

## SAMPLING COMPONENTS

To direct the content of the NIRPC Household Travel Inventory, the sampling plan includes the following components: (1) Population definition, (2) Sampling frame, (3) Sampling method, (4) Stratification Plan, and (5) Calculation of sample size. Each of these is discussed below.

### Survey Universe

The survey population will represent all households residing in the NIRPC study area, currently defined by three Indiana counties: Lake, Porter, and La Porte. The population or the study universe is thus comprised of over 250,000 households, distributed across the counties as shown in Table 1.

**TABLE 1: COUNTIES IN THE STUDY AREA**

County	Total Households	% of Total Households in Study Area
Lake	181,589	65.5%
Porter	41,086	14.8%
La Porte	54,721	19.7%

Source: Census 2000.

### Sampling Frame

A dual frame sampling approach will be used for this study. Dual Frame sampling combines the strengths of Random Digit Dialing (RDD) and Directory/Address-based samples. Specifically, dual frame sampling combines the 100% coverage provided by RDD frame of the listed and unlisted households with landline telephones, and the coverage of households with no telephones or cell-phone only households provided by address-based frame. Thus, a dual frame sample provides a comprehensive coverage of the study area, more accuracy in locating the survey universe and higher efficiency in contacting the households in the survey universe.

A dual frame sampling approach is necessitated in the NIRPC region due to the growth in cellular only households, particularly among the younger single-person households as well as the low-income and immigrant populations in the region. The address-based portion of the sampling frame provides access to these populations. An address-based sample is a random sample of all residential addresses that receive US Mail delivery. This sample may or may not have the resident's name or phone number. Its main advantage is its reach into population groups that typically participate at lower-than-average levels, largely due to coverage bias (i.e., most of these households do not have traditional telephone service). The main disadvantage is that the recruitment of households without traditional telephone service is passive – respondents must open the mailing and respond via mail, web, or telephone (calling in) in order to participate in the survey. The use of address-based sample requires additional mailings of reminder postcards and attractive, eye-catching packaging of the initial mailing of survey information.

The RDD portion of the sampling frame includes a random sample of all residential telephone numbers in the region and provides access to the majority of residents in the region. RDD sample includes both “listed” and “unlisted” sample. The “listed” sample includes all telephone numbers for which the name and address associated with that telephone number are known. The “unlisted” sample is comprised of telephone numbers for which there is no associated name or address. The advantage to RDD sample is its efficiency in conducting the survey effort – being able to directly reach households and secure their participation in the survey in a direct and active approach. The disadvantages of RDD sample are that it does not include households with non-traditional telephone service (i.e., cellular-only service) and, for the unlisted sample, the geographic location of the household is not known until after the household has been contacted and agrees to participate.

As indicated by reviewing the advantages and disadvantages of each sample type, it can be seen that a dual sampling frame (RDD and address-based) provides the greatest reach in terms of including all residents in the study area, which supports the overall sampling objective of achieving a mix of residents such that model validity is achieved.

## **Sampling Method**

In this study, we will employ a stratified probability sample of households. Stratified probability sampling is a common technique for household travel inventories as it ensures high levels of coverage, accuracy, and efficiency compared to non-probability samples. A strictly random sample from throughout the study area would result in under-representation of households with specific travel characteristics, thereby reducing the anticipated model validity. By stratifying the sample, survey goals can be allocated to specific portions of the region in order to maximize the inclusion of different travel characteristics. The stratified sampling method thus results in over-samples for some strata to ensure that we capture the diversity of the population according to specific geographic and behavioral factors affecting travel activity in the NIRPC study area. Thus, within strata and frame, households will be selected with equal probabilities but the combined sample (across strata and frames) will comprise an unequal probability sample of households.

## **Sample Stratification**

As NIRPC is currently updating its regional travel demand model, it is important to capture the behaviors of interest as part of this household travel inventory. A sampling strategy to maximize the capture of behaviors of interest is therefore needed. The following is a description of a recommended strategy that should yield unbiased results, with an adequate representation of the behaviors of interest by market segment desired for modeling and policy analysis.

In particular, a census tract stratification variable was developed that is a compilation of key measures highly relevant to and compliant with the stratification objectives of the survey. This stratification variable takes into account the environment in which travel takes place (defined by population and job densities – with higher densities reflecting the more urbanized portions of the region) and the level of transit services (both bus and rail) available. These measures include:

- Population density – inhabitants per square mile
- Job density – jobs per square mile

- Train stations within tract
- Bus Stops within tract

Stratification that considers the environment in which travel takes place is highly relevant to the development of a valid model. Studies had shown that levels of non-motorized travel are higher in higher density areas, as there are more destinations within walking or biking distance. In addition, travel in the lower density areas tends to be predominantly by auto and include higher proportions of trip chaining. In addition, the types of households found in the different settings are related to differences in travel patterns as well. Households with children tend to settle in the lower density areas (suburban housing) while those comprised only of workers might be found closer to the areas with high densities of jobs. To capture the environment of travel, two standardized measures reflecting population and job densities were developed for each census tract, proving measures of 0 to 100 for each indicator. To minimize the skew associated with resulting low means and standard deviations, the measures were “capped” at the 95<sup>th</sup> percentile unstandardized value (thus all values of 95% or above were assigned a value of 100 and the remaining tracts were scaled accordingly from 0 to 100.

In a region with a full range of transportation options (from non-motorized travel to auto travel to several transit options), model validity requires sufficient samples from travelers using each mode. Transit service in the region includes both bus (Gary Public Transit, East Chicago Transit, Hammond Transit and Michigan City Municipal Coach) and rail (South Shore Railroad). To identify the availability of the different transit options, the transit bus stops and rail stations were imported into VISUM. For each census tract, an access to transit variable was created. In particular, access to transit was determined by calculating the fraction of the area of the census tract intersected by buffers of size 0.5 miles around bus stops and 15 mile around the rail stations.

Thus, each census tract in the region received values of 0 to 100 in each of the four variables of interest. The following table shows the four variables standardized with the capping of the maximum value at the 95<sup>th</sup> percentile.

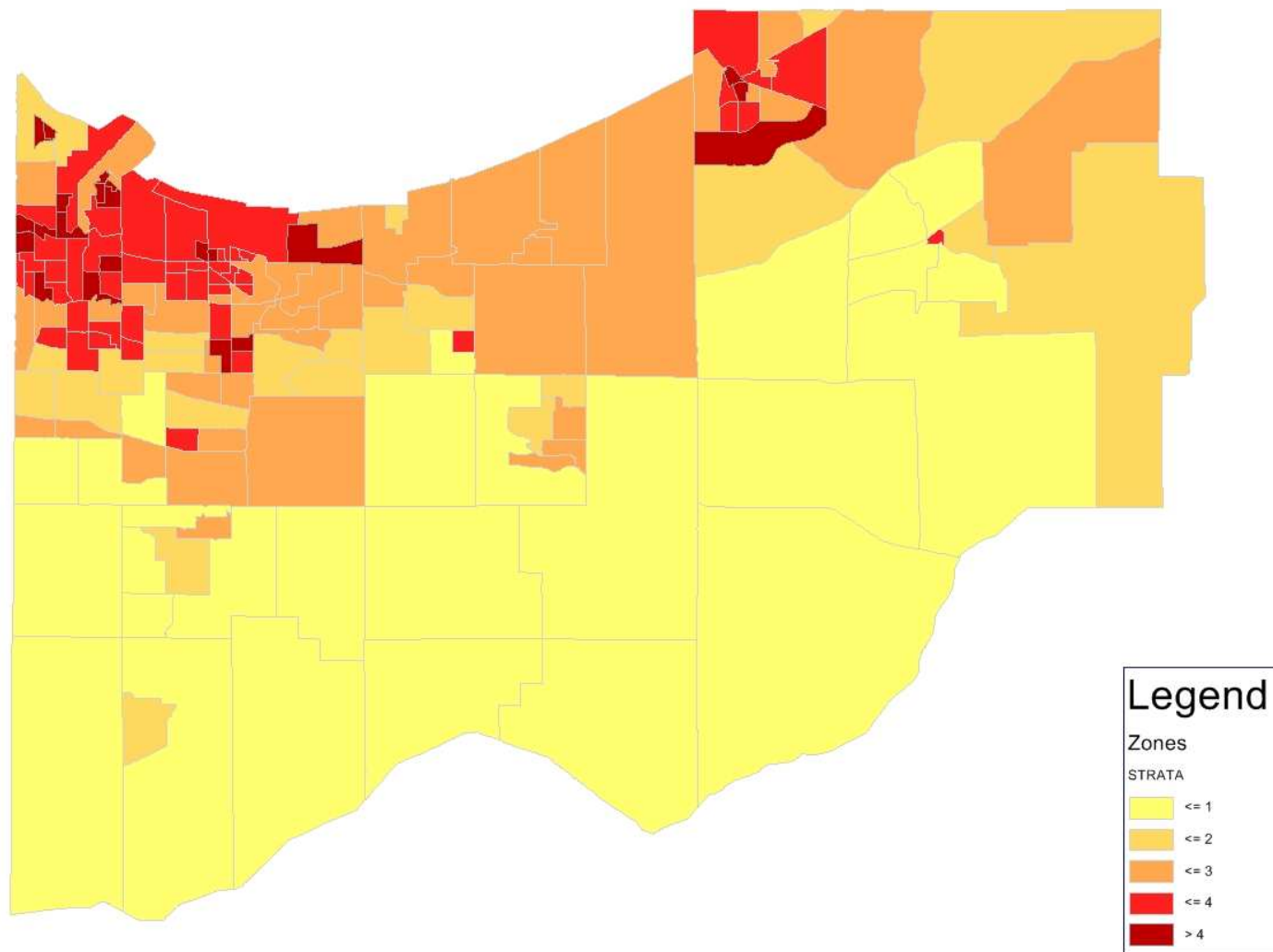
**TABLE 2: STRATIFICATION VARIABLES**

Variable	Minimum	Maximum	Mean	St. Dev.
Population density	0.4	100	33.4	29.0
Job density	0.3	100	34.2	28.5
Access to bus service	0	100	21.1	30.6
Access to train service	0	100	16.2	25.4

Using the standardized scales, an overall density scale that combined population and job density with equal weights was created and standardized to the 0 to 100 scale. Similarly, an overall service scale to combine the level of bus transit service and the access to rail transit service scales was also created. Finally, a final stratification index was assigned to each census tract that reflected the combined influence of both the overall density scale and the overall service scale.

This final stratification index was then divided into five categories, which were then color-coded mapped (see Figure 1). The resulting index has five levels, reflecting the combined influence of densities and transit service access, with level 1 having the lowest levels of densities and transit service and level 5 having the highest. As to be expected, the highest levels are concentrated primarily in the urban core of Lake County.

FIGURE 1: STRATIFICATION OF NIRPC STUDY AREA



## SAMPLE SIZE AND DISTRIBUTION

The final sample size of 3,750 surveys includes 1,875 surveys where households will record travel details for a 24-hour period and 1,875 surveys where households will record travel details for a 48-hour period. These samples will be drawn from the region in proportion to the population within each of the 5 stratification levels. Table 3 shows the distribution of surveys across the stratification levels, while Table 4 shows the resulting number of surveys for each county.

**TABLE 3: SURVEY GOALS BY STRATA**

Strata	Total Households	% of Total Households	# of Surveys	% of Surveys
1	48,142	17.4%	651	17.4%
2	47,476	17.1%	643	17.1%
3	83,561	30.1%	1,128	30.1%
4	64,910	23.4%	876	23.4%
5	33,307	12.0%	452	12.1%
Total	277,396	100%	3,750	100%

**TABLE 4: SURVEY GOALS BY COUNTY**

County	Total Households	% of Total Households	# of Surveys	% of Surveys
Lake	181,589	65.5%	2,455	65.5%
Porter	41,086	14.8%	555	14.8%
La Porte	54,721	19.7%	740	19.7%
Total	277,396	100%	3,750	100%

## Sample Coverage of Specific Travel Patterns

Of particular interest in this survey is travel by specific population subgroups. In this section, the relative level of inclusion of these groups is reviewed. This includes estimated surveys from low-income, minority, and young residents (age 18-24) in the region by strata and county.

### Low Income Travelers

Low-income travelers are important to the survey effort because they typically have lower vehicle ownership rates and often rely on the public transit system for most of their transportation needs. Assuming a proportionate sample, the stratification should yield approximately 1,015 surveys of low-income travelers or 27% of all surveys. We anticipate higher levels of non-response among these respondents as compared to the general population. While they will be tagged for incentives, we also anticipate the need for more address-based sample and other focused data collection techniques. Given past experience, we anticipate a minimum of 100 surveys from low-income households in each strata, for a total of at least 500 surveys (or 13% of the sample).

**TABLE 5: SURVEY GOALS BY INCOME BY STRATA**

Strata	Total Households	Total Households with incomes < \$25,000	% Low Income HH of All Households	Total Surveys for Strata	# of Surveys for incomes < \$25,000
1	48,142	8,598	17.9%	651	116
2	47,476	9,470	19.9%	643	128
3	83,561	21,563	25.8%	1,128	292
4	64,910	22,693	35.0%	876	306
5	33,307	12,721	38.2%	452	173
Total	277,396	75,045	27.1%	3,750	1,015

**TABLE 6: SURVEY GOALS BY INCOME BY COUNTY**

County	Total Households	Total Households with incomes < \$25,000	% Low Income HH of All Households	Total Surveys for Strata	# of Surveys for incomes < \$25,000
Lake	181,589	52,876	29.1%	2,455	715
Porter	41,086	11,311	27.5%	555	153
La Porte	54,721	10,858	19.8%	740	147
Total	277,396	75,045	27.1%	3,750	1015

## African American Travelers

African American travelers are important to the survey effort because their transportation needs have been found to vary based on income levels and where the household is located. Assuming a proportionate sample, the stratification should yield approximately 660 surveys of African American travelers or 18% of all surveys. We anticipate higher levels of non-response among these respondents as compared to the general population. While they will be tagged for incentives, we also anticipate the need for more address-based sample and other focused data collection techniques. Given past experience, we anticipate the final inventory will include at least 300 surveys from African American households at a minimum (about half the proportionate goals for each strata).

**TABLE 7: SURVEY GOALS BY MINORITY STATUS BY STRATA**

Strata	Total Households	Total African American HH	% Afr Am HH of All Households	Total Surveys for Strata	# of Surveys for Minority HH
1	48,142	268	0.6%	651	4
2	47,476	842	1.8%	643	11
3	83,561	10,798	12.9%	1,128	146
4	64,910	25,234	38.9%	876	341
5	33,307	11,643	35.0%	452	158
Total	277,396	48,785	17.6%	3,750	660

**TABLE 8: SURVEY GOALS BY MINORITY STATUS BY COUNTY**

County	Total Households	Total African American HH	% Afr Am HH of All Households	Total Surveys for Strata	# of Surveys for Minority HH
Lake	181,589	45,242	24.9%	2,455	612
Porter	41,086	3,033	7.4%	555	41
La Porte	54,721	510	0.9%	740	7
Total	277,396	48,785	17.6%	3,750	660

## Young Travelers

Young travelers, defined in the census data as households with the age of the “householder” (main reference person) being between 15 and 24 years old, are important to the survey effort because their participation in surveys tends to be lower than average. Assuming a proportionate sample, the stratification should yield approximately 156 surveys of young households or 4% of all surveys. We anticipate higher levels of non-response among these respondents as compared to the general population. While they will be tagged for incentives, we also anticipate the need for more address-based sample and other focused data collection techniques.

**TABLE 9: SURVEY GOALS BY YOUNG RESIDENCE STATUS BY STRATA**

Strata	Total Households	Total Young HH	% Young HH of All Households	Total Surveys for Strata	# of Surveys for Young HH
1	48,142	1,070	2.2%	651	14
2	47,476	1,537	3.2%	643	21
3	83,561	3,839	4.6%	1,128	52
4	64,910	3,025	4.7%	876	41
5	33,307	2,090	6.3%	452	28
Total	277,396	11,561	4.2%	3,750	156

**TABLE 10: SURVEY GOALS BY YOUNG RESIDENCE STATUS BY COUNTY**

County	Total Households	Total Young HH	% Young HH of All Households	Total Surveys for Strata	# of Surveys for Young HH
Lake	181,589	7,511	4.1%	2,455	102
Porter	41,086	1,731	4.2%	555	23
La Porte	54,721	2,319	4.2%	740	31
Total	277,396	11,561	4.2%	3,750	156