

**INFORMATION COLLECTION
SUPPORTING JUSTIFICATION**

**Survey of Northeast Regional and Intercity Household Travel Attitudes
and Behavior**

This ICR is to request the Office of Management and Budget's (OMB) one year approved clearance for the information collection entitled, "Survey of Northeast Regional and Intercity Household Travel Attitudes and Behavior."

Part A. Justification

The Federal Railroad Administration (FRA) proposes to collect information from the public to determine current intercity and regional travel behavior of northeast residents. The information collected will include frequency of trips, origin and destination, modes of travel (and class of service if applicable), trip purpose, party size, trip costs, and other trip characteristics. It will also ask for travel preferences under alternative choice scenarios that include different and new modes, classes of service, costs, and amenities.

1. Explain the circumstances that make the collection of information necessary. Identify any Legal or administrative requirements that necessitate the collection. Attach a copy of the appropriate section of each statute and regulation mandating or authorizing the collection of information.

a. Circumstances making the collection necessary

The Northeast region has one of the most extensive multi-modal passenger and freight transportation systems in the world—highways, airports, ports, intercity and commuter rail, and public transit serving all major cities and many intermediate markets. However, despite significant investment over decades in all modes, the region still faces major congestion and capacity constraints. These constraints, if not addressed, have the potential to curtail future mobility and economic growth and place the Northeast at a competitive disadvantage to other regions of the U.S. and the world.

With these issues in mind, FRA awarded a collective grant to the northeast states through which the NEC passes for planning investments in the NEC. FRA is the designated lead agency for the resulting planning program called NEC Future. The goal of the NEC FUTURE program is to prepare a Passenger Rail Corridor Investment Plan (PRCIP) for the northeast region. The PRCIP, consisting of a Service Development Plan (SDP) that articulates the overall scope and approach for proposed service and Tier 1 Environmental Impact Statement (EIS), when complete will define an integrated, comprehensive passenger rail transportation solution for the Northeast. The purpose of this solution is to improve mobility, effectively serve travel demand due to population and jobs growth, support economic development, reduce growth in carbon emissions and dependence on foreign oil, and contribute to improved land utilization and investment in both urban and non-urban communities in the region.

The results of the survey will be used to develop a new model for forecasting future travel behavior in response to future services provided by different modes of travel in the Northeast.¹ The primary use of the model will be to analyze the ridership impacts of alternative rail investments plans for northeast corridor (NEC) as part of the aforementioned PRCIP for the northeast region. In addition, the model and the data underlying the model will be available to FRA for use in future projects involving the NEC.

The NEC Future planning program has three phases, as shown in Figure A.1.

The recently completed Phase I involved the early service planning and alternatives development and evaluation. Technical work completed in Phase I includes: data collection, development of the project's Purpose and Need, preparation of a Public/Stakeholder Involvement Plan, analysis of existing ridership and currently available ridership forecasts, performance of operations analysis, identification and evaluation of service alternatives, identification of infrastructure requirements, and the initiation of Early Engagement with Resource Agencies and National Environmental Policy Act (NEPA) Scoping process. The market analysis and forecasting supporting the identification and coarse screening of alternatives in Phase 1 was based on existing available market data, forecast results from prior studies, and the use of existing models to forecast ridership and revenue results from alternatives.

Phase 2, which has just started, will involve further refinement of the alternatives and the preparation of the Draft Tier 1 EIS and the Draft SDP. Detailed alternatives analyses will be conducted in Phase 2, supported by ridership and revenue forecasting based on the new survey data and new model developed as part of this project. Phase 3 results in the preparation of the Final Tier 1 EIS and Record of Decision, as well as the final SDP.

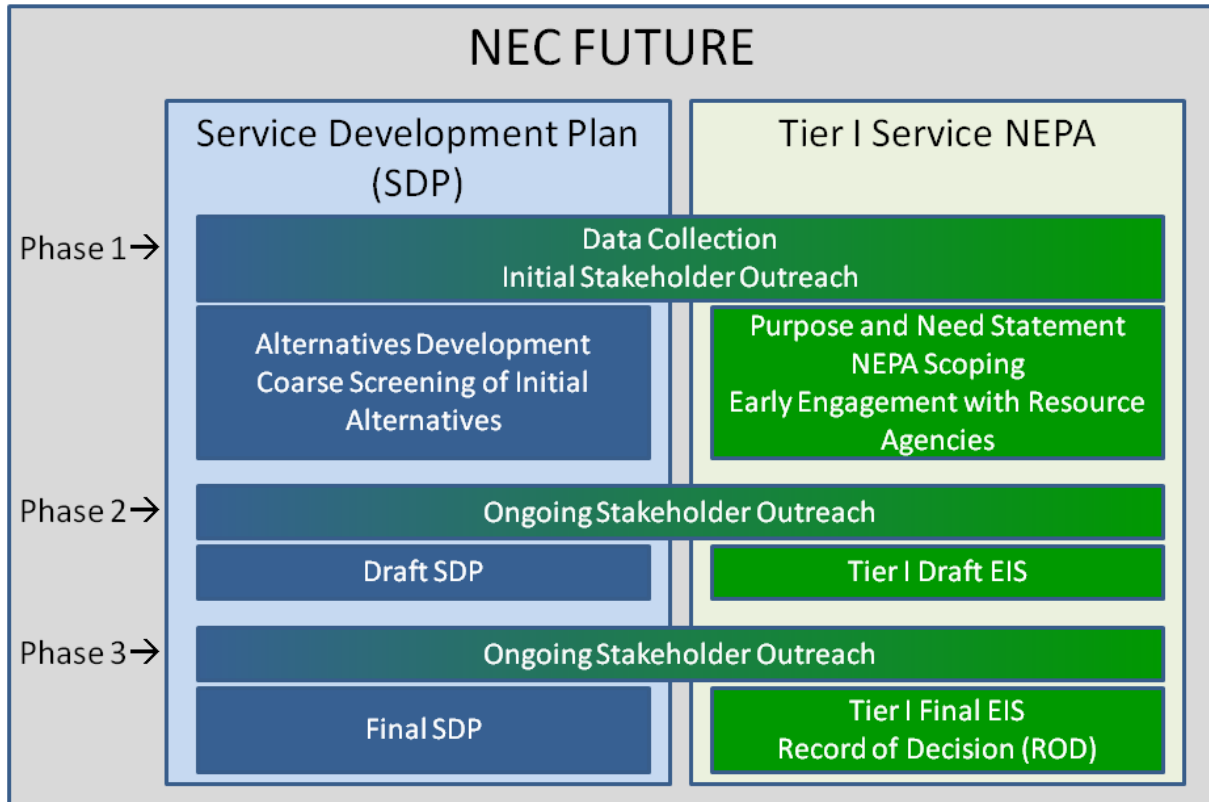
The project team has divided Phase 2 into two subparts. During Phase 2A, the analysis of alternatives will be supported by existing data and models. Phase 2B will begin once the data collection is complete and the new model has been developed and is available to for the detailed analysis of alternatives. Pending OMB approval, the survey work will be planned and programmed starting in mid-June 2013 and the surveys will subsequently be put into the field in August, beginning with an initial sample of 300 respondents constituting a "pilot phase". The results of the "pilot phase" will be summarized for review with OMB in September. Based on the observed "pilot phase" results, the survey will be revised as necessary, and the field work resumed – most likely in October unless major revisions and/or restructuring of the survey is necessary. It is expected that the survey will need to be in the field for about 3 months to generate the specified number of completed surveys. This schedule will allow for the data to be available (after processing) by January 2014. The delivery of data no later than January 2014 is essential to complete the model estimation so that Phase 2B can be evaluated in the Spring of 2014. Further slippage in schedule would lead to an equivalent delay of the project which would delay the EIS process, selection of an alternative, and the ultimately impact when the

¹ The Northeast includes all MSAs along the existing rail transportation corridor linking Washington, Baltimore, Philadelphia, Newark, New York, New Haven, Providence, and Boston – known as the NEC spine – as well as MSAs that can be reached by train directly or via a single transfer to connecting corridors from the NEC spine, including Richmond, Harrisburg, Albany, and Springfield (see Supporting Statement B Table B.1 for a detailed listing). This definition of the relevant market for NEC service was developed by the project team based on the NEC FUTURE alternatives under consideration.

investments can start. Or, otherwise, the NEC FUTURE study would proceed on its current schedule without the full benefit of the new model.

Further development and implementation of a number of major Northeast Corridor projects must await completion of the NEC FUTURE Tier 1 NEPA Record of Decision. This includes advancing portions of the Gateway project, perhaps the most important program on the Northeast Corridor at this time, where resolution of key issues impacting the number and location of platforms at Penn Station must await completion of NEC FUTURE. The size and alignment for various other critical improvement projects currently pending on the NEC – including the Susquehanna Bridge replacement in Maryland – similarly are dependent on the findings coming from the NEC FUTURE project. Final design and construction of these projects must hold until completion of NEC FUTURE.

FIGURE A.1 PROJECT STRUCTURE



As described above, the PRCIP requires the development and evaluation of improved transportation alternatives for the Northeast. In support of preparing the PRCIP, this data collection is needed to build a model for estimating market demand for transportation in the northeast and to evaluate how travelers would respond to alternative transportation service options. While there are certain publically available data that can assist in analyzing the northeast travel market during the early phases of the project when alternatives are evaluated using coarse screening tools, more detailed data will be needed to support development of the detailed PRCIP. Currently available data include airline ticket data, Amtrak ticket data, and commuter rail ticket data. But these data sources do not contain information on the characteristics of the traveler (such as age, income, or vehicle ownership), trip purpose, detail on party size, or actual origins and destinations. Importantly, there is no current information on the number of intercity trips taken by automobile in the northeast. Further, there is not currently a sufficient information source for traveler preferences regarding new transportation services that might be developed as part of the PRCIP.

b. Statute authorizing the collection of information

FRA has the authority to conduct this information collection through the National Environmental Policy Act (NEPA) (42 U.S.C. § 4321 et seq.) and 49 U.S.C. § 103 (FRA’s authorizing statute). FRA’s authorizing statute gives the FRA Administrator the authority to “support rail intermodal development and high-speed rail development, including high speed rail planning” and “ensure that the programs and initiatives . . . benefit the public and work toward achieving regional and national transportation goals.” 49 U.S.C. 103(j)(5)-(6). NEPA directs Federal agencies to study and consider the environmental impacts of any “major Federal actions significantly affecting the quality of the human environment.” 42 U.S.C. § 4332(2)(C). Federal agencies shall “utilize a systematic, interdisciplinary approach which will insure the integrated use of the natural and social sciences and the environmental design arts in planning and in decision-making which may have an impact on man’s environment” in any NEPA study. 42 U.S.C. § 4332(2)(A); 40 C.F.R. 1501.2(a) and 1502.6.

This information collection is instrumental to FRA’s rail service planning efforts and Tier 1 environmental impact statement (EIS) for the Northeast Corridor (NEC), known as the NEC FUTURE program. The survey will provide the FRA with information on where and how people travel along and within the NEC as well as their preference for how they would like to travel in the future. FRA will then use the gathered information to develop a new model to forecast travel demand within the NEC by mode (rail, air, intercity bus and car). The travel demand modeling tool will provide FRA with valuable information on the characteristics of NEC travelers (e.g., age, income or vehicle ownership), trip purpose (e.g., business, leisure or commute), detail on party size, trip origins and destinations, number of intercity trips taken by car and traveler preference. FRA will use this information to make better planning and investment decisions on a regional basis by providing the tools to evaluate changes in travel demand generated by the way in which train service is operated on the Northeast Corridor and by the types of investments made to enhance the NEC infrastructure. In addition, the information collection will provide the data necessary for FRA to ensure that its planning initiative and Tier 1 EIS benefit the public and achieves regional and national transportation goals, as set forth in FRA’s authorizing statute. See 49 U.S.C. § 103(j)(5)-(6). Moreover, the information collection is exactly the sort of

“interdisciplinary approach” required by NEPA. See 42 U.S.C. § 4332(2)(A); 40 C.F.R. 1501.2(a) and 1502.6. The survey data will provide insights into travel patterns and traveler preferences that will inform the definition of alternatives, ensuring a sufficiently broad spectrum of alternatives are considered in the Tier 1 EIS. This data will further be valuable input to the analysis of several environmental factors including noise and vibration, air quality, and transportation.

2. Indicate how, by whom, and for what purpose the information is to be used. Except for a new collection, indicate the actual use the agency has made of the information received from the current collection.

The information collection will support the development of a new model to forecast travel demand by mode within the Northeast Corridor (NEC) as part of the NEC FUTURE program for developing a Passenger Rail Corridor Investment Plan (PRCIP) for the northeast region. The FRA is named as the lead agency for the effort. The PRCIP, consisting of a Service Development Plan (SDP) that articulates the overall scope and approach for proposed service and Tier 1 Environmental Impact Statement (EIS), when complete will define an integrated, comprehensive passenger rail transportation solution for the Northeast.

NEC FUTURE requires a travel demand forecasting model capable of forecasting future travel demand by mode, including different intercity and regional/commuter modes of travel, within the NEC. None of the existing available travel demand models maintained by Amtrak, regional agencies, or others within the NEC address the full extent of the NEC geography and the existing and future available intercity and regional/commuter modes of travel. For this reason, the NEC FUTURE study scope includes developing a new travel demand forecasting model for the NEC.

Specifics of the Model

The new NEC model will be a trip-based model similar in structure to other existing models, including Amtrak’s NEC travel demand forecasting model. The new NEC model consists of two major components, addressing:

- Total Travel Market Size, based on existing travel market size and estimated growth
- Mode Shares, based on the characteristics of the competing modes

These two components are actually applied in reverse order (i.e., mode share before total travel) so that the mode share model results can be incorporated within the total demand model structure. This linkage provides the total travel model with sensitivity to changes in the level of service provided by all modes, where each mode’s contribution to the overall “impedance” or “ease of travel” within a geographic market is effectively weighted by mode share.

Total Travel demand forecasts define the total market size to which the modal shares are applied to forecast demand by mode. In general, there are two distinct types of factors that influence total travel demand between geographic areas:

- Population growth and changes in economic activity in the geographic areas
- Changes in the modal levels of service provided between the geographic areas

Measures used to represent the impacts of these changes respectively include:

- Socio-economic data and forecasts, provided by Moody's:
 - Population
 - Household Income
 - Employment
- Composite modal level of service, defined by the mode share model structure and equivalent to summing across all top level choices, as follows:

$$LOS = \sum_i \lambda \exp(U_i) \lambda$$

where:

i - mode

U_i - utility of choice i

The total travel volumes are estimated using a ratio formulation that relates total travel market growth to growth in the independent variables, computed as the ratio of the forecast year to the base year values. This is illustrated by the following equation:

$$\frac{TRP_F(i,j)}{TRP_B(i,j)} = \left(\frac{POP_F(i)}{POP_B(i)} \right)^a \times \left(\frac{POP_F(j)}{POP_B(j)} \right)^b \times \left(\frac{INC_F(i)}{INC_B(i)} \right)^c \times \left(\frac{EMP_F(j)}{EMP_B(j)} \right)^d \times \left(\frac{LOS_F(i,j)}{LOS_B(i,j)} \right)^e$$

where:

TRP - trips

POP - population

INC - household income

EMP - employment

LOS - level of service

F - future year

B - base year

i - home zone

j - attraction zone

That is, inter-zonal trips (TRP) are projected to grow in proportion to population in the home zone (POP_i), adjusted for its estimated effect, a; in proportion to population changes in the attraction (non-home) zone (POP_j), adjusted for its estimated effect, b; in proportion to changes in household income (INC_i) in the home zone, adjusted for its estimated effect, c; in proportion to the employment changes in the attraction zone (EMP_j); adjusted for its estimated effect, d; and in proportion to changes in the overall level of service, adjusted for its estimated effect, e.

The total travel demand models will be estimated from base year travel data, using non-linear regression techniques, with total trips as the dependent variable and population, income, employment, and level of service as the dependent variables. Given that this is cross-sectional data from a single point in time, the actual model form to be estimated is given by:

$$TRP(i, j) = Constant \times POP(i)^a \times POP(j)^b \times INC(i)^c \times EMP(j)^d \times LOS(i, j)^e$$

where:

TRP - trips
POP - population
INC - household income
EMP - employment
LOS - level of service

i - home zone
j - attraction zone

The estimated constant drops out when the model is applied in the ratio formulation, shown earlier. Separate models will be estimated for each of the trip purpose market segments reflected in the mode share models, described below.

An essential input to the total travel model estimation and application is provided by base year travel demand – existing trips by geographic market. NEC FUTURE will develop estimates of existing trips by market and by mode from a variety of sources, including:

- Passenger car/truck/van trips, from a parallel study by the NEC Commission, which is collecting auto data from toll records and surveys to build a NEC auto mode trip table
- Plane trips, from the FAA 10% ticket sample
- Rail trips, from:
 - Amtrak, for intercity rail
 - Commuter Rail agencies, for regional/commuter rail
- Bus trips, from current service offerings and estimated load factor (although the NEC study continues to seek other sources of intercity bus data, if they exist) and from regional agencies that also provide longer haul bus service

In addition to the above modal data sources, the new NEC FUTURE survey will collect data from a mode-neutral household-based NEC sample, which will be utilized to provide additional market detail, such as a basis for allocating trips by trip purpose, which is not provided by many of the above modal sources.

The **Mode Share** component estimates the share of total person travel by mode. This model component addresses travel by the following modes:

- Passenger car/truck/van
- Plane
- Bus
- Train, addressing the following types of train service separately:
 - Premium high-speed rail (similar to Amtrak's Acela train service)
 - Regional intercity rail (similar to Amtrak's Regional train service)
 - Commuter rail (similar to the train service provided by MBTA, MNR, LIRR, NJ Transit, SEPTA, MARC, and VRE within specific regions)
 - Connecting Regional/Commuter rail

The new model will estimate shares among these as a function of the following key independent variables describing the service characteristics:

- Travel time, expressed separately for line haul, terminal, and access/egress portions of air, rail, and bus trips
- Travel cost or fare, taking account of the cost implications of travel by group and individuals and also including parking charges

- Schedule of service provided by air, rail, and bus; taking account of the time of day and spacing of the departures/arrivals offered
- Separate from travel considerations, the impact of connecting versus one-seat ride service
- Reliability of the service, expressed as the “expected” delay time at the 80th percentile (for example, “20% Chance of a 30 minute longer total travel time”)
- Mode specific constants reflecting the differences between modes not directly measured by other independent variables in the model (factors and traveler perceptions such as the comfort and convenience provided by each mode would be reflected here)
- Income and/or occupation variables to account for differences in value of time and possibly included as modifiers of the trip cost to specifically account for the differential sensitivity of travelers’ income levels to travel costs

The mode share model will reflect the following trip purpose market segmentation:

- Business trips
- Non-business/non-work trips
- Commute (journey to work) trips

Models of travel choice can be based on revealed (RP) or stated (SP) data. Each type of data provides certain advantages over the other. Combining the two sets of data to estimate a single model can produce a model that retains the advantages of both RP and SP models and eliminate or dramatically reduce the disadvantages of each. The NEC FUTURE travel survey is intended to collect data which can be used to study travel patterns and travel behavior along the Northeast Corridor. This information will be used to estimate a forecasting model of travel mode choice in the Northeast Corridor.

Over a number of years, it has been recognized that a combination of revealed preference (RP) and stated preference (SP) questions can yield a rich dataset which is capable of retaining the advantages of both types of data and minimize the limitations. The resulting models obtain a strong connection with real behavior as the RP data takes account of the real world constraints the respondents face and the SP data takes account of the wider range of alternatives and alternative attributes. The SP data also enables the survey designer to create alternatives in which the explanatory variables have a larger range of variability within and between alternatives and break the correlation between explanatory variables within each alternative. (Louviere, et al, 230). In addition, there are also several situations where they can provide insight into a future traveler’s choice behavior. These situations include the introduction of a new choice alternative with new attributes or include features where estimation results are limited due to little variability for a subset of explanatory variables or a subset of explanatory variables are highly collinear, such as time and cost (Louviere, et al, 21).

Models estimated solely from SP data require careful calibration to match base conditions in order to produce reliable results. A major advantage to utilizing SP data in addition to RP data is that the survey design can dictate much wider ranges of attributes and control the relationships between attributes, which increases the robustness of models estimated with SP data over models estimated with RP data (Louviere, et al, 231). Including stated preference questions in a survey also increases the number of records in the survey dataset substantially, as revealed preference by definition is one response per respondent, while stated preference questions can return multiple responses per respondent, enhancing the dataset with the ability to test for individual

tipping points (Louviere, et al, 24). In the NEC FUTURE survey, each respondent will answer 12 SP questions, in addition to the questions about their actual trip.

One way to improve the realism of the SP data responses is to use the pivot method (Train and Wilson, 192). This method specifically references the RP response in the SP questions to encourage the respondent to think about their response with a higher level of real world constraints, and assuring the alternatives are similar to what the respondent might actually experience. In the NEC FUTURE survey, the SP questions use the same origin and destinations as the RP response, two of the three modes being tested are the chosen mode from respondent's own choice set (as indicated survey questions 25-26), and the characteristics of these modes are scaled to the respondent's trip length. This ensures that the modes being tested are viable options for the respondent and the stated response will more likely mimic what they would actually chose in a real-world situation.

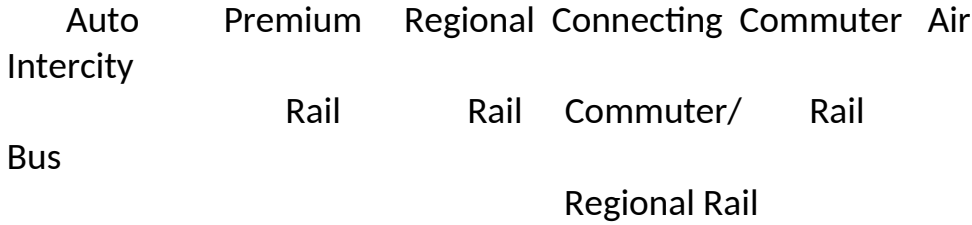
Unlike other choice situations where stated preference data may have been used unsuccessfully, transportation mode choice is an exercise that respondents are familiar with as they consider the available options to make choices that satisfy their own individual travel requirements. We believe that the stated preference experiments in the survey mimic this process by providing respondents with similar information that they may find from published schedule or timetables, published fares or pricing, and their own experiences about travel by their chosen mode.

While models can be estimated with each type of data separately, the most robust models combine RP and SP data in order to take advantage of the unique characteristics of each type. The NEC FUTURE mode choice models will use a nested logit structure, to reflect the differential substitution that exists between different modes of travel. There will be three models in total for each trip type: journey to work, business, and other non-work/non-business trips.

The nested logit structure is preferable for mode choice modeling over MNL because of the Independence of Irrelevant Alternatives (IIA) property inherent in the MNL model. The IIA property is an issue because any changes or additions to the alternatives cause a proportional change to the probabilities of all other alternatives. In other words, there is no differentiation among choices to account for similarities between modes and the potentially higher propensity for respondents to switch to a similar alternative. The nested logit model, on the other hand, allows for grouping similar alternatives, so that they are more competitive within the nest versus than with other alternatives outside of the nest.

Ultimately, the survey data itself will dictate the nested logit structure used in the final mode choice model, but examples of possible nested logit structures to be tested in the model estimation are shown in Figure A.2. These nesting structures show that common carrier modes have similarities to each other while auto is quite different in very important ways such as accommodating larger group sizes with minimal additional cost, flexible departure times, and flexible stopping patterns for multi-destination trips. The appropriate nesting structure for the rail alternatives will be an important consideration and will be data driven.

FIGURE A.2 CANDIDATE MODE CHOICE NESTED LOGIT STRUCTURES



The utility equations for the nested logit structure all follow the same formulation, which is shown below with the anticipated variables to be tested.

$$U_{Mode} = ASC_{Mode} + \beta_{TT} * Travel\ Time + \beta_{Cost} * Travel\ Cost + \beta_{Freq} * Schedule\ Delay + \beta_{Reliability} * Reliability + \beta_{Connection} * N$$

The formulation of each mode's probability is dependent on its location in the nesting structure. The following formulas show the probabilities for the first nesting structure in Figure A.2.

Final Probabilities – RP Model:

$$PR_{Auto} = \frac{\exp(U_{Auto})}{\exp(U_{Auto}) + \exp(\mu_{Common\ Carrier} * \Gamma_{Common\ Carrier})}$$

$$PR_{Prem\ Rail} = PR_{Prem\ Rail \vee Prem \wedge Reg\ Rail} * PR_{Prem \wedge Reg\ Rail \vee Common\ Carrier} * PR_{Common\ Carrier}$$

$$PR_{Reg\ Rail} = PR_{Reg\ Rail \vee Prem \wedge Reg\ Rail} * PR_{Prem \wedge Reg\ Rail \vee Common\ Carrier} * PR_{Common\ Carrier}$$

$$PR_{Comm-Oth\ Rail} = PR_{Comm-Oth\ Rail \vee Common\ Carrier} * PR_{Common\ Carrier}$$

$$PR_{Comm\ Rail} = PR_{Comm\ Rail \vee Common\ Carrier} * PR_{Common\ Carrier}$$

$$PR_{Air} = PR_{Air \vee Common\ Carrier} * PR_{Common\ Carrier}$$

$$PR_{Bus} = PR_{Bus \vee Common\ Carrier} * PR_{Common\ Carrier}$$

Conditional Probabilities, Logsums, and Logsum Parameters – RP Model:

$$PR_{Prem\ Rail \vee Prem \wedge Reg\ Rail} = \frac{\exp\left(\frac{U_{Prem\ Rail}}{\mu_{Prem \wedge Reg\ Rail}}\right)}{\exp\left(\frac{U_{Prem\ Rail}}{\mu_{Prem \wedge Reg\ Rail}}\right) + \exp\left(\frac{U_{Reg\ Rail}}{\mu_{Prem \wedge Reg\ Rail}}\right)}$$

$$PR_{Reg\ Rail \vee Prem \wedge Reg\ Rail} \dot{=} \frac{\exp\left(\frac{U_{Prem\ Rail}}{\mu_{Prem \wedge Reg\ Rail}}\right)}{\exp\left(\frac{U_{Prem\ Rail}}{\mu_{Prem \wedge Reg\ Rail}}\right) + \exp\left(\frac{U_{Reg\ Rail}}{\mu_{Prem \wedge Reg\ Rail}}\right)}$$

$$PR_{P \wedge R Rail \vee ComCar} = \frac{\exp\left(\frac{\mu_{P \wedge R Rail} * \Gamma_{P \wedge R Rail}}{\mu_{ComCar}}\right)}{\exp\left(\frac{\mu_{P \wedge R Rail} * \Gamma_{P \wedge R Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{C-O Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{C Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{Air}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{Bus}}{\mu_{ComCar}}\right)}$$

$$PR_{C-O Rail \vee ComCar} = \frac{\exp\left(\frac{U_{C-O Rail}}{\mu_{ComCar}}\right)}{\exp\left(\frac{\mu_{P \wedge R Rail} * \Gamma_{P \wedge R Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{C-O Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{C Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{Air}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{Bus}}{\mu_{ComCar}}\right)}$$

$$PR_{C Rail \vee ComCar} = \frac{\exp\left(\frac{U_{C Rail}}{\mu_{ComCar}}\right)}{\exp\left(\frac{\mu_{P \wedge R Rail} * \Gamma_{P \wedge R Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{C-O Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{C Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{Air}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{Bus}}{\mu_{ComCar}}\right)}$$

$$PR_{Air|ComCar} = \frac{\exp\left(\frac{U_{Air}}{\mu_{ComCar}}\right)}{\exp\left(\frac{\mu_{P \wedge R Rail} * \Gamma_{P \wedge R Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{C-O Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{C Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{Air}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{Bus}}{\mu_{ComCar}}\right)}$$

$$PR_{Bus \vee ComCar} = \frac{\exp\left(\frac{U_{Bus}}{\mu_{ComCar}}\right)}{\exp\left(\frac{\mu_{P \wedge R Rail} * \Gamma_{P \wedge R Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{C-O Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{C Rail}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{Air}}{\mu_{ComCar}}\right) + \exp\left(\frac{U_{Bus}}{\mu_{ComCar}}\right)}$$

$$PR_{Common carrier} = \frac{\exp(\mu_{Common Carrier} * \Gamma_{Common Carrier})}{\exp(U_{\ddot{i}\ddot{i} Auto}) + \exp(\mu_{Common Carrier} * \Gamma_{Common Carrier})}$$

$$\Gamma_{Prem \wedge Reg Rail} = \ln \left[\exp\left(\frac{U_{Prem Rail}}{\mu_{Prem \wedge Reg Rail}}\right) + \exp\left(\frac{U_{Reg Rail}}{\mu_{Prem \wedge Reg Rail}}\right) \right]$$

$$\Gamma_{Common Carrier} = \ln \left[\exp\left(\frac{\mu_{Prem \wedge Reg Rail} * \Gamma_{Prem \wedge Reg Rail}}{\mu_{Common Carrier}}\right) + \exp\left(\frac{U_{Comm-Oth Rail}}{\mu_{Common Carrier}}\right) + \exp\left(\frac{U_{Comm Rail}}{\mu_{Common Carrier}}\right) + \exp\left(\frac{U_{Air}}{\mu_{Common Carrier}}\right) + \exp\left(\frac{U_{Bus}}{\mu_{Common Carrier}}\right) \right]$$

$$\mu_{Nest} = \log \sum \text{parameter for nest (estimated)}$$

The combined RP-SP model can be structured similarly to any model structure. For example, a nested logit structure can be used with either RP or SP data alone as shown in Figure A.2.

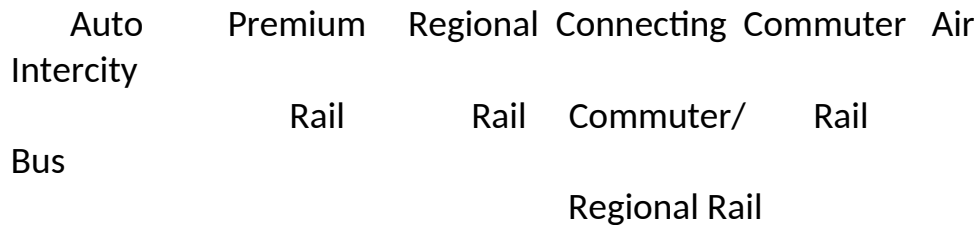
Because of the differences between the two types of data though, this model structure has to be modified to combine RP and SP data. The use of a scaling factor applied to the SP data allows for the combined estimation of the choice model, to account for greater uncertainty and possible biases in the SP data. (Ben-Akiva, et al, 339). Figure A.3 shows a revised structure which incorporates the scaling factor between the RP and SP models.

Because each respondent provides multiple SP responses, the SP questions are typically weighted at a lower value than the RP questions (Ben-Akiva, et al, 345). Judgmental approaches can use weights which can range from equal weight between each SP and RP question or equal weight between the set of SP questions and each RP question. The twelve SP questions in the NEC FUTURE survey are split into two groups, six questions which hold values for mode 3 (out of three modes) constant, and the other six questions hold values for mode 2 (out of three modes) constant. Because the SP questions offer a more limited choice set than the RP, the SP question weights will probably not be set at 1/12 of the revealed preference, but will still be much lower per question than the RP.

The choices present in the RP nest versus the SP nest can differ based on the choices presented to the respondents (i.e. a new mode would only be present in the SP data nest). The scaling factor can be found in the SP data nest, which is estimated during the model estimation process. Other model parameters can be constrained to be equal between the RP and SP data.

For the NEC FUTURE model, it is anticipated that some initial testing will be done using the RP data alone, which can utilize the nesting structure shown in Figure A.2. Most estimation testing though, including the final model estimation runs, will use the richer SP data and combined RP-SP data, which requires modifications to the model structure to accommodate the differences between RP and SP data.

FIGURE A.3 EXAMPLE NESTING STRUCTURE WITH SCALING FACTOR



The utility equations for the RP modes do not change for the combined RP-SP model. In addition to the original utility equations, there is now a separate equation for each SP mode, with the only difference being an additional alternative-specific constant. The other coefficients are the same and are thus jointly estimated from the RP and SP. This can be seen in the equations below.

Utility Equations – RP-SP model

$$U_{Mode-RP} = ASC_{Mode} + \beta_{TT} * Travel\ Time + \beta_{Cost} * Travel\ Cost + \beta_{Income} * Income + \beta_{Freq} * Frequency + \beta_{Reliability} * Reliability$$

$$U_{Mode-SP} = ASC_{Mode} + ASC_{Mode-SP} + \beta_{TT} * Travel\ Time + \beta_{Cost} * Travel\ Cost + \beta_{Income} * Income + \beta_{Freq} * Frequency + \beta_{Reliability} * Reliability$$

The probabilities for the RP modes are all the same as previously described for the previous structure, with the addition of the SP nest term in the denominator. The SP probabilities all incorporate the scaling factor into the probability equations, with the new terms shown below.

Final Probabilities – RP-SP Model:

$$PR_{Auto} = \frac{\exp(U_{Auto})}{\exp(U_{Auto}) + \exp(\mu_{CommonCarrier} * \Gamma_{CommonCarrier}) + \exp(\mu_{SP} * \Gamma_{SP})}$$

$$PR_{PremRail} = PR_{PremRail \vee Prem \wedge RegRail} * PR_{Prem \wedge RegRail \vee CommonCarrier} * PR_{CommonCarrier}$$

$$PR_{RegRail} = PR_{RegRail \vee Prem \wedge RegRail} * PR_{Prem \wedge RegRail \vee CommonCarrier} * PR_{CommonCarrier}$$

$$PR_{Comm-OthRail} = PR_{Comm-OthRail \vee CommonCarrier} * PR_{CommonCarrier}$$

$$PR_{CommRail} = PR_{CommRail \vee CommonCarrier} * PR_{CommonCarrier}$$

$$PR_{Air} = PR_{Air \vee CommonCarrier} * PR_{CommonCarrier}$$

$$PR_{Bus} = PR_{Bus \vee CommonCarrier} * PR_{CommonCarrier}$$

$$PR_{Auto-SP} = PR_{Auto-SP \vee SP} * PR_{SP}$$

$$PR_{PremRail-SP} = PR_{PremRail-SP \vee Prem \wedge RegRail-SP} * PR_{Prem \wedge RegRail-SP \vee CommonCarrier-SP} * PR_{CommonCarrier-SP \vee SP} * PR_{SP}$$

$$PR_{RegRail-SP} = PR_{RegRail-SP \vee Prem \wedge RegRail-SP} * PR_{Prem \wedge RegRail-SP \vee CommonCarrier-SP} * PR_{CommonCarrier-SP \vee SP} * PR_{SP}$$

$$PR_{Comm-OthRail-SP} = PR_{Comm-OthRail-SP \vee CommonCarrier-SP} * PR_{CommonCarrier-SP \vee SP} * PR_{SP}$$

$$PR_{CommRail-SP} = PR_{CommRail-SP \vee CommonCarrier-SP} * PR_{CommonCarrier-SP \vee SP} * PR_{SP}$$

$$PR_{Air-SP} = PR_{Air-SP \vee CommonCarrier-SP} * PR_{CommonCarrier-SP \vee SP} * PR_{SP}$$

$$PR_{Bus-SP} = PR_{Bus-SP \vee CommonCarrier-SP} * PR_{CommonCarrier-SP \vee SP} * PR_{SP}$$

$$PR_{SP} = \frac{\exp(\mu_{SP} * \Gamma_{SP})}{\exp(U_{Auto}) + \exp(\mu_{CommonCarrier} * \Gamma_{CommonCarrier}) + \exp(\mu_{SP} * \Gamma_{SP})}$$

$$\Gamma_{SP} = \ln \left[\exp\left(\frac{U_{Auto-SP}}{\mu_{SP}}\right) + \exp\left(\frac{\Gamma_{CommonCarrier-SP}}{\mu_{SP}}\right) \right]$$

As described in Whitehead, et al (37), there are two tests which can be done to test the predictive validity of the combined RP-SP model. These are the within sample test (testing how well the model predicts behavior of respondents in the model) and the out-of-sample test (testing how well the model predicts behavior of individuals outside of the model). Whitehead, et al (38) finds that in the literature jointly estimated RP-SP models have much greater predictive validity

than independently estimated RP or SP models with the within sample test, and that the out-of-sample test tends to show similar results among the three model types.

As described above, the proposed model structure for the NEC mode choice model is the Nested Logit (NL) model. If during model estimation, there is difficulty getting nesting parameters or convergence for the models, or if key parameters do not have statistically significant coefficients, it may be necessary to investigate alternative model forms. Additional flexibility can be obtained in formulated discrete choice models relative to the MNL or NL models. Three of these are:

- The Paired Combinatorial Logit (PCL) model, a two level model that includes nests equal to the number of pairs of alternatives. Each nest includes a distinct pair of two alternatives; each alternative appears in the number of nests equal to the number of alternatives less one. In addition to the utility function parameters, these models require estimation of a nesting parameter for each pair. The alternatives are equally proportionally assigned to each nest in which they appear (i.e. they are equally similar to each alternative they share a nest with). [F.S. Koppelman and C-H Wen, The Paired Combinatorial Logit Model: Properties, Estimation and Application, Transportation Research-B, V.34, N.2, 2000, pp.75-89.]
- The Cross Nested Logit Model (CNL) is a model derived from the Generalized Extreme Value (GEV) model, like the others discussed here, which allows alternatives to belong to multiple nests. The CNL allows different proportions of each alternative to be assigned to nests, relaxing the constraint of the PCL, but each nest has the same structural parameter, unlike the PCL model. [P. Vovshva, Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area, Transportation Research Record, V. 1607, 1997, pp.6-15.]
- The Generalized Nested Logit (GNL) is also a two level model but it includes nests selected by the analyst, which combines the flexibilities of the PCL and CNL models. In addition to the utility function parameters, and the nesting parameter for each nest, a set of allocation parameters is estimated to represent the degree to which each alternative is allocated to each nest. This allows maximum competitive flexibility among pairs of alternatives. [C-H Wen and F. S. Koppelman, The Generalized Nested Logit Model, Transportation Research-B, v. 35, N. 7, 2001, pp. 627-641.] It includes all two level extreme value (logit type) models as special cases.

A further level of model flexibility can be obtained by adopting the Network Generalized Extreme Value Model (NGEV) [A. Daly and M. Bierlaire. A general and operational representation of generalized extreme value models. Transportation Research Part B: Methodological, 40:285–305, 2006] which is a multi-level extension of the Generalized Nested Logit Model Structure. The use of this model has been limited, to date, as its complexity may overwhelm its advantages. Newman (2008) identified and demonstrated the need for constraints to ensure that the model is identifiable. [J. P. Newman. Normalization of network generalized extreme value models. Transportation Research Part B: Methodological, 42(10):958–969, 2008]. The constraint imposed is arbitrary; that is, the choice of which constraint to impose does not change the goodness of fit of the model; but the nesting interpretation differs depending on the constraint adopted.

Additional model flexibility in any of the above models can be obtained by combining the concept of the Mixed Logit Model [D. McFadden and K. Train, Mixed MNL Models for Discrete Response, Journal of Applied Econometrics, 15:447-470 (2000)] with any Logit Model (MNL, NL, PCL, GNL, and NGEV). The concept of the mixed logit model is to add random parameters to the already structured model. While originally developed for use with the multinomial logit model, it can be used with any member of the logit family or Generalized Extreme Value model. There are four different ways in which random parameters can be used.

The first two contribute to an improved estimate of the behavior under study; the second two can be used to offset problems associated with the use of SP data based models or combined SP and RP data based models. The four methods, all of which are relevant to the study of intercity travel, are:

The first use of mixed logit is to take account of variable values of a level of service value. E.g., in the case of variability of the value of travel time, the equation

$$\begin{aligned} V_{1t} &= \dots + \beta_{\zeta} x_{1t} \zeta + \dots \\ V_{2t} &= \dots + \beta_{\zeta} x_{2t} \zeta + \dots \\ V_{3t} &= \dots + \beta_{\zeta} x_{3t} \zeta + \dots \end{aligned}$$

can be modified to add a random value as follows,

$$\begin{aligned} V_{1t} &= \dots + (\beta_{\zeta} + \delta_{\zeta}) x_{1t} \zeta + \dots \\ V_{2t} &= \dots + (\beta_{\zeta} + \delta_{\zeta}) x_{2t} \zeta + \dots \\ V_{3t} &= \dots + (\beta_{\zeta} + \delta_{\zeta}) x_{3t} \zeta + \dots \end{aligned}$$

This allows the parameter β_{ζ} to be increased by a draw from the random distribution of δ_{ζ} . Any distribution can be used but it is generally desirable to use a bounded distribution; this can be accomplished by truncating a distribution with an infinite tail of tails.

The second use of mixed logit is to take account of random preference bias for a particular alternative. This might be appropriate in the case when a new alternative is considered and it is expected that the people will have variable values of the new alternative. E.g., the constant for a new alternative can be modified from

$$\begin{aligned} V_{1t} &= \alpha_{1t} + \dots \\ V_{2t} &= \alpha_{2t} + \dots \\ V_{3t} &= \alpha_{3t} + \dots \end{aligned}$$

to

$$\begin{aligned} V_{1t} &= \alpha_{1t} + \dots \\ V_{2t} &= \alpha_{2t} + \gamma_{2t} + \dots \\ V_{3t} &= \alpha_{3t} + \dots \end{aligned}$$

in the case of alternative 2 being a new alternative.

The third use of mixed logit is to represent correlation between sets of common alternatives in a series of experiments presented to a single individual. It is possible that an individual's preferences among alternatives are common for the series of experiments. This can be represented by introducing random parameters, added to the constant, for each alternative. For example, the bias parameter changes to

$$\begin{aligned} V_{1t} &= \alpha_{1t} + \dots \\ V_{2t} &= \alpha_{2t} + \eta_{2t} + \dots \\ V_{3t} &= \alpha_{3t} + \eta_{3t} + \dots \end{aligned}$$

or

$$\begin{aligned} V_{2t} &= \alpha_{2t} + \eta_{2t} + \dots \\ V_{3t} &= \alpha_{3t} + \eta_{3t} + \dots \end{aligned}$$

$$V_{4t} = \alpha_{4t} + \dots$$

Finally, the fourth use of mixed logit is to represent the expected bias in favor of the chosen alternative in the reported trip by adding a common random variable to that alternative in every case in which the real chosen alternative is an alternative in the stated preference experiment. Thus, the basic equation

$$\begin{aligned} V_{1t} &= \alpha_{1t} + \dots \\ V_{2t} &= \alpha_{2t} + \dots \\ V_{3t} &= \alpha_{3t} + \dots \end{aligned}$$

can be modified to

$$\begin{aligned} V_{1t} &= \alpha_{1t} + \lambda_{1t} + \dots \\ V_{2t} &= \alpha_{2t} + \dots \\ V_{3t} &= \alpha_{3t} + \dots \end{aligned}$$

because alternative one was chosen in the RP case.

Regardless of the model structure used in estimation, the data requirements are the same, as they belong to the same family of models. The proposed NEC survey will provide adequate data to estimate any of the model structures discussed above.

Questions particularly relevant to mode choice of the respondent include those about the specific one-way trip (questions 1-24), which provide trip details to estimate revealed preference models, and mode choice trade-off stated preference (SP) questions (questions 27-38), which provide data that can be used to estimate the model. In addition, respondents are asked to indicate their second and third choices of mode (questions 25-26), which provides insight into the competitive relationship among alternatives, and “pivot” questions (39-42) that address additional independent variables that cannot be included in the SP experimental design without adding too much complexity.

The “pivot” questions have been structured to directly assess the impacts of three variables – Transfers, Station/Airport availability, and Access Mode availability – on mode choice probabilities. These variables are of secondary interest to the travel time, cost, schedule, and reliability variables included in the SP experiments. In discussions among the project technical teams, including those responsible for alternatives development and evaluation, and the Technical Working Group (TWG), there was a strong desire to have these variables addressed in the model if at all possible.

The underlying relationships between demand and Transfers, Station/Airport availability, and Access Mode availability will be estimated directly by cross-tabulating the survey results. As an example, Q39A asks the respondent to review one of the previous SP choice exercises (randomly chosen from the Q27-Q38). Instead of asking for a singular choice of a mode, the question asks the respondent to state their probability of choosing that mode (from 1-“definitely would not use” to 5 – “definitely would use”). Then, Q40A asks the respondent to state their probability of choosing that same mode if a transfer (that required no additional travel time) was involved. The cross tabulation of these two questions will show whether a transfer (requiring no additional travel time) significantly impacts the probability of choosing a mode. If significant, these secondary mode characteristics will be taken into account when estimating the probability of each mode being chosen.

Model development does not end with model estimation, however. The estimated model will be implemented within an NEC model application package that includes procedures for processing input data and summarizing output results. Current input data will be applied to the model and outputs compared to actual observed current travel volumes to validate the model and confirm its accuracy. This will include reviewing not only overall NEC-wide results, but also individual markets to confirm that the model is addressing the broad range of different markets that exist in the NEC.

As may be necessary, the model will be further adjusted to better match the existing data. However, significant adjustments are not expected to be necessary because the model estimation was based in part on observed RP data addressing all modes. This is not always the case in other corridors throughout the US where, for example, rail service and travel volumes may not currently be significant.

In contrast to existing available models, the new NEC model will provide NEC FUTURE with a demand forecasting tool that addresses the full range of the available modes of rail travel, including intercity and regional commuter services. Although Amtrak and regional agencies maintain forecasting models that collectively address the Northeast, they do not address the interaction of the intercity and regional commuter services. For example, Amtrak's model focuses exclusively on intercity services that operate between regions, even though some such trips may be served by commuter rail, and the most of regional models ignore intercity service that may serve local markets too.

The new model is designed to address NEC travel markets served by future integrated operations and service planning alternatives under development in NEC FUTURE, encompassing both intercity and commuter services along the NEC, including shared markets (where intercity and commuter services operate in parallel) and linked markets (where intercity and commuter services connect to each other, with the commuter service essentially serving as a feeder/distributor from/to local markets not directly served by the intercity service). As described above, none of the existing models used by Amtrak and/or regional agencies within the NEC adequately addresses all of these potential markets and services, which go beyond the existing and historical institutional and geographic limitations of Amtrak and commuter rail services.

Because it will address a broad range of modes and geography throughout the Northeast, the development of a new NEC model will require new surveys designed to address these dimensions. These new surveys include questions addressing existing travel by intercity and regional commuter modes of travel between and within the Northeast (see footnote 1). All of the existing available survey data is tied to specific existing models developed by Amtrak and regional agencies that address more limited geography and/or modes of travel. For example, Amtrak's survey data focuses exclusively on intercity travel modes and survey data collected by regional agencies is limited to a specific region and does not address intercity modes. Although they collectively address all of the major NEC markets, these existing data and models do not provide a consistent integrated analysis and forecasting basis that spans all geographies and modes throughout the NEC.

Data Collection for Model

The information collection will be conducted in two phases. The first phase will be conducted by telephone, using computer-assisted telephone interviewing (CATI). The questionnaire for Phase I is titled "NEC FUTURE Recruit Survey – Screener" and is included in this ICR package. A dual frame sample design will be used including both landlines and cell phones. This first survey will obtain basic travel information and invite just those who experienced a qualifying intercity or

regional travel to provide more detailed travel information and choice preferences in a second phase. The phase 1 survey interview is estimated to take 5 minutes to complete. The second phase, which will immediately follow the first, will be administered by web. The second phase will ask more detailed questions about one randomly selected trip that the respondent reported in the first phase of the survey. In addition, it will ask the travel preference questions. For people who do not have easy web access, a packet of travel scenarios will be first mailed to them. Being able to visually review the travel scenarios is very important because of the complexity of the information being presented. These respondents will be called back via a follow-up phone survey that will reference the mailed materials. The phase 2 survey is estimated to take 15 minutes to complete for both web and mail/phone administration modes. The web-based questionnaire is titled “NEC FUTURE – Internet Follow-up Survey”. Screenshots of how the web-based questionnaire will look are included in the file “HSR_F”. The phone follow-up version is essentially the same as the internet version with a few minor differences that are also noted in the questionnaire document. The mailed visual materials for the telephone follow-up are titled “SampleSurveyScreens” and the introductory letter for the mailed materials is titled “High Speed Rail Mailing Introductory Letter”. The reference files mentioned in the follow-up surveys are titled “Area-Area Matrix” and “Places NEC.”

The following text explains in further detail what information each of the questionnaires collects and how it will be used.

Recruit Survey - Screener

Question C.1 allows for cell phone respondents to be called later at a better, safer time.

Questions S.1- S.1b identify a random member of the household to participate in the survey.

Question S.1c asks cell phone users for their home area (this is already known for land lines).

Questions S.2 & S.3 ask if non-commute trips were taken and, if so, how many trips were taken to randomly ordered eligible areas for that respondent’s home area. Eligible areas exclude the respondent’s home area, other nearby areas (typically less than 50 miles away from the home), and areas where the trip would be entirely outside of the NEC spine. Questions S.2 & S.3 terminate when trips are found to an area or no trips are found to any of the eligible areas.

Questions S.2a, S.2b & S.3a ask if commute trips were taken to any of the eligible areas and, if so, how many were taken in a typical week.

If no trips were found in the above series of questions, the survey skips to collect demographic information in Questions D-1 through D-12. In this case, the respondent does not qualify for the follow-up survey and is not counted as a completed survey.

Question S.4 is directed to all respondents who made a trip (as determined through the previous questions) and thus qualify for the survey. It explains that the survey consists of two parts and asks if respondent is willing to participate. If not willing to participate, the survey skips to collect demographic information in Questions D-1 through D-12.

Questions 1A-1B ask for the number of trips made by mode and non-commute trip purpose within the past year within one of the areas identified (in S.2) above (randomly selected if there were trips to more than one area). In addition to providing a basis for selecting the respondent’s “reference trip”, this information will also be used with other available data to estimate existing travel volumes by mode and purpose in the NEC.

Question 1C asks for the number of commute trips made by mode in a typical week to the area identified (in S.2b) above. In addition to providing a basis for selecting the respondent’s “reference trip”, this information will also be used with other available data to estimate existing travel volumes by mode and purpose in the NEC.

Next, the recruit survey randomly selects a specific mode and trip purpose for those identified above. This is the “reference trip” for the respondent. Then, the recruit survey asks for an email address for a link to complete the follow-up survey. If no email address, the respondent is told they can participate by mail/phone follow-up. For those that participate by mail/phone follow-up, the following questions are asked to collect information so that the mail materials can be customized for the respondent.

Question 2 asks for the origin of the reference trip. This provides more detailed geographic information necessary for a more precise definition of the trip and its key characteristics.

Question 3 asks for the destination of the reference trip. This provides more detailed geographic information necessary for a more precise definition of the trip and its key characteristics.

Question 4 asks for the type of train service that was used (if respondent’s “reference trip” was by train). This is important since we need to distinguish among different types of train service in the model and the survey.

Questions 5A-5C ask for the fare paid by the respondents if the trip was by plane, train, or bus respectively. This self-reported fare will be used as the base fare value for use in SP choice exercises. If respondent does not remember the fare paid, or if the value they provide is unreasonable, default values will be used based on published fares for travel between the origin place and destination.

Questions 6A-7D ask what modes the respondent would have used instead if the chosen mode was not available. These questions provide important insights on respondent’s alternative modes of travel, which is important to understanding mode choice structures that define different levels of substitutability among modes. In addition, the mode selected by the respondent in this question will be used as an alternative mode for that respondent in the SP choice exercises.²

Questions D-1 through D-12 collect demographic information from respondents who do not qualify for the survey or are not willing to participate in the follow-up survey. This information provides a demographic profile of those without qualifying trips and those who did not agree to participate in the full survey, which will be compared to demographic profile of respondents and the NEC. These are not counted as completed surveys.

Procedures for Follow-up Survey – Mailed Visual Aid /Phone

Those respondents who chose to participate by phone will be mailed the visual aids depicting the 12 SP choice exercises which are customized to the respondent’s answers to questions 2 through 7 in the Recruit Survey-Screener. The respondent will then be called by the telephone interview and asked the Follow-up Survey Questions (below). The telephone interviewer will be using the same web-based interview and supplying the answers on behalf of the respondent. A few of the questions (which are identified below) will be skipped by the phone interview because the answers to those questions have already been provided by the respondent during the Recruit Survey-Screener.

Follow-up Survey - Internet

Question 1 asks for the type of train service that was used (if respondent’s “reference trip” was by train). This is important since we need to distinguish among different types of train service in the model and the survey. Since this information is collected in the screener for phone follow-up respondents, this question is skipped in the phone follow-up.

² As explained in Supporting Statement B, the mode actually used for the reference trip and a randomly selected third mode will be used as the other alternatives for the respondent’s SP choice exercises. The respondent also has the option to take no trip at all.

Question 2 asks for the origin of the reference trip. This provides more detailed geographic information necessary for a more precise definition of the trip and its key characteristics. Since this information is collected in the screener for phone follow-up respondents, this question is skipped in the phone follow-up.

Question 3 asks for the type of place for the origin of the reference trip. The type of place can be important to the choice of mode. For example, most travelers are more likely to have access to a car at their home. The response to this question in combination with Question 8 also supplements the trip purpose information collected in Question 15 and can be used to infer trip purpose if it is not provided when Question 15 is asked.

Questions 4A, 5A, and 6A apply only if respondent's "reference trip" was by train and ask for the rail station used to board the train, the mode of access used to get from the origin to the station, and the time spent at the station prior to boarding the train. Again, this provides more detailed information necessary for a more precise definition of the trip and its key characteristics.

Questions 4B, 5B, and 6B apply only if respondent's "reference trip" was by plane and ask for the airport used to board the plane, the mode of access used to get from the origin to the airport, and the time spent at the airport prior to boarding the plane. Again, this provides more detailed information necessary for a more precise definition of the trip and its key characteristics.

Question 5C applies only if respondent's "reference trip" was by bus and asks for the mode of access used to get from the origin to the bus. Again, this provides more detailed information necessary for a more precise definition of the trip. However, given that there are many bus stops, and many are not at formal terminals like airports and rail stations, we do not attempt to ask about specific stops (thus, there is no Question 4C or 6C).

Question 7 asks for the destination of the reference trip. This provides more detailed geographic information necessary for a more precise definition of the trip and its key characteristics. Since this information is collected in the screener for phone follow-up respondents, this question is skipped in the phone follow-up.

Question 8 asks for the type of place for the destination of the reference trip. The type of place can be important to the choice of mode. For example, most travelers are more likely to have access to a car at their home. The response to this question in combination with Question 3 also supplements the trip purpose information collected in Question 15 and can be used to infer trip purpose if it is not provided when Question 15 is asked.

Questions 9A, 10A, 11A, 12A, and 13A apply only if respondent's "reference trip" was by train and ask for the rail station used to get off the train, the mode of access used to get from the station to the destination, if a connection from one train to another was required, and information about the fare paid by the respondent for the train. Again, this provides more detailed information necessary for a more precise definition of the trip and its key characteristics. This self-reported fare will be used as the base fare value for use in SP choice exercises. If respondent does not remember the fare paid, or if the value they provide is unreasonable, default values will be used based on published train fares for travel between the origin place and destination. Since train fare information is collected in the screener for phone follow-up respondents, 12A and 13A are skipped in the phone follow-up.

Questions 9B, 10B, 11B, 12B, and 13B apply only if respondent's "reference trip" was by plane and ask for the airport used to get off the plane, the mode of access used to get from the airport to the destination, if a connection from one plane to another was required, and information about the fare paid by the respondent for the plane. Again, this provides more detailed information necessary for a more precise definition of the trip and its key characteristics. This self-reported fare will be used as the base fare value for use in SP choice exercises. If respondent does not remember the fare paid, or if the value they provide is unreasonable, default values will be used

based on published air fares for travel between the origin place and destination. Since plane fare information is collected in the screener for phone follow-up respondents, 12B and 13B are skipped in the phone follow-up.

Questions 10C, 11C, and 12C apply only if respondent's "reference trip" was by bus and ask for the mode of access used to get from the bus to the destination, if a connection from one bus to another was required, and information about the fare paid by the respondent for the bus. As for the origin end, given that there are many bus stops, and many are not at formal terminals like airports and rail stations, we do not attempt to ask about specific stops (thus, there is no Question 9C). Again, this provides more detailed information necessary for a more precise definition of the trip and its key characteristics. This self-reported fare will be used as the base fare value for use in SP choice exercises. If respondent does not remember the fare paid, or if the value they provide is unreasonable, default values will be used based on published bus fares for travel between the origin place and destination. Since bus fare information is collected in the screener for phone follow-up respondents, 12C is skipped in the phone follow-up.

Questions 14A, 14B, 14C and 14D apply only if respondent's "reference trip" was by passenger car/truck/van and ask for the estimated one-way travel time, from a range of network times, and the estimated cost for tolls, parking, and fuel. Again, this provides more detailed information necessary for a more precise definition of the trip and its key characteristics. This self-reported cost estimate will be used as the base travel cost value for use in SP choice exercises for internet follow-up respondents. To the extent that the respondent does not remember these costs or if they are unreasonable, default values will be used based on published tolls likely to be encountered along the highway network between the origin and destination and per-mile costs applied to the trip distance along the highway network. Default values will also be used for phone follow-up respondents, since this information is not being collected in the screener.

Question 15 asks for the overall purpose of the respondent's trip. Trip purpose is one of the most important trip characteristics defining different market segments. In cases where trip purpose is not provided here, it can often be inferred from the responses to questions 3 and 8.

Questions 16 through 19 ask if the trip was made alone or with others and, if it was not alone, collect information on the composition of the group. Mode choice is often impacted by group size and composition. Group size and composition also impact travel costs.

Questions 20 through 23B address the respondent's travel schedule, including the relative importance of departure or arrival times, their ideal departure or arrival time, and their actual departure or arrival time. This information defines key travel schedule dimensions of the respondent's reference trip which is used in the SP questions.

Questions 24 through 24B ask if the respondent made a round trip and, if so, the overall duration of the trip (expressed as nights away from home). The duration of the trip can also have important implications with respect to scheduling and pricing. Respondents making day trips or single overnight trips tend to have less schedule flexibility. Trip duration can also impact costs.

Questions 25A through 26D ask what modes the respondent would have used instead if the chosen mode was not available. These questions provide important insights on respondent's alternative modes of travel, which is important to understanding mode choice structures that define different levels of substitutability among modes. In addition, the mode selected by the respondent in this question will be used as an alternative mode for that respondent in the SP choice exercises.³

³ As explained in Supporting Statement B, the mode actually used for the reference trip and a randomly selected third mode will be used as the other alternatives for the respondent's SP choice exercises. The respondent also has the option to take no trip at all.

Questions 27 through 38 are the 12 Stated Preference (SP) choice exercises that really represent the “core” of the survey and provide the primary basis for estimating the new mode choice model. These SP questions ask respondents to choose from among three modes of travel, each with specific characteristics. These modal characteristics vary across the questions, with values developed from an experimental design that minimizes correlations among variables which can be problematic when seeking to estimate sensitivities to variables independently. Supporting Statement B provides details on the experimental design.

Questions 39A through 42C are “pivot” questions that supplement the SP questions by addressing some additional independent variables that cannot be included in the SP experimental design without adding too much complexity. There are three groups of questions addressing the impact of Transfers (39A & 40A), Station/Airport availability (39B & 40B), and Access Mode availability (39C through 42C). Respondents will be randomly assigned to one of these three groups and receive that set of questions. This information will be used to estimate the sensitivity of mode choice to transfers, station/airport availability, and access mode availability.

Questions D-1 through D-12 collect demographic information from respondents. This information serves several purposes. It provides a demographic profile of the respondents which can be compared to other information to confirm the survey sample is representative of the NEC. Some of these specific question responses, like income, are used to account for differences in value of time and are included as modifiers of the trip cost to specifically account for the differential sensitivity of travelers’ income levels to travel costs in the estimated model.

Table 1 below summarizes the questions that apply to specific respondent groups.

**TABLE 1
RESPONDENT GROUPS AND QUESTIONS ASKED OF EACH GROUP IN RECRUIT
AND FOLLOW_UP SURVEYS**

Respondent Group	Recruit Survey Questions	Follow-up Survey Questions
1. Did not travel in study areas – not qualified	C.1 through S.3a and D-1 through D-12	None
2. Qualified based on travel areas but refused to participate in follow-up survey during recruit survey	C.1 through S.4 and D-1 through D-12	None
3. Qualified, willing to do follow-up survey and opted for Internet follow-up	C.1 through “Email Recruit” section	Entire follow-up survey (1 through D-12)
4. Qualified, willing to do follow-up survey and opted for mail/phone follow-up	C.1 through 7D	All questions in follow-up survey except: 1, 2, 7, 12A, 13A, 12B, 13B, 12C, 25A, 26A, 25B, 26B, 25C, 26C, 25D, 26D (Note: these excluded questions were already asked of this group in Recruit Survey)

3. Describe whether, and to what extent, the collection of information involves the use of automated, electronic, mechanical or other technological collection techniques or other information technology. Also describe any consideration of using information technology to reduce burden.

The data during phase 1 of data collection (the Recruitment Phase) will be collected electronically through the use of Computer Assisted Telephone Interviewing (CATI). The CATI system allows a computer to perform a number of functions prone to error when done manually by interviewers, including:

- Providing correct question sequence;
- Automatically executing skip patterns based on prior answers to questions (which decreases overall interview time and consequently the burden on respondents);
- Recalling answers to prior questions and displaying the information in the text of later questions;
- Providing random rotation of specified questions or response categories (to avoid bias);
- Ensuring that questions cannot be skipped; and
- Rejecting invalid responses or data entries.

The CATI system lists questions and corresponding response categories automatically on the screen, eliminating the need for interviewers to track skip patterns and flip pages. Moreover, the interviewers enter responses directly from their keyboards, and the information is automatically recorded in the computer's memory.

CATI systems typically include safeguards to reduce interviewer error in direct key entry of survey responses. CATI also allows the computer to perform a number of critical assurance routines that are monitored by survey supervisors, including tracking average interview length, refusal rate, and termination rate by interviewer; and performing consistency checks for inappropriate combination of answers.

The data collected during phase 2 of the study will be collected primarily by Web (self-administered). The Web incorporates the programming logic which is utilized by CATI to ensure responses are validated within a given range and skip patterns are followed correctly. The respondent will receive the link to the web site via e-mail so he/she simply has to click on the link in order to access the survey. For most answers the respondent simply has to click on the response category and click the next. The programming logic will determine the path the respondent takes through the survey based on his/her responses. For those respondents without Web access, we will mail a packet of travel scenarios to them and follow up with a phone survey. Reminders will be done by email or phone. Up to five weekly reminders will be attempted among those who do not respond to the initial contact attempts in phase 2.

4. Describe efforts to identify duplication. Show specifically why any similar information, already available cannot be used or modified for use for the purposes described in Item 2 above.

Presently there is no information on consumer preference available regarding the feasibility of High Speed Rail in the Northeast Corridor. To FRA's knowledge, data do not exist anywhere dealing with the specific High Speed Rail project that is being proposed, and certainly no data which will offer the degree of specificity which will be obtained from a data collection effort of this size and scope. A sample of up to 15,000 follow-up respondents (expected from 22,500 recruits) will allow the FRA to analyze the data by subgroups and regions which will not be possible using smaller sample sizes.

5. If the collection of information involves small businesses or other small entities, describe the methods used to minimize burden.

The collection of information involves randomly selected individuals, not small businesses.

6. Describe the consequence to Federal program or policy activities if the collection is not conducted or is conducted less frequently, as well as any technical or legal obstacles to reducing burden.

Without this information from travel surveys, the new NEC travel demand forecasting model cannot be developed as described above. The new model has been designed to address the needs of NEC FUTURE by providing a basis forecasting response to a broad range of intercity and regional modes of travel throughout the Northeast. The survey program has been designed specifically to support the development of a new NEC forecasting model. Without this new model, there will be inadequate travel demand forecasting capabilities to fully address the range of future intercity and commuter services to be evaluated as part of the NEC FUTURE Program. As described above, none of the existing models address the availability of both intercity and commuter services, which do compete with each other in some NEC markets. Instead, these transportation modes are currently addressed by separate models maintained by Amtrak and the regional agencies respectively.

7. Explain any special circumstances that require the collection to be conducted in a manner inconsistent with the guidelines set forth in 5 CFR 1320.6.

No special circumstances require the collection to be conducted in a manner inconsistent with the guidelines in 5 CFR 1320.6.

8. Provide a copy of the FEDERAL REGISTER document soliciting comments on extending the collection of information, a summary of all public comments responding to the notice, and a description of the agency's actions in response to the comments. Describe efforts to consult with persons outside the agency to obtain their views.

As required by the Paperwork Reduction Act of 1995, FRA published a 60-day notice in the Federal Register on September 20, 12, soliciting comments on the particular collection of information. See 76 FR 7116. FRA received no comments from the railroad industry, the general public, or any other interested party regarding this information collection.

On November 21, 2012, FRA published in the Federal Register a 30-day Notice Regarding Collection Information from the Public to Determine Current Intercity and Regional Behavior of Northeast Resident. See 77 FR 225.

9. Explain any decision to provide any payment or gift to respondents, other than remuneration of contractors or grantees.

Respondents will receive a \$5 check for their participation in the survey. The \$5 check is mentioned during the recruitment phase and awarded after the respondent completes the follow-up survey online or through a follow-up phone survey. The \$5 is a token of appreciation for the respondents' effort and will help maximize response rates to the phase 2 follow-up survey. A \$5 check should be sufficient based upon the outcome of two comparable two-phase studies conducted for Amtrak where participants were recruited by phone and asked to complete the follow-up by web (or be willing to complete the follow-up by phone when web access was not

possible) for a token of appreciation of \$5. In 2011, the follow-up rate was 67% using a 9 minute follow-up survey instrument and up to 3 reminders for completion. In 2007, the follow-up rate was 60% using a 10 minute follow-up survey instrument and up to 3 reminders for completion. In this study, the follow-up survey instrument is 15 minutes with more reminders - up to 5 reminders for completion. Up to 500 respondents who responded during phase 1 but failed to respond to the phase 2 follow-up survey will be targeted for a non-response follow-up (NRFU) survey. These respondents will receive a \$20 check as token of appreciation for their time and to help maximize the NRFU survey.

10. Describe any assurance of confidentiality provided to respondents

In the survey’s introduction, respondents are informed that participation is voluntary, and their answers will be kept private and will be used only for statistical purposes. The only personal information which will be collected will be name, and address, so the \$5 check can be mailed to the responding household. Name and address, along with phone number will be stripped from the data file that the FRA will receive.

11. Provide additional justification for any questions of a sensitive nature, such as sexual behavior or attitudes, religious beliefs, and other matters that are commonly considered private.

The survey does not contain any questions related to matters that are commonly considered sensitive or private. The survey questions are directed at consumer preference for traveling in the Northeast Corridor.

12. Provide estimates of the hour burden of the collection of information on the respondents.

Data collection will involve two phases. Phase 1 is the recruitment phase and we will attempt to recruit 22,500 respondents in order to yield 15,000 who will complete the phase 2 survey. The phase 1 survey instrument (FRA Form F 222) is time tested to average 5 minutes in length. Phase 2 survey web instrument (FRA Form F 222a) is time tested to take the respondents 15 minutes to complete. For those respondents without Web access they will be sent a packet of travel scenarios and followed up with a phone survey; the follow-up phone version is also expected to take 15 minutes to complete. Before completing the phase 2 follow-up survey instrument each respondent will have completed the CATI phone survey in phase 1, however not all respondents who complete the CATI phone survey will complete the phase 2 follow-up survey. Additionally, there will be a 5 minute non-response follow-up (NRFU) survey with a sub-sample (n=500) of households that responded during phase 1, but failed to respond to the phase 2 interview. In Table 2, below, we show the maximum expected number of responses and calculate the total burden hours based on these assumptions.

**TABLE 2
ESTIMATED BURDEN HOURS**

Phase	Minutes	Respondents	Burden Hours
Phase 1 (CATI)	5 min.	22,500	1,875
Phase 2 (Web/Mail-CATI)	15 min.	15,000	3,750
Non-Response Follow-Up (CATI)	5 min.	500	42

Total			5,667
--------------	--	--	--------------

13. Provide an estimate of the total annual cost to the respondents or record keepers resulting from the collection of information.

There are no record keeping or reporting costs to respondents. Respondents will be contacted randomly, and asked questions about their recent travel as well as preferences for travel throughout the Northeast Corridor. All responses are provided at the time of the survey while viewing the travel scenarios; no prior preparation is needed. Each respondent only participates once in each phase of data collection. Thus there is no preparation of data required or expected of respondents. Respondents do not incur: (a) capital and start up costs, or (b) operation, maintenance, and purchase costs as a result of participating in the survey.

14. Provide estimates of the annualized cost to the Federal Government.

The estimated cost to the government for conducting the survey is as follows:

Number of completed interviews	22,500 recruit/15,000 follow-up/500 non-response follow-up
Total estimated cost of conducting survey	\$1,185M
Cost per completed follow-up interview	\$79.00

This estimate is based on the total cost of the awarded survey contract divided by the specified number of completed survey interviews. Costs of conducting the survey will be concentrated within a one year period, making the annual cost to the government the full \$1,185M.

The table below presents the estimated cost for federal oversight of this project.

Position	Grade/Step	Cost per Hour (Pay and Benefits)	Hours	Cost
Project Manager	15/6	\$ 110	8	\$ 880
Deputy Project Manager	13/6	\$ 75	16	\$ 1,200
Technical Reviewer	13/6	\$ 75	40	\$ 3,000
Total				\$ 5,080

Thus the total annual cost to the government including both conducting the survey and federal oversight will be \$1,190M.

15. Explain the reasons for any program changes or adjustments in Items 13 or 14 of the OMB 83-I.

This is a new information collection. As such, it requires a program change to add the estimated 5,625 hours for the new information collection to existing burden.

16. For collection of information whose results will be published, outline plans for tabulation and publication.

Weighted frequencies will be computed for each of the questions in the survey. Cross-tabular analyses of the survey data by population subgroups and key analytical variables will also be conducted. The key analysis activity, the model development itself, will rely on a maximum value estimation procedure for estimating parameter values from the survey data. Simply stated, a maximum likelihood estimator is the value of the parameters, on the independent variables, for which the observed choice, the dependent variable, is most likely to have occurred. Several statistical tests are used to evaluate the significance of the estimators, including “t” tests of individual parameter significance and likelihood ratio tests used to evaluate a set of parameters.

Findings will be disseminated through internal briefings to FRA managers who must make strategic planning decisions, as well as through printed technical reports distributed to stakeholders at the national, State and local levels.

17. If seeking approval to not display the expiration date for OMB approval of the information collection, explain the reasons that display would be inappropriate.

FRA will display the expiration date for OMB approval on the Web survey instrument and the hard copy survey instrument. The interviewer will provide the OMB number, if requested by the respondent, during Phase I data collection.

18. Explain each exception to the certification statement identified in Item 19, “Certification for Paperwork Reduction Act Submissions” of the OMB Form 83-I.

No exceptions to the certification are made.