48	2.02	0.03	0.18	0.09	0.29
	6	117	6.19	4.10	0.92	1.15	0.92	1.15	1.93	2.14	0.04	0.20	0.02	0.14
	7	70	7.64	4.68	1.28	1.13	1.28	1.13	2.07	2.59	0.14	0.35	0.02	0.14
	8	20	7.59	4.88	0.64	0.71	0.64	0.71	2.06	1.46	0.35	0.51	0.00	0.00
	9	246	4.96	2.64	0.05	0.25	0.05	0.25	1.75	1.57	0.01	0.08	0.07	0.25
	10	137	7.75	4.40	0.49	0.77	0.49	0.77	2.66	2.49	0.06	0.27	0.04	0.20
6	1	0
	2	977	10.02	5.40	2.03	1.98	2.03	1.98	2.52	2.78	0.16	0.49	0.08	0.36
	3	6	10.49	9.83	1.52	1.53	1.52	1.53	2.93	3.68	0.28	0.45	0.00	0.00
	4	117	13.30	7.98	2.13	1.89	2.13	1.89	3.58	3.50	0.16	0.42	0.01	0.09
	5	12	14.66	10.93	2.43	1.51	2.43	1.51	4.13	3.56	0.11	0.31	0.15	0.36
	6	456	17.65	9.31	2.89	2.54	2.89	2.54	4.56	3.96	0.16	0.44	0.08	0.36
	7	83	11.00	4.89	2.09	1.84	2.09	1.84	3.01	2.41	0.26	0.52	0.03	0.17
	8	645	14.76	7.50	2.80	2.41	2.80	2.41	3.61	3.35	0.22	0.54	0.07	0.27
	9	0
	10	102	10.67	5.65	1.48	1.57	1.48	1.57	3.15	2.99	0.15	0.50	0.01	0.14

TABLE 4 Travel Characteristics for Various Lifecycles in Each Cluster, cont.

Cluster	Life Cycle	Count	No. of HH Trips		No. of Mandatory		No. of Maintenance		No. of Discretionary		% of Transit		% of Carpool	
			Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
7	1	59	4.54	2.65	0.90	1.13	0.90	1.13	0.97	0.97	0.05	0.22	0.05	0.22
	2	1282	8.90	5.17	1.70	1.85	1.70	1.85	2.33	2.43	0.08	0.33	0.04	0.21
	3	0												
	4	38	10.70	7.48	2.09	2.72	2.09	2.72	2.67	2.45	0.08	0.27	0.07	0.26
	5	6	12.82	1.81	2.79	2.15	2.79	2.15	1.67	1.59	0.00	0.00	0.20	0.40
	6	66	11.27	6.75	1.56	1.85	1.56	1.85	3.41	3.17	0.10	0.36	0.11	0.39
	7	10	11.30	5.76	0.85	1.10	0.85	1.10	4.39	3.01	0.00	0.00	0.25	0.66
	8	77	12.38	6.36	1.69	1.59	1.69	1.59	3.88	3.08	0.17	0.44	0.05	0.22
	9	7	4.68	1.88	0.00	0.00	0.00	0.00	1.81	0.95	0.00	0.00	0.00	0.00
	10	916	9.30	4.91	0.66	1.27	0.66	1.27	3.04	2.61	0.04	0.19	0.04	0.21
8	1	5	5.03	3.29	1.04	1.57	1.04	1.57	1.63	1.22	0.37	0.48	0.14	0.35
	2	317	8.61	5.54	1.69	1.63	1.69	1.63	2.24	2.45	0.31	0.69	0.13	0.40
	3	71	13.88	9.54	1.68	1.85	1.68	1.85	4.33	4.60	0.10	0.30	0.33	0.82
	4	522	14.95	8.99	1.94	2.12	1.94	2.12	4.40	4.38	0.20	0.47	0.22	0.66
	5	75	11.60	7.35	1.59	1.58	1.59	1.58	2.91	3.15	0.13	0.34	0.30	0.62
	6	494	15.12	8.64	2.55	2.16	2.55	2.16	3.93	4.01	0.16	0.44	0.22	0.65
	7	32	9.28	4.73	1.59	1.83	1.59	1.83	2.36	2.20	0.21	0.41	0.33	0.60
	8	145	12.96	6.96	2.26	2.44	2.26	2.44	3.41	2.86	0.25	0.55	0.13	0.45
	9	1	4.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
	10	440	7.54	4.69	0.50	0.89	0.50	0.89	2.41	2.32	0.05	0.24	0.07	0.28
9	1	0												
	2	127	9.09	5.18	1.24	1.72	1.24	1.72	2.59	2.70	0.07	0.29	0.00	0.00
	3	0												
	4	11	13.27	7.94	1.13	1.28	1.13	1.28	3.77	3.82	0.00	0.00	0.00	0.00
	5	0												
	6	26	12.11	7.12	1.16	1.31	1.16	1.31	4.18	4.12	0.13	0.34	0.17	0.38
	7	0												
	8	14	11.78	4.50	0.72	0.70	0.72	0.70	2.94	1.81	0.14	0.34	0.10	0.30
	9	1	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	10	1869	8.79	4.83	0.23	0.74	0.23	0.74	2.97	2.59	0.01	0.11	0.03	0.19
10	1	274	4.21	2.52	0.47	0.91	0.47	0.91	1.12	1.47	0.01	0.10	0.08	0.27
	2	34	4.71	2.74	0.75	0.97	0.75	0.97	1.15	1.28	0.00	0.00	0.11	0.41
	3	5	8.59	8.13	0.48	0.86	0.48	0.86	2.66	2.53	0.00	0.00	0.00	0.00
	4	8	4.57	3.92	0.20	0.40	0.20	0.40	2.13	2.82	0.20	0.40	0.00	0.00
	5	23	6.27	4.54	0.91	0.92	0.91	0.92	1.43	1.91	0.00	0.00	0.07	0.25
	6	4	8.03	3.06	1.54	2.02	1.54	2.02	0.72	0.45	0.00	0.00	0.00	0.00
	7	8	5.00	3.13	0.44	0.62	0.44	0.62	2.06	1.67	0.00	0.00	0.00	0.00
	8	1	2.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
	9	1205	4.57	2.47	0.02	0.18	0.02	0.18	1.62	1.53	0.00	0.00	0.03	0.17
	10	133	5.27	3.71	0.28	0.75	0.28	0.75	1.97	2.01	0.02	0.12	0.02	0.17
11	1	182	4.92	2.59	0.81	0.94	0.81	0.94	1.38	1.47	0.01	0.09	0.36	0.48
	2	189	7.62	4.18	1.43	1.55	1.43	1.55	2.03	1.99	0.17	0.44	0.50	0.83
	3	14	7.15	3.37	0.79	1.02	0.79	1.02	1.06	0.99	0.00	0.00	0.40	0.60
	4	50	8.31	5.05	1.54	1.36	1.54	1.36	2.66	2.42	0.16	0.44	0.74	0.88
	5	20	8.94	5.04	1.04	1.08	1.04	1.08	2.88	2.00	0.00	0.00	1.12	0.99
	6	43	8.40	5.59	1.45	1.73	1.45	1.73	1.97	2.30	0.08	0.33	0.62	0.71
	7	9	5.74	3.27	1.57	0.77	1.57	0.77	0.61	1.00	0.12	0.33	0.60	0.74
	8	23	11.16	4.51	1.99	1.89	1.99	1.89	3.13	2.24	0.10	0.30	0.86	0.91
	9	56	3.83	2.65	0.05	0.28	0.05	0.28	1.31	1.48	0.00	0.00	0.30	0.46
	10	46	6.01	4.05	0.70	1.06	0.70	1.06	1.79	1.68	0.00	0.00	0.36	0.50

TABLE 5 Comparisons of Observed vs. Transferred Travel Statistics

Cluster	Trip Counts per Person								Mandatory Trips per Person							
	National		Wisconsin		New York		Baltimore		National		Wisconsin		New York		Baltimore	
	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.
1	4.54	0.1%	4.54	-6.6%	4.24	-10.6%	4.06	0.68	0.65	-4.6%	0.61	-10.5%	0.72	5.0%		
2	3.90	-8.4%	3.57	-8.2%	3.58	-2.5%	3.80	0.63	0.60	-4.1%	0.59	-5.9%	0.65	3.0%		
3	4.32	-8.9%	3.93	-13.3%	3.74	-6.0%	4.06	0.62	0.60	-3.2%	0.54	-13.1%	0.57	-6.9%		
4	3.96	-11.8%	3.49	-13.8%	3.41	-6.7%	3.69	0.60	0.54	-9.6%	0.56	-7.3%	0.58	-3.4%		
5	4.38	-2.2%	4.29	-3.7%	4.22	-15.1%	3.72	0.77	0.82	6.5%	0.68	-12.1%	0.80	3.7%		
6	4.25	-6.4%	3.98	-6.8%	3.96	-11.7%	3.75	0.79	0.73	-7.9%	0.79	-1.0%	0.80	0.9%		
7	4.13	-6.4%	3.87	-1.6%	4.07	-1.4%	4.08	0.59	0.55	-5.3%	0.54	-7.1%	0.65	10.7%		
8	3.43	-8.1%	3.15	-12.3%	3.01	-10.0%	3.09	0.49	0.42	-14.0%	0.44	-10.2%	0.48	-1.6%		
9	4.21	-3.0%	4.08	-5.5%	3.98	-6.2%	3.95	0.14	0.15	8.2%	0.13	-3.6%	0.18	30.9%		
10	4.30	-0.8%	4.26	0.7%	4.33	-10.5%	3.85	0.12	0.12	-4.0%	0.16	32.8%	0.07	-37.8%		
11	3.72	-	-	-10.9%	3.32	0.6%	3.74	0.56	-	-	0.54	-4.5%	0.64	14.0%		

Cluster	Maintenance Trips per Person								Discretionary Trips per Person							
	National		Wisconsin		New York		Baltimore		National		Wisconsin		New York		Baltimore	
	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.
1	1.27	-5.9%	1.19	-12.5%	1.11	-33.1%	0.85	1.10	1.23	11.3%	1.04	-5.4%	1.03	-6.4%		
2	1.09	-17.4%	0.90	-9.0%	1.00	-13.4%	0.95	0.91	0.87	-4.3%	0.81	-11.0%	0.88	-3.1%		
3	1.20	-14.0%	1.04	-16.1%	1.01	-19.6%	0.97	1.04	0.95	-8.7%	0.88	-15.4%	1.09	4.7%		
4	1.08	-23.9%	0.82	-11.2%	0.96	-14.7%	0.92	0.93	0.90	-3.0%	0.71	-23.7%	0.80	-13.6%		
5	1.21	-8.4%	1.11	-0.3%	1.20	-28.5%	0.86	1.05	1.04	-1.0%	1.01	-4.2%	0.81	-22.9%		
6	1.09	-9.2%	0.99	-13.1%	0.95	-34.3%	0.71	0.95	0.90	-6.2%	0.86	-10.1%	0.79	-17.4%		
7	1.19	-14.3%	1.02	-6.3%	1.12	-9.3%	1.08	1.01	1.01	0.2%	1.02	1.0%	0.95	-5.7%		
8	0.97	5.9%	1.03	-14.6%	0.83	-23.0%	0.75	0.77	0.56	-27.3%	0.66	-14.3%	0.73	-5.3%		
9	1.41	-7.2%	1.31	-6.1%	1.32	-14.1%	1.21	1.21	1.18	-2.0%	1.11	-8.2%	1.19	-1.5%		
10	1.45	0.7%	1.46	0.5%	1.46	-20.4%	1.16	1.25	1.22	-2.5%	1.25	0.3%	1.17	-6.6%		
11	1.05	-	-	-13.9%	0.90	-4.1%	1.01	0.90	-	-	0.73	-18.5%	0.83	-8.1%		

Cluster	Tours per Person								Average Trips per Tour							
	National		Wisconsin		New York		Baltimore		National		Wisconsin		New York		Baltimore	
	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.
1	3.28	-0.5%	3.26	-6.0%	3.08	-4.9%	3.12	1.41	1.41	0.2%	1.39	-1.2%	1.30	-7.5%		
2	2.87	-7.2%	2.66	-10.5%	2.57	-2.5%	2.80	1.39	1.37	-1.9%	1.43	2.7%	1.38	-1.0%		
3	3.11	-8.9%	2.84	-12.0%	2.74	-2.0%	3.05	1.41	1.43	2.1%	1.39	-1.5%	1.36	-3.6%		
4	2.89	-8.7%	2.64	-15.1%	2.45	-2.0%	2.83	1.40	1.36	-2.9%	1.41	0.4%	1.29	-7.9%		
5	3.12	-1.3%	3.08	-5.1%	2.96	-11.3%	2.77	1.45	1.44	-0.5%	1.48	2.6%	1.34	-7.2%		
6	3.14	-6.4%	2.94	-7.6%	2.90	-5.7%	2.97	1.36	1.36	0.1%	1.38	1.2%	1.28	-6.4%		
7	2.96	-6.8%	2.76	-2.6%	2.88	0.6%	2.98	1.44	1.47	1.7%	1.47	2.1%	1.43	-1.0%		
8	2.51	-5.2%	2.38	-11.4%	2.23	-9.1%	2.28	1.39	1.32	-5.0%	1.37	-1.7%	1.39	0.0%		
9	2.86	-1.1%	2.83	-4.9%	2.72	-8.0%	2.63	1.57	1.53	-2.5%	1.54	-1.7%	1.66	5.7%		
10	2.99	-1.7%	2.94	0.1%	2.99	-7.8%	2.76	1.54	1.59	2.9%	1.54	0.1%	1.45	-5.9%		
11	2.72	-	-	-7.2%	2.53	2.8%	2.80	1.41	-	-	1.32	-6.2%	1.38	-2.6%		

Cluster	Auto Trips per Person								Average Daily Travel Distance							
	National		Wisconsin		New York		Baltimore		National		Wisconsin		New York		Baltimore	
	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.	Mean	Diff.
1	3.96	-1.3%	3.91	-11.6%	3.50	-11.3%	3.52	51.88	39.78	-23.3%	36.77	-29.1%	47.47	-8.5%		
2	3.47	-6.4%	3.25	-10.9%	3.09	-2.5%	3.39	40.00	35.19	-12.0%	31.39	-21.5%	47.81	19.5%		
3	3.79	-9.3%	3.44	-16.9%	3.15	-9.6%	3.43	42.50	35.22	-17.1%	31.07	-26.9%	39.24	-7.7%		
4	3.47	-13.1%	3.02	-20.8%	2.75	-10.7%	3.10	36.79	36.52	-0.7%	32.70	-11.1%	30.08	-18.2%		
5	3.87	-2.0%	3.79	-10.1%	3.48	-20.5%	3.08	44.73	38.58	-13.8%	44.34	-0.9%	44.09	-1.4%		
6	3.91	-7.0%	3.63	-12.9%	3.40	-12.2%	3.43	43.90	49.49	12.7%	41.78	-4.8%	39.40	-10.3%		
7	3.80	-5.4%	3.59	-4.8%	3.61	0.2%	3.80	46.33	53.31	15.1%	40.15	-13.3%	41.91	-9.5%		
8	2.92	-9.8%	2.63	-37.3%	1.83	-28.0%	2.10	31.54	24.94	-20.9%	16.82	-46.7%	22.27	-29.4%		
9	3.84	-3.7%	3.63	-5.5%	3.63	-5.0%	3.65	37.67	40.13	6.5%	29.15	-22.6%	30.35	-19.4%		
10	3.69	-1.9%	3.62	0.0%	3.69	-22.5%	2.86	28.99	35.58	22.7%	29.97	3.4%	21.11	-27.2%		
11	1.88	-	-	-40.7%	1.12	20.1%	2.26	39.16	-	-	16.89	-56.9%	22.28	-43.1%		