New York Best Practice Model

Model Validation Plan

final

report

prepared for

New York Metropolitan Transportation Council

prepared by

Cambridge Systematics, Inc.

with

EA Harper Consulting

August 7, 2017

www.camsys.com

report

New York Best Practice Model

Model Validation Plan

prepared for

New York Metropolitan Transportation Council

prepared by

Cambridge Systematics, Inc.
100 CambridgePark Drive, Suite 400
Cambridge, MA 02140

with

EA Harper Consulting

date

August 7, 2017

Table of Contents

List of Tables

List of Figures

1.0 Overview

This document presents the model validation plan for the updated activity-based model, the New York Best Practices Model (NYBPM), that is being developed for the New York Metropolitan Transportation Council (NYMTC). The development of this model is being led by a consultant team consisting of Cambridge Systematics, Inc. (CS), University of Texas, Austin, Arizona State University, CDM Smith, Inc., Watchung Transportation, Inc., Gallop Corporation, Florida International University, EA Harper Consulting, and David Rubin. This report presents the plan for validating the updated activity-based model that will be developed by the CS team.

The development of this validation plan and the validation procedures and recommendations presented herein are based on guidance provided in the 2011 Travel Model Improvement Program Travel Model Validation and Reasonableness Checking Manual – Second Edition (Cambridge Systematics, Inc., 2010), hereafter referred to as the “Validation Manual.” This validation plan also builds on experience gained with other activity based models developed by the CS team.

The report is organized as follows:

- **Chapter 1.0** provides an overview of the validation process and describes the types of validation checks to be performed during the process.
- **Chapter 2.0** documents guidelines for model validation and summarizes validation efforts of the existing version of the NYBPM.
- **Chapter 3.0** presents an overview of the model development sequence and describes how model validation fits into this process.
- **Chapter 4.0** discusses the data available for use in model validation.
- **Chapter 5.0** presents the details of the validation tests for all model components, including the model input data, the components of PopGen, CEMSELTS, and CEMDAP, and the highway and transit assignments. Also discussed in this chapter are the temporal and sensitivity tests that are proposed to be part of the model validation process.
- **Chapter 6.0** provides citations for literature referenced in this report.

Because transportation planners may use the terms “estimation,” “validation,” and “calibration” differently, the terms are defined here as they are used in this document. These definitions are from the Validation Manual.

- **Estimation** is the use of statistical analysis techniques and observed data to develop model parameters or coefficients. While model estimation typically occurs at a disaggregate level without bias or correction factors, model estimation may also use statistical analysis procedures to analyze more aggregate data.

- **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 year or otherwise produce more reasonable results. Model calibration is often incorrectly considered to be model validation.

- **Validation** is the application of the models and comparison of the results against observed data. Validation should also include sensitivity testing.

1.1 Validation Process

The overall activity based model system is defined by the integration of three key components:

- **PopGen**, the synthetic population generator;
- **CEMSELTS**, the socioeconomic modeling system; and
- **CEMDAP**, the activity-based modeling engine.

The other important components of the overall NYBPM include the highway and transit assignment processes, and the non-activity based components including external travel, special generators, and commercial vehicle travel. The model design is documented by Cambridge Systematics, Inc. et al (2017). The updated NYBPM is being implemented using **TourCast**, Cambridge Systematics' activity based modeling platform as described in Cambridge Systematics, Inc. (2017).

The Validation Manual recommends that a validation plan be developed in conjunction with the model design plan whenever a travel model is estimated or updated. The validation plan should assess the available validation data, determine additional validation data that might reasonably be collected, determine what can be validated, set priorities, and specify any guidelines or standards that might be necessary.

The validation plan should also set the stage for quality model validation documentation. The validation documentation should be an honest assessment of how the model performs. Thus, rather than a blanket statement that the model is valid since it has met some artificial standard such as "the R^2 for assigned traffic volumes versus observed traffic counts exceeds 0.89," the validation documentation will summarize how closely the various model components reproduce observed data and the sensitivities of the various model components. The model validation and model validation documentation are intended to demonstrate levels of confidence that can reasonably be placed in model results as well as providing information to help set priorities for future data collection and model improvement efforts.

Finally, based on experience gained with other ABM validations, individual model component validations should focus on how well the component reproduces distributions and data that will be passed to subsequent modeling components. While other checks may also be useful, it is quite possible for the amount of information produced for the various tests to be overwhelming.

1.2 Types of Validation Checks

Four primary types of validation checks can be specified:

- **Base year comparisons**. Base year comparisons are, in some ways, the weakest validation checks since they are often performed using the same data that were used for model estimation, or because they are results of calibration efforts (e.g. changing model constants or model parameters) to match observed base year conditions. The usefulness of base year comparisons can be enhanced by splitting observed data into estimation and validation data sets (provided sufficient data exist) or by validating models to observed data using stratifications different from those used for estimation.

- **Temporal comparisons**. Estimating a model using data from one year and validating the model using data from a different time period is a strong validation process. It is especially strong if there are substantial socioeconomic, demographic, or transportation supply changes that take place between the two time periods used for the estimation and validation.

- **Sensitivity testing**. The application of the models and the model set using alternative input data or assumptions is especially important for models such as the NYBPM, which are designed to model traveler behavior, not travel patterns. Sensitivity testing of individual model components may include the estimation of the elasticities and cross-elasticities of model coefficients. However, sensitivity testing should also include the application of the entire model set using alternative assumptions regarding the input demographic data, socioeconomic data, or transportation system to determine if the model results are plausible and reasonable.

- **Reasonableness and logic testing**. These tests 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.”

Accuracy Requirements and Guidelines

The Validation Manual strongly supports the notion that matching specified standards is neither necessary nor sufficient to prove model validity. It avoids the specification of validation standards for this reason. Nevertheless, it is recognized that past standards may have been set by agencies such as NYMTC or others regarding model validation statistics. Such standards will be acknowledged and the necessary model statistics will be calculated. However, the meaning and implication of achieving (or not achieving) the standard will also be discussed.

Aggregation Level

Validation may be performed at two levels of aggregation:

- **Disaggregate validation**. As used in this document, disaggregate validation refers to comparisons performed at the household or individual level. Validation and reasonableness checking measures such as elasticities, prediction success tables, or R^2 are all examples of disaggregate measures providing the base unit for producing the measure is an individual or household. Individual information is required for both the modeled and the observed travel behavior being compared.

- **Aggregate validation**. Aggregate validation refers to comparisons performed after individuals or households have been aggregated over some common variable such as a geographic unit (zone, district, or community) or a socioeconomic unit (e.g. household size, income group, or auto ownership). Aggregate validation allows error for one trip making unit to cancel error for another similar unit in the aggregation scheme. By definition, measures based on traffic counts or transit boardings are aggregate measures since the counts or boardings are aggregations of individuals on roadway links or transit lines.

The accuracy of the validation data coupled with the ability to assimilate the information being provided by a validation test at any specific aggregation level must be considered. For example, any model that has a distribution component (e.g. regular workplace location, regular school location, work tour destination choice, etc.) could include validation tests at the district level, provided reasonable validation data can be developed and the results can be reasonably interpreted. Two issues must be considered:

- If the “observed” data are developed from a summary of the household survey, the estimate of the “observed” interchange might not be statistically significant. Smith (1979) provides a good summary of the impossibility of building statistically significant trip tables at anything beyond the most rudimentary level. For regular workplace location, it might be possible to construct reasonable home versus workplace matrices using data from the American Community Survey (ACS).

- Reviewing raw validation results for a large number of interchanges might be overwhelming (“you can’t see the forest for the trees”). Thus, innovative approaches might be required such as calculating percent differences and shading the differences in the tables to look for outliers or patterns, if performing comparisons at a larger aggregation level, such as county.

Disaggregate validation tests are typically performed in conjunction with model estimation. Review of the results coupled with stratifying the results by different strata can help guide model estimation. While the tests are typically performed as part of model estimation, it is important to summarize, review and document the results as part of the validation process.

Disaggregate validation tests are expected to reveal, at best, moderate success; low prediction success is the more common result. If high prediction success results are obtained when low or moderate results are expected, a review of the tests should be made. Unexpected results should lead to increased investigation of the model, the validation data, or the tests: either the results are providing information that is important to know, or there is an error somewhere.

The disaggregate tests should also be reviewed in light of the aggregate tests. The following matrix shows how various outcomes of disaggregate and aggregate validation results might be interpreted:

		Aggregate Validation Success	
		Low	**High**
Disaggregate Prediction Success	**Low**	Caution: model improvement might be possible	A likely result, but model might not be as sensitive as desired
	High	Not an expected result; model sensitivity might be overstated	Possibly too good to be true–double and triple check the results

Sources of Error

As documented in the Validation Manual, there are several types of error that can affect models and model validation, including:

- Model specification error;
- Model estimation error;
- Model aggregation error;
- Input data error; and
- Model validation data error.

The model development process seeks to minimize the first four types of error listed above and the model validation process seeks to measure the success of the model development process in minimizing the errors. However, it is often overlooked that model validation data also are subject to error. For example, “observed” average daily traffic counts are, in effect, based on surveys of traffic. Most are estimated from actual traffic counts performed on one or two days over the course of a year and factored to "average daily traffic."

All of the above types of error will affect validation results. At some levels of aggregation, the impacts might be significant. For example, at a disaggregate validation of individual mode choice behavior, the impact might be substantial and a rho-square value of 0.2 might represent a very reasonable model. For other, more aggregate comparisons, an R^2 value in that range might suggest that the model is not reasonably reproducing the observed data. However, in either case, it must be remembered that the observed data being used for the validation might be a source of some of the error.

2.0 Model Validation Guidelines

This section provides a brief review validation procedures suggested in the Validation Manual. This review is intended to provide context for the validation plan outlined in subsequent chapters.

2.1 Validation of Input Data

The Validation Manual recommends that the validation process start with the inputs to the modeling process, namely the transportation networks and socioeconomic data. The recommended validation approach for the transportation networks includes careful review of coded distances, network connectivity, and network characteristics. The Validation Manual also suggests that example paths be built through the highway and transit networks and checked for reasonableness both in terms of path and travel times.

Recommended validation of socioeconomic data includes review of data at both transportation analysis zone (TAZ) and county levels. Comparisons to independently collected data such as Census data, American Community Survey (ACS) data, data from the Quarterly Census of Employment and Wages (QCEW), and school enrollment data are recommended.

2.2 Validation of Activity Patterns/Amount of Travel

For trip-based models, the Validation Manual focuses the validation of trip generation models on measures such as:

- Person trip productions per household or per person;
- Proportion of person trips by trip purpose;
- Correlation of average modeled and observed (from a travel survey) trip rates for different geographies such as districts or area types; and
- Correlation of trips on a household basis (this may be a disaggregate test by applying the model on observed household data or an aggregate measure by applying the model to households aggregated by the independent variables used for the model).

These types of checks are also valid for activity-based models. Since activity-based models produce numbers of activities and tours by purpose, stops on each tour, numbers of work-based subtours, and possibly joint travel, each of these items can be summarized by market segment and compared to observed data (albeit, typically the same data that were used for model estimation). In addition, since activity-based models produce individual household and person records analogous to those found in a household survey, model results can be processed to produce traditional trip-based summaries such as person trip productions per capita and the proportions of person trips by trip purpose.

2.3 Validation of Trip Distribution/Destination Choice

For trip distribution-type models, the aggregate checks recommended in the Validation Manual are focused on averages and frequency distributions of travel or travel related information by different length measures. For an activity based model, the relevant models with length/distance components include regular workplace location, regular school location, tour primary destination choice, and intermediate stop location choice. Recommended validation measures include:

- Comparison of modeled and observed average lengths/distances by trip purpose and income group (or other socioeconomic group);
- Comparison of modeled and observed length/distance frequency distributions by trip purpose and, possibly, calculation of the "coincidence ratio[1];"
- Checking the percent of intrazonal location choices; and
- Comparison of modeled and observed location choices by purpose at an aggregate level such as county to county, or Public Use Microdata Area (PUMA) to PUMA.

2.4 Validation of Mode Choice

The Validation Manual emphasizes the disaggregate validation of mode choice models via application of the estimated choice models to observed choice data from a survey. If possible, the survey data used for the disaggregate validation should be an independent subset of the survey data used to estimate the models – in other words, the survey data would be randomly divided into two groups with one being used for model estimation and the other being used for disaggregate model validation. Practically, there is rarely sufficient data available to perform such an estimation/validation process. Because of this, the Validation Manual suggests validating using the data used for model estimation, but stratifying the validation by different socioeconomic or impedance values such as:

- Household characteristics such as household size, income level, number of workers, and auto ownership;
- Traveler characteristics such as age, gender, driver license status, and employment status;
- Zonal characteristics such as geographical location, area type, population density, and parking costs; and
- Trip/tour characteristics such as trip/tour distance, time, cost, and purpose.

The Validation Manual also suggests that model sensitivities be checked by reviewing direct and cross-elasticities of the model coefficients. The elasticities can be compared to those reported elsewhere, are derived from other models, or have been determined empirically.

Similar types of checks can be made for auto ownership models included in activity based models.

2.5 Validation of Time of Day of Travel

For time of day choice, the Validation Manual recommends that modeled percentages of tours and intermediate stops by purpose and time period be compared to observed percentages. Such checks can also be performed by market segments. In addition, time of day for tours is represented by an arrival and

[1] See Table 5-4 for more details.

departure time. The combination of arrival and departure times implies a duration, and comparisons of observed versus modeled tour durations should also be checked.

Disaggregate validation is also recommended for time of day choice models in the Validation Manual, although such tests are typically limited to reapplication of the choice models to the data used for model estimation. Using such an approach, aggregations of modeled and observed trips by time of day for different market/traveler segments such as worker status are recommended. In addition, the Validation Manual recommends sensitivity tests of the time of day model to verify its sensitivity to travel times and costs during specific periods in the day (e.g., morning and evening peak periods).

2.6 Validation of Assignment Procedures

Traffic and transit assignments represent the culmination of the modeling process. From the standpoint that the inputs to the assignment processes are based on the previous steps in the process, the assignment validations have often been used to represent a validation of the entire modeling process. Validation has traditionally been focused on the reproduction of traffic volumes and transit line boardings. More recent validations have also focused on the reproductions of reasonable speeds on roadway facilities.

The Validation Manual suggests that vehicle-miles of travel (VMT) be summarized for the region, per household, and per capita. The region-wide summaries should be by facility type with comparisons to VMT summaries obtained from regional traffic count programs.

The Validation Manual suggests comparing modeled to observed traffic volumes on a more disaggregate basis after the regional VMT and per household or per capita VMT estimates have been deemed acceptable. The following measures are suggested:

- Modeled versus observed volumes by screenline;
- Modeled versus observed volumes for all links with counts;
- Coefficients of determination (R^2) by link type (e.g. Functional class or volume group); and
- Root mean squared errors (RMSE) or percent RMSE (%RMSE) by link type.

The Validation Manual also suggests that speeds be reviewed for reasonableness. The Validation Manual suggests summarizing link speeds by facility type and area type, showing the minimum, maximum, and average speeds for each category. It also suggests comparing the assigned speeds with speeds used for distribution and mode choice and comparing estimated to observed speeds by highway segments, if the observed data are available.

For transit assignments, the Validation Manual states that the primary validation checks are modeled versus observed boardings for the region, by mode and, possibly, sub-mode, and by trip length. Optional, additional checks include modeled versus observed:

- Boardings per trip (transfer rates);
- Screenline volumes;
- Boardings by route or group of routes; and
- District-to-district transit trips.

3.0 Model Estimation-Calibration-Validation Sequence Overview

The validation process of an activity based model must be comprehensive. The CS team proposes to conduct the NYBPM validation task in various steps as described below.

3.1 Estimation of Model Components – Regional Household Survey/Regional Establishment Survey Data

The estimation of parameters for the various components of the updated NYBPM takes place under Task 8 of the project scope of work. In this task, the values of the various parameters for all of the CEMSELTS and CEMDAP components will be estimated using statistical model estimation procedures. The model design plan (Cambridge Systematics, Inc. et al, 2017) presents the details of these model components.

The main data sources for the model estimation are the Regional Household Travel Survey (RHTS) and the Regional Establishment Survey (RES). The validation process described later in this chapter includes running the model using the initially estimated parameters with inputs representing base year conditions, including the 2012 synthetic population produced by PopGen. These results will be compared to statistics generated from the RHTS and RES (mainly for measures for which there are no available independent data sources) and to other available data, such as the ACS and traffic and transit ridership counts (see Chapter 2). The comparisons will undoubtedly demonstrate areas where the estimated parameters should be modified to better reflect observed travel conditions, as described below.

3.2 Single-Pass Validation of Model Components on 2012 Base Year

First, the estimated models will be applied in sequence in a "single pass" using the congested speed network data used for the model estimation process. In other words, single-pass application will not involve any feedback loops (the model will be tested using feedback loops later, as described in Section 3.3). Model parameters and constants will be adjusted as necessary to better replicate expanded data from the household and transit on-board survey data. The single-pass process involves the following steps:

- **Apply each estimated model component** using the skims and socioeconomic data used for model estimation. This sequenced model run will be used to set a benchmark for the subsequent model validation work. The applied model results will be compared to the validation targets for every model component to see how well the model is performing.

- **Calibrate individual CEMSELTS and CEMDAP components**. Based on the above step, each component will be calibrated by adjusting parameter values and as necessary re-estimating certain models to include new variables. This process is described in more detail in Chapter 4 of this report.

- **Examine error propagations**. As all of the activity based model components are linked to one another and applied in sequence, each subsequent model component is affected by models upstream. Doing a single-pass validation therefore helps understand the magnitude and direction of error propagation through the model system.

This task will be done for a base year of 2012 with all 2012 inputs – synthetic population, zonal socioeconomic data, networks and skims.

3.3 Full Feedback Validation to 2012 Base Year

After all individual model components are validated, a full model run with feedback loops will be run to evaluate the impact of feedback loops on the calibrated components. Feedback loops enable the model system to use congested characteristics on travel behavior; that is, the input data (times, etc.) are as realistic as possible. This will affect the number, destination, mode, timing and routing of tours and trips. The NYBPM model region includes not only New York City, but other large cities and busy travel corridors. So having adequate feedback loops in the model system is essential to obtain accurate estimates of such measures as air quality, transit demand, and peak spreading.

The number of feedback loops that are necessary will be determined based on checks of model convergence. An automated process will be set up so that the whole model run can be executed using a batch process, and the convergence criteria will be specified to state of practice standards. The gap closure between subsequent feedback loops will be based on optimal weighting procedures such as mathematical program-based weighting among iterations, method of successive averages (MSA), or equalized weighting among iterations. The goal is to minimize the difference between input initial values and resulting output values, and in the process, it will ensure model stability.

There are several effects of feedback loops on the ABM components. For example:

- **Tour generation** – The feedback of congested skims affects accessibility measures that influence each person's tour generating behavior. This could either result in fewer or more trips depending upon the tour origin and destination. It could also affect trip chaining and tour formation.
- **Daily activity patterns** – The accessibility to various activities and opportunities is affected by congested travel times, which will in turn affect the daily activity patterns of travelers.
- **Destination/location choices** – The congested skims influence the trip lengths directly, which affects the distribution patterns of tours and trips as well as tour and trip lengths.
- **Mode choices** – Mode choice is directly affected by the feedback of congested times.
- **Time-of-day** – Congestion causes people to alter time of travel, which induces peak spreading. So these models also are affected by the feedback loops.
- **Highway and transit assignments** – These are affected by the equilibration process of the iterations within each loop as well as between subsequent loops.

Once the number of feedback loops is determined, the key results from each model component will be compared to the observed (survey) data; the results of this comparison may warrant additional calibration. The validation of the whole model will be performed and the assignment results will be compared against traffic counts and transit boardings.

3.4 Sensitivity Testing

As part of the validation process, a series of sensitivity tests will be performed using the updated NYBPM. This will include executing a number of different scenarios with varying input parameters for several of its components. These tests will be focused on the most important factors that will have a direct impact on the projection of future travel behavior in the region. The tests are meant to examine how vulnerable the forecasts are for certain plausible scenarios and how sensitive is the forecasting model to various level-of-service parameters and other variables. A subset of the following tests will be undertaken:

- Socioeconomic and demographic factors:
  - Alternate growth rates of population and/or employment
  - Alternate growth rates of different market segments such as aging of population, presence of more females in the workforce, increase in low income households, etc.
- Cost inputs:
  - Adjustments to fuel costs
  - Adjustments to toll levels, perhaps by time of day
  - Adjustments to parking costs
  - Adjustments to transit fares
- Impact of new highway projects:
  - New managed lanes or roadway pricing scenarios
  - Widening/narrowing of highways
- Impact of new transit projects:
  - Extension of rail lines

Elasticity is a convenient, quantitative measure of travel demand response to price and service changes which influence demand (McCollom et al, 2004). For elasticity measures to be applicable, the transportation system change must be based on a relative measure. In other words, it must involve a quantifiable percentage increase or decrease in the system parameter involved. For example, while elasticity measures can be used to describe the response to a change in the overall amount of transit service, they cannot be used to describe the response to a new transit system. Transportation elasticities are informally adopted from the economist's measure "price elasticity." The price elasticity of demand is loosely defined as the percentage change in quantity of service demand in response to a 1 percent change in price. For instance, a price elasticity of -0.3 indicates that for a 1 percent increase (or decrease) in the price of a service, there is 0.3 percent decrease (or increase) in the demand for that service.

A model is said to be inelastic if the calculated elasticity is less than 1.0 and elastic if 1.0 or greater. Roughly, an elasticity of 1.0 implies that a 1 percent increase in "price" of a "product" will produce a 1 percent increase in demand for the product. A negative 1.0 elasticity simply implies that a 1 percent increase in "price" will produce a 1 percent decrease in demand. For the NYBPM validation effort, both direct and cross-elasticities will be estimated for certain key explanatory variables. Direct elasticities are based on the change in demand for a mode based on a change in cost (price) for that mode, while cross-elasticities are based on the change in demand for an alternate mode based on the change in cost for a given mode.

Backcasting

Backcasting provides a good validation test as it provides a second point in time for which model results can be compared to observed data. It 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.

The main factor in choosing a backcast year is the availability of model input data and observed data to which model results can be compared. It is desirable to choose a backcast year that precedes some demographic, transportation system, or travel cost changes so that the model's sensitivity to these items may be considered. This means that the backcast year should not be too close to the base year; for example, 2010 would not be a good choice for a backcast year given that 2012 is the model's base year. However, the backcast year should not be too far in the past due to the difficulty in developing model input and observed travel data sets. For the updated NYBPM, a backcast year around 2005 would be reasonable, assuming the availability of data.

It should be noted that backcasting is a time and resource consuming process, and there is a risk of spending too much 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 usually requires substantial work to provide consistency with the new model. There are many changes in the NYBPM update from the previous model version that will affect the development of backcast year data, including:

- Changes to TAZ boundaries, especially in New Jersey;
- The use of micro-analysis zones (MAZ);
- The use of General Transit Feed Specification (GTFS) data in the development of transit networks;
- Other transportation network changes;
- Migration to the CEMSELTS/CEMDAP structure for the activity based model components; and
- Use of PopGen for the development of the synthetic population.

After discussions with NYMTC, it was agreed that rather than a backcast, a forecast will be performed from the base year of 2012 to the year 2016 (or to 2015 if 2016 data are not available in time). Model inputs for 2016 will be developed, including highway and transit networks, changes to employment data and a new 2016 synthetic population reflecting 2016 population and household characteristics. Model results from the 2016 scenario will be compared to observed travel data for 2016, including traffic counts and transit ridership. The comparison of the overall model results will be similar to those done for the base year.

4.0 Validation Data Assessment

This section identifies the data available for validation of the activity-based model. A variety of data are needed to perform the validation tests described in Chapter 5.0, but those tests are limited by the available data. While some tests identified in Chapter 5.0 require data that are not currently available, identifying and assessing the existing data help focus Chapter 5.0 toward meaningful tests that can be readily performed.

4.1 Existing Data Sources

This section describes the existing data and data sources available for use in validation of the updated NYBPM.

Socioeconomic Data

Typically, input socioeconomic data validation sources are the same as those used to develop the data. Few regions have multiple sources of the same socioeconomic data for a particular year. The main socioeconomic data available are the Census, American Community Survey (ACS) (including the Census Transportation Planning Products (CTPP) and Public Use Microdata Sample (PUMS)), and Longitudinal Employment Household Dynamics (LEHD).

2010 Census Data

The decennial U.S. Census provides information on all persons and households in the country and can be viewed at census block level geographic resolution. The 2010 Census focused on population and housing questions. Questions that were asked in the 2000 Census regarding income, auto ownership, and employment are now part of the ACS and were not asked in the 2010 Census, and are therefore no longer available on SF1. The 2010 SF1 data will be used to examine univariate distributions of households and persons across particular variables (e.g., households by household size and persons by sex or age).

American Community Survey Data

The ACS is conducted continuously by the Census Bureau and provides a great deal of information that can be used for validation. Because the ACS is conducted continuously, the Census Bureau can make data available every year rather than every 10 years like the decennial census (though for a smaller samples of the population).

The ACS samples about one in every 40 addresses every year, or 250,000 addresses every month. For areas with large populations (65,000 or more), survey estimates are based on 12 months of ACS data. For all areas with populations of 20,000 or more, the survey estimates are based on three years of ACS data. The Census Bureau produces estimates for all areas, down to the census tract and block group levels, based on five years of ACS data. One, three, and five year estimates based on survey data are currently available.

The ACS provides data on housing and population not available from the 2010 Census. Information includes:

- Population characteristics
    - Age

- Sex
- Relationship to head of household
- Income
- Employment information including labor force status, industry and occupation
- Journey to work information

- Household characteristics
  - Vehicles available
  - Income
  - Tenure
  - Housing value
  - Rent

The Census Transportation Planning Products (CTPP) are based on data from the ACS. In addition to providing information on the place of work of residents and the journey to work noted above for the ACS, the CTPP also provide zonal level information in the form of cross-classifications across variables. For instance, one could examine households by size, number of vehicles, and income rather than univariate distributions provided in the data sources described above.

The Public Use Microdata Sample (PUMS) from the ACS show the full range of population and housing unit responses collected on individual ACS questionnaires. For example, they show how respondents answered questions on occupation, place of work, and so forth. The records contain information from the completed ACS questionnaires for most questions for the selected subsample of housing units and group quarters persons including questions on age, sex, tenure, income, education, language spoken at home, journey to work, occupation, condominium status, shelter costs, vehicles available, and other subjects. Many multi-variate customized tabulations can be summarized from the PUMS files.

Longitudinal Employment Household Dynamics

The Longitudinal Employment Household Dynamics (LEHD) provides even more detailed data on workers and employers than the CTPP. Employer characteristics are reported for each employer with geocoded addresses. Worker/individual characteristics include wage records, personal characteristics (e.g., gender, age, ethnicity), and location of residence. Compared with the worker flow and journey to work data from the CTPP, the LEHD data enumerates the workforce population (rather than a small sample), includes a more comprehensive geographic coverage including counties with low population, and includes second jobs for workers.

Travel Survey Data

The NYMTC region has a rich array of relevant travel survey data available for the timeframe surrounding 2012. The Regional Household Travel Survey (RHTS) and the Regional Establishment Survey (RES) were initiated by NYMTC in anticipation of the 2012 update, and the major transit operators in the region have data collection programs designed to understand and serve their customer's needs.

Regional Household Survey

The RHTS is the primary source of observed choice preference data that will be used for estimation of most model components. It is also an important source of model validation data since much of the information on specific travel behavior of specific travel types is unavailable elsewhere.

The RHTS was conducted by NYMTC and North Jersey Transportation Planning Authority (NJTPA) in 2010-2011. Like most household travel surveys, the survey collected information specific to each household, including information related to each person living in the household and each vehicle owned by the household. In addition, each household was assigned a travel day, when household members were asked to record all travel and the characteristics of that travel for a 24-hour period. In total, nearly 19,000 households completed the travel diary information and made nearly 144,000 (linked) trips during their designated travel days. Household were surveyed from each of the 28 counties within the New York Metro area, including 12 New York counties, 14 New Jersey counties, and two Connecticut counties.

The data are organized in four relational databases described as the household file, person file, vehicle file, and trip file. The trip file will be further processed to extract tour, joint travel, and intermediate stop-making data items. Table 4-1 shows the variables that are available or will be derived from the data and used in model estimation. Other variables may be derived from those listed in the table, but the basic information is fully contained in the table.

Table 4-1. Data Items from the RHTS

DESCRIPTION	DETAILS
Basic Person and Household Variables	
Household ID number	Survey ID field
Person ID number	Survey ID field
# people in household	
# vehicles in household	
Household income	Categorical household income
Residence type	1 = single-family, 2 = multi-family, 3 = other
Gender	1 = male, 2 = female
Age	Categorical age range
Race/ethnicity	
Worker status	1 = yes, 2 = no
Hours worked	
Telecommute hours	
Work location type	1 = fixed, 2 = home, 3 = varies
Work start time variation	Categorical start time flexibility
Work end time variation	Categorical end time flexibility
Student status	1 = full time, 2 = part time, 3 = no
Derived Person and Household Variables	
# employed household members	Derived by adding across household members
# student household members	Derived by adding across household members
School day	Derived from day activities (for students)
Work day	Derived from day activities (for workers)
Mode escorting to/from school	Derived from day activities
Joint activity decisions	Number, purpose, participation, & duration of joint activity episodes for household
Individual activity decisions	Number, purpose, participation, & duration of individual activity episodes for household
Commute mode	
# before work, work-based, & after work tours	For worker scheduling
# independent tours	For non-worker scheduling

Table 4-1: Data Items from the RHTS (continued)

DESCRIPTION	DETAILS
Person/Household Location Variables	
Household residence ID number	Survey ID field
Household residence X coordinate	Geocode
Household residence Y coordinate	Geocode
Household zone	Geocode
Regular work location ID	Survey ID field
Regular work X coordinate	Geocode
Regular work Y coordinate	Geocode
Regular work zone	Geocode
School X coordinate	Geocode
School Y coordinate	Geocode
School zone	Geocode
Vehicle Variables	
Vehicle ID number	Survey ID field
Household ID number	Survey ID field
Vehicle year	
Vehicle type	Categorical vehicle body type
Fuel type	Categorical fuel type
E-ZPass tag	1 = yes, 2 = no
Tour Level Variables	
Tour ID number	Created ID field
Subtour parent tour ID (work-based subtours)	Created ID field
Tour mode	
Home or work stay duration before tour	Continuous time
Number of stops	
Vehicle ID number used	Survey ID field
Stop Level Variables	
Activity purpose	Categorical purpose
Activity duration	Continuous time
Stop X coordinate	Geocode
Stop Y coordinate	Geocode
Stop zone	Geocode
Stop mode	Categorical mode

The data in Table 4-1 is split into several categories.

1. **Basic person and household variables**. These variables are taken directly from the survey data.

2. **Derived person and household variables**. These variables will be derived from the data. Most of these variables deal with the activity participation of individuals in the household.

3. **Person/Household Location Variables**. These variables describe the locations of each household and the key locations of household members, such as the regular workplace for workers and regular school location for students.

4. **Vehicle Variables**. These variables are drawn directly from the survey data and describe the information related to vehicles owned by a household.

5. **Tour Level Variables**. These variables describe information at the tour level. The variables include tour mode, stay duration previous to the tour, number of stops, vehicle information, and information related to work-based subtour formation.

6. **Stop Level Variables**. These variables include activity purpose, duration, and location.

Regional Establishment Survey

The RES, like the RHTS, was designed to collect key information from establishments from the same counties in New York, New Jersey, and Connecticut for use in the regional model development. Like the RHTS, the RES data set is being used in model estimation, but is also valuable for model validation as the only available source for travel behavior for specific types of travelers at the establishment type level.

Two main components of the RES will be used in the development of the NYBPM:

- **Core sample** – The core sample includes establishments in the retail, food services, and recreational categories and is further stratified by establishment size and location within the region. The target number of responses is 7,840 visitors to establishments. The core sample provides more detailed information about travel to these types of establishments than is available from the RHTS. The plan is to pool data from the RHTS and the RES for the estimation of key components of the activity based model. This information can be used for the estimation of destination choice (parameters for size and impedance variables), time of day choice, and mode choice models.

- **Hotel sample** – The hotel sample data set will be the main estimation data set for the new visitor component of the NYBPM (see Section 7.4). Each response includes one day's worth of travel for a visitor to the region staying at an area hotel, similar to diary data collected in the RHTS but with less detail. The target number of responses is 900 hotel guests.

Transit On-Board Surveys

The region has many transit operators that collect survey data for their systems. The existing NYBPM includes the services of 65 transit operators and 4,170 routes of various modes. Many of these operators are smaller companies and/or private operators with no obligation to report information about their operations or passengers.

However, the major carriers are public, and many have collected data on patronage for planning and customer service purposes. The extent of data collected varies considerably across operators, but many of the major operators have recent, usable, full on-board surveys that are sufficient for validation at the boarding station, access mode, and time-of-day level. Major operators/systems that have provided this ideal level of data include (year of survey shown in parentheses):

- Long Island Rail Road (2012-2014)
- Metro-North Rail Road (East of Hudson, 2007)
- PATH (Port Authority of New York an New Jersey) (2012)
- New York City Transit (NYCT) subway/bus (2008)
- Newark City Subway (2008)
- Hudson-Bergen Light Rail (2008)
- NY Waterways ferries (2013)
- Nassau Inter-County Express (NICE, formerly MTA Long Island Bus) (2013) (in report form only)
- New Jersey Transit (NJ Transit) Bus (2013)

In all of the on-board surveys listed above the time of day is recorded in hours, and in some cases to the minute of boarding, so that aggregations into time periods is possible. In addition, all fixed guideway surveys capture the boarding station or terminal and access modes and have been expanded to counts at the boarding location for the year of the survey.

Three other NJ Transit commuter rail full on-board surveys have been requested. These include:

- Main-Bergen Line and Pascack Valley Line (jointly with Metro-North West of Hudson Lines, 2013)
- Northeast Corridor Line and New Jersey Coastline (2014)
- Raritan Valley Line (2014)

Survey data from 2013 is available in a merged private bus survey for the following bus companies.

- Suburban Lines (2005)
- Olympia Trails/ONE Bus (2013)
- Rockland Transit (2013)
- Academy (2013)
- Coach USA (2013)
- DeCamp (2013)
- Lakeland Bus (2013)
- Martz (2013)
- Monsey Trails (2013)
- Short Line (2013)
- Transbridge Lines (2013)

There are another 45 smaller bus, ferry and tram services, both private and public, for which there is no available survey data. The largest bus operators for which we have no data include Westport and Connecticut Transit. The most consequential omissions are three NJ Transit Commuter Rail lines for which we are aware of no available survey data. These include the Montclair-Boonton Line, the Morristown Line, and the Gladstone Branch. Efforts to identify and obtain data for these services is underway as a part of Task 5.

Passively Collected Travel Data

NYMTC is in the process of purchasing Streetlight data, which provides a large amount of travel “traces” from GPS data. Once the data has become available, CS will explore using this data set for model validation.

Transportation System Utilization

A variety of available count data will be used for validation. This includes both highway utilization data (i.e., traffic counts) and transit ridership data (i.e., boarding counts). Both traffic and transit counts are being assembled as part of Task 5 of this project, which is scheduled to be completed in September 2017. This section discusses the data that has been assembled so far as well as data that is expected to be assembled later in Task 5.

Another key source of count data that includes both auto and transit users is the Hub Bound Travel data developed by NYMTC. This data set includes person and vehicle travel to the Manhattan Central Business District (defined as south of 60th Street). The archived 2012 data[2] will be used for the base year validation effort.

Traffic Counts

A variety of available traffic count data will be used for validation. Screenline count locations have been identified and review of these data, available by hour, is underway as part of Task 5. Other count data, including classification counts, from New York City, New York State, river crossings, and major New Jersey screenlines from the NJTPA travel demand model, are being reviewed for use in model validation.

Screenline count locations have been identified and review is underway. Additional counts, including classification counts, may be requested. Count data from New York City and State, river crossings, and major New Jersey screenlines from the NJTPA travel demand model, will be used.

The Task 5 technical memorandum will provide the details of the assembly of the traffic count data.

Transit Boarding Counts

For those services which were surveyed in a year other than 2012, we will need boarding counts for both the year of the survey and 2012 in order to adjust transit surveys to the model base year.

All of the on-board surveys described above have been expanded to counts appropriate for the year of the survey. For purposes of adjusting the survey data to the 2012 base year, additional counts are required. Counts by route and, in the case of fixed guideways and ferries, by station will be obtained from the operators. The objective is to obtain as much count detail as possible for the bus systems, recognizing that for some of the smaller bus services and for many of the private carriers estimates of total boardings by system will be the most detail that can be obtained. Data gathering of count data and the creation of comprehensive datasets for validation is underway as a part of Task 5 and will be documented in the Task 5 technical memorandum.

Highway Speed Data

The National Performance Management Research Data Set (NPMRDS) provides vehicle probe-based travel time data for passenger autos and trucks on the National Highway System (NHS). The real-time probe data

[2] https://www.nymtc.org/Data-and-Modeling/Transportation-Data-and-Statistics/Publications/Hub-Bound-Travel
https://www.nymtc.org/Utility-Menu/Archive/Data-and-Modeling-Archive/Hub-Bound-Travel-Archive

are collected from a variety of sources including mobile devices, connected autos, portable navigation devices, commercial fleet and sensors. NPMRDS includes historical average travel times in 5 minute increments on daily basis. The NPMRDS travel times are available for the existing NYBPM and can be accessed via NYMTC's Transportation Information Gateway (TIG). We will compare modeled travel times in a select number of representative corridors to the NPMRDS data during the validation process.

4.2 Model Parameters

A key piece of information that will be used in model validation is the set of model parameters estimated in other regions. As discussed in more detail in Section 5.0, comparison of model parameters to other regions provides a relatively simple reasonableness check. The most likely comparison source for model parameters are the CEMDAP/CEMSELTS models developed for the Southern California Association of Governments regions, which have similar structures to the NYBPM update. Other activity based models developed for large U.S. metropolitan areas may have some components with somewhat similar structures and may also be used for parameter comparisons. Additionally, model parameters from the existing NYBPM may be used for comparison in cases where the model structures are similar enough.

It is important to note that parameter comparisons are done as reasonableness checks only. We do not expect the NYBPM parameters to exactly match those from other regions since the New York region has many unique characteristics.

4.3 Temporal Validation Year Survey and Model Inputs and Counts

Forecast Year Data

NYMTC has identified the forecast year as 2045. While the forecast year model results cannot be compared to observed travel data, the results can be checked for reasonableness and proper sensitivity to the things that change between the base and forecast years. The main data needed are therefore the input data for the forecast year run. These include the following:

- **Highway and transit networks**, consistent with the base year networks. It will probably be most efficient to develop these networks by creating copies of the base year networks and editing them to reflect expected forecast year conditions.

- **Socioeconomic data** (households, population, and employment) at the TAZ and MAZ level, at a segmentation level consistent with the base year data. Other zone level data, such as parking costs, will also be needed for the forecast year. We note that forecasts of population, household, and employment may not be available at the same level of segmentation as for the base year; it may therefore be necessary to make assumptions in segmenting the available forecasts.

- A **synthetic population** for the forecast year. This would be created by running PopGen using control totals consistent with the forecast year's socioeconomic data. It would probably be most efficient to use the same ACS-based seed distributions as for the base year rather than developing new seeds from more recent ACS data, which will be at most only a few years closer to the forecast year.

Backcast Year Data

If a backcast is performed as part of the model validation, several data items will be needed, including the following.

The RHTS data used in the development of the updated NYBPM were collected in 2010-2011. The previous household survey for the region was performed in 1997-1998. Depending on the backcast year chosen, the older survey may not represent travel behavior for the backcast year better than the more recent survey. Given the expense of preparing the household survey data set for backcasting, we recommend continuing to use the 2011-2012 RHTS data set for backcast year validation. Note that the use of the household survey data is not to measure the overall amount of travel but rather the shares and distributions of various factors such as time of day splits, trip length frequencies, and mode shares.

There may be other surveys (for example transit on-board surveys) conducted for periods closer to the backcast year than the surveys used for base year validation. These survey data sets must be examined to determine whether they can be processed efficiently for use in the backcasting.

Traffic counts for the period around the backcast year must be processed. Transit boarding counts for this period must also be obtained and processed. Ideally, the data should be developed at the same level of detail as for the base year validation. For example, the number of traffic counts by segment (facility type, geographic subregion, etc.) should be about the same as for the base year validation, and if possible the screenline definitions should be the same.

A complete set of model inputs for the backcast year will be needed. These include the following:

- **Highway and transit networks**, consistent with the base year networks. Centroid connectors must be consistent with the updated NYBPM TAZ system (in nearly every case, they should be the same as in the base year). Toll and fare values should be consistent with what was in place in the backcast year.

- **Socioeconomic data** (households, population, and employment) at the TAZ and MAZ level, at a segmentation level consistent with the base year data. Other zone level data, such as parking costs, will also be needed for the backcast year. Depending on the backcast year chosen, data from the 2000 U.S. Census or the ACS from years earlier than the base year might be useful. (Note that there is a gap between the 2000 Census and the start of the ACS in 2005.)

- A **synthetic population** for the backcast year. This would be created by running PopGen using control totals consistent with the backcast year's socioeconomic data. It would probably be most efficient to use the same ACS-based seed distributions as for the base year rather than developing new seeds from earlier ACS or 2000 U.S. Census data.

5.0 Planned Validation Tests

This section of the model validation plan discusses specific validation tests that are planned. These include tests of the socioeconomic and network input data used, tests for ABM component model, system-level validation tests for highways and transit, and temporal validation and sensitivity tests.

Because the validation effort represents a major undertaking, each proposed test was assigned a priority describing its importance. Priorities are based on the following considerations:

- **Level 1 priority tests** include those tests that can be produced with a relatively low level of effort and provide good measures of the validity of the models. This priority level includes aggregate validation measures like vehicle miles traveled (VMT) and vehicle-hours traveled (VHT) summaries, tours per person and per household, and average tour durations and distances.

- **Level 2 priority tests** include more detailed and non-standard validation tests. These tests can be considered to be more directly focused on the ABM components. Since the tests are more detailed and/or non-traditional, they may be more difficult to produce or interpret, or it may be more difficult to acquire data or information from other regions for comparison.

The priorities were assigned to each validation test with the goal of producing the most return for the investment in the validation testing.

In addition to priorities, the tests are also classified into (a) basic tests, and (b) debugging tests. The "basic" tests will be conducted for all ABM components and it will provide enough insights into the performance of each component. The "debugging" tests are designed to provide more information at disaggregate levels and are intended to catch any errors or examine any anomalies found in the "basic" test results.

The validation tests for inputs to the ABM are discussed in Section 5.1. Section 5.2 discusses validation tests specifically for the population synthesis procedure and each of the estimated model components. System level validation tests discussed in Section 5.3 will rely on the final outputs of the model. These include highway and transit trip tables, highway volumes, and transit boardings by time of day.

5.1 Input Data Validation

Two main types of data are required for any travel model process: socioeconomic data and transportation network data. These data represent the basic building blocks used with travel models to forecast regional travel. Not only do these inputs affect forecasts, but they are also used in model estimation, calibration, and validation, which can have important consequences for each step. Thus, ensuring the reasonableness of socioeconomic and network data is critical to the overall modeling process.

Socioeconomic Data

The zonal-level data available as inputs into the model include:

- Number of households
- Population
- Employment by type (retail, office, other)

- K-12 educational enrollment
- University enrollment

The CS team will rely on the accuracy of inputs provided by NYMTC. Since the household-based data will form the control totals for PopGen, and the employment-based and educational enrollment data will be instrumental in several model components, the data should be checked if NYMTC has not previously validated the data.

Table 5-1 presents a list of validation checks by level of aggregation that should be considered. Validation checks at the TAZ level should be based on the compatibility of the TAZ system with 2010 census block definitions. No specific criteria have been specified for the validation tests; rather, the reasonableness of the data should be gauged by potential impacts on model results.

Table 5-1. Socioeconomic Data Validation Tests

AGGREGATION LEVEL	DATA ITEM	VALIDATION TEST	PRIORITY

County	• Number of households
	• Population in households
	• Median household income
	• Employment by Type
	  • Retail
	  • Office
	  • Other
		• Compare to 2010 Census SF1 for households and population – data should match at county level
		• Estimate median household income from TAZ-level data and compare to available ACS 5-year estimates – results should be "close"
		• Compare employment by type to LEHD data – results should be "close"
			Level 1 (Basic Test)

TAZ	• Color coded GIS plots of percent differences between TAZ data and 2010 Census data, ACS data, and LEHD for:
	  • Number of households
	  • Population in households
	  • Median household income
	  • Employment by type
	• Color coded plots of TAZ data for:
	  • K-12 educational enrollment
	  • Area type
		• Review trends and look for outliers in term of large percentage differences	Level 1 (Debugging Test)

Transportation Network Data

Transportation network data (highway and transit) are the other key input to travel demand forecasting models. The transportation network data and the path-building procedures that use these data must be verified. Verification will include the following steps:

Highway Network Verification

Highway networks will be spot-checked to confirm accuracy.

- On a systematic level, estimated highway travel times for a series of highway journeys will be compared to observed travel times to confirm approximate agreement. This test will help to confirm the appropriateness of the entire network processing procedures including assignment of free flow speeds and capacities, and the how volume-delay functions relate traffic to reduced (congested) travel speeds.

- Toll and HOV facilities and skims will also be reviewed for accuracy (e.g. toll charges) and reasonableness. As with the validation of general purpose highway skims noted above, estimated travel times for a series of journeys using toll or HOV facilities will be compared to observed travel times to confirm approximate agreement.

For model estimation, reasonable speeds will need to be used to build highway skims. It was decided to build highway skims for the model estimation at the outset and develop the speed processing procedures in parallel with the model estimation process, with checks performed of the highway skims.

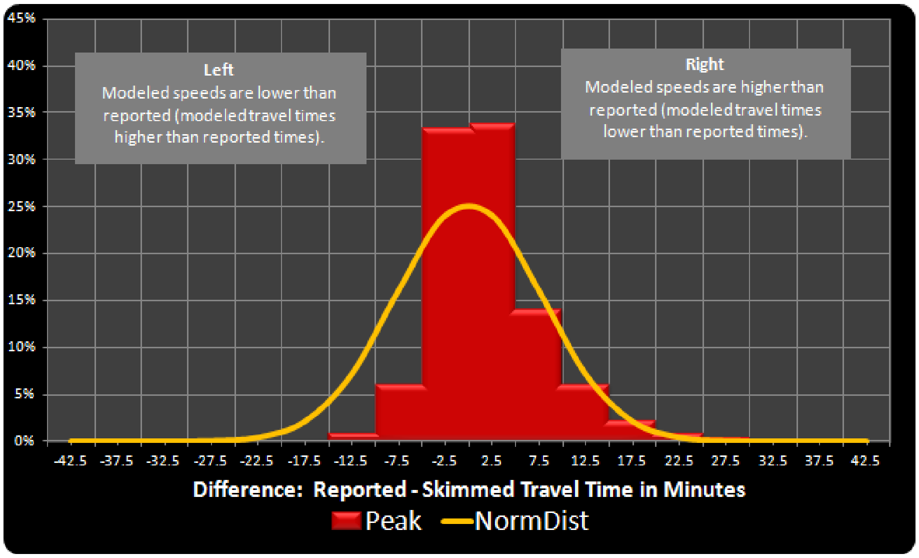
Validation of the highway skims will rely on observed travel time data. In addition to identified data sources (see Section 4.1), we will look to the reported travel time data from the household survey. Of course, these data have well documented reporting issues of rounding starting and ending times. In addition, depending on how activities were defined, reported departure and arrival times may include “terminal times.” Nevertheless, understanding these issues will help us interpret the results.

The suggested validation process will involve posting the modeled interchange specific travel times on the household survey trip data for auto drivers and producing scatter plots and trip length frequency distributions of the modeled versus reported travel times. The following household survey trip data are required for each auto driver trip:

- Origin TAZ
- Destination TAZ
- Time period
- Expansion factor
- Reported travel time

The skim data, which will be posted on the survey records, will be based on the final “equilibrium” speeds from the time of day assignments. In addition to the modeled travel time for the appropriate time of day, the modeled distance will also be posted. This will allow us to calculate “reported” interchange travel speeds to filter out outliers (e.g. implied speeds greater than, say, 80 miles per hour). The trip length frequency distributions for both the modeled and reported travel times will be produced for each time of day as shown in Figure 5-1.

Figure 5-1. Example of Modeled and Observed Trip Length Frequency Distributions



Transit Network Verification

Transit networks will be examined using the following tests:

- Review of rules for developing walk-access, park-ride access links.
- Review of rules for estimating mixed-flow (bus) running times as a function of congested highway times.
- Review of resulting estimates of bus running time on a time point-to-time point basis by time of day.
- Review coded fixed guideway running time estimates by time of day against timetables.
- Review rules and coded headways by time-of-day for routes against timetable values.
- Aggregate path-building test through an assignment of a transit trip table estimated from the on-board survey to network to confirm that line loads and station on-offs by mode of access and time of day reasonably match observed values.
- Disaggregate path-building test of selected individual paths from the on-board survey to confirm that reported and path-builder paths are similar. If possible, this will be automated so that prediction success tables of modeled versus reported boardings can be compared for the numerous selected paths. Previous experience indicates that about 60 percent of the modeled paths can be expected to have the same number of boardings per linked trip as the reported paths.
- Examine resulting estimated travel times (by time of day) contained in the level-of-service matrices to confirm that travel times estimates are properly scaled (e.g., minutes) and are located in the proper file and table.

As with the socioeconomic data tests, there will not be any specific criteria by which these network tests will be measured. Reasonableness of the data will be gauged by potential impacts on model results.

5.2 Component Validation Tests

PopGen

The updated NYBPM will use the PopGen synthetic population generator. Table 5-2 summarizes the reasonableness tests for household data that will be produced for base year 2012. The primary reasonableness tests will be socioeconomic distributions stratified by various geographic strata. In addition, color-coded plots showing changes between selected variable values by TAZ will be produced. Table 5-2 also shows the priority level for the tests.

Table 5-2. Synthetic Population Generator Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOME	PRIORITY
Regional	• Two dimensional cross-tabulations: • Persons by age and sex • Average household size	• Confirm that control variables from 2010 census data have been maintained (±3 percent) • Review trends of non-control variables for reasonableness over time	Level 1 (Basic Test)
County	• Same as for regional	• Review trends over time for reasonableness	Level 1 (Debugging Test)

TAZ	• Color coded GIS plots of percent differences between TAZ data and 2010 Census data, ACS data for:
	  • Average household size
	  • Median age
		• Review trends and look for outliers in term of large percentage changes	Level 2 (Debugging Test)

It should be noted that some person and household characteristics are estimated later by CEMSELTS. The tests of the model components that estimate these characteristics are discussed later during the discussion of CEMSELTS.

CEMSELTS Components

The following are the main components of CEMSELTS to be validated:

- Person level characteristics
  - Student status (by age group)
  - Labor force participation
  - Occupation Industry
- Location choices
  - School location
  - College location
  - Work location
- Work activity characteristics
  - Work duration
  - Work flexibility

  - Mobility choices
- Household characteristics
  - Household income
  - Residential tenure
  - Housing unit type
  - Annual household mileage
  - Vehicle fleet and primary driver allocation

Table 5-3 presents the validation tests for the person level characteristics. The school and college location choice tests are shown in Table 5-4, and Table 5-5 presents the tests for workplace location choice. Table 5-6 shows the validation tests for the work activity characteristics, and the tests for the household characteristics are presented in Table 5-7.

Table 5-3. CEMSELTS Person Level Characteristics Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOME	PRIORITY

Regional	• Two dimensional cross-tabulations:
	  • Persons by age and employment status
	  • Workers by age and industry
	  • Persons by age by school grade
	• Population/employment ratio
	• Percentage of retired households (e.g., all one or two person households with all adults over 55 and no working adults)
		• Confirm that control variables from 2010 census data have been maintained (±3 percent)
		• Review trends of non-control variables for reasonableness
			Level 1 (Basic Test)

County	• Same as for regional	• Review trends for reasonableness	Level 1 (Debugging Test)

TAZ	• Color coded GIS plots of percent differences between TAZ data and 2010 Census data, ACS data for:
	  • Average number of workers
	  • Proportions of population in pre-school, K-8, or 9-12
	  • Proportions of retired households
		• Review trends and look for outliers in term of large percentage changes	Level 2 (Debugging Test)

Table 5-4. CEMSELTS School/College Location Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY

Aggregate	• Modeled versus observed (from household survey) home-to-school impedance histograms with impedances based on:
	  • Congested auto travel time
	  • Straight-line travel distance
	• Modeled versus observed (from household survey) average impedances (same as above) stratified by:
	  • Household income level
	  • Grade level of student
		• Modeled to observed averages should be ±5 percent
		• Modeled to observed impedance histogram coincidence ratios[1] (reasonableness test only)
			Level 2 (Basic Test)

	• Modeled versus observed school-at-home by:
	  • Geographic location (county)
	  • Household income level
	  • Grade level of student
		• Comparisons may be made to observed (expanded) household survey data (if survey contains sufficient observations to include this as a regular school location choice)	Level 2 (Basic Test)

	• Modeled versus observed (from household survey) home-to-school flows for:
	  • County-to-county for elementary and high school grade levels
	  • If applicable, school district of residence-to-school district of school for elementary and high school grade levels
	  • Major colleges and universities
		• Look for anomalies	Level 2 (Debugging Test)

Note:

1. Coincidence ratio is a measure of fit between two distributions, in this case the observed and modeled distance frequency distributions. Mathematically, it is the area under both curves divided by the area under either curve.

Table 5-5. CEMSELTS Workplace Location Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY
Aggregate	• Modeled versus observed (from household survey) home-to-regular workplace impedance histograms with impedances based on: • Congested auto travel time • Straight-line travel distance • Modeled versus observed (from household survey) average impedances (same as above) stratified by: • Household income level • Number of workers in the household • Age of worker (under 18 and 18 or older) • Gender of worker	• Modeled to observed averages should be ±5 percent • Modeled to observed impedance histogram coincidence ratios (reasonableness test only)	Level 1 (Basic Test)
	• Modeled versus ACS home-to-regular workplace based on impedance histograms and average impedances for: • Congested auto travel time • Straight-line travel distance • Modeled versus ACS-based average impedances (same as above) stratified by: • Age of worker • Industry of worker	• Modeled to observed averages should be ±5 percent • Modeled to ACS-based impedance histogram coincidence ratios (reasonableness test only) • If ACS data differ substantially from household survey, will not be able to match both survey and ACS results	Level 2 (Debugging Test)
	• Modeled versus observed work-at-home by: • Geographic location (county) • Household income level • Number of workers in the household • Age of worker (under 18 and 18 or older) • Gender of worker • Industry of worker	• Comparisons may be made to observed (expanded) household survey data and/or ACS data	Level 2 (Debugging Test)
	• Modeled workers choosing location as regular workplace versus employment by: • Geographic location (TAZ, county) • Worker industry	• Compare modeled to expanded observed proportions at county level • Review for reasonable patterns • Workers choosing location as regular work location per employee should be around 0.8	(Debugging Test)
	• Modeled versus observed home-to-workplace flows for: • County-to- county	• Look for anomalies • Comparisons may be made to expanded household survey data, ACS data, and 2010 LEHD data	Level 2 (Basic Test)

Table 5-6. CEMSELTS Work Activity Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOME	PRIORITY
Regional	• Average work duration • Percentage of workers by work flexibility category	• Review trends of non-control variables for reasonableness	Level 1 (Basic Test)
County	• Same as for regional	• Review trends for reasonableness	Level 1 (Debugging Test)

TAZ	• Color coded GIS plots of percent differences between TAZ data and 2010 Census data, ACS data for:
	  • Average number of workers
	  • Proportions of population in pre-school, K-8, or 9-12
	  • Proportions of retired households
		• Review trends and look for outliers in term of large percentage changes	Level 2 (Debugging Test)

Table 5-7. CEMSELTS Household Level Characteristics Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOME	PRIORITY

Regional	• Two dimensional cross-tabulations:
	  • Households by household income and household size
	  • Households by household income and number of employed household members (full- or part-time)
	• Median household income
	• Median residential tenure
	• Percentage of housing units by type
	• Cross-tabulations of households by vehicle availability (0, 1, 2, 3+ autos) by:
	  • Household income and household size
	  • Household income and number of workers
		• Confirm that control variables from 2010 census data have been maintained (±3 percent)
		• Review trends of non-control variables for reasonableness
		• Cross-tabulations should be compared to:
		  • Expanded household survey data
		  • ACS data
			Level 1 (Basic Test)

County	• Same as for regional	• Review trends for reasonableness	Level 1 (Debugging Test)
TAZ	• Color coded GIS plots of percent differences between TAZ data and 2010 Census data, ACS data for: • Median household income	• Review trends and look for outliers in term of large percentage changes	Level 2 (Debugging Test)

CEMDAP Components

There are many components of CEMDAP that need to be validated. These components are (with component codes from the model design plan shown):

- Generation-allocation model system
    - GA1, GA6 – Decision to go to school

  - GA2, GA3, GA7, GA8 – School start/end times
  - GA4 – Decision to go to work
  - GA5 – Work start and end times
  - GA9, GA10 – Travel mode to/from school
  - GA11, GA12 – Allocation of school drop-off/pick-up
  - GA13, GA14 – Out-of-home duration
  - GA15, GA16 – Independent/joint activity participation
  - GA17 – Decision to undertake serve-passenger activities
- Activity scheduling model system – workers
  - WSCH1 – Commute mode
  - WSCH2, WSCH3, WSCH4 – Number of before-work, work-based, after work tours
  - WSCH5, WSCH6, WSCH7– Before-work, work-based, after work tour modes
  - WSCH8 – Number of stops in a tour
  - WSCH9 – Home or work stay duration before the tour
  - WSCH10 – Activity type at a stop
  - WSCH11 – Activity duration at stop
  - WSCH12 – Travel time to a stop
  - WSCH13 – Location of a stop
- Activity Scheduling Model System – Non-Workers
  - NWSCH1 – Number of independent tours
  - NWSCH2, NWSCH3 – Decision to undertake an independent tour before/after pick-up or joint discretionary tour
  - NWSCH4 – Tour mode
  - NWSCH5 – Number of stops in a tour
  - NWSCH6 – Number of stops following a pick-up/drop-off
  - NWSCH7 – Home stay duration before a tour
  - NWSCH8 – Activity type at stop
  - NWSCH9 – Activity duration at stop
  - NWSCH10 – Travel time to stop
  - NWSCH11 – Stop location
- Joint discretionary tour scheduling model system
  - JASCH1 – Decision of joint or separate travel
  - JASCH2 – Joint activity start time
  - JASCH3 – Joint activity travel time to stop
  - JASCH4 – Joint activity location
  - JASCH5 – Vehicle used for joint home-based tour
- Activity scheduling model system – children
  - CSCH1 – School to home commute time
  - CSCH2 – Home to school commute time
  - CSCH3 – Mode for independent discretionary tour
  - CSCH4 – Departure time from home for independent discretionary tour (time from 3 a.m.)
  - CSCH5 – Activity duration at independent discretionary stop
  - CSCH6 – Travel time to independent discretionary stop
  - CSCH7 – Location of independent discretionary stop

We have organized the types of validation tests to reflect the focus of each particular model component and to group components by the types of outputs they produce. The following typology is used to describe the validation tests:

- **Activity choices**
    - GA1, GA6 – Decision to go to school
    - GA4 – Decision to go to work
    - GA13, GA14 – Out-of-home duration
    - GA15, GA16 – Independent/joint activity participation
    - GA17 – Decision to undertake serve-passenger activities
    - GA11, GA12 – Allocation of school drop-off/pick-up
    - WSCH2, WSCH3, WSCH4 – Number of before-work, work-based, after-work tours
    - WSCH8 – Number of stops in a tour (worker)
    - WSCH10 – Activity type at a stop (worker)
    - NWSCH1 – Number of independent tours
    - NWSCH2, NWSCH3 – Decision to undertake an independent tour before/after pick-up or joint discretionary tour
    - NWSCH5 – Number of stops in a tour (non-worker)
    - NWSCH6 – Number of stops following a pick-up/drop-off
    - NWSCH8 – Activity type at stop (non-worker)
    - JASCH1 – Decision of joint or separate travel
- **Timing/scheduling choices**
    - GA2, GA3, GA7, GA8 – School start/end times
    - GA5 – Work start and end times
    - WSCH9 – Home or work stay duration before the tour (worker)
    - WSCH11 – Activity duration at stop (worker)
    - NWSCH7 – Home stay duration before a tour (non-worker)
    - NWSCH9 – Activity duration at stop (non-worker)
    - JASCH2 – Joint activity start time
    - CSCH4 – Departure time from home for independent discretionary tour (time from 3 a.m.)
    - CSCH5 – Activity duration at independent discretionary stop
- **Location choice/travel time models** (note that work and school locations have already been simulated in CEMSELTS)
    - WSCH13 – Location of a stop (worker)
    - NWSCH11 – Stop location (non-worker)
    - JASCH4 – Joint activity location
    - CSCH7 – Location of independent discretionary stop
    - WSCH12 – Travel time to a stop (worker)
    - NWSCH10 – Travel time to stop (non-worker)
    - CSCH1 – School to home commute time
    - CSCH2 – Home to school commute time
    - CSCH6 – Travel time to independent discretionary stop
    - JASCH3 – Joint activity travel time to stop
- **Mode choices**
    - GA9, GA10 – Travel mode to/from school
    - WSCH1 – Commute mode
    - WSCH5, WSCH6, WSCH7– Before-work, work-based, after work tour modes
    - NWSCH4 – Tour mode (non-worker)

- Trip mode choice
- JASCH5 – Vehicle used for joint home-based tour
- CSCH3 – Mode for independent discretionary tour

Table 5-8 presents the validation tests for the activity choices group. The tests for the timing/scheduling choices group are presented in Table 5-9. Table 5-10 presents the location choice tests while Table 5-11 shows the mode choice tests.

Table 5-8. CEMDAP Activity Choice Model Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY
Aggregate	• Numbers or percentages of residents making tours by purpose: • For the region • By county • By household size/income group • By household size/vehicle availability • By gender and age group • By employment status • By student status	• Compare modeled to observed numbers or percentages from the expanded household survey • Review for reasonable patterns	Level 2 (Basic Test)
	• Percent of "immobiles" (persons with no out-of-home activities during the day) by: • By household size/income group • By household size/vehicle availability • By gender and age group • By employment status • By student status	• Compare modeled to observed percentages from the expanded household survey	Level 2 (Debugging Test)
	• Numbers or percentages of school half-tours by escort decision by child age group: • For the region • By county • By household size/income group • By vehicle availability/number of workers • By household size/vehicle availability • By gender	• Compare modeled to expanded observed numbers or percentages from the household survey • Review for reasonable patterns	Level 2 (Basic Test)
	• Numbers or percentages of escorts by person type by: • By household size/income group • By household size/vehicle availability • By gender and age group • By employment status • By student status	• Compare modeled to expanded observed numbers or percentages from the household survey • Review for reasonable patterns	Level 2 (Debugging Test)
	• Numbers or percentages of households making joint travel tours by activity type and party size: • For the region • By county • By household size/income group • By household size/vehicle availability • By household size/number of workers	• Compare modeled to expanded observed numbers or percentages from the household survey • Review for reasonable patterns	Level 2 (Basic Test)

Table 5-8. CEMDAP Activity Choice Model Validation Tests (continued)

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY
Aggregate	• Average number of joint travel tours per household: • For the region • By county • By household size/income group • By household size/vehicle availability • By household size/number of workers	• Compare modeled to expanded observed numbers or percentages from the household survey • Review for reasonable patterns	Level 2 (Debugging Test)
	• Mean number of intermediate stops per half-tour by primary tour purpose: • For the region • By time of day (time of day based on arrival or departure time from primary tour destination for each appropriate half-tour)	• Compare means of modeled numbers of stops to unexpanded means from surveyed numbers • Review for reasonable patterns	Level 2 (Basic Test)
	• Mean number of intermediate stops regardless of half-tour by primary tour purpose for each: • Gender/age group combination • Employment status • Student status • Income level • Tour mode choice (non-motorized, auto driver, auto passenger, transit) • Area type of tour origin by area type of tour primary destination	• Compare modeled to expanded observed average numbers of intermediate stops • Review for reasonable patterns	Level 2 (Debugging Test)

Table 5-9. CEMDAP Timing/Scheduling Choice Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY
Aggregate (tour level)	• Histograms of tour arrival and departure times to/from primary tour destination by tour purpose: • Repeat above for departure times	• Compare modeled to expanded observed temporal distributions • Review for reasonable patterns	Level 2 (Basic Test)
	• Activity durations by traveler characteristics: • Gender and age group • Employment status • Student status • Income level • Tour mode choice (non-motorized, auto driver, auto passenger, transit)	• Compare modeled to expanded observed temporal distributions • Review for reasonable patterns	Level 2 (Debugging Test)
Aggregate (trip level)	• Histograms of stop arrival times by tour purpose • Repeat above for departure times	• Compare modeled to expanded observed temporal distributions • Review for reasonable patterns	Level 2 (Basic Test)

Table 5-10. CEMDAP Location Choice Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY

Aggregate (tour level)	• Modeled versus observed (from household survey) home-to-other tour impedance histograms with impedances based on:
	  • Congested auto travel time
	  • Straight-line travel distance
	• Modeled versus observed (from household survey) average impedances (same as above) stratified by:
	  • Household income level
	  • Number of workers in the household
	  • County
	  • Age group of traveler
		• Modeled to observed averages should be ±5 percent
		• Modeled to observed impedance histogram coincidence ratios (reasonableness test only)
			Level 1 (Basic Test)
			
			(Debugging Test)

Aggregate (trip level)	• Modeled versus observed trip impedance histograms and average impedances for:
	  • Congested auto travel time
	  • Straight-line travel distance
	• Modeled versus observed average impedances (same as above) stratified by:
	  • Household income level
		• Modeled to observed averages should be ±5 percent
		• Modeled to observed impedance histogram coincidence ratios (reasonableness test only)
			Level 2 (Basic Test)

Table 5-11. CEMDAP Mode Choice Validation Tests

AGGREGATION LEVEL	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY

Aggregate (tour level)	• Numbers of purpose-specific tours by mode and mode shares:
	  • For the region
	  • By county (at both tour origin and tour destination)
	  • By area type (at both tour origin and tour destination)
	• Purpose-specific tour mode shares by transit in-vehicle time, walk time, or bicycle time
	• Purpose-specific tour transit mode shares by walk time to transit
	• County-to-county purpose-specific tour mode shares
		• Compare modeled to expanded observed numbers of tours and tour mode shares
		• Review for reasonable patterns
		• Compare to ACS home-to-regular workplace by means of transportation (for work tours)
			Level 2 (Basic Test)

	• Purpose-specific tour mode shares:
	  • Household size by income group
	  • Autos available per worker less than 1.0, equal to 1.0, and greater than 1.0
	  • Autos available per household member of driving age less than 1.0, equal to 1.0, and greater than 1.0
	  • Gender by age group
		• Compare modeled to expanded observed numbers of tours and tour mode shares
		• Review for reasonable patterns
			Level 2 (Debugging Test)

Aggregate (trip level)	• Numbers of purpose-specific trips and shares by mode, access type, transfer frequency and socioeconomic class:
	  • For the region
	  • By county (at both trip origin and trip destination)
	• Trip mode shares by tour mode share to validate level of mode switching (i.e. trip mode is not the same as tour mode)
	• Purpose-specific trip mode shares by transit in-vehicle time, walk time, or bicycle time
	• Purpose-specific trip transit mode shares by walk time to transit
		• Compare modeled to expanded observed numbers of trips and trip mode shares
		• Review for reasonable patterns
			Level 2 (Basic Test)

	• Tour purpose-specific trip mode shares:
	  • Household size by income group
	  • Autos available per worker less than 1.0, equal to 1.0, and greater than 1.0
	  • Autos available per household member of driving age less than 1.0, equal to 1.0, and greater than 1.0
	  • Gender by age group
		• Compare modeled to expanded observed numbers of trips and trip mode shares
		• Review for reasonable patterns
			Level 2 (Debugging Test)

5.3 System Level Validation

System level validation tests are related in many ways to highway and transit assignment procedures. They represent some of the most traditional validation measures.

For highways, validation tests focus on vehicular volumes and vehicle miles traveled across screenlines, along corridors, and any other geographic representation of the network. For transit, validation tests are focused more on boardings and alightings for stations and routes.

Initially, the system level validation tests of highway and transit assignment will be performed without model feedback loops. This is to allow us to find model issues that can be discovered and addressed without the run time needed to run multiple feedback loops. Once the validation results are satisfactory for assignments for a single model iteration, the model will be run with feedback loops to confirm that the results are still valid at model convergence.

The remainder of this section describes the planned system level validation tests for highways and transit in greater detail.

Highway Assignment

The system level highway validation is an overall validation of the travel modeling process with a very specific focus on the reproduction of observed traffic volumes in a region. Experience with other models has shown that even a poorly specified model can be made to reproduce observed traffic volumes for a base year. The validations of the individual modeling procedures described above are intended to help ensure that the regional travel model is, in fact, reasonable. However, if the individual model components are all deemed to be reasonable but the modeled traffic volumes do not reasonably reproduce observed traffic volumes for a base year, the model is not valid.

The system level highway validation will focus on several classes of measures:

- Vehicle-miles of travel (VMT);
- Individual link traffic volumes;
- Intra-regional traffic flows as defined by screenlines; and
- Congested roadway speeds.

Table 5-12 summarizes the system level highway validation measures for the auto modes, which will focus much more specifically on numerical guidelines than the validation of the individual model components. A primary reason for this is the availability of independently collected data such as traffic counts for the system level validation.

Table 5-13 presents the targets for VMT by functional class and area type, which is the state of practice guidance. Table 5-14 shows the targets for RMSE and RMSE% by functional class. Figure 5-2 presents the assignment results by RMSE% by volume group, and Figure 5-3 shows the percent difference by volume group for screenlines.

Table 5-12. System Level Highway Validation Tests

VALIDATION FOCUS	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY
Vehicle-miles of travel	• Comparison of modeled VMT to VMT estimated from traffic counts by: • Region • County • Functional class (with particular attention to toll and HOV facilities)	• Modeled regional VMT should be within the following percentages of estimated VMT: • ±1 percent for the region • Percentages shown in Table 5-23 by functional class, area type, and county on links with counts	Level 1 (Basic Test)
Individual link volumes	• Plot of absolute and percent difference between modeled and observed volumes for links with counts • Scatterplot of modeled versus observed daily traffic volumes by link • Root mean squared error (RMSE) and percent RMSE by: • Region • Functional class (with particular attention to toll and HOV facilities) • Volume group • Anomalous links • Links with 0 volumes • Links with very high v/c ratios	• Visual inspection for large errors in modeled link volumes or for general trends in errors • RMSE and %RMSE by functional class should be within targets shown in Table 5-14; %RMSE by volume group should be within the targets shown in Figure 5-2	Level 1 (Basic Test)
Screenlines	• Percent deviation by screenline	• Percent deviation should be within the targets shown in Figure 5-3	Level 1 (Basic test)
Congested Speeds	• Comparison of modeled and observed average congested speeds by functional class and county for each time of day period used in highway assignment • Scatterplots of modeled to observed congested speeds by time of day for links with speed data	• Average assigned congested speeds should be within ±5 miles per hour of average observed congested speeds • 80 percent of the link specific assigned congested speeds should be within ±5 miles per hour of average observed congested speeds	Level 1 (Basic Test)

Figure 5-2. Traffic Assignment %RMSE Targets by Volume Group

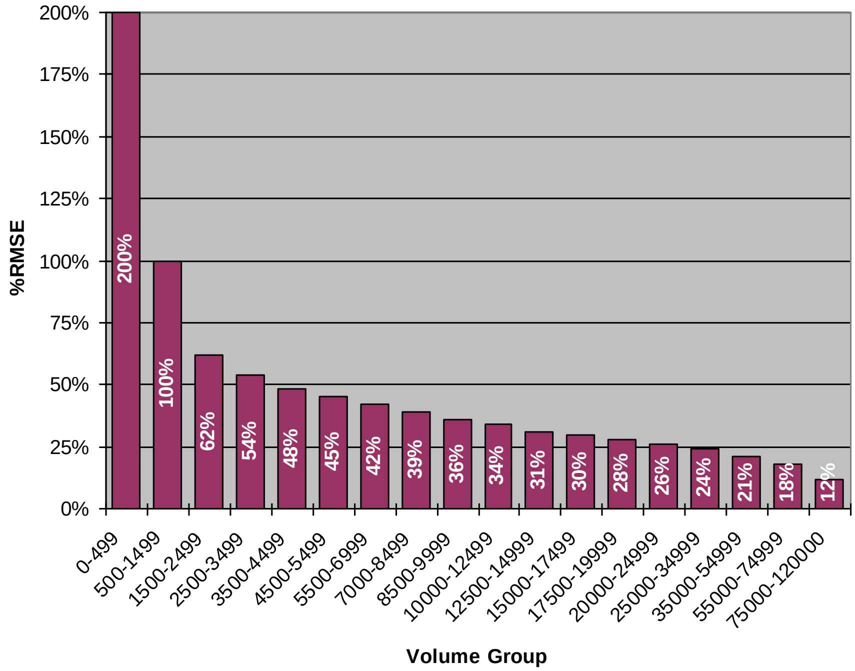
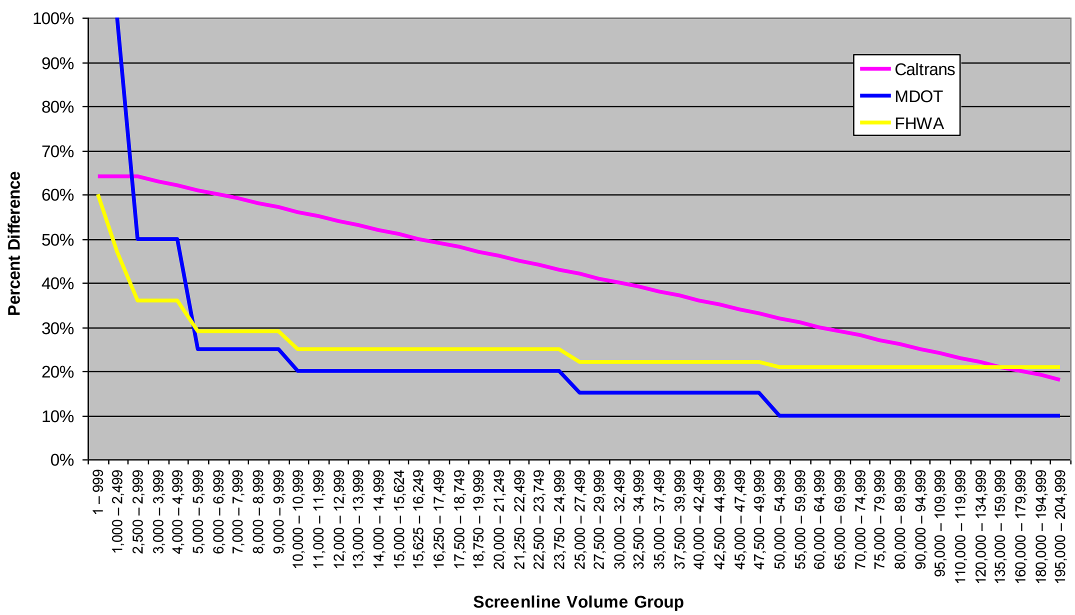


Figure 5-3. Screenline Crossing Percent Difference by Volume Group



Source: Adapted from *Las Vegas Travel Demand Model Guidelines for Estimation, Calibration, & Validation*, prepared for Regional Transportation Commission of Southern Nevada, prepared by Fehr & Peers Transportation Consultants, March 25, 2005, page 28. Note that individual state guidelines may have changed since this document was published.

Table 5-13. Traffic Assignment VMT Targets by Functional Class

Stratification	VMT Target
Functional Class	
Freeways	±7%
Expressways	±7%
Principal Arterials	±10%
Minor Arterials	±10%
Collectors	±15%
All Links	±1%

Source: Adapted from Giaimo, Gregory, Travel Demand Forecasting Manual 1 – Traffic Assignment Procedures, Ohio Department of Transportation, Division of Planning, Office of Technical Services, August 2001.

Table 5-14. RMSE and %RMSE Targets by Functional Class

Functional Class	Target Values	
	RMSE[2]	%RMSE[3]
Freeways	<12,500	<20%
Expressways	<7,500	<30%
Principal Arterials	<3,750	<30%
Minor Arterials	<3,000	<40%
Collectors	<2,250	<70%
All Links	n/a	<40%

1 $\text{RMSE} = \sqrt{\frac{\sum(\text{Count}-\text{Assigned})^2}{(\text{Number of Observations}-1)}}$

$\%\text{RMSE} = \frac{100\times\text{RMSE}}{\text{Avg. Count}}$

2 Based on one-half lane of capacity and assumption of 8 percent peak hour factors for interstates, freeways, system ramps, expressways, and external connectors; based on 10 percent peak hour factor for other functional classes

3 Rules of thumb from model validation efforts for several regions.

Transit Assignment

The system level transit validation is an overall validation of the travel modeling process with a very specific focus on the reproduction of observed transit boardings and transit volumes in the region.

The system level transit validation will focus on several types of measures, including:

- Transit boardings by route, station, and access mode;
- Boardings per linked trip; and
- Park-and-ride lot utilization.

Table 5-15 summarizes the aggregate validation measures that are proposed for the system level transit validation.

Table 5-15. System Level Transit Validation

VALIDATION FOCUS	VALIDATION MEASURES	EXPECTED OUTCOMES	PRIORITY
Boardings by mode by time of day	• Comparison of modeled to observed boardings by: • Region by time of day • Mode by time of day • Line and station by time of day and access mode • Local vs. express by line (where applicable)	• Modeled boardings for the region should be within ±10 percent of the estimated boardings • Modeled boardings by mode by time of day should be within ±20 percent of the estimated boardings • Modeled boardings by line by time of day should be within ±30 percent of the estimated boardings	Level 2 (Basic test)
Park-and-Ride lot utilization	• Modeled daily drive access vehicle trips to park-and ride lots to observed parking at lots	• Modeled drive access trips to lots should be in the range of 1.0 to 2.0 to account for turnover of parking spaces	Level 2 (Basic Test)

5.4 Temporal Validation and Sensitivity Tests

Temporal Validation

Temporal validation involves using the estimated/calibrated model to make forecasts for a year other than the base year, and requires data for (at least) one other year. Validation tests for the forecast year could include any tests discussed above, but are usually limited to system level tests.

As discussed in Section 3.4, NYMTC has decided not to perform a backcast as part of the model validation. A forecast to the year 2016 (2015 if data for 2016 are not available in time). A review of data will be undertaken before making the final decision on the forecast year.

At a minimum, system level validation to the forecast year will be performed, which will be considered priority level 1 for temporal validation. Priority level 2 temporal validation tests to the backcast year will include aggregate validation tests of the model components described in the previous subsections. Since the 2016 model inputs will likely be very similar to the 2012 base year inputs, it is expected that the checks will yield similar results to the base year validation.

After the base year validation and 2016 forecast analysis are complete, the model will be run for a forecast year of 2045. While the forecast year results cannot be compared to observed data, the complete set of results, including the activity based component results and the highway and transit assignment results can be checked for reasonableness and consistency with the validated base year results.

Sensitivity Testing

One goal of activity-based models is an increased sensitivity to model inputs that are known or believed to affect travel behavior. Because of the nature of activity based models, a richer set of such input variables can be considered, and the sensitivity of the model results to these inputs can be checked. As discussed in Section 3.4, sensitivity testing involves revising key factors and observing the effects on forecasted travel. These revisions can be made to model parameter values (e.g., the mode choice cost coefficient) and to model inputs (e.g., land use variables, socioeconomic conditions, fuel costs, etc.).

Like temporal validation, sensitivity tests can be performed for any of the validation measures described in the previous chapters. Typically, however, data is not available to compare the results of sensitivity tests to. Instead, sensitivity tests should be reviewed for reasonableness, with expected outcomes of the tests noted beforehand. Any unexpected outcomes observed from the tests should be explainable. Examples of model sensitivity tests that have been successfully used elsewhere include the following:

- Land use changes (for example, development of a brownfield site)
- Changes in population characteristics (for example, aging of the population)
- Changes in highway capacities (such as reducing the number of lanes on a key roadway, or removing a traffic bottleneck)
- Changes in transit fares or service frequencies
- Changes in parking costs
- Changes in roadway tolls (perhaps varying by time of day)

The specific sensitivity tests should reflect the uses of the model for planning purposes. Therefore, the definitions for these tests will be developed in consultation with NYMTC staff. Depending on the resources allocated to other parts of the validation process (for example, whether a backcast is performed) and the complexity of the type of sensitivity tests, it is expected that between three and six sensitivity tests can be performed.

5.5 System Integrity Tests

The system integrity tests involve running the software to ensure that it works as documented. Actual tests cannot be designed until the software design and coding are complete. However, the following general tasks will be performed:

- Verify that the software can be successfully installed on NYMTC's servers and runs with consistent outputs.

- Verify that program modules can be invoked as documented.

- Perform selected calculation verifications.

- Review the output database for unexpected missing or inaccurate information.

6.0 References

Cambridge Systematics, Inc. (2017). *New York Best Practice Model 2012 Base Update: Model Implementation Plan*. Prepared for NYMTC.

Cambridge Systematics, Inc., University of Texas, Austin, and Arizona State University (2017). *New York Best Practice Model 2012 Base Update: Model Design Plan*. Prepared for NYMTC.

McCollom, B.E., R.H. Pratt, Richard H. Pratt, Consultant, Inc., Texas Transportation Institute, Jay Evans Consulting LLC, Parsons Brinckerhoff Quade & Douglas, Inc., Cambridge Systematics, Inc., J. Richard Kuzmyak, L.L.C., BMI-SG, Gallop Corporation, McCollom Management Consulting, Inc., Herbert S. Levinson, Transportation Consultant, and K.T. Analytics, Inc. *Traveler Response to Transportation System Changes: Chapter 12—Transit Pricing and Fares*. TCRP Report 95, Transportation Research Board, Washington, D.C., 2004.

Smith, M.E. (1979). Design of Small Sample Home Interview Travel Surveys. Transportation Research Record 701, Transportation Research Board.