Summary Report for the Peer Exchange on Data Transferability

Held December 16, 2004
Washington, DC
Keck Center of the National Academies

Sponsored by the
Federal Highway Administration Travel Model Improvement Program (TMIP) and
TRB Committees on Urban Transportation Data and Information Systems (ABJ30), Traveler Behavior and Values (ADB10) and National Transportation Data Requirements and Programs (ABJ10)

Prepared by the Volpe National Transportation Systems Center
U.S. Department of Transportation

Table of Contents

I. Executive Summary

The following report summarizes the results of a peer exchange on data transferability. The exchange was organized and sponsored by the Travel Model Improvement Program (TMIP) and co-sponsored by the TRB Committees on Urban Transportation Data and Information Systems (ABJ30), Traveler Behavior and Values (ADB10) and National Transportation Data Requirements and Programs (ABJ10). The exchange brought together representatives of state and local departments of transportation (DOTs), metropolitan planning organizations (MPOs), academics, and transportation consultants. Related work in the area of data transferability was presented. Participants then discussed data transferability issues structured around a set of questions prepared prior to the meeting. A common theme throughout the day was that a great deal of further research is needed to progress the field of data transferability so that more widespread data sharing can occur. To this end, the group produced a set of research topics which would be beneficial to data transferability for USDOT to further pursue.

II. The Travel Model Improvement Program Peer Exchange

The Travel Model Improvement Program (TMIP), begun in 1992 under the sponsorship of the U.S. DOT and the EPA, provides resources to planning agencies in their efforts to improve their travel analysis techniques. TMIP has three goals:

1. Help build the institutional capacity of planning agencies to perform technical analysis,
2. Support development of analytical methods that respond to the needs of planning and environmental decision making processes, and
3. Support mechanisms to ensure the quality of technical analysis used to meet local, state and federal program requirements.

As part of its effort to meet these goals, TMIP sponsors peer exchanges that bring together representatives of state and local departments of transportation (DOTs), metropolitan planning organizations (MPOs), academics, and transportation consultants to discuss important issues being faced in the travel demand modeling field. The goal of the current peer exchange was to lay down the issues associated with data transferability for travel demand modeling and identify ways the FHWA can help improve the transferability of data across the country. The gathering, held on December 16, 2004 in Washington, DC, brought together fifteen individuals who have experience with or are in need of transferable data. Tom Rossi of Cambridge Systematics moderated the peer exchange, assisted by three representatives from TMIP. A representative from the U.S. DOT Volpe National Transportation Systems Center assisted in documenting the exchange proceedings. A list of all participants is provided in the Appendix C.

Four participants made presentations on their past and current research relating to data transferability. The presentations were followed by a panel discussion of issues relating to and affecting data transferability. At the end of the day, the group developed a set of research topics that FHWA could pursue for further development.

The peer exchange had three objectives:

1. To present existing research in the area of data transferability

2. To discuss issues in data transferability
3. To write scopes of work for research which would help progress the field of data transferability

III. Background on Data Transferability

This peer exchange assembled experts in travel demand modeling and data collection and analysis to discuss and debate the issue of data transferability for four-step travel demand models and activity-based models. Data transferability is important because primary data collection for household surveys for travel demand modeling is expensive and time consuming. Planning agencies are eager to minimize their costs by borrowing or transferring data from others, to the extent possible. Additionally, emerging practices like activity- and tour-based models and micro-simulation approaches require additional and more specialized data on travel and activity patterns. Data transferability is currently a term used loosely to refer to using data or products derived from data (models, equations, parameters) for modeling situations other than those the data was collected in such using existing data to forecast into the future (temporal transfer) or using a model from one region to model aspects of travel behavior in another region (spatial transfer). It might mean applying nationally collected data to local areas, applying regional survey data to other regions, or using data collected in the past to model the future.

The primary goal of this meeting was to discuss ways that data collected in one region can be used for travel analysis in other regions. A broader objective was to identify an action plan for research and develop project scopes for advancing the concepts of data transferability. This, in turn, will support improvement of travel models; facilitate the implementation of activity, tour-based and micro-simulation models; and stretch limited data collection resources. Another objective was to solicit input for the National Household Travel Survey (NHTS) Transferability project to be conducted by Oak Ridge National Laboratory and to review the work of others in the field.

IV. Presentations

Presentation: National Household Travel Survey Data for Benchmarking and Transferred Data

Nancy McGuckin

Nancy McGuckin is one of the authors of *NCHRP Report 365, "Travel Estimation Techniques for Urban Planning"* which updated *NCHRP Report 187*, "Quick Response Urban Travel Estimation Techniques and Transferable Parameters" when micro-computers replaced large computers, at which time travel demand modeling completely changed. NCHRP Report 365, which reviews the fundamentals of the four-step travel demand process and provides transferable parameters for use, is an invaluable resource to travel demand modelers.

The NCHRP team was tasked in 1994 to look at existing work in travel demand modeling by surveying planning agencies, review existing manuals, and collect data from different sources. The team discovered that there exists a tremendous amount of difference in the way surveying and modeling was done from one model to the next. Definitions were different, weighting

techniques were different, etc. Thus, the NCHRP 365 team determined that it was best to use the 1990 National Personal Transportation Survey (NPTS) for calculating household trip rates by metropolitan area size and trip purpose, vehicle occupancy and trip purposes by time of day. *NCHRP Report 365* outlined step-by-step the best practices for four-step travel demand modeling, and walked the user through a case study by transferring parameters, many of which derived from the 1990 NPTS, provided in the document.

The target market for data transferability is small areas which are not able to collect their own data. However, it is believed that even the large areas are not collecting the data that fully explains travel, because models are still not able to explain even half of the difference in, for example, household trip purposes. Thus until more research is conducted on the data that should be collected, the success of data transfer is still limited. The first step to this research is to collect data on more variables, which would allow for study of additional relationships in travel demand modeling.

Presentation: Can Nationwide Personal Transportation Survey Data be Used by Small Communities?
Pat Hu, Center for Transportation Analysis - Oak Ridge National Laboratory

The NPTS is designed as a national survey, therefore using it indiscriminately to represent an area smaller than the U.S. total can result in unreliable estimates. To address this weakness, Oak Ridge National Laboratories developed a system to use NPTS data to estimate travel behavior for areas smaller than the U.S. total.

The method uses predictor variables (variables shown to have significant explanatory power) for household daily trips, income limits, and area type classifications to form homogenous census tract clusters. The census tract data can then be used to bridge to the NPTS data using household size and number of vehicles. Vehicle trips, vehicle miles traveled, person trips, and person miles traveled for each cluster can then be calculated using NPTS estimated rates for household size and number of vehicles.

This census tract clustering approach was tested using the 1995 NPTS and applied to four regions and states, New York, Massachusetts, Oklahoma, and Baton Rouge. The tract-level results were compared to NPTS add-on data for that specific location or to independent survey data. For these particular add-on surveys, the performance of the ORNL census tract method was better than the competing methods which used larger geographic definitions such as Metropolitan Statistical Area (MSA), Census Division, or Census Regionto differentiate household travel demand.

The tool they have created for calculating cluster-based estimates is available at http://cta.ornl.gov/NPTS.

The new ORNL project will use the 2001 NHTS dataset and see if the results are similar and if they can be improved.

Presentation: Travel Related Inputs Model for Mobile 6.x (TRIMM)

Mohan Venigalla, University of Illinois at Chicago

Dr. Venigalla presented his research on using NPTS data as inputs to MOBILE6. When modeling regional emissions, modelers prefer actual local data over default MOBILE6 parameters because of the implications for affect air quality attainment status. The team at George Mason University was able to derive some of the local variables from NPTS that MOBILE6 requires. The subsamples are limited to large geographic areas, such as Census Region, MSA Size, and State (if sample size is sufficient). They found that in general, NPTS data looks very similar to MOBILE6 default data. However, there are significant regional variations in several variables of the NPTS which have been found to affect emission factors. Because some data gaps were found and data analysis was complex, Dr. Venigalla's team created a tool, Travel Related Inputs Model for Mobile 6.x (TRIMM), which automates the process of using NPTS data to generate MOBILE6 inputs.

TRIMM mines the NPTS data and provides as many MOBILE6 inputs as possible from the data. It then allows the user to compare each variable to those of other geographic regions and to MOBILE6 defaults, and gives the user the capability of choosing which variables to write to the MOBILE6 input format.

It was noted that National Household Travel Survey (NHTS) data does not have location or distance information, so GPS survey information would be extremely useful in the next NHTS survey.

Presentation: Data Transferability: Idealism or Realism
Kouros Mohammadian

Dr. Mohammadian presented the results of his literature review on data transferability.

The main goals of transferability are:

- To use data collected in one context in a new context.
- To reduce or eliminate the need for a large data collection effort for model development in the application context.

Data transferability has been a subject of numerous studies. He found two existing methods for transferring travel data. The first group of studies address the transferability of travel demand models in either a spatial context or temporally. A second group of studies address transferring of data, of which the ITE trip generation data is the most commonly transferred. These two types of data transfer are aggregate data transfer methods.

A third method of data transferability--data simulation—is being performed at the University of Illinois at Chicago (UIC). Data simulation can be used at a local level where NHTS and NPTS sampling is small and inadequate. Data simulation combines local socio-demographic data for individuals and households from sources such as a census with probability distributions of activity and travel patterns from other travel surveys such as NHTS to simulate local travel survey data. Essentially, data simulation expands any existing data (with all the inherent biases) to a wider population. The team at UIC plans to develop a framework to facilitate transferability

of household survey data for calibrating and validating travel forecast models. To achieve this goal, they will be researching more detailed classifications of NHTS and census data using advanced clustering schemas since homogeneous groups will improve the simulation.

V. Peer Exchange Discussions

Topics for discussions were developed by FHWA and prioritized by the panel members prior to the meeting. Thirty minutes were allotted for discussion on each topic.

1. What are your ideas about data transferability for travel demand/activity models? How do you define "transferability" spatially and temporally?

Panelists agreed that there are several **"layers" of transferability** each of which have the potential to be transferred. These layers are:

1) A conceptual layer (which consists of the modeling structure or mechanisms)
2) The parameters layer
3) The outcomes layer (e.g. trip rates)

Data is typically transferred due to a lack of resources (financial or expertise limitations) to develop models that are unique to local circumstances. However, data transfer also requires resources. Unfortunately, variables are often transferred with very little understanding or insufficient analysis. Although it is always possible to transfer data from one model to another, it may not always be technically valid to do so. It is not clear how often inappropriate transfer of data occurs.

Models vary across regions and, subsequently, so do their data requirements. Smaller areas may only require **aggregate numbers or average values** for transfer, such as aggregate trip rate, while other models—typically those done for large urban areas with complex tour-based models—require extremely disaggregate data. What is of greater interest to planning agencies in larger regions are whether the determinants of **travel behavior** (e.g. household composition, disposable income, proximity to shopping opportunities) **are really getting captured** and **whether models from different areas are similar enough for transferability** to other regions.

Participants felt that data transferability guidelines would be helpful for the entire travel demand modeling community for preventing technically invalid data transfers while encouraging proper data transfer. Useful guidance would be **standards for transferability of data**, which lay out criteria and guidelines on what data is transferable, define a correct method to conduct data transfer, and provide a method for measuring whether data transfer was performed successfully and correctly (beyond data matching).

Some areas of travel demand modeling and data transferability will require more research before a "correct" or recommended method can be identified. One area that needs more research is identifying "**pre-conditions**" that signal the appropriateness of data transfer, e.g. the existence of data on the availability of transportation systems and alternatives or transportation system measures, such as data on transit or pedestrian friendliness. Since there are situations where data

transfer should not be applied, identifying these pre-conditions would help to identify these situations.

Although data such as trip rates, trip generation, and process tables are transferred frequently, there is little research that shows that such transferability is actually valid. Some participants related their own experiences with models as anecdotal support of the concept of model transferability. For example, a participant who is a consultant to many large MPOs said that from his experience in several cases two models or regions that initially look substantially different, and therefore do not appear to be good candidates for transferability, have actually become sufficiently similar or almost identical when properly scaled. But an important task for the travel demand modeling community is to **validate the current state-of-practice of data transferability, both spatially and temporally**. This is discussed in more detail under question #3.

Research that helps to explain travel behavior will also facilitate data transferability analyses. For example, it maybe possible to **determine whether certain variables or types of variables are more transferable than others**. Some participants speculated that variables which dominate travel behavior, may be more transferable. This is discussed in more detail in question #3.

Research should also be conducted to determine **whether the model constant is capturing (i.e. hiding) some of the regional characteristics.** If so, explicitly breaking them out would facilitate transferability. Such research would require combining data from multiple regions which is discussed under question #7. Participants suggested adding the following variables to capture regional variability: seasonal weather, condition of the transportation system (e.g. level of congestion, road condition), and land use variables (e.g. distance to employment, entertainment, and retail stores).

Land use variables are very important in travel demand modeling. However different regions have different definitions of the variables so these variables are typically not directly transferable. The first step to capturing the differences is to introduce quantifiable variables into the model rather than using central business district (CBD) dummies.

Travel demand modeler are generally interested in transferability of three different types of variables:

1) Headline Variables – variables which are easy to categorize (e.g. household size)
2) Underlying Variables – e.g. household composition, presence of children of a certain age
3) **Variables which represent Hypothetical Measurements of Demand** – e.g. household score representing maintenance required for children

This third type of variable requires more work to generate. But when identified and defined correctly, they can be highly explanatory variables that capture a great deal of travel behavior. If these hypothetical measurements of demand are properly identified, it is possible that in addition to improving travel demand modeling across the board, models would become more transferable.

However, it should be noted whenever adding variables to a model is being considered, that it is necessary that the variable will be available or can be estimated for the region where or time period when the transferred model will be applied. Therefore, the additional **variables should be limited as often as possible to universal variables, particularly those that are obtainable through census data**.

2. **For which types of applications does data transferability already occur, and how has the transfer been achieved? Was it successful? What applications have data/parameters that are not typically transferred currently, but might be difficult to originally estimate due to future data limitations? What are some new/different applications for which transferability of data/parameters might work?**

Data transfer is often applied when forecasting the effect of **policies such as carpool lane requirements or infrastructure such as new modes that do not currently exist in a region**. **Pricing and toll** policies are currently being considered in most new travel demand models, even though very few cities have real pricing data. Although it is useful to transfer data from other regions for modeling these new effects, it is important to also collect local stated preference (SP) data to capture any regional characteristics.

Another prerequisite for successful data transfer is that the **data set and the target data set be comparable**. To determine if two data sets are comparable, one should combine the data sets and perform usability and reasonability tests, such as testing whether a variable works the same way before and after the data set combination. Although one goal of transferring data is to save money, agencies must invest some resources into **analyzing data before transferring** it to another application. For example, one common mistake is to overlook scaling of the data [XX - someone please clarify what kind of scaling is meant here—This is what I think it means—For example, Population density is a nice variable, but what if you just use "low, medium and high" Then, this could mean different values for different metropolitan areas. For example 2000 persons per sq mil in Manhattan would be LOW, but would be HIGH in Kansas City. At least population density is easily quantifiable, but what if you have a variable called "pedestrian accessibility" what does it mean?]. It would beneficial to have an outline of some **basic requirements for modelers to test whether data is comparable**.

To get a sense of what data are currently available, participants were asked to list household surveys of which they are aware. Table 1 contains the list of surveys.

Table 1 - Agency Household Surveys

Agency	Last household survey	Comments
Capitol Region Council of Governments, CT	1976	Trip generation model adjusted based on 1976 data collection effort.
Chicago	1970 and 1990	Would require $20M to do another survey

Iowa DOT	early 1970s (Des Moines add-on)	
Salt Lake City	1970s and ??	
Oregon DOT	1996	
New York Metropolitan Transportation Council	1963, 1995 and 2001 NPTS add-ons, 1996, and one scheduled for 2005/6	
North Central Texas Council of Governments	Dallas-Fort Worth home interview in 1964, true household survey in 1984 and 1996, next one will be 2007/8	
St. Louis	Every 10 years	
San Francisco (Purvis?)	Surveys regularly on the scale of 15,000 households ??, 1990, 1981, 1965	
Montreal	Every 5 years	
Tampa	1996, 2000/1, 2005/6	

The participants noted that very few agencies have the resources to conduct a household survey every 10 years, while most small agencies are lucky to have the opportunity to conduct any surveys at all.

A new application for transferability is to the **emerging markets of statewide models.** Statewide models will require additional data collection on rural, small urban, and long distance trip information, creating additional opportunities for data transfer between states and regions. Subregional models are also growing although they will require analysis methods that can be done with minimal resources.

Another important potential use for transferred data is for **air quality modeling**. Air quality modeling currently uses a great deal of data from state department of motor vehicles (DMVs). One participant suggested validating the DMV data with NPTS data and reconciling any discrepancies.

3. What types of data/parameters can be transferred or should <u>not</u> be transferred?

Some participants felt that, although **temporal transferability is currently used regularly, its validity has not been has not been questioned enough** by modelers. Rather than assuming temporal transferability, dynamic or longitudinal data and temporal studies should be conducted to test whether the data set allows for temporal transfer. A **controlled study of temporal transferability** would help the industry learn how to model temporal changes such as increases in trip rate. Trends over time should also be analyzed to determine whether the context of the data is changing. Temporal data context includes **household characteristics and structure** (e.g. change in the number of multi-worker households, smaller households, increase in the amount of eating out, extra stops for coffee). These changes in lifestyle may not be captured in existing household survey data. Large but unmeasured changes in the context will affect the temporal

transferability of the travel behavior modeling. This type of study is difficult to perform retroactively because one cannot go back in time to collect additional information from previous years. These studies need to be well planned in advance.

Disaggregate trends also affect temporal transferability. The reasons behind changes in travel behavior are important to modeling. Detailed travel time expenditure and travel time budget information is available in a paper by Polzin, Chu and Toole-Holt, "Case for Moderate Growth in Vehicle Miles of Travel A Critical Junction in U.S. Travel Behavior Trends" available at http://nhts.ornl.gov/2001/articles/moderateGrowth/moderateGrowth.html. For example, the NPTS and NHTS series of data show that there is a 1.9 minute per day increase in travel time per year. Questions such as whether people are spending more or less time traveling are complicated by the combination of activities with traveling, for example eating and talking on the cell phone may have made travel time less of a disutility. It has been shown that non-work travel has been increasing steadily while work travel has stayed reasonably constant. The growth in VMT per person has been linear, but this may change as speeds in urban areas deteriorate.

Temporal transferability is also affected by the **changing urban form**. Specifically, low density areas are beginning to form their own structure. Some of this effect is currently being captured in travel models in the density variables, but the variable also needs to be interactive. If this trend continues, these low density areas will depend less on large urban areas.

The most difficult component to transfer, is the destination choice model due to the high level of calibration needed. If one were to **rank components of the travel model by their ease of transferability**, trip production and tour-generation seem easier to transfer (as long as urban service variables are available) due to the ease of calibration. The second most transferable component would be the time-of-day choice model as it does not require geographic constants.

Panelists all believed that there are probably **key or core variables** in travel demand modeling that are transferable and then there are **context sensitive variables** which are less transferable. But it is not yet clear which variables belong to which group nor which variables can be transferred in a valid manner. Some participants suggested that for the time being, until further research is performed, the **more stable variables could be transferred with more confidence**. This would allow certain steps of the modeling process, such as trip generation, to be transferred so that resource-limited agencies could concentrate on more specific modeling issues of other steps such as mode choice. An example of a variable which has proven to give good results is the rule that in-vehicle time is about half of the value of out-of-vehicle time. A great deal of existing data is available and could be used to determine what is transferable.

One variable of particular interest is travel time reliability. Some attempts have been made to better understand travel time reliability, but a more solid understanding of the effect of the variable would immediately add explanation to models and make models more transferable.

There are currently some **things which are beyond understanding and are difficult to transfer**. For example, the observed positive travel distance utility in Los Angeles is probably very unique to the region. Also, cross-cultural variables currently are not very transferable because not enough is known about which variables are the most explanatory, such as whether a

person is English-language speaking, the number of years a person has been in the U.S., the level of "assimilation to the dominant culture", location of residence, such as living in a neighborhood where one can maintain your immigrant culture. An intrametropolitan variable is also most likely not transferable. And finally, constants should not be transferred as they represent factors that the data do explicitly not explain.

The participants also discussed the **Federal Transit Administration (FTA) New Starts guidance on travel demand modeling**, which can also be considered a data transfer**.** [XX-is there a document to reference here?]. Participants emphasized that since modelers everywhere are attempting to adhere to these guidelines, it is important that these guidelines contain valid ranges for coefficient values and that the context for these guidelines should be conveyed by FTA to the modeling community clearly. Examples of these guidelines are: in-vehicle time must be between the values of -0.2 and -0.3; auto and transit highway in-vehicle coefficients must be the same; or the acceptable in-vehicle to out-of-vehicle time ratio must be between 2 and 3.

Auto occupancy values are not transferable from area to area, unless the system configurations are similar (e.g. both cities have similar employer sponsored van pool activity). An auto occupancy equation or model may be more transferable. A good model would account for the differences in configurations and allow for transferability. Large MPOs use mode choice instead of auto occupancy, while auto occupancy data is only used for carpooling. Capturing auto occupancy by time has been found to be valuable. An alternative to transferring auto occupancy in a trip assignment model would be to include "carpool" separate from "drive alone" and "transit" in a mode choice model. Then treat HOV as a sub-division of mode choice, since it has been proven to be nested with transit (as it is a strong alternative to transit). With this model structure, in regions with transit, transit will be used more heavily. In regions without transit, then a high HOV usage will be forecasted.

4. How are data/parameters transferred in current applications? What are correct methods for data/parameter transfer?

In most cases, agencies just borrow variables from the best source they can find. At this point, it is still not clear how to determine whether it is acceptable to transfer data without any supporting data. There are some general guidelines for data transfer.

Ideally, modelers should **have data from both the source and recipient areas** to determine the suitability of the data transfer for each specific case. It is very important that there is a basic understanding of the source and recipient circumstances before transferring data (e.g. understanding of what type of errors are associated with the source data, how the recipient agency will treat values for alternatives of unobserved modes). Before transferring out-of-vehicle time, modelers should also carefully examine the way it is specified in the source and recipient model, as there are many different ways that it is specified.

Modelers should also perform a "goodness of fit" test to determine whether data can be transferred. However, a goodness of fit test is not sufficient on its own since a simple goodness of fit does not mean that the coefficient makes sense. The sign and value of the coefficient

should also be reasonable. Verifying the reasonableness of assumptions, parameters, and results is important to any travel demand modeling process, especially when transferring data. To aid users with identifying transferability, it would be beneficial to come up with a set of supplementary model specification tests for transferability.

Two camps of opinion exist for how to model **travel cost**. One camp believes that perceived data should be used for travel cost as this is what the traveler uses to make decisions. The second camp believes that perceived data is not consistent enough for modeling since people's perceptions are often wrong. Therefore, this second camp believes that travel costs should be calculated using network data. These two methods of modeling travel cost could result in significant model differences.

Transferability between the four-steps within a single model is currently assumed, but the issue is in need of further research. For example, NTCOG found that the value of time was different for transit and toll users and therefore must be calculated separately. They believe that because people with toll tags are not thinking about the cost of the trip and neither does someone who is using a monthly transit pass, the derived value within their model that each vehicle hour is worth $10, is not a true value of time and should not be transferred to another model without the understanding of where it came from. A clarification of the different value of times would be a good first step. For example, a $1 fare is not the same as a $1 toll because sensitivity to travel time and cost is not the same for all modes.

Agencies should be careful when transferring results from models written using different software. The **software being used** for modeling can affect the resulting value of the coefficients. For example, some software do not consider fares when performing the skims in the path modeling process. A little known fact is that EMME/2 allows for specific boarding penalties which is used to compute the cost of a path. Therefore given the same network inputs, different software may produce different path assignments. This leads to promising research question of whether a model could be adapted for use on several different software programs. An often debated topic is whether a **"universal model"** could be created. Such a universal model would require that the exact same procedure be applied to different sets of data, including the skims. Such a universal model, applied to multiple regions, could be used to answer many transferability questions.

Data is also transferred sometimes as distributions instead of averages. This method would not work for four-step modeling, but could be used in simulations and in an activity-based model. An example where transferring a distribution is more appropriate than an average is in the variable free parking, where some people are reimbursed for parking and others have to pay for parking. This variable is better represented as a distribution than as an average.

In general, average **trip rates are used for attraction and production models.** Most attraction models assume that attractions and productions for different retail businesses are similar. It has been discovered that there can be large differences trip rates can vary significantly. A suggestion for accounting for this difference is to use NAICS for ITE Standard Industrial Classification (SIC) codes.

Special generators are also sometimes transferred, but they are not very well understood. They are relied upon because of outliers. Smaller areas can use special generators because it is easy for them to deal with a few special cases, but large areas need to avoid using too many special generators.

5. What are the implications in using transferred data (e.g. need to use same input variables)?

Typically, parameters or model structure is transferred. Data itself is not generally transferred; the estimates generated from data are transferred. Currently, **data are not transferable because most data do not include the context in which it is gathered.** For transferability to be successful, modelers must understand the context in which transferred data were gathered and the context in which models and parameters were estimated. Lacking an understanding of the context, data are often transferred incorrectly or inappropriately.

Context may account for a large proportion of information. For example, no travel data can be transferred without network O-D information. Even differences in the survey design will affect the data. Without context, utility functions for the data cannot be developed and the data cannot be weighted (since the weighting factor is part of the survey records). The **data generation process could be standardized** to include the required context so that data from different regions can be pooled and exchanged. Standardization should include standardization of the surveys, building of the skims, the preferred sources for land use data, and how the data is to be summarized.

6. What can be done with existing datasets such as NHTS, past travel surveys, etc.? What new data should be captured in future surveys and data collection efforts?

A list of "missing mysterious variables" is needed so the NHTS can consider collecting additional variables next time around. Variables that were discussed were:

- Anything that captures the information process. For example, some research in Europe is underway on a **"social network" effect** and things that are not seen or included in the data. They are finding that there is a convergence on personal propensities but also that people are affected by how/what everyone else around them is doing. So behavior may be dependent not only on traditional demand modeling variables, but may also depend on what the person's circle of acquaintances are doing. Are there variables that could capture some of this "social network" effect?
- Variables which capture the fundamental decision making process: How and whys of a decision. How did you come to this decision, why did you use this mode for the trip? What were the most important factors in your decision? For example, have people rank cost and time as factors in their decision. This would ideally not be stated preference data but process data about revealed data for specific trips. How did you chose the path? Why did you choose the path that you did?

- Data to help better understand the difference in how a household may make choices differently from individuals.
- Sequence questions: Which decision was made first? Destination, TOD, mode, etc.
- Technology information (internet)
- Data on the difference between stated preference and actual travel behavior.
- Variables that could **relate home based work trips to trips for other purposes**, since-home-based work data are available in census data. This is a topic that should be researched.
- Auto occupancy rates in a region can be gathered from state accident reports and adjusted, but it is difficult to estimate auto occupancy for trip purpose. Currently, auto occupancy by trip purpose exists pretty uniformly in NCHRP 365, which were calculated using 1990 NPTS results. An improvement for the next generation of surveys would be to **mark transit availability into census tracts** so that it would be easy to model transit availability.

Further **research is also needed on how to model non-existing alternatives, alternatives which are physically available but not actually practical alternatives,** e.g. a trip requiring 4 transfers. One suggestion on how to capture the real-world viability of alternatives when performing the travel survey is to use the following method:

1) Perform behavior data collection: Ask what way something was done
2) Quickly shift the question to ask "Would you have rather done it another way?"

This would capture which alternatives were actually considered by the person in the decision making process.

One caveat is that process information needs a process model, which does not currently exist. Therefore the NHTS survey can only add some of these questions in to help develop a process model, but a complete set of questions will not be done until it is shown that this data is needed.

7. What types of models could reasonably and correctly be applied using default national parameters? What information would be needed to estimate these national parameters?

A useful exercise would be to **identify three MPOs with good existing survey data or to gather the exact same data**, collected the same way, for three MPOs, and then perform the skims to calibrate a sophisticated four-step or activity-based model. This would allow for analysis not only of what parameters are consistent at a national level but of whether the right questions are being asked for data transferability. Agreeing on the questions, contract management, and network coding for the participating MPOs would all be difficult tasks. All decisions would have to be agreed on at each step. Fewer participating MPOs would obviously make this easier. This project may be ideal for smaller areas as it may be possible to collect fewer data than would be needed for a single MPO to perform a complete survey.

Although four-step models are easier to estimate, **activity-based models may have a better chance of getting transferred**. Unfortunately, small areas cannot support activity models and most areas are not ready for it yet. Therefore it would be wise to design the survey to allow for either model to be estimated from the data. If it can be shown that activity-based models are

more transferable between regions, the movement towards activity-based models may pick-up. But rigorous testing of activity-based sensitivities is necessary to ensure that activity-based models are proven to be sensitive and logical in the same way that four-step model have been proven.

VI. Research Topics/Scopes

An important output of the peer exchange was a list of research topics for advancing the concepts of data transferability. These research topics are presented below.

A. Can a data and network format be agreed upon so that add-on surveys can be performed in a consistent manner to facilitate data transfer? What things need to be agreed upon (e.g. scale: parcel, zone, neighborhood) for data to be transferable. What contextual data needs to be available for data transfer? Part of this research includes determining the status or state-of-the-practice of land use data around the country and whether surveys need to gather additional land use data to facilitate data transfer.

B. Can a national travel demand model metadata database be created and maintained? How should metadata be defined so that all necessary details for sharing data are available for reference (e.g. description of survey and model data)? What data currently exist (e.g. self-describing data databases)? Who should be charged with maintaining the database? What data collection standardization can be done and what are the obstacles to data collection standardization? Can a metadata database be applied to small local areas?

C. What cross-sectional and temporal data are transferable? For how long is cross-sectional data valid? Analyze long-time temporal data to see if there is consistency of tour-patterns over a long period of time (e.g. using San Francisco data, NPTS data, and Puget Sound data).

D. What are the drivers of travel behavior dynamics? Are there main variables that explain the majority of travel behavior? What are these variables? Are these variables being collected in the current travel surveys. Is there a need to collect additional process data to better understand the drivers of travel behavior? Explore market research models. Can transferable variables be determined through comparing utility functions?

E. Further explore the regional impacts of travel behavior. Design a single model to cover two dissimilar regions. Determine whether regional behavior can be separated from generic variables and captured in new variables which explain regional differences so that models will be transferable. This research includes conducting surveys beyond NPTS/NHTS to capture parallel information for different regions to better understand the main reasons for regional differences.

F. Create a guidebook which outlines data transferability issues and guides a user step-by-step through evaluating data transferability (checklist: MSA size, etc.). The guidebook would discuss, for example, data management issues, how to evaluate data quality, what

is meant by transferability. The deliverable would include a Web-based tool where you can go through a check-list.

G. NCHRP 365 is a very important document to states, especially smaller states. Update the document to include new technologies and modeling methodologies.

H. How can disaggregate data be used to simulate data in a new context? Assumptions need to be explored. For example, what distributions are observed in travel behavior? Are different rates found in different clusters? Can trip rates be expanded to a larger scale?

I. Is it possible to develop a relationship between work trip rates and non-work trips? If so, what data support this relationship and what variables are needed in the relationship (e.g. work trip flows, TOD, occupancy, trip length). Does this relationship vary between regions?

J. The ability to adjust models for regional and temporal variation is essential for transferability. Research of the invariance, or constants, of travel would assist with determining whether there are additional variables which are not currently being included in travel demand modeling that can provide more explanatory power and facilitate transferability, for example:
   - What are people's upper-limits/tolerances in terms of the number of trips, number of tours, activity episodes, and time allocation?
   - Is there a variability or persistence of travel by individuals per household by day, day of week, or by month?
   - Would expanding surveys to include not just work location but also "WHY people stopped where they did" produce better travel demand models?
   - Use land use data at a disaggregate level which is currently not used in the four-step process because of computational complexity and previous data unavailability to explore extra land use and employment categories.

[Please help to elaborate:]

K. Acquiring good employment data is always difficult in travel demand modeling. Is lack of employment data an impediment to transferability? What are the current sources for employment data and methods for modeling employment? How is employment currently being broken down for activity-based modeling?

Appendix A. Presentations and Handouts

Presentations and handouts are available as links on the TMIP website. They can be accessed at: http://tmip.fhwa.dot.gov/

Presentation #1: (Nancy McGuckin)
Presentation #2: (Pat Hu)
Presentation #3: (Mohan Venigalla)
Presentation #4: (Kouros Mohammadian)

Appendix B. Data Transferability Resources

Greaves, Stephen. "Simulating Household Travel Survey Data." Presented at International Conference on Travel Survey Methods. Punta Renas, Costa Rica, August 2004.
http://www.its.usyd.edu.au/isctsc/costarica_papers/resource/B8%20-%20Resource%20Greaves.pdf

Xu, Min, P. Stopher, and S. Greaves. "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.
http://www.trb.org/conferences/nhts/Xu.pdf

Ruescher, T., R. Schmoyer, and P. Hu. "Transferability Of Nationwide Personal Transportation Survey Data To Regional And Local Scales" draft paper submitted to XX in 2001.
http://npts.ornl.gov/npts/1995/doc/transfer.pdf

Urban Travel Demand Forecasting Use website
http://nhts.ornl.gov/2001/presentations/index.shtml (this is list of presentations only)

Martin, WA, McGuckin, NA. "Travel Estimation Techniques for Urban Planning." NCHRP Report 365. Transportation Research Board Barton-Aschman Associates, Inc., Washington, DC, 1998.
http://nationalacademies.org/trb/bookstore

Wilmot, Chester G. and Peter R. Stopher. "Transferability of Transportation Planning Data." TRR 1768, pp 36 –42. National Academy Press. Washington, D.C. 2001.

Venigalla, Mohan, S., Chalumuri. "Vehicle Activity and Personal Travel Inputs to Emission Models." Final report to the Transportation and Environment Research Program (TERP), Office of Planning, Environment, & Realty (HEP), FHWA. George Mason University, Fairfax, VA, March 2004.
http://www.edthefed.com/xferability/TRIMM-FinalReport.pdf

Beckman, R.J., K.A. Baggerly and M.D. McKay. "Creating Synthetic Baseline Populations," *Transportation Research A*, Volume 30A, Number 64, 1995, pp. 415-429.

Los Alamos National Laboratory, "TRANSIMS: Transportation Analysis Simulation System, Version: TRANSIMS - 2.0." Volume 3 (Modules), Chapter 2 (Population Synthesizer), March 1, 2001.
http://transims.tsasa.lanl.gov/IOC-3.html

Appendix C. List of Participants

Names	Affiliation	Telephone	E-mail
Carlos Arce	NuStats	512.306.9065	carce@nustats.com
Richard Arnold	Oregon DOT	503.986.4218	Richard.ARNOLD@odot.state.or.us
John Britting	Wasatch Front Regional Council	801.363.4250	jbritting@wfrc.org
Ken Cervenka	North Central Texas Council of Governments	817.695.9266	kcervenka@nctcog.org
Kuo-Ann Chiao	New York Metropolitan Transportation Council	212.383.7212	kchiao@gw.dot.state.ny.us
Ed Christopher	FHWA	708.283.3534	edc@berwyned.com
Brian Gardner	FHWA	202.366.4061	brian.gardner@fhwa.dot.gov
Edward F. Granzow	CH2MHill Engineers	530.546.3113	egranzow@ch2m.com
Pat Hu	ORNL	865.946.1349	psh@ornl.gov
Ging Ging Liu	Volpe National Transportation Systems Center/Cambridge Systematics	617.494.2039	liu@volpe.dot.gov
Tom Maziarz	Capitol Region Council of Governments, CT	860.522.2217	tmaziarz@crcog.org
Nancy McGuckin	Oak Ridge National Laboratories	202.248.8640	n_mcguckin@rocketmail.com
Phil Mescher	Iowa DOT	515.239.1629	Phil.Mescher@dot.state.ia.us
Kouros Mohammadian	UIC	312.996.9840	kouros@uic.edu
Elaine Murakami	FHWA	206.220.4460	elaine.murakami@fhwa.dot.gov
Ram Pendyala	University of Southern Florida	813.974.1084	pendyala@eng.usf.edu
Thomas Rossi	Cambridge Systematics	617.354.0167	trossi@camsys.com
Jim Ryan (not present)	FTA	206.366.0954	james.ryan@fta.dot.gov
Mohan Venigalla	George Mason University	703.993.1630	mvenigal@gmu.edu
Peter Vovsha	PB	212.613.8807	vovsha@pbworld.com

Appendix D. Agenda

Peer Exchange on Data Transferability
Thursday, December 16, 2004
Keck Center of the National Academies
Transportation Research Board
500 Fifth Street, NW, ROOM 206
Washington, DC 20001

AGENDA

8:00 AM	CONTINENTAL BREAKFAST
8:30	**Introduction and Goals for the Peer Exchange** – *Ed Christopher, FHWA and Brian Gardner, FHWA*
8:45	**Presentations** – *Tom Rossi, Cambridge Systematics (moderator):* NCHRP 365 – *Nancy McGuckin* 1995 NPTS Transferability – *Pat Hu, ORNL* TRIMM – *Mohan Venigalla, George Mason University* Other transferability research including Stopher/Greaves – *Kouros Mohammadian, University of Illinois at Chicago*
10:00	COFFEE BREAK
10:15	**Facilitated Discussion Part A**: **Topics 1, 2 and 3**– *Tom Rossi, Cambridge Systematics (moderator); Elaine Murakami, FHWA (recorder):*
11:45	LUNCH
1:00 PM	**Facilitated Discussion Part B**: **Topics 4, 5, 6 and 7**– *Tom Rossi, Cambridge Systematics (moderator):*
3:00	BREAK
3:15	**Breakout Session**: **Developing research scopes of work** – *Tom Rossi, Brian Gardner, Elaine Murakami, Ed Christopher (notetakers)*
4:45	**Conclusion and Next Steps** – *Ed Christopher, FHWA and Brian Gardner, FHWA*