VALIDATION GUIDELINES

Validation Guidance

1.0 Introduction

The travel demand modeling community in Florida has recognized the need to enhance the understanding and usage of Florida's travel demand modeling framework. To achieve this goal, there is a need for common technical standards and guidelines in developing regional travel demand models in Florida, particularly regarding terminology, data formats, model structure, scripting, model outputs, and validation guidelines. The purpose of this document is not to provide a comprehensive manual on the validation of Florida's travel demand models.

The guidelines are drawn from a variety of resources, including standard technical documents such as the Federal Highway Administration (FHWA) *Model Validation and Reasonableness Checking Manual, Second Edition* (Cambridge Systematics, Inc., 2010) (hereafter referred to as the "FHWA Validation Manual") and NCHRP Report 716 - Travel Demand Forecasting: Parameters and Techniques (Cambridge Systematics, Inc., et al, 2012). Recent experience in validation of models throughout Florida and elsewhere was also considered, especially as they pertain to activity-based models, which were not as widespread when the aforementioned documents were prepared years ago.

2.0 The Model Validation Process

Sometimes, there is confusion regarding the terms “validation” and “calibration” with respect to travel demand models. To provide clear guidance, these terms are defined in this document in the same way that they are defined in the FHWA Validation Manual. The definitions are as follows:

- **Calibration** is the adjustment of constants and other model parameters in estimated or asserted models in an effort to make the models replicate observed data for a base (calibration) year or otherwise produce more reasonable results.

- **Validation** is the application of the calibrated models and comparison of the results against observed data. Ideally, the observed data are data not used for the model estimation or calibration but, practically, this is not always feasible. Validation data may include additional data collected for the same year as the estimation or calibration of the model or data collected for an alternative year. Validation should also include sensitivity testing.

Note that the term “calibration” is used by some practitioners to refer to the initial creation of the parameters of a model. In this document, this process of creating parameters is referred to as model **estimation**.

Considerations and Best Practices

The FHWA Validation Manual lists five primary elements of travel model validation:

1. Model Validation Plan Specification
2. Collection and Assessment of Validation Data
3. Validation of Model Components
4. Validation of Model System
5. Documentation of Validation Results

This document addresses elements 3 and 4 as they relate to Florida's travel demand models. That said, it is strongly recommended that the validation process of regional travel models include the development of a validation plan (#1 above) and complete documentation of model validation (#5 above) (of course, item #2, collection of validation data, is a given). The FHWA Validation Manual provides guidance for the preparation of this documentation.

The primary consideration in travel model validation—indeed, in travel model development itself—is that the model should be tailored to the specific planning analysis needs of the region and the agency. These needs can vary depending on the size and nature of the region or agency, and the specific planning needs should be defined before model development. These needs may include support for long-range transportation plan development, project-specific analysis, analysis of proposed developments, policy analysis, or special studies. Consideration of the uses of the model is important in determining the type and scale of model validation.

Most people have heard the term "garbage in, garbage out." It is important to recognize that erroneous model results can often be due to flawed input data, rather than model parameters that require calibration. It is therefore necessary to validate the model input data before commencing the validation process for the model itself. This is discussed more at the end of this chapter.

Since travel demand models include many components, it is important to check that every component is validated. This is especially critical for more complex models, such as activity-based models, that have a large number of components. A common method is to run each component of the model in sequence, calibrating the parameters of each before moving onto the next component, and checking to ensure that any differences between model results and observations are not due to errors propagating from a previous component.

Part (but not all) of model validation is checking to ensure that the model can adequately reflect known travel conditions. It has been standard practice for validation to include base year checks, where model results are compared to observed data for a period in the recent past. These base year checks must cover all model components and are discussed further in the next section.

Types of Validation Checks

The FHWA Validation Manual lists four broad categories of validation checks:

1. **Comparisons of base year model results to observations** might be considered "traditional" validation. The comparisons might be of model results to disaggregate data, such as data from a supplementary survey not used for model estimation, or to aggregate data, such as traffic counts or transit boardings. Comparing base year model results to different aggregations of the same data used to estimate a model (such as household travel survey data) is not as sound of a validation practice as comparing to independent data. However, for many validation tests, the data used for model estimation or calibration are the only data available. These comparisons can be thought of as "static" checks.

2. **Temporal validation** is an important aspect of model validation since, by definition, it implies comparing model results to data not used in model estimation. Either backcasts or forecasts may be used for model validation. For example, if the model base year is 2010, the model could be used to backcast to 2000 conditions, or it could be used to forecast 2015 conditions, in either case comparing the model results to observed data for the backcast or forecast year. These tests are considered "dynamic" checks.

3. **Sensitivity testing** includes several important types of checks, including both disaggregate and aggregate checks. Disaggregate checks, such as the determination of model elasticities, are performed during model estimation. Aggregate sensitivity testing results from temporal validation. Sensitivity testing can also include model application using alternative demographic,

socioeconomic, transportation supply, or policy assumptions to determine the reasonableness of the resulting travel forecasts. Sensitivity tests are also considered dynamic checks.

4. **Reasonableness and logic checks** include the types of checks that might be made under model sensitivity testing. These checks also include the comparison of estimated (or calibrated) model parameters against those estimated in other regions with similar models. Reasonableness and logic checks may also include "components of change" analyses and an evaluation of whether or not the models "tell a coherent story."

Use of Targets and Guidelines

Accuracy requirements and guidelines for model validation depend on the intended use of the model being validated. These requirements must be weighed against economic realities affecting planning agencies. Reasonable validation thresholds may be important in helping establish the credibility of a model and helping model developers and users determine when the model is "close enough." The definition of acceptable thresholds should balance the resources and time available for model development with the decisions that will be supported by the travel forecast obtained using the model.

The term "threshold" rather than "standard" is used in the FHWA Validation Manual, and this convention is followed in this document. The term standard connotes a formal definition of acceptance: "The standard has been met; therefore, the model is valid." While it is important to match base year observations for validation, simple matching of traffic counts, for instance, is insufficient to establish the validity of a travel model. Furthermore, adjusting model parameters to achieve a better match with observed data may make the model worse for its intended uses of forecasting and analysis, by changing its sensitivity to important input variables. Setting standards that must be met increases the temptation for modelers to make these types of adjustments to favor base year data matching over model sensitivity.

So, while this documentation presents some numeric guidelines, or thresholds, these should be considered targets to help planners gauge the accuracy of model results. There may be reasons why a particular model's results do not match some specific observed data, including differing levels of errors in the collected data itself. It is also important to remember that the types of changes made to model parameters to achieve a closer fit to observed data can sometimes reduce the model's predictive abilities, especially in cases where constants that are not sensitive to model inputs are revised.

The following guidelines are used to determine acceptable methods for achieving an improved match between modeled and observed travel characteristics:

- The adjustments should reflect transportation supply or traveler behavior rather than simple arithmetic;
- The adjustments should be reproducible; and

- The reasons for adjustments should be clearly documented.

Development of a Model Validation Plan

It is recommended to develop a model validation plan as early in the process as feasible, usually before model estimation begins. The model validation plan should include the following components:

- Identification of validation tests,
- Identification of validation data, and
- Identification of validation costs

The reader is referred to Chapter 2 of the FHWA Validation Manual, which provides a detailed discussion of validation plan development. Basically, the specific validation data sources should be documented, and the validation tests should be defined as specifically as possible. This definition includes identifying the comparisons between model results and observed data that will be made for each model component, the segmentation and levels of disaggregation for comparison, the thresholds for comparisons, and expected outcomes.

Validation of Model Input Data

Most model input data for any scenario can be considered either transportation system data, represented in the transportation networks, and socioeconomic data, usually at the zone level, although some models may use finer levels of disaggregation (e.g., micro-analysis zone). Network data is generally maintained within the travel demand modeling software. Socioeconomic data can be stored in various formats for input to the models. For activity-based models, this includes the synthetic population data.

Socioeconomic Data

Population and Households

Base year zone-level population and household data usually come from the decennial U.S. Census, although the data may have to be processed when the base year is not a census year. The decennial census provides substantial detail about person and household characteristics that are usually sufficient to provide the detailed zone-level inputs for the model base year (or a backcast year). However, U.S. Census estimates between decennial census years are much more aggregate, and allocations to zones and distributions of person and household characteristics require additional information. A typical source for this is the American Community Survey (ACS), which is conducted continuously but only for a small sample of the population.

For a base, backcast, or forecast year that is the same as a decennial census year, the main checks of zone-level population and household data are to verify that the census data has been accurately reflected in the zone-level data. This involves comparing zone-level model input data files to census data summaries for the variables used in the model. It also makes sense to check the totals against census data for aggregate geographic levels such as city or county. Maps can be created to perform visual checks of variables, transformations of variables (e.g., population density), or combinations of variables (e.g., vehicles per household). Maps can be useful in spotting anomalies that may result from incorrectly entered data.

The same types of checks should be performed for base, forecast, or backcast years that are not decennial census years. However, the appropriate basis for checking may not be the decennial census, especially when the scenario year is far removed from the census year in fast-growing areas of Florida. The ACS data may be the best data source in such cases (though the error ranges resulting from the sampling must be considered). For forecast years that are well into the future, the most recent ACS data is the best available data source, but modelers should consider changes that are expected to take place between the ACS data collection period and the forecast year. In Florida, the Bureau of Economic and Business Research (BEBR), housed within the University of Florida, is the official source for population forecasts. It is prudent to consult these sources for an understanding of the projected population in Florida.

Synthetic populations for base, backcast, or forecast year scenarios require additional checking since mathematical processes are used to synthesize the data using control totals from sources such as ACS data (usually the Public Use Microdata Sample, or PUMS). Checks should consist of verifying consistency with all control totals at the various levels at which they are used (e.g., regional, county, district, zone) and comparing cross-classifications of the synthetic population to available data sources (most likely ACS). Forecast year synthetic populations can be compared to the base year to check whether differences in aggregate totals and cross-classifications make sense.

Employment

Employment data validation is a bit more problematic than population validation since there is no equivalent source to the U.S. Census with a 100 percent sample of the population. Proprietary employment data sources are often used in developing employment data, but these sources must be checked for accuracy of both location data and size (number of employees). The locations and sizes of the major employers in the region need to be individually checked manually; other checks can be done at more aggregate levels.

The number of workers in a region is correlated with the number of employees at establishments, but these are not exactly the same measure. Workers may hold more than one job, and individual proprietors and those working from home may not be adequately represented in employment data sources. Additionally, people who live in the region may work outside the region, or, as is more often the case, workers within the region may live outside the region in less densely developed areas. It is common for the number of employees to exceed the number of workers in a region by a modest amount, but the difference may be greater depending on the size, location, and nature of the model region.

A key check for employment forecasts is the consistency between the growth rates in employment and the number of workers. The growth rates may not be exactly the same, especially if areas just outside the region are growing at a different rate than the region itself. But the differences should be relatively modest.

Transportation Networks

As is the case with socioeconomic input data, usually the networks are developed using the best available data. For highway networks, a GIS base map is the usual source. For transit networks, the General Transit Feed Specification (GTFS) is often used in regions where the transit systems are represented in GTFS; in other cases, especially for smaller transit systems, networks may be coded directly from route and schedule information. Rather than comparing the networks to information from independent data sources, the checks are typically performed to ensure that errors in translating the information into the model networks were not made.

Basic types of network checks include:

- Mapping the networks and making visual checks, often by color coding variables of interest (e.g., number of lanes) and checking for differences among similar links;
- Summarizing network data and sorting or aggregating the data by variables of interest (for example, miles of roadway by functional class and free flow speed);
- Building paths and checking for reasonableness; and
- Checking assignment results. While full assignments cannot be completed until the demand model components are finished, "dummy" assignments can be done using synthetic trip tables (or trip tables from an earlier model version, if the zone structures are consistent or can be made so). Even if dummy assignments are done, the full assignments done later will undoubtedly reveal additional network errors, but the dummy assignments may catch some errors earlier, which can be helpful not only in saving time later but also in providing more accurate skim data for model estimation.

3.0 Demand Model Component Validation

For the purposes of this document, "demand model components" are those that estimate the amount of travel between different parts of the region, potentially by mode and time of day. They may be considered as the components that are part of the process of creating trip tables for assignment, i.e., auto vehicle trip tables by time period and transit person trip tables by time period. This also would include, in models where walk and bicycle travel are included, the components that create walk and bicycle person trip tables (though these are rarely assigned). The assignment process itself is not considered part of the demand model components in this document; validation of highway and transit assignment is discussed in Chapter 4.

In conventional (i.e., four-step) travel models, the demand model components include trip generation, trip distribution/destination choice, mode choice, and time of day. A section in this chapter is devoted to each of these components. In activity-based models, the exact types of demand model components depend on the specific model formulation, but generally include components that are analogous to these four choices: the amount of travel, destination choice, mode choice, and time of day choice. Since activity-based models are tour-based, there are tour-level and stop (trip)-level models in each of these areas. Activity-based model demand component validation is discussed in the final section of this chapter.

Trip Generation

Trip Productions

Following the usual convention, trip productions refer to the home end of home-based trips and the origin end of non-home-based trips. Trip production models are commonly two- or three-dimensional cross-classification models with input variables such as household size, dwelling type, number of workers, income level, and number of vehicles.

The main focus of static validation checks of trip productions is on whether the number of trips produced by household by trip purpose are reasonable. These trip rates are computed by dividing the total number of trip productions by the total number of households. These rates can vary substantially among transportation analysis zones (TAZs), as zonal household characteristics can differ significantly. It therefore makes sense to examine trip rates for more aggregate segments as well as the entire region. Typical segments might be districts consisting of several dozen to several hundred zones, or corresponding to other subdivisions such as counties, depending on the size of the region, its geography, and political subdivisions.

It should be noted that some of Florida's travel demand models include all trips, including those by walk and bike modes, while others include only "motorized" trips. Naturally, trip rates are lower when they exclude non-motorized trips.

The FHWA Model Validation Manual reports average household trip rates from the 2001 NHTS in the range of 10 to 11, with higher rates in more urbanized areas. Also reported is the percentage of trips that are non-motorized in the 2001 NHTS, which is about 10 percent, varying by trip purpose. NCHRP Report 716, which was written more recently when the 2009 NHTS data were available, indicated trip rates of 9.5 for smaller areas and 10.0 for larger areas. It therefore makes sense to assume that a reasonable range for the regional trip rate is about 8 to 11 for models with motorized trips only and 9 to 12 for models that include non-motorized trips. For some geographic segments, rates outside these ranges might be reasonable; for example, the rate for an area with more affluent residents and higher average household sizes might be above 12.

Another trip production check is the percentage of trips by purpose. The FHWA Model Validation Manual provides average trip rates from the 2001 NHTS for home-based work (HBW), home-based non-work (HBNW), and non-home-based (NHB) trips, varying by Metropolitan Statistical Area (MSA) size. NCHRP Report 716 provides the average trip rates for the HBW, home-based school (HBSc), HBNW, and NHB purposes from the 2009 NHTS. This information is summarized in Table 1.

Table 1. Trip Percentages by Purpose from the Literature

		FHWA Validation Manual	**NCHRP Report 716**
Trip Purpose	**Description**	**(2001 NHTS)**	**(2009 NHTS)**
HBW	Home-Based Work	14-15%	14-15%
HBSc	Home-Based School	-	6%
HBNW	Home-Based Non-Work	52-55%	54-56%
NHB	Non-Home-Based	31-34%	30-32%

Trip Attractions

Trip attraction models are typically linear functions of zonal totals of employment by type, number of households, and school enrollment. The main general reasonableness check for trip attractions is that the regional totals by purpose should be close to the totals for trip productions. For most trip purposes, there is no single variable check analogous to the check of trip production rates per household.

The sole trip purpose for which a single variable check exists is HBW, where the number of trips per employee can be checked. As noted in the FHWA Validation Manual, one might expect that an upper bound on the ratio of HBW attractions to employees would be 2.0 since HBW attractions include trips both to and from work, summarized at the workplace. However, some workers are not scheduled to work on every weekday, and many workers do not make two HBW trips on an average weekday. Some workers are absent due to vacations, personal days, sick leave, telecommuting, or work-related travel. In addition, journeys to or from work that include stops on the way are usually considered as combinations of HBNW and NHB trips.

Since NCHRP Report 716 presents an average HBW trip attraction rate per employee of 1.2 for 16 models in the U.S., it is therefore suggested that a reasonable range for travel demand models in Florida be 1.2 to 1.5 HBW attractions per employee.

Summary of Trip Generation Model Validation Checks

Checks

- Check number of trips per household for reasonableness.
- Check percentages of trips by purpose.
- Check number of HBW trips per employee for reasonableness.

Guidelines

- Total person trip productions per household:
    - o 9 to 12 if non-motorized trips are included
    - o 8 to 11 if non-motorized trips are excluded
- Percentages of trips by purpose: As presented in Table 2
- HBW attractions per employee: 1.2 to 1.5

Table 2. Reasonable Ranges for Trip Percentages by Purpose

Trip Purpose	Description	Low	High
HBW	Home-Based Work	12%	20%
HBSc	Home-Based School	5%	8%
HBSh	Home-Based Shopping	10%	20%
HBSR	Home-Based Social/Recreational	9%	12%
HBO	Home-Based Other	14%	28%
HBNW	Home-Based Non-Work	45%	60%
NHB	Non-Home-Based	25%	35%

Trip Distribution

Regardless of the formulation of the trip distribution model (gravity model or logit destination choice), the outputs are the same—person trip tables by trip purpose. Validation, therefore, includes comparisons of the modeled origin-destination patterns to observed data. These comparisons may be performed for segments of the population, such as households by income level, where the model includes such segmentation. There are various sources available for the observed data, including the following:

- **Household travel surveys** conducted in the model region. If such a survey is available for the region, it may be the best overall source for observed origin-destination data because it is specific to the model region and includes information on all travel purposes. The survey includes observations for only a sample of travelers (perhaps one percent, or less for larger regions), and so the sampling error must be considered in doing the comparisons, especially for relatively small segments.

- The **National Household Travel Survey** (NHTS) is released every five years and is available for all parts of Florida. The 2009 NHTS included an oversample for Florida, which reduces sampling error. However, the base years for models are now after 2009, and travel patterns likely have shifted since then, especially in faster-growing regions. Also, the 2009 NHTS is the last time an oversample was conducted in the state of Florida. The 2022 NHTS data is the latest release, providing a more up-to-date set of observed travel data. However, due to the lack of an oversample in Florida in the more recent NHTS, sample sizes are smaller (and sampling errors are larger).

- The **American Community Survey** (ACS) from the U.S. Census Bureau includes a large sample of work travel for all regions in Florida. The sample rate for the ACS is much higher than in household surveys, and so sampling error is less in the ACS; however, only work travel is covered. Additionally, it is important to note that there are differences between the information gathered from the ACS compared to household surveys, which more closely reflect how models simulate travel. The ACS asks about where a respondent worked last week, while travel demand models simulate where people work on an average weekday. This means there can be differences between what is modeled and what is observed from the ACS data. Household travel surveys capture intermediate stops on the way to or from work, whereas the ACS only records home and work locations.

There are two general types of aggregate checks of base year trip distribution model results: Trip length checks and origin-destination pattern checks.

Trip Length Checks

Note that "trip length" can refer to either trip travel time (e.g., in minutes) or trip distance (e.g., in miles). It makes sense to perform checks for both time and distance measures, although it is important to use a consistent basis for the observed and modeled trip length summaries. This is most easily achieved through the use of the same distance or time skims applied to the observed trips (on a zone level origin-destination basis). It is risky to use reported travel times or distances from surveys, as respondents can provide inaccurate estimates, and they tend to round estimates to the nearest 5, 10, 15, or even 30 minutes (or similar rounding for distances). Since skims are usually available by time of day, and trip distribution results often represent an entire day, it may be necessary to decide which skims to use by applying rules. For example, HBW trip comparisons may use peak skims while non-work trips may use off-peak skims. This means that both the "observed" or modeled daily times or distances may not reflect the true values. However, the comparisons will still be valid as long as the same skims are used for both observed and modeled data.

Comparing average trip lengths – The modeled average trip length by trip purpose can be compared to observed data from a household travel survey, the NHTS, or (for HBW trips) the ACS. The FHWA Validation Manual suggests, as a guideline, that the modeled average trip lengths for each trip purpose should be within five percent of observed for each trip purpose. In models with many trip purposes, some purposes may have relatively few trips, and so the five percent guideline can be relaxed in these cases.

Comparing trip length frequency distributions – Even if observed and modeled average trip lengths are consistent, the trip length frequency distributions may not be. The most common method used to compare observed and modeled trip length frequency distributions is the coincidence ratio. This represents the area under both of the plotted frequency distribution curves divided by the area under at least one of the curves:

The procedure to calculate the coincidence of distributions is as follows:

$$CR = \left\{\frac{\sum_T [min(PM_T, PO_T)]}{\sum_T [max(PM_T, PO_T)]}\right\}$$

where:

CR	=	Coincidence Ratio
PM_T	=	Proportion of modeled distribution in interval T
PO_T	=	Proportion of observed distribution in interval T
T	=	Histogram interval for time, distance, or other impedance measure (e.g., 0-4.9 minutes, 5.0-9.9 minutes…)

The coincidence ratio lies between 0 and 1.0, where a ratio of 1.0 indicates identical distributions. The FHWA Validation Manual states that it is preferable for the coincidence ratio for each trip purpose to be at least 70 percent. The guideline can be relaxed in models with many trip purposes, as some purposes may have relatively few trips.

Origin-Destination Pattern Checks

Two typical origin-destination pattern checks are:

- Comparisons of modeled and observed trip tables by purpose at a district level; and
- Comparisons of modeled and observed intrazonal trips by purpose.

As described above in the trip generation summary, districts may consist of several dozen to several hundred zones or correspond to subdivisions such as counties, depending on the size of the region and its geography and political subdivisions. Again, the observed data may be from a household travel survey, the NHTS, or (for HBW trips) the ACS. Because districts can vary greatly in size, and the number of district-to-district trips can differ substantially across a model, it does not make sense to set percentage difference guidelines for these comparisons (the FHWA Validation Manual presents no such guidelines). However, these checks are critical in determining whether the model accurately reflects travel patterns, and so analysts should examine any significant differences in the comparison, taking note of the effects of small sample sizes on error rates.

The modeled percentage of trips for each purpose that are intrazonal can be compared to observed percentages from the same data used for other trip distribution checks. The FHWA Validation Manual presents an example range of three percentage points difference between observed and modeled percentages (for example, if a trip purpose had an observed intrazonal trip percentage of seven percent, the modeled percentage should be between 4 and 10 percent).

Summary of Trip Distribution Model Validation Checks

Checks

- Compare modeled average trip lengths by trip purpose to observed.
- Compare trip length frequency distributions by trip purpose to observed, calculating coincidence ratios.
- Compare modeled and observed trip tables by purpose at a district level and check for reasonableness of match.
- Compare modeled and observed intrazonal trips by purpose.

Guidelines

The suggested guidelines for trip distribution come from the FHWA Validation Manual:

- Average modeled trip lengths by trip purpose are within five percent of observed.
- Coincidence ratios by trip purpose of at least 70 percent (lower values acceptable for trip purposes with relatively small numbers of trips).
- Modeled percentages of intrazonal trips with three percentage points of observed percentages.

Mode Choice

The main aggregate checks of mode choice model results would be shares of trips by mode by trip purpose, perhaps additionally segmented by income level, auto ownership level, or geographic subregion. However, it is important to recognize that the main outputs of travel by mode are the outputs of the assignment step. For example, it is more important for the model to accurately represent transit ridership, as output from the transit assignment, than transit mode shares. With this in mind, it makes more sense to focus on the accuracy of transit assignment results compared to observed ridership than on transit share comparisons to observed data.

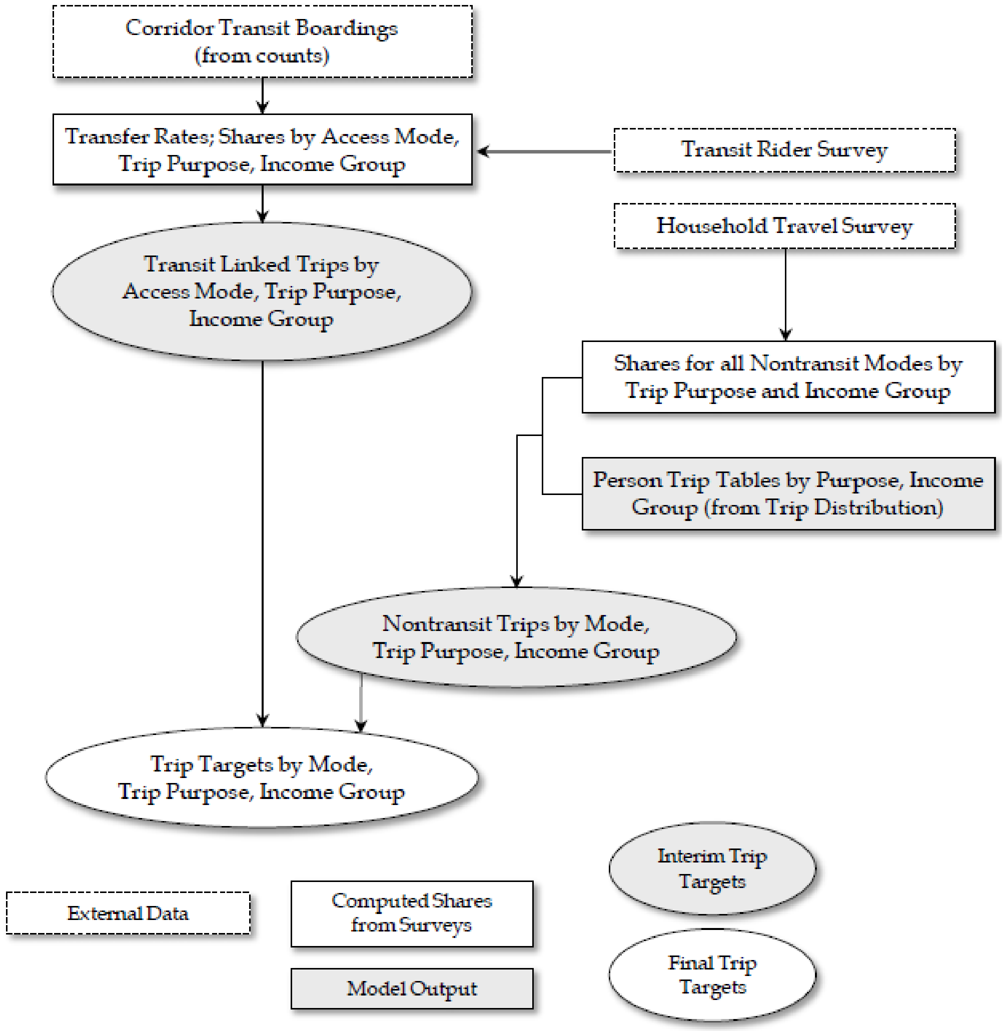
Of course, ideally, the transit assignment results should be accurate because the transit shares by trip purpose and other segments are also accurate. However, the available observed mode share information is often based on data sources that are dominated by auto travel. Household travel surveys, for example, may have only a handful of reported transit trips, especially for segments or trip purposes with low transit use.

The recommended procedure, which is documented in the FHWA Validation Manual, is to develop targets for total trips by purpose and segment. Transit targets are created by starting from total transit boardings, which are then converted to linked trips and segmented by trip purpose, as well as other segments used in the mode choice model. Segmentation is typically performed by applying percentages of transit trips by purpose from a transit rider survey, if available. Percentages of transit trips by segment can also be obtained from a household travel survey, the NHTS, or (for HBW trips) the ACS, but since the number of transit trip observations can be small in such data sets, the associated error can be substantial, and so trying to precisely match targets based on such percentages can be problematic.

Once the total transit trips by segment are estimated, they can be subtracted from the total trips generated in earlier model steps to obtain total auto trips (and non-motorized trips if they are included in the model). The household survey data (or ACS for HBW trips) can then be used to obtain percentages to segment the trips for these modes. Figure 1, which is from the FHWA Validation Manual, illustrates the process for developing the mode choice validation targets.

The FHWA Validation Manual does not present any criteria guidelines for checks of mode choice models, recognizing that it would not make sense to provide a guideline for the modeled number of trips to be within X% of observed trips when, for many segments, the error in the observed data could be much greater than X%. The true checks of travel by mode can be done after trip assignment. Trip assignment checks are presented in Chapter 4 of this document.

Figure 1. Example of Transit Mode Choice Validation Target Development



Source: Cambridge Systematics, Inc., 2010.

Summary of Mode Choice Model Validation Checks

Checks

- Create targets for observed trips by mode and purpose and compare model results for reasonable match.

Time of Day

It is typical for models to start by estimating daily travel, but in many models, especially for regions with peak period traffic congestion, daily travel measures are converted to measures by time of day at some point in the modeling process. Time periods often include morning and afternoon peak periods as well as one or more off-peak periods. Generally, four-step models use fixed factors to segment daily trips by purpose into trips by time periods. In some models, peak-hour travel estimates are computed using peak-period totals, again usually through the use of fixed factors.

Probably the most important checks of travel by time period are done following highway and transit assignment, where results can be compared to traffic or transit ridership counts by time period. These assignment-level checks are discussed in Chapter 4. The main check that can be done before trip assignment is to compare the modeled percentages of trips by time period for each trip purpose (and perhaps other segments) to observed percentages, from a household travel survey or the NHTS. One way to do this comparison is to use the coincidence ratio measure.

The FHWA Validation Manual does not provide any specific guidelines for the comparison of modeled and observed travel by time period. The factors used to convert daily trips to trips by time period typically come from the same observed data sources as those used for the comparisons. Time-of-day factors can be derived from passive origin-destination data, but not by trip purpose. Comparisons to percentages in other regions could be problematic since these percentages can vary widely even among regions of similar size, as noted in the FHWA Validation Manual.

Summary of Time-of-Day Validation Checks

Checks

- Compare modeled percentages of trips by time period for each trip purpose to observed percentages.

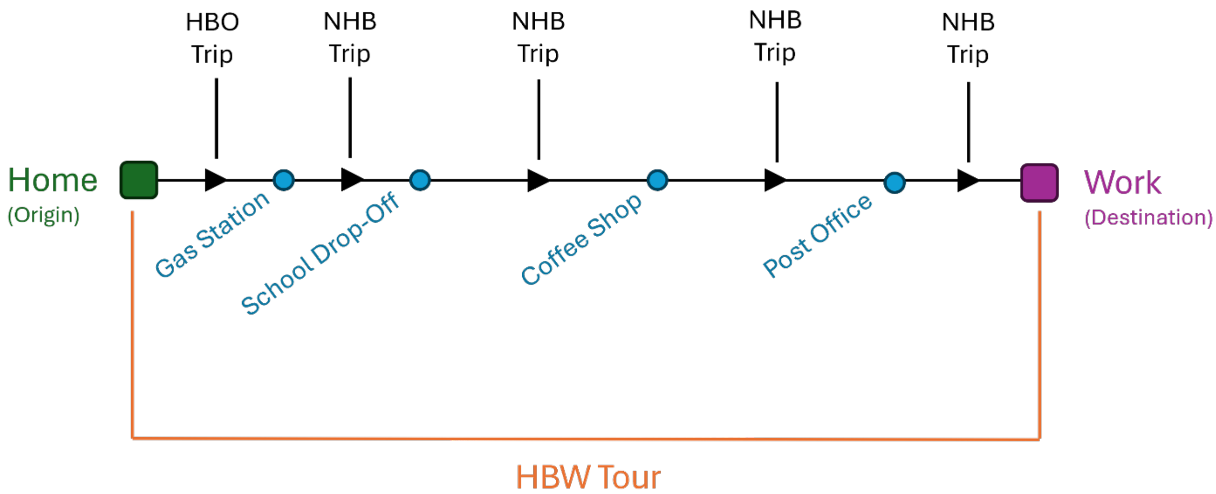
Activity-Based Model Components

The main practical differences between the activity-based model applications in Florida and the trip-based model applications are:

- The activity-based models produce **disaggregate results**, including rosters of tours and trips. These rosters are aggregated to create trip tables for the same assignment procedures used with the trip-based models.

- The activity-based models produce **tours** as well as trips, which means that another level of checks is required. Figure 2 provides a visual example of the relationship between tours and trips.

Figure 2. Diagram of a tour.



Checks Similar to Trip-Based Model Checks

Since activity-based models produce at least as much information as trip-based models, all of the checks described earlier in this chapter for trip-based models should also be done for activity-based models. An advantage of disaggregate models is that checks can be done for many more segmentation schemes than can be done for aggregate trip-based models, where segment checks can be done only for segments defined in the model. Since the outputs are disaggregate, they can be segmented using any variables that are used in the model. For example, any person characteristics in the synthetic population, such as age, gender, or worker/student status can be used as segmentation variables. Household characteristics such as household size, vehicle ownership, and income level can also be used for segmentation.

The trip-based model checks described earlier that apply to activity-based models are summarized below, with additional comments pertaining to their use in activity-based models. The guidelines for each check are the same as for trip-based models.

Trip Generation Checks

- Check the number of trips per household for reasonableness.
- Check percentages of trips/tours by purpose. Note that trip/tour purposes are different in activity-based models. For example, many non-home-based trips are part of home-based tours, and so the percentage of tours that are for, say, the purpose of work would not be the same as the percentage of HBW trips. The percentages of tours by purpose can be compared to observed data form the household survey or NHTS.

- Check number of HBW trips per employee for reasonableness. Again, HBW trips are not the same as work tours, and so the guidelines for tours would not apply. (However, additional checks can be done—see "demand rate" below.)

Trip Distribution Checks

These checks would include not only tour and trip origin-destination and tour/trip length checks, but also checks of activity-based model components such as usual workplace and school location.

- Compare modeled average trip/tour lengths by trip purpose to observed.
- Compare trip/tour length frequency distributions by trip purpose to observed, calculating coincidence ratios.
- Compare modeled and observed trip/tour tables by purpose at a district level and check for reasonableness of match.
- Compare modeled and observed intrazonal trips/tours by purpose.

Mode Choice Checks

- Create targets for observed trips/tours by mode and purpose and compare model results for reasonable match.

Time of Day Checks

- Compare modeled percentages of trips/tours by time period for each trip purpose to observed percentages. (Note that there are additional checks for tour time of day described below.)

Additional Activity-Based Model Component Checks

While not all activity-based models have the same structure, they all have components that need to be checked that are not covered in the list of checks similar to the four-step models above. Examples of additional checks that pertain to most activity-based models are listed below. Again, person and household characteristics should be used as segmentation variables for these checks. There are no specific guidelines for how "close" model results should be to observed data since the error associated with the observed data is highly dependent on the sampling plan for the observed survey data and the sizes of the various segments.

- **Daily activity patterns**. Activity-based models have components that simulate the types of activities that people perform during the day, including work, school, and other "non-mandatory" activities. The specific checks to be done depend on the model structure, but the following types of checks are typical:
    - <u>Worker/student incidence of travel to work/school</u> – The modeled percentages of workers who go to work and students who go to school on the travel day can be compared to observed data, segmented by person characteristics.

- “No travel” activity patterns – The modeled percentages of people who do not travel (stay home or are not in the region on the travel day) on the travel day can be compared to observed data, segmented by person characteristics.

- **Workplaces per employee**. Many activity-based models simulate the regular work locations of workers. The simulated workplace locations can be compared to the employment data.

- **Tours per person by activity type**. Similar to trip generation checks, the modeled number of tours per person, segmented by person and household characteristics, can be compared to observed data.

- **Tour time of day/duration**. Not only can the departure times for tours be compared to observed data, as noted above, but the combinations of tour departure time from home and return time to home can also be compared to observed data. The modeled durations of the activities performed on the tours can also be compared to the observed data.

- **Stops per tour by activity type**. The modeled number of stops per tour can be compared to observed data. This comparison should be done by tour purpose and segmented by person and household characteristics.

4.0 Highway and Transit Assignment Validation

Traffic volumes on the highway network are key outputs of travel demand models. They are used in a variety of analyses, including long-range transportation plans, highway project studies, site impact studies, and others. In regions where transit demand is significant, modeled transit ridership is also a key output for many types of studies. The assignment components often produce the ultimate results from models as opposed to interim results used as inputs to subsequent components.

While validation is important for all model components, validation of highway and transit assignment is especially critical given the wide-ranging use of the assignment results. The assignment components also provide a unique opportunity for model validation using data that is completely independent of the data used in estimating the models, namely traffic counts and transit ridership counts.

The validation guidance for highway and transit assignment is provided below.

Highway Assignment

Traffic count data are the primary source used for the validation of highway assignment procedures. Since models typically reflect weekday travel, the counts used for validation should be for weekdays. Ideally, the counts should reflect average annualized weekday travel (AAWDT) as opposed to average annualized daily travel (AADT), which includes weekend data. Those involved in validating models should be aware of whether their models reflect a true average condition, as opposed to a spring/fall or summer period, and use traffic count data that reflects the period being modeled.

Count data may be collected on a continuing basis (i.e., permanent traffic recording stations) or for a fixed period of time, often ranging from 24 or 48 hours to a week or two. Data from permanent counting stations show that counts can vary substantially over the course of the year, or even among different days of the same week. Agencies that collect traffic count data often provide counts that have been factored to represent average conditions by using seasonal and/or day of week adjustments, and may provide the specific factors used for the adjustments. This information can be valuable in preparing the count data to be used for model validation.

Even when the count data have been thoroughly checked to ensure that they reflect the weekday period being modeled, there is still error associated with the data that must be considered. The counting devices themselves are not 100 percent accurate, and there can be errors related to the processing of the raw data and the factoring of the data to represent the appropriate validation period. Other than locations with permanent count stations, many locations might not be counted every year, and it is sometimes necessary to use data that may be up to three years before or after the model's base year. An important point is that data collected from non-permanent counting devices reflects only the period when the device was active. In such cases, the data can be factored using information from permanent stations, but the temporal variation in volume for the location with a temporary count may not match the variation at the permanent station.

The main point of this discussion is to note that there can be substantial variation between the “best” base year volume that can be obtained or processed for a location and its “true” average volume for the period being modeled. It is therefore unwise to attempt to force the model to precisely match such an “observed” volume. (It is often the case that the best observed volumes for different nearby locations are inconsistent due to differences in the periods when the counts were performed, and that it will be impossible for the model to “match” both counts.)

A general rule for assignment checks is that higher percentage differences between modeled and observed volumes are expected when the volumes are lower.

VMT Checks

Some of the basic aggregate highway assignment checks are comparisons of modeled vehicle-miles traveled (VMT) to observed VMT, computed from the count data. At the most aggregate level—for the entire model region over all links for an average weekday—the model can be expected to match observed VMT well.

The FHWA Validation Manual presents guidelines for differences between modeled and observed daily VMT by functional class from several sources, ranging from six to seven percent for freeways and expressways to 20 to 25 percent for collectors. Other common comparisons of modeled and observed VMT are by geographic subregions, such as counties and by area type.

VMT checks by time period are less common but can be valuable. The expected percentage differences between modeled and observed VMT for time periods shorter than an entire day can be expected to be higher than those for daily VMT, since the volumes are significantly lower.

Screenline and Cutline Checks

Screenline and cutline checks may be considered “semi-aggregate” since they represent volumes across groups of highway network links, but are relatively small numbers of links focused on specific geographic areas. The term screenline is used to describe a group of links that splits the region into two subregions, where vehicles must use one of the screenline links to pass from one subregion to the other. Cutlines are subsets of screenlines that do not extend to the regional boundary.

The FHWA Validation Manual presents a graph of the percentage “maximum desirable deviation” by volume level for screenlines. A sample of points on this graph is shown in Table 3.

Table 3. "Maximum Desirable Deviation" by Volume Level for Screenlines

Screenline Volume	Percentage Difference
10,000	55%
20,000	47%
30,000	40%
40,000	35%
50,000	32%
75,000	28%
100,000	25%
125,000	22%
150,000	21%
200,000	19%

Source: FHWA Validation Manual

Individual Link Volume Checks

The aggregate and semi-aggregate checks described above can provide information on overall or average errors or biases, but if an average modeled value is close to the observed, there could still be offsetting errors in individual link volumes. A commonly used measure for checking the general accuracy of an entire group of modeled link volumes is root mean square error (RMSE). The formula for RMSE, as given by the FHWA Validation Manual, is:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}[(Count_i - Model_i)^2]}{N}} \qquad \%RMSE = \frac{RMSE}{\left[\frac{\sum_{i=1}^{N} Count_i}{N}\right]} * 100$$

It is appropriate to examine the percentage RMSE not only for the entire set of links with observed counts but also for several specific subsets of these links. A common approach to segmentation is by volume group.

Summary of Highway Assignment Validation Checks

Checks

- Compare modeled VMT on links with counts to observed VMT, summarized by functional class.
- Compare volumes on screenlines and cutlines to observed.
- Compute RMSE between modeled and observed link volumes, summarizing by volume group.

Guidelines

The suggested guidelines for highway assignment validation checks are derived from information in the FHWA Validation Manual:

- Percentage difference between modeled and observed VMT by functional class according to the guidelines in Table 4
- Percentage difference between modeled and observed screenline volumes according to the guidelines in Table 5
- Percentage RMSE by volume group according to the guidelines in Table 6

Table 4. Guidelines for Percentage Difference Between Modeled and Observed VMT by Functional Class

Functional Class	Percentage VMT Difference
Freeways/Expressways	7%
Principal Arterials	10%
Minor Arterials	15%
Collectors	20%
All Links	2%

Table 5. Guidelines for Percentage Difference Between Modeled and Observed Screenline Volumes

Volume Range	Percentage Volume Difference
0-20,000	45%
20,000-50,000	30%
50,000-100,000	25%
100,000-200,000	20%
200,000+	15%

Table 6. Guidelines for Percentage RMSE by Volume Group

Volume Range	Percentage Volume Difference
0-5,000	100%
5,000-10,000	45%
10,000-15,000	35%
15,000-20,000	30%
20,000-50,000	25%
50,000+	20%
All Links	40%

Transit Assignment

Transit assignment routes the trips from the transit trip table outputs of the mode choice model onto the transit network. This involves determining at which stops the trips from each origin zone to each destination zone board the transit system, which routes they use, at which stops they transfer to other routes (if applicable), and at which stops they alight. The main transit assignment outputs are therefore total ridership on each route and boardings at each stop. If transit assignments are performed by time period, the outputs are available for each period.

There are several issues that can complicate the transit assignment validation process, as discussed below.

- Ridership count data are the observed data to which assignment results are compared. The availability and quality of data can vary. Some regions may have several transit operators that collect and manage ridership data differently. Data may not be available at the stop level everywhere or even at the individual route level when multiple routes overlap. The time periods for which data are available may not match those used in travel demand models.

- Because bus (and sometimes rail) routes often overlap, using the same roadways or rights of way, the assignment process often is not sophisticated enough to allocate riders among overlapping routes, even when multi-path assignment processes are used.

- The relatively low numbers involved in transit assignment (some stops or even routes might have only a handful of riders) mean that percentage differences between model results and counts can be quite large.

- Many assignment processes assign transit trips in a “production to attraction” direction, which is the opposite direction to actual travel for trips where a traveler is returning home. This can confound comparisons involving directionality of travel.

- Because nearly all transit trips require walking at the access end or egress end (or both), the aggregate geographic level of zones can produce significant aggregation error affecting transit route choice decisions. The distances between zone centroids and transit stops are fixed, but the actual distances experienced by riders can differ substantially within a zone. Some riders may have very short walks, while others have much longer walks.

- While the assignment process uses average values to weight different time and cost components of transit paths, the relative values for individual riders may vary substantially. For example, people value time differently, or some people who have more difficulty walking may have a greater preference for paths that minimize walking.

- To a greater extent than other model components, the quality of modeled transit assignment results is highly dependent on the accuracy of the other model components. Network coding, especially for access and egress to transit, is critical to building accurate paths. The assumptions used in path building, including the weights of different time and cost components and maximum allowable walk and auto access distances/times, have a large impact on path choices. Trip distribution models, which produce mainly auto trips, may not be validated sufficiently to produce accurate origin-destination patterns for transit riders. Relatively small inaccuracies in mode choice models can have substantial impacts on transit assignment results.

These issues are likely reasons why there are few published specific transit assignment validation guidelines. It is difficult to establish expected differences between relatively small amounts of modeled and observed transit ridership, especially when considering segments such as routes and stops (perhaps by time period). Ideally, any numeric guidelines would be relatable to specific transit planning needs, but short-term planning efforts often rely on observed data rather than model results, and longer-term planning efforts, such as planning for major new transit projects, may require information on markets not well served by existing transit services for which observed data exist.

The FHWA Validation Manual does not specify any numeric transit assignment guidelines. Given the uncertainty surrounding the transit assignment validation process and the small numbers associated with this model component, no specific numeric guidelines are provided here. However, models that are used to provide transit ridership information to planners need to have the transit assignment component validated. With this in mind, the transit assignment checks listed below are recommended.

Summary of Transit Assignment Validation Checks

Checks

- Compare modeled transit ridership to observed ridership, summarized by groups of routes. It is suggested that groups include routes with the combined daily ridership of several hundred to a few thousand, which have similar characteristics and are located nearby. For example, a set of bus routes in a corridor could be used as a group. If transit trips are assigned by time period, summaries should be done by time period as well as for the entire day.
- Compare modeled volumes to observed volumes for major transit rider subgroups, for example, auto versus walk access, bus versus light rail, or express bus versus local bus.
- Compare modeled volumes on “major” screenlines to observed volumes. “Major” screenlines could be created across major barriers (i.e., rivers) or jurisdictional boundaries.

5.0 Temporal Validation and Sensitivity Testing

As noted in Chapter 1, temporal validation and sensitivity testing are important parts of model validation since models are mainly used to provide travel demand estimates for alternative or forecast scenarios that are, by definition, different than the base year scenario. The base year validation process shows that the model can accurately represent observed travel demand, while temporal validation and sensitivity testing can show that the model will accurately reflect the effects of changes in conditions or policies.

This chapter generally describes the temporal validation and sensitivity testing process. Unlike the base year validation process described in Chapters 2 and 3, the process does not focus on comparisons of model results to observed data, although for some types of temporal validation, these comparisons are still part of the process. Thus, there are no guidelines provided for expected differences between model results and observed data. Of course, when such comparisons are made, significant differences between model results and observed data should be examined and addressed.

Temporal Validation

Most travel models are based on "snapshot" data, such as household survey data collected in a periodic, but infrequent, survey effort. The model relationships, parameters, and coefficients might be significant and accurately reproduce travel for the point in time represented by the model estimation data. However, the relationships may not hold true over time; the further one moves from the base year for validation, the more uncertain one should be regarding the veracity of the models. For this reason, good validation practice should include temporal validation for at least one year other than the base year for model estimation or calibration.

The objective of temporal validation is to provide another point in time, besides the base year, for which model results can be checked. There are three basic options for another point in time (of which more than one can be chosen, depending on the resources available for validation):

- A year prior to the base year
- A year after the base year but before the present
- A year in the future

Each of these options has benefits and disadvantages, as described below.

Modeling a year before the base year (often referred to as "backcasting") is appealing if all of the necessary model input data and observed travel data to compare the model results to can be gathered. Often, the base year of a previous model version is used since much of the necessary data has already been assembled. Backcasting also provides a way of testing the sensitivity of the model to input variables related to items that may have been changing in the recent past. These items may include changes in demographics and development patterns; transit system changes, including new, improved, or discontinued services; highway system changes, including new or improved facilities; and travel cost changes such as toll, parking cost, and transit fare changes.

However, proper backcasting is a time and resource-consuming process, and there is a risk of spending a large portion of the available resources for model validation on this single item. The main costs are in the development of all of the necessary data for a complete model run, in a manner consistent with the requirements of the updated model. Using data from a previous model version could require substantial work to provide consistency with the new model if the model has changed significantly. Some of the changes from the previous model version that can affect the development of backcast year data include:

- Changes to TAZ boundaries or expansion of the model region;
- Transportation network changes;
- Changes to model structure, such as changes in model functional forms (for example, gravity model to logit destination choice), changes in modal alternative definitions in mode choice, and changes in time periods;
- Changes in the input variables used in various model components;
- Changes in the data input definitions (for example, income groups); and
- The need to adjust cost data and income levels from the dollars of the original model's base year to the new model's base year.

Choosing a year after the model's base year, but before the present, is also appealing since the necessary data is recent and can sometimes be collected anew if necessary. Unlike a backcast, there is no need to try to use a previous model version. However, that also means that the data for the new "forecast" year has not yet been assembled and will require some time to do so. Additionally, since the model's base year is not too far in the past, the forecast year in this case is likely to be only a few years after the base year, which may be limiting in terms of the amount of change to test the sensitivity of the model against.

Using a forecast year in the future can be valuable since changes to the transportation system, demographic and land use characteristics, policies, and technology for which there are little or no observed data can be considered. For example, the sensitivity of the model results to levels of congestion, or amounts of development, that are not experienced presently can be considered. The main disadvantage, of course, is that there are no observed data to which model results can be compared (though this also means that no resources are required to assemble such data).

Sensitivity Testing

Sensitivity testing can include relatively simple tests of the sensitivity of the model to a single variable or more sophisticated tests of a particular scenario that encompasses one or more types of changes in conditions that the model might be asked to help analyze. The model is run with changed inputs, and the results are compared to a base scenario. Often, the base scenario can simply be the model's base year, although for some tests, a future "no build" scenario might serve as the basis for comparison.

Single variable tests are limited in that they test the sensitivity of the model to only one input (or type of input). But they can be performed with minimal setup time and are easier to interpret than some more complex tests. Some examples of single variable sensitivity tests include:

- Change in auto operating costs (for example, to represent fuel price changes);
- Change in transit fares;
- Change in transit headways;
- Change in roadway capacity; and
- Change in auto ownership.

Some areas have done more sophisticated sensitivity tests to examine the model's sensitivity to the types of changes in conditions that might be expected in the future. These might involve changing several different model inputs, perhaps both in the socioeconomic data and the transportation networks. Some examples of these types of sensitivity tests include the following:

- **Aging of the population** – In a model where person characteristics such as age are represented, such as an activity-based model, the ages can be changed to represent an older (or younger) population. In a four-step model, this type of change might be approximated by decreasing home-based work trips and increasing non-work trips.

- **Brownfield development** – The land use (and possibly transportation network) changes associated with the redevelopment of a brownfield site could be coded into the model inputs.

- **Reallocation of population or employment within the region** – For a forecast year scenario, more (or less) of the population or employment can be assumed to be located in suburban areas as opposed to city centers.

- **Increased (or decreased) levels of transit service** – The transit network can be revised to reflect an assumed increase or decrease in transit system investment.

- **Strategic road pricing** – A system of road pricing, perhaps varying by time of day, could be coded into the highway network.

6.0 References

Cambridge Systematics, Inc., 2010. Federal Highway Administration (FHWA) Model Validation and Reasonableness Checking Manual, Second Edition. Prepared for Federal Highway Administration.

Cambridge Systematics, Inc., Vanasse Hangen Brustlin, Inc., Gallop Corporation, Bhat, C.R., Shapiro Transportation Consulting, LLC, Martin/Alexiou/Bryson, PLLC, 2012. NCHRP Report 716 - Travel Demand Forecasting: Parameters and Techniques. Transportation Research Board.