NuStats

3006 Bee Caves Road
Suite A-300
Austin, Texas 78746

Stacey Bricka
512.306.9065 ext. 2240
Fax: 512.306.9077
e-mail: sbricka@nustats.com

MEMORANDUM

To: Kermit Wies, CMAP **Date:** January 3, 2007

From: Bricka, Sen, Arce - NuStats **CC:** Project Team

Re: Chicago Regional Household Travel Inventory – Sampling Plan

The purpose of this memo is to summarize the proposed sampling approach work plan for the Chicago Regional Household Travel Inventory. This includes an introduction to the key sampling issues, a review of white paper and related issues discussed at the expert panel meetings, the presentation and review of a stratification plan, and the operational details necessary to execute the sampling plan.

INTRODUCTION

The Chicago Regional Household Travel Inventory is a comprehensive study of demographic and travel behavior characteristics of residents in the greater Chicago area. The study universe is defined as households residing in Cook, DuPage, Grundy, Kane, Kendall, Lake, McHenry, and Will Counties (all in Illinois). The objective of the Chicago Regional Household Travel Inventory is to provide data for the continuing development and refinement of the Chicago regional travel demand forecast models. Thus, from a modeling standpoint, it is important that the data reflects the full diversity of the behavioral determinants of travel activity and provide for a statistically valid model. This technical memo discusses the sampling approach for collecting the data for use in conventional and new generation modeling efforts.

Sampling is a consideration made to draw inferences about the population based upon the inferences from the sample. Sampling a population, rather than conducting a census on the study population, saves time and money and hence results in an effective use of the resources. Also, it is a more cost effective approach compared to data collected from a full population census. Ideally, developing a statistically reliable sample includes identification of the survey population or the universe, identification of the sampling frame, designation of sampling stratification, calculation of sample size, and estimation of necessary resources. These were discussed in the white paper on Sampling and are summarized again here.

The purpose of this document is to outline the sampling approach that will yield the travel behavior details of regional households such that a valid model can be attained. This includes detailing the components of the sampling plan as well as specifying how the sample will be distributed across the region. The remainder of this document focuses on those details.

SAMPLING COMPONENTS

To direct the content of the Chicago Regional 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 CMAP modeling area, currently defined by eight Illinois counties: Cook, DuPage, Grundy, Kane, Kendall, Lake, McHenry, and Will. The population or the study universe is thus comprised of over 2.9 million 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
Cook	1,974,181	67.1%
DuPage	325,601	11.1%
Grundy	14,293	0.5%
Kane	133,901	4.6%
Kendall	18,789	0.6%
Lake	216,297	7.4%
McHenry	89,403	3.0%
Will	167,542	5.7%

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 Chicago 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-then-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 CMAP 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 activity- and tour-based models are considered for future model development in the Chicago region, 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
- CTA train stations within tract
- METRA stations within tract
- CTA bus miles of service in tract
- PACE bus miles of service in 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 95th 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 (Pace and CTA) and rail (CTA and Metra), with significant overlap of services in the census tracts nearest to downtown Chicago and very little overlap in the outlying census tracts. To identify the availability of the different transit options, the transit bus routes and rail stations were imported into VISUM. For each census tract, two variables were created: a level of service variable and an access to transit variable. These were created as follows:

- Level of transit service variable was created by calculating the length of the Pace and CTA bus lines that are located within each census tract. This measure of length was divided by the size of the census tract to provide a level of service measure for each census tract reflective of the size of the census tract.
- 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 CTA rail stops and 1 mile around the METRA 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 95th percentile.

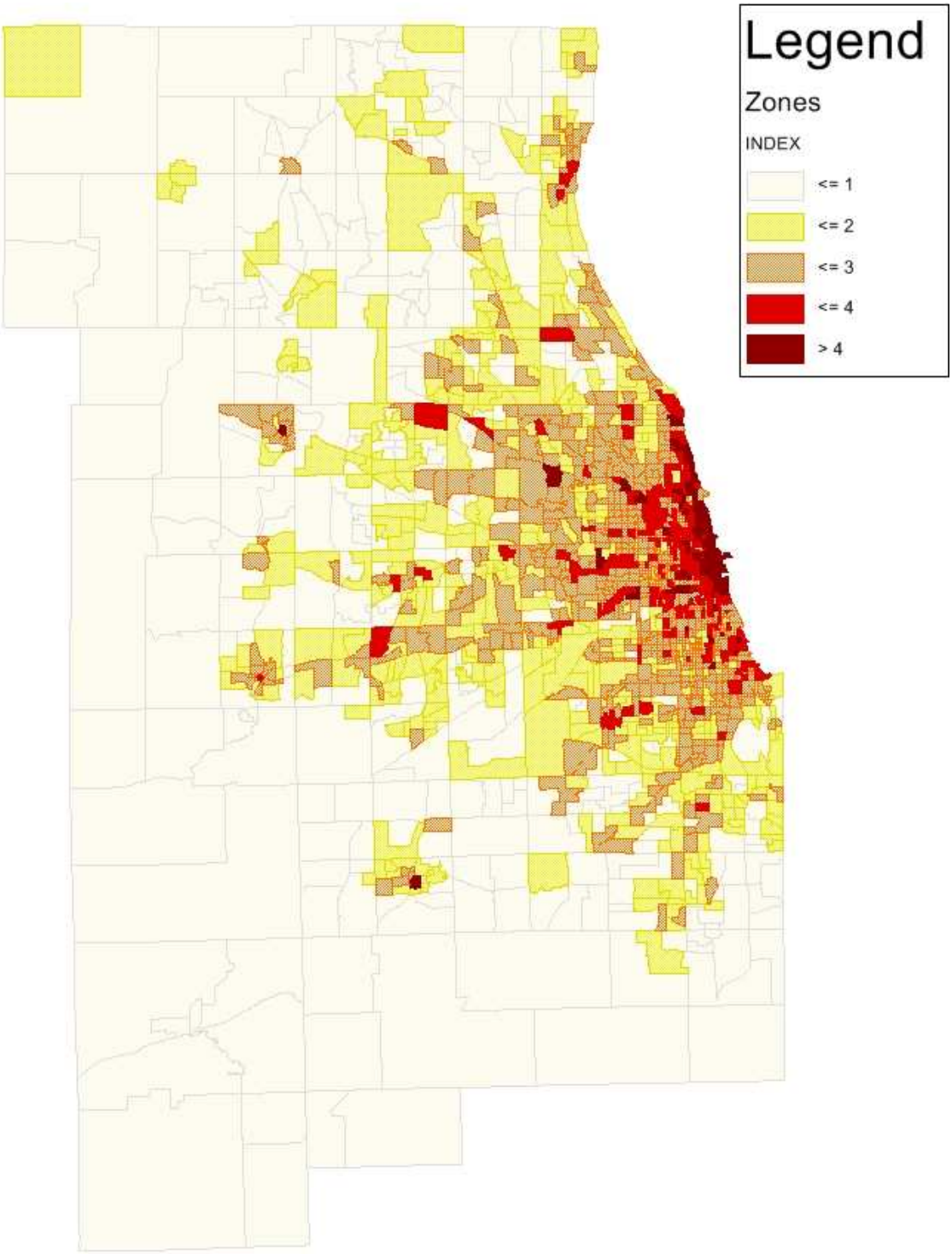
TABLE 2: STRATIFICATION VARIABLES

Variable	Minimum	Maximum	Mean	St. Dev.
Population density	0	100	31.6	28.9
Job density	0	100	26.5	26.3
Level of service	0	100	21.0	19.4
Access to service	0	100	13.3	19.0

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 transit service and the access to 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 availability and 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 with radials that follow the rail service lines out into the surrounding census tracts.

FIGURE 1: STRATIFICATION OF REGION



SAMPLE SIZE AND DISTRIBUTION

The final sample size of 11,600 surveys includes 5,800 surveys where households will record travel details for a 24-hour period and 5,800 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	667,099	22.7%	2,630	22.7%
2	772,894	26.3%	3,049	26.3%
3	900,915	30.6%	3,557	30.6%
4	367,311	12.5%	1,447	12.5%
5	232,485	7.9%	917	7.9%
Total	2,940,704	100%	11,600	100%

TABLE 4: SURVEY GOALS BY COUNTY

County	Total Households	% of Total Households	# of Surveys	% of Surveys
Cook	1,974,181	67.1%	8061	69.5%
DuPage	325,601	11.1%	975	8.4%
Grundy	14,293	0.5%	68	0.6%
Kane	133,901	4.6%	454	3.9%
Kendall	18,789	0.6%	61	0.5%
Lake	216,297	7.4%	1044	9.0%
McHenry	89,403	3.0%	345	3.0%
Will	167,542	5.7%	592	5.1%
Total	167,542	100%	11,600	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 travel "flows" by commuters, as well as estimated surveys from low-income, minority, and young residents (age 18-24) in the region.

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 2,543 surveys of low-income travelers or 22% 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 200 surveys from low-income households in each strata (400 from strata 3), for a total of at least 1200 surveys (or 10% of the sample).

TABLE 6: SURVEY GOALS BY INCOME

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	667,099	78,548	11.8%	2,630	310
2	772,894	137,200	17.8%	3,049	542
3	900,915	241,249	26.8%	3,557	953
4	367,311	121,526	33.1%	1,447	479
5	232,485	65,609	28.2%	917	259
Total	2,940,704	644,132	21.9%	11,600	2,543

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. In addition, they were under-represented in the 1990 survey effort. Assuming a proportionate sample, the stratification should yield approximately 2,090 surveys of low-income 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 1,000 surveys from African American households at a minimum (about half the proportionate goals for each strata).

TABLE 7: SURVEY GOALS BY MINORITY STATUS

Strata	Total Households	Total African American HH	% Afr Am HH of All Households	Total Surveys for Strata	# of Surveys for Minority HH
1	667,099	42,723	6.4%	2,630	168
2	772,894	99,608	12.9%	3,049	393
3	900,915	253,779	28.2%	3,557	1003
4	367,311	98,341	26.8%	1,447	388
5	232,485	34,817	15.0%	917	138
Total	2,940,704	529,268	18.0%	11,600	2,090

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 472 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 8: SURVEY GOALS BY YOUNG RESIDENCE STATUS

Strata	Total Households	Total Young HH	% Young HH of All Households	Total Surveys for Strata	# of Surveys for Young HH
1	667,099	15,090	2.3%	2,630	59
2	772,894	23,492	3.0%	3,049	93
3	900,915	35,404	3.9%	3,557	140
4	367,311	23,722	6.5%	1,447	93
5	232,485	22,068	9.5%	917	87
Total	2,940,704	119,776	4.1%	11,600	472

Travel Flows

For modeling purposes, it is important to ensure the 11,600 surveys capture sufficient travel throughout the region, both within specific geographies (the rural or outlying areas, suburban areas, and the urban areas) as well as between geographies. Using the Census Transportation Planning Package (CTPP) Part 3 data, which focuses on the "flow" of commute trips within and across areas, the number of surveys to be expected from this stratification was estimated for the three geographic areas (Table 9) as well as capturing travel within and between areas (Tables 10 and 11). While the census data focuses only on the work trip, we are assuming that the non-work travel will mimic the work travel in terms of proportions. (This is an assumption that will be monitored weekly and adjustments will be made as necessary). As indicated in Table 11, the smallest number of surveys (proportionately) would be those typifying reverse commuters.

TABLE 9: SURVEY GOALS BY GEOGRAPHIC AREA

Geographic Area	Total Households	Total Workers[1]	Total Workers (Residing and Working in the Study Area)	Total Surveys
Urban	1,780,205	2,246,230	2,459,590	7,501
Suburban	904,300	1,329,450	975,710	3,170
Rural	256,199	304,700	276,270	928
Total	2,940,704	3,880,380	3,711,570	11,600

TABLE 10: TOTAL NUMBER OF WORKERS BY AREA TYPE OF WORK PLACE AND RESIDENCE

		Area Type of Residence			Total Workers by Area Type
		Urban	Suburban	Rural	
Area Type of Work Place	Urban	2,077,800	312,510	69,280	2,459,590
	Suburban	234,820	664,395	76,495	975,710
	Rural	24,895	33,360	218,015	276,270
	Total	2,337,515	1,010,265	363,790	3,711,570

TABLE 11: SURVEY GOALS BY AREA TYPE OF WORK PLACE AND RESIDENCE

		Area Type of Residence			Total Surveys by Area Type
		Urban	Suburban	Rural	
Area Type of Work Place	Urban	6,337	953	211	7,501
	Suburban	763	2,159	249	3,170
	Rural	84	112	733	9,28
	Total	7,184	3,224	1,192	11,600

[1] The total number of workers includes workers that work in the study area but may live outside the study area.