



Chicago Metropolitan
Agency for Planning

233 S. Wacker Drive, Suite 800
Chicago, IL 60606

www.cmap.illinois.gov

voice 312-454-0400

fax 312-454-0411

Weighting the Chicago Regional Household Travel Inventory Survey

With 2005-2007 American Community Survey Data
and an Eleven Zone Geographic System

December, 2009

CMAP Congestion Management Process

Author: Parry Frank

Table of Contents

Table of Contents.....	2
List of Figures.....	2
List of Tables.....	3
DRAFT Weighting the Chicago Regional Household Travel Inventory with 2005-2007 American Community Survey Data and an Eleven Zone Geographic System.....	5
Executive Summary.....	5
Introduction.....	7
Section 1 Characteristics used to Weight the Travel Surveys.....	8
Surveys were separately weighted in eleven zones.....	8
The eleven analysis zones vary in travel characteristics.....	9
Distance traveled per day varies across the region.....	9
Mode choice varies across the region.....	9
Workforce participation varies across the region.....	10
Survey sample rates vary across the region.....	11
The surveys have been weighted based on six household characteristics.....	12
The survey data has been balanced to match the 2005-2007 ACS data.....	12
Advantages and disadvantages of the ACS data.....	13
The survey data is weighted at the household level as opposed to the person level.....	13
The weighting based on the ACS data produces a household weight and a population weight.....	14
Section 2 Raking Control Totals and Weighting Results.....	15
Vehicle availability.....	15
Workers per household.....	16
Household Size (one, two, three, four, five, six or more).....	16
Ethnicity (Hispanic, non-Hispanic).....	17
Race (White, African-American, Other).....	18
Lifecycle.....	19
Section 3 Weighting Verification.....	22
Verifying the effects of weighting of the survey data: A comparison of the weighted survey data to the 2000 Census data.....	22
Conclusion.....	23

List of Figures

FIGURE 1 THE NORTHEASTERN ILLINOIS REGION WAS SEPARATED INTO 11 ANALYSIS ZONES.....	8
FIGURE 2 AGE DISTRIBUTION.....	21

List of Tables

TABLE 1 AVERAGE DAILY MILES OF TRAVEL BY ZONE	9
TABLE 2 CTPP JOURNEY TO WORK MODE-USE RATE FOR THE ELEVEN ANALYSIS ZONES	10
TABLE 3 SHARE OF PEOPLE IN THE SURVEY WHO WORK IN THE ELEVEN ANALYSIS ZONES	10
TABLE 4 DISTRIBUTION OF POPULATION BY SURVEY, CENSUS, AND ACS	11
TABLE 5 SURVEY SAMPLE RATE FOR THE ELEVEN ANALYSIS ZONES	12
TABLE 6 THE POPULATION VARIATION BETWEEN THE 2000 CENSUS AND ACS FOR THE ELEVEN ANALYSIS ZONES	13
TABLE 7 VEHICLE OWNERSHIP. MATCHING HOUSEHOLD SURVEY DATA TO ACS CONTROL TOTALS	15
TABLE 8 WORKERS PER HOUSEHOLD. MATCHING HOUSEHOLD SURVEY DATA TO ACS CONTROL TOTALS	16
TABLE 9 HOUSEHOLD SIZE. MATCHING HOUSEHOLD SURVEY DATA TO ACS CONTROL TOTALS	17
TABLE 10 ETHNICITY. MATCHING HOUSEHOLD SURVEY DATA TO ACS CONTROL TOTALS	17
TABLE 11 RACIAL DISTRIBUTION. MATCHING HOUSEHOLD SURVEY DATA TO ACS CONTROL TOTALS	18
TABLE 12 RELATIONSHIPS BETWEEN AGE AND AVERAGE MILES OF TRAVEL IN THE SURVEY DATA	19
TABLE 13 LIFECYCLE CHARACTERISTICS. MATCHING HOUSEHOLD SURVEY DATA TO ACS CONTROL TOTALS	20
TABLE 14 WORK TRIP MODE SHARES FOR THE 2000 CENSUS, AND THE HOUSEHOLD SURVEY (WEIGHTED AND RAW DATA)	22

Weighting the Chicago Regional Household Travel Inventory with 2005-2007 American Community Survey Data and an Eleven Zone Geographic System

Executive Summary

The Chicago Regional Household Travel Inventory (Travel Tracker Survey) collected travel information during 2007 and 2008 for 23,808 individuals who resided in 10,552 households in the northeastern Illinois region. The surveys have been weighted to represent the travel made by the households and the population in the region.

The mix of people who participated in the Travel Tracker Survey do not exactly match the socio-economic profile for the region, so the weights of individual households were adjusted so that the final modeled composition of the survey data matched the region's characteristics in the available Census data.

The surveys weights were balanced so that the totals for the following characteristics remained consistent with Census estimates: household population, the number of workers in the household, the number of vehicles per adult for one person households and also for larger households, the household's race, the household's ethnicity, and the household's lifecycle status

The region was divided into 11 sub-regions. Each of these areas was weighted separately because travel in each of these sub regions is unique and it is important that the weights of households with similar travel patterns are increased to represent an under sampled characteristic.

This analysis of survey data was weighted to match the 3-year American Community Survey (2005-2007) which surveyed about 2.5% of the households in the region. This data represents 3,027,301 households and 8,366,434 individuals living in non-group quarters households.

The preliminary results show that the weighted survey data has an increase in the share of transit, pedalcyclist, and walking work trips and a decrease in the share of driving work trips, compared to the 2000 census data.

Introduction

This paper summarizes steps that were taken to create a weighting scheme for the household surveys that were collected during the 2008 CMAP household travel survey. Weighting the surveys allows a limited number of surveys to be expanded so that they represent all of the travel in the region. The travel survey has been collected in the region every 10 to 15 years to provide information that is used in transportation modeling for the northeastern Illinois region.

The Chicago Regional Household Travel Inventory (Travel Tracker Survey¹) collected travel information for 23,808 individuals who resided in 10,552 households in the northeastern Illinois region.² The surveys have been weighted to represent the travel made by the households and the population in the region.

The mix of people who participated in the Travel Tracker Survey do not exactly match the socio-economic profile for the region, so the weights of individual households were adjusted so that the final modeled composition of the survey data matched the region's characteristics in the available Census data.

The survey's weights were balanced using Iterative Proportional Fitting, or raking, so that the totals for the following characteristics remained consistent with Census estimates. The household surveys were weighted on household population, the number of workers in the household, the number of vehicles per adult for one person households and also for larger households, the household's race, the household's ethnicity, and the household's lifecycle status (there are eight lifecycle categories and these are based on the age of people, the household size and the presence of preschool and older children).

The region was divided into 11 sub-regions.³ Each of these areas was weighted separately because travel in each of these sub-regions is unique and it is important that the weights of households with similar travel patterns are increased to represent an under-sampled characteristic.

The analysis of survey data was weighted to match the 3-year American Community Survey 2005-2007 (ACS) which surveyed about 2.5% of the households in the region. This data represents 3,027,301 households and 8,366,434 individuals living in non-group quarters households. When using the weights for the Travel Tracker Survey, the WGTHH value should be used when the number of household needs to be correct and the WGTP value should be used if the total population needs to be correct.

¹Report location <http://www.cmap.illinois.gov/TravelTrackerData.aspx>

² Cook County, DuPage County, Grundy County, Kane County, Kendall County, Lake County, McHenry County, and Will County

³ Central Chicago, North Chicago, South Chicago, North Cook County, West Cook County, South Cook County, Lake County, DuPage County, McHenry and Kendall Counties (and Western Kane County, East Kane County, and Will County (and Grundy County

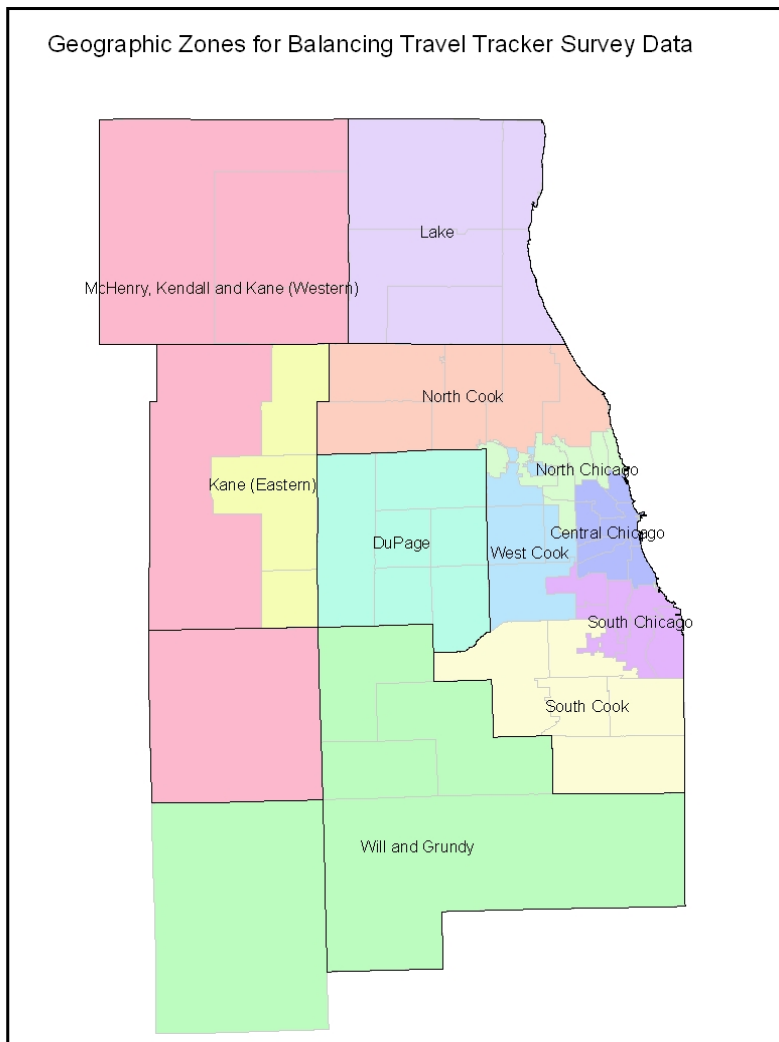
Section 1 Characteristics used to Weight the Travel Surveys

Surveys were separately weighted in eleven zones

The surveys were balanced based on an eleven-zone weighting analysis. This approach has three important characteristics. It is based on 11 geographic sub-areas that are each weighted separately (Shown in Figure 1). Each zone has a separate value for the demographic control numbers. The data that is used for the majority of the weighting is from the 2005-2007 American Community Survey. There are six variables used to rake the surveys.

The zones are based on the census Public Use Micro Sample (PUMS) geography and match up with census tract data. Smaller sub-regional zones are used instead of one regional zone because travel patterns are not uniform across the region, as will be shown in the following sections. Using eleven zones was a compromise between having many smaller zones which would have more similarities in travel patterns and keeping the sample sizes large enough so that the survey data could be balanced and weights remain within a reasonable range.

Figure 1 The Northeastern Illinois Region separated into 11 Analysis Zones



The eleven analysis zones vary in travel characteristics

Distance traveled per day varies across the region

Within the northeastern Illinois region there is a great deal of variation in miles of travel per person. Listed in Table 1 are the survey results for the daily average travel per person for each of the eleven analysis zones. The travel values are quite varied. The average travel per person in the 7 zones with the highest daily travel, have 50% more travel than the average traveler in the zones with the 4 lowest values.

Table 1 Average Daily Miles of Travel by Zone

Sub-Region	Average Travel Distance (Miles) Day 1 of Survey
Central Chicago	11.55
North Chicago	15.13
South Chicago	13.82
North Cook County	18.22
West Cook County	15.84
South Cook County	19.12
Lake County	20.24
DuPage County	21.91
McHenry, Kendall and western Kane Counties	26.66
Eastern Kane County	21.42
Will County and Grundy County	23.89

Mode choice varies across the region

Not only do the areas of the region have different travel patterns, but they also vary greatly in their mode choice for the work trip. The following data (Table 2) is from the 2000 CTPP and shows the mode choice for the journey to work. The City of Chicago has the lowest automobile use, but the city is not uniform in mode use, the southern area of the city uses vehicles for the commute to work at a rate that is about 30% higher than the central area of Chicago. The commute by railroad is twice as prevalent in DuPage County when compared to the western and southern areas. These trends support the eleven zone analysis of the region because households with common travel patterns should be weighted more heavily to compensate for households that are under-sampled.

Table 2 CTPP Journey to Work Mode-use Rate for the Eleven Analysis Zones

CTPP 2000	Auto- Truck	Bus or Trolley Bus	Subway or Elevated	Railroad	Bicycle	Walked	Taxicab	Worked at Home	Percentage of Vehicle users Carpooling
Central Chicago	55.5%	15.9%	11.9%	0.9%	0.8%	9.4%	1.6%	2.9%	26.7%
North Chicago	69.6%	10.5%	10.4%	1.6%	0.5%	3.8%	0.4%	2.2%	19.5%
South Chicago	72.3%	14.7%	5.1%	2.9%	0.1%	2.3%	0.2%	1.8%	20.9%
North Cook County	86.1%	1.0%	1.3%	4.8%	0.4%	2.3%	0.1%	3.5%	10.3%
West Cook County	83.9%	2.6%	3.4%	3.4%	0.3%	3.1%	0.2%	2.3%	14.8%
South Cook County	86.7%	1.6%	1.0%	6.1%	0.1%	1.7%	0.1%	2.0%	11.4%
Lake County	86.5%	0.6%	0.1%	3.7%	0.1%	2.9%	0.2%	4.2%	11.8%
DuPage County	87.1%	0.3%	0.2%	6.1%	0.3%	1.8%	0.1%	3.5%	8.7%
McHenry, Kendall and western Kane Counties	91.0%	0.3%	0.0%	2.5%	0.2%	1.3%	0.0%	4.2%	9.2%
Eastern Kane County	91.4%	0.8%	0.1%	1.8%	0.2%	1.6%	0.1%	3.1%	13.2%
Will County and Grundy County	91.4%	0.4%	0.1%	3.2%	0.1%	1.2%	0.0%	2.9%	9.1%
Region	80.2%	5.0%	3.6%	3.5%	0.3%	3.2%	0.3%	2.9%	13.8%

Workforce participation varies across the region

Participating in the workforce has a significant effect on travel behavior. In the survey data, people who work travel an average of 23.5 miles per day while adults who did not work traveled an average of 12.8 miles. People under the age of 18 who did not work traveled an average of 10.7 miles. The differences in workforce participation are shown in Table 3. For the CTPP data at the zone level, the zone with the highest worker participation rate was almost 50% higher than the lowest zone. In every instance, the workforce participation is higher in the survey data than from 2000 census and the weighted and un-weighted ACS.

Table 3 Share of People in the Survey who Work in the Eleven Analysis Zones

Population	Survey Work Share	Raw ACS	Weighted ACS	2000 Census CTPP
Central Chicago	57.0%	48.4%	47.1%	42.6%
North Chicago	57.2%	47.7%	47.6%	45.4%
South Chicago	43.9%	34.9%	35.6%	35.0%
North Cook County	56.5%	48.4%	48.9%	50.7%
West Cook County	54.3%	45.0%	45.9%	45.3%
South Cook County	52.2%	41.1%	42.9%	45.1%
Lake County	55.1%	45.8%	48.5%	49.3%
DuPage County	57.8%	48.9%	50.1%	51.9%
McHenry, Kendall and western Kane Counties	54.7%	52.2%	50.4%	51.4%
Eastern Kane County	55.4%	48.9%	47.2%	47.3%
Will County and Grundy County	52.1%	45.4%	48.3%	51.8%

Survey sample rates vary across the region

The collection of surveys has not sampled the households of the eleven zones at the same rate (Table 5). Some zones are under sampled and others have been over sampled. At the county level, Cook County might have a reasonable sample rate, but as is evident in Table 4, North Cook County was over-sampled while the City of Chicago was under-sampled. These areas have different travel patterns and should not “balance” each other out in a region-wide weighting analysis.

Table 4 Distribution of Population by Survey, Census, and ACS

Sub-Region	Census	Survey	ACS Raw	ACS Weighted	Average of Census/ ACS Weighted
Central Chicago	13.9%	11.2%	11.4%	12.3%	13.1%
North Chicago	11.0%	9.5%	8.6%	10.1%	10.5%
South Chicago	10.7%	9.1%	8.8%	9.9%	10.3%
North Cook County	12.9%	17.2%	12.9%	12.5%	12.7%
West Cook County	7.9%	8.2%	7.5%	7.7%	7.8%
South Cook County	9.7%	8.0%	9.5%	9.7%	9.7%
Lake County	7.9%	9.9%	9.9%	8.3%	8.1%
DuPage County	11.1%	9.8%	11.2%	10.9%	11.0%
McHenry, Kendall and western Kane Counties	4.3%	5.7%	6.4%	5.4%	4.8%
Eastern Kane County	4.5%	3.9%	4.7%	5.0%	4.8%
Will County and Grundy County	6.2%	7.5%	8.9%	8.2%	7.2%

In the past, surveys were balanced across the entire region. This might result in some undesirable outcomes. If there were an over sample of a characteristic in an outlying area and an under sampling of the characteristic in a highly urban area, the two might balance each other and there would not be any adjustments to the weights for this characteristic. Furthermore, if the characteristic for the region is different than the control value, populations in areas with different travel profiles would have their weights altered to compensate for under/over-representation in a different zone. For example, a single person household in the center of Chicago might have their weight increased to compensate for an under representation of a single person household in an outlying area. To lessen this effect, the Chicago region was broken down into 11 geographic areas and each area was analyzed and weighted individually. As a result of this eleven zone approach, household with characteristics that were underrepresented in specific geographic areas have their weights increased just in that area, so that the travel characteristics of the zone were retained. Table 5 shows the number of surveys from each zone, the number of households in the zone and the sample rate.

Table 5 Survey Sample Rate for the Eleven Analysis Zones

Sub-Region	Suvey Households	Total Households	Sample Rate per 1000 Households
Central Chicago	1,371	413,329	3.32
North Chicago	1,086	321,710	3.38
South Chicago	1,038	279,080	3.72
North Cook County	1,809	398,160	4.54
West Cook County	818	226,392	3.61
South Cook County	864	297,086	2.91
Lake County	988	233,618	4.23
DuPage County	994	335,270	2.96
McHenry, Kendall and western Kane Counties	519	155,409	3.34
Eastern Kane County	386	139,705	2.76
Will County and Grundy County	679	227,596	2.98
Region	10,552	3,027,301	3.49

The surveys have been weighted based on six household characteristics

The Travel Tracker survey, as the name implies, is first and foremost about travel. Elements that affect travel patterns need to be emphasized while those that do not affect travel, such as one’s political affiliation, can be safely excluded. The weighting scheme is based on six variables that are used for balancing the survey data and the eleven geographic regions are weighted individually;

- Household size (one, two, three, four, five, six or more)
- Ethnicity (Hispanic, non-Hispanic)
- Race (White, African-American, Other)
- Eight lifecycle categories are used in the analysis.
- The availability of a vehicle was tracked and weighted
- The numbers of workers per household was tracked and weighted.

The survey data has been balanced to match the 2005-2007 ACS data

The 2005-2007 American Community Survey is the source of the control values for the weighting analysis. It was decided that the most recent demographic data would be used since the population size and structure may change over time (Table 6). For example, the zone covering Will and Grundy Counties has increased in population by 39% between the 2000 census and the 2005-2007 ACS estimates. The Central Chicago zone might have lost nearly 8% of its population during this period.

Table 6 The Population Variation between the 2000 Census and ACS for the Eleven Analysis Zones

Sub-Region	Estimated Population 2005-7 ACS	Population Share 2005-7 ACS	2000 Census Population	2000 Census Population Share	Change from 2000 Census to 2005-7 ACS
Central Chicago	1,042,360	12.3%	1,128,411	13.9%	-7.6%
North Chicago	855,451	10.1%	893,618	11.0%	-4.3%
South Chicago	842,918	9.9%	873,163	10.7%	-3.5%
North Cook	1,066,626	12.5%	1,048,947	12.9%	1.7%
West Cook	651,787	7.7%	642,427	7.9%	1.5%
South Cook	829,324	9.7%	789,168	9.7%	5.1%
Lake	704,102	8.3%	644,356	7.9%	9.3%
DuPage	927,680	10.9%	904,161	11.1%	2.6%
McHenry, Kendall and western Kane Counties	458,851	5.4%	350,425	4.3%	30.9%
Eastern Kane County	428,207	5.0%	368,315	4.5%	16.3%
Will County and Grundy County	699,997	8.2%	502,266	6.2%	39.4%
Grand Total	8,507,303		8,145,257		4.4%
Institutional	140,869		147,712		
Household Pop	8,366,434		7,997,545		

Advantages and disadvantages of the ACS data

The choices for the control data are the 2000 decennial census, the 2005-2007 American Community Survey (ACS), and the 2007 American Community Survey. The 2005-2007 ACS was chosen over the 2007 ACS because of the larger sample size. The 2005 -2007 ACS surveyed nearly 2.5% of the households in the northeastern Illinois region. Compared to the 2000 Census, the ACS has the advantage of providing detailed characteristics on specific households and individuals, and the population information is more current. The census has the advantage of having an exact relationship between the number of households and the population. The ACS is not accurate in this regard.

A second issue with using the ACS as the source for weighting the survey data is that it is itself a survey result that has been weighted. To be used properly, it should be recognized that the ACS represents estimates for the entire population and should only be assumed to be accurate within the standard errors associated with the data. It was decided that it was better to use the newest data available, even though the information was an estimate of current conditions, as opposed to using older data that is known to be not representative of the population totals by county.

The survey data is weighted at the household level as opposed to the person level

The goal of this travel survey and analysis is to aid in modeling travel in the region, not just the number of people. In weighting the survey data, the data could be analyzed at the household level or the person level. The survey data has been weighted at the household level.

Individuals travel, but the household characteristics of the individual affect the person's travel patterns. While it is possible to weight the surveys at the person level, and therefore have the correct number of people in the region, it is more defensible if the households are weighted properly and the resulting populations in the households represent the region's population. Weighting only at the person level

could have some undesirable effects. For example, comparing household and person level data, seven people in a household could use a single vehicle to make a trip, but seven individuals may not be modeled in a single vehicle due to variations in weights and trips. If seven people live in a single household, this is important information and should be retained.

The weighting based on the ACS data produces a household weight and a population weight

The 2005-2007 ACS household file is survey data that has been weighted to represent all of the households in the region. Each household has a weight included that in sum represents the estimate of all of the households in the region. The 2005-2007 ACS person file also has a weight included that in sum represents the estimate of the population in the region. Unfortunately, the household file and the person file are not equivalent. They contain different values for the population. If the weight in the household file is applied to the each person in the household (HHSIZE) the resulting population is different and lower than the estimated population.

The ACS person file is related to the ACS household file in that each person belongs to one of the households in the household file. They are not a different sample set. These 208,567 person records for the region produce a weighted total of 8,365,721 people (not in group quarters/institutions). When the number of households that are used to produce the 8,365,721 people is calculated, the total is 3,135,510 households. This is 108,209 more households than are estimated to exist in the region according to the household file (3,027,301). Based on the household size profile for the region, it takes this many additional households to account for the additional population that is thought to live in the region.

Because of this discrepancy in the ACS data, there are two weights listed for each household in the Travel Tracker dataset. When using the weights, the WGTHH value should be used when the number of household needs to be correct and the WGTP value should be used if the total population needs to be correct.

Section 2 Raking Control Totals and Weighting Results

The eleven zones were each processed and analyzed separately, based on zone specific control totals that were derived from the 2005-2007 ACS data. The surveys were balanced for the following characteristics: household population, the number of workers in the household, the number of vehicles per adult per household, the household's race, the household's ethnicity, and the household's lifecycle status. All variables within each of the eleven zones were balanced (all values within 0.05% of control totals using Iterative Proportional Fitting (raking)), with a maximum weight value for a household of 3,619 and a minimum household weight of 36. The following tables and analysis aggregates the eleven zones into a region-wide total in order to more easily examine the results.

Vehicle availability

Vehicle availability is central to how people choose to travel. If a vehicle is available, a person may choose an alternate mode, but if no vehicle is available, then transit, biking, walking or car pooling are the modes that will most likely be chosen for a trip. If no car is available the trip may not be made.

This variable was broken down into four vehicle availability categories for households with more than one person and two vehicle availability categories for one-person households

- At least one vehicle per adult for households with more than one person
- Between one half and one vehicle per adult for households with more than one person
- Between one-quarter and one-half vehicle per adult for households with more than one person
- Less than one-quarter vehicle per adult for households with more than one person
- At least one vehicle per adult for one-person households
- No vehicle per adult for one-person households

The raw survey data had a too high response rate from single adult households that had a vehicle and the survey under represented the population with more than one person, but very limited access to a vehicle.

After the surveys were raked, the weighted data was a very close match to the control totals (Table 7).

Table 7 Vehicle Ownership. Matching Household Survey Data to ACS Control Totals

	ACS Control Totals	ACS Control Percentage	Survey Final Weighted Totals	Survey Final Percentage	Difference: Total Control - Final Weights	Raw Survey Data	Raw Survey Percentage
One person, no vehicle	212,344	7.0%	212,344	7.0%	1	686	6.5%
More than one person, less than 0.25 cars per adult	156,181	5.2%	156,144	5.2%	37	299	2.8%
More than one person, between 0.25 and 0.5 vehicles per adult	472,756	15.6%	472,672	15.6%	84	1,510	14.3%
More than one person, between 0.5 and 1.0 cars per adult	195,023	6.4%	195,215	6.4%	-192	351	3.3%
More than one person, one or more cars per adult	1,371,669	45.3%	1,371,597	45.3%	71	5,002	47.4%
One person, access to at least one vehicle	619,329	20.5%	619,329	20.5%	-1	2,704	25.6%

Workers per household

The survey data was balanced on the number of worker per household. The categories for workers are: no workers, one worker, two workers or three or more workers in the household.

There is a strong correlation between being in the workforce and an increase in travel. The information for the control values was the summation of persons who worked, for each household in the ACS. Across the region, there was a reasonably good fit between the survey data and the control values. Within zones, there was a larger variation between the survey data and the control. For instance, in the southern Chicago zone, 36.4% of the households did not have any workers while the control data was listed at 31.7% for this category.

After the surveys were raked, the weighted data was a very close match to the control totals (Table 8).

Table 8 Workers per Household. Matching Household Survey Data to ACS Control Totals

	ACS Control Totals	ACS Control Percentage	Survey Final Weighted Totals	Survey Final Percentage	Difference: Total Control - Final Weights	Raw Survey Data	Raw Survey Percentage
No workers in household	643,821	21.3%	643,761	21.3%	60	2,379	22.5%
One worker in houusehold	1,235,542	40.8%	1,235,335	40.8%	207	4,082	38.7%
Two workers in househols	906,602	29.9%	906,677	30.0%	-75	3,528	33.4%
Three or more workers in household	241,336	8.0%	241,528	8.0%	-192	563	5.3%
Total	3,027,301		3,027,301			10,552	

Household Size (one, two, three, four, five, six or more)

The survey data was balanced on the number of people per household. The categories for household size are: one person, two people, three people, four people, five people, and six or more people.

The survey data for the number of people in each household was over-sampled for one and two-person households and under-sampled for the larger households. It should be noted that the distribution of household size data in the ACS does not have enough large households to account for all of the people that are estimated to reside in the region. The Travel Tracker data has very few household with more than six people. The lack of larger households results in a smaller modeled population (0.6% smaller).

After the surveys were raked the weighted data was a very close match to the control totals (Table 9).

Table 9 Household Size. Matching Household Survey Data to ACS Control Totals

	ACS Control Totals	ACS Control Percentage	Survey Final Weighted Totals	Survey Final Percentage	Difference: Total Control - Final Weights	Raw Survey Data	Raw Survey Percentage
One person Households	831,673	27.5%	831,673	27.5%	0	3,390	32.1%
Two person Households	885,182	29.2%	885,182	29.2%	0	3,874	36.7%
Three person Households	475,350	15.7%	475,350	15.7%	0	1,401	13.3%
Four person Households	452,891	15.0%	452,891	15.0%	0	1,211	11.5%
Five person Households	237,764	7.9%	237,764	7.9%	0	492	4.7%
Six or more person Households	144,441	4.8%	144,441	4.8%	0	184	1.7%
Total	3,027,301		3,027,301			10,552	

Ethnicity (Hispanic, non-Hispanic)

The survey data was balanced on the ethnicity of the household. This category tracks if a household is Hispanic.

The region was severely under sampled for Hispanic households. Within the region, all of the sub-zones were also under sampled for Hispanic households. The data for Hispanic ancestry is difficult information to gather consistently. The 2000 Census and 2005-2007 ACS have different values for this data and the question in the Travel Tracker Survey was slightly different from the census and the ACS. The value for the control is an average of the 2000 Census and the ACS data.

After the surveys were raked the weighted data was a very close match to the control totals (Table 10).

Keeping the NA responses separate allows the households to be raked on characteristic that are known and to not mistakenly alter the household weighting on a wrong estimate of the value.

Table 10 Ethnicity. Matching Household Survey Data to ACS Control Totals

	ACS Control Totals	ACS Control Percentage	Survey Final Weighted Totals	Survey Final Percentage	Difference: Total Control - Final Weights	Raw Survey Data	Raw Survey Percentage
Hispanic	560,130	18.5%	560,220	18.5%	-90	585	5.5%
Not Hispanic	2,447,481	80.8%	2,447,392	80.8%	89	9,900	93.8%
NA	19,690	0.7%	19,689	0.7%	1	67	0.6%
					0		0.0%
Total	3,027,301	100.0%	3,027,301			10,552	100.0%

Race (White, Black, Other)

The survey data was balanced on the major race categories of households in the region. The categories are white, black or other.

The following table shows that there was a major over sampling of white households and a severe under sampling of households that are described as “Other” (not black or white). The 2000 Census and 2005-2007 ACS have different values for this data and the question in the Travel Tracker Survey was different from either source. The value for the control is an average of the 2000 Census and the ACS data.

The final weighting has balanced the racial mixes close to the control totals (Table 11).

Keeping the NA responses separate allows the households to be raked on characteristic that are known and to not mistakenly alter the household weighting on a wrong estimate of the value.

Table 11 Racial Distribution. Matching Household Survey Data to ACS Control Totals

	ACS Control Totals	ACS Control Percentage	Survey Final Weighted Totals	Survey Final Percentage	Difference: Total Control - Final Weights	Raw Survey Data	Raw Survey Percentage
White	1,927,971	63.7%	1,927,844	63.7%	127	8,231	78.0%
Black	539,194	17.8%	539,026	17.8%	168	1,516	14.4%
Other	516,343	17.1%	516,617	17.1%	-274	653	6.2%
NA	43,793	1.4%	43,814	1.4%	-21	152	1.4%
Total	3,027,301		3,027,301			10,552	

Lifecycle

The lifecycle variable was used in an attempt to keep the functional types of households consistent even if the age categories cannot be made to match control data.⁴ This variable has been used extensively in the Puget Sound Regional Council (PSRC) survey work⁵ over the past couple of decades.

The lifecycle variable is used instead of an age variable because a household does not have a specific age. There is a general correlation between a person’s age and the amount that they travel as is shown in the following chart. Increases in age are associated with additional travel up to the age category 40-49 and then travel diminishes at older ages. Much of the increase in travel between younger adults and older adults is associated with transporting and caring for children in a household. The lifecycle variable captures this characteristic.

Table 12 Relationships between Age and Average Miles of Travel in the Survey Data

	ACS	2000 Census	Survey	Number of People in the Survey	Total Distance in Survey	Average Miles of Daily Travel from Survey
Age - Under 5	7%	7%	6%	1,359	20,563	15.1
Age 5 to 17	19%	19%	15%	3,490	54,965	15.7
Age 18 to 21	5%	5%	3%	625	14,039	22.5
Age 22 to 29	10%	12%	5%	1,296	37,529	29.0
Age 30 to 39	14%	16%	12%	2,787	88,976	31.9
Age 40 to 49	15%	15%	15%	3,506	120,491	34.4
Age 50 to 64	18%	14%	25%	5,985	185,363	31.0
Age 65 and Higher	12%	11%	20%	4,760	91,546	19.2

The Lifecycle variable details the presence or absence of preschool children and school aged children since these have a significant impact on household travel. The categories are as follows: eight distinct life-cycle categories: household with children under age 6; household with children age 6-18; household with a single adult under age 35; household with a single adult age 35-64; household with a single adult age 65 and older; household with 2 or more adults: oldest under 35; household with 2 or more adults: oldest age 35-64; and household with 2 or more adults: oldest age 65 or older

In order to create a lifecycle value for households, the raw survey data was analyzed and individuals who withheld their age, had the age category estimated based age related information for that they did provide. The age related information was contained in fields such as school locations or names, professions, work status, and presence of young children. This data helped to place individuals into one of the five age categories and allowed for the lifecycle categories to be estimated.

⁴ Households have a mix of people with different ages, not a single age for the household, and weighting the households so that the age distribution of the region is achieved is not useful for this analysis because it would unbalance many of the other variables.

⁵The PSRC has used the 2005-2007 ACS to balance their most recent household travel survey For more information see http://www.psrc.org/assets/2128/09-44_Household_Travel_Surveys.pdf

Across the region, households with preschool children were under sampled in the survey as were households with school-aged children. In general, younger people were under-sampled and older people were over-sampled.

After the surveys were raked the weighted data was a very close match to the control totals (Table 13).

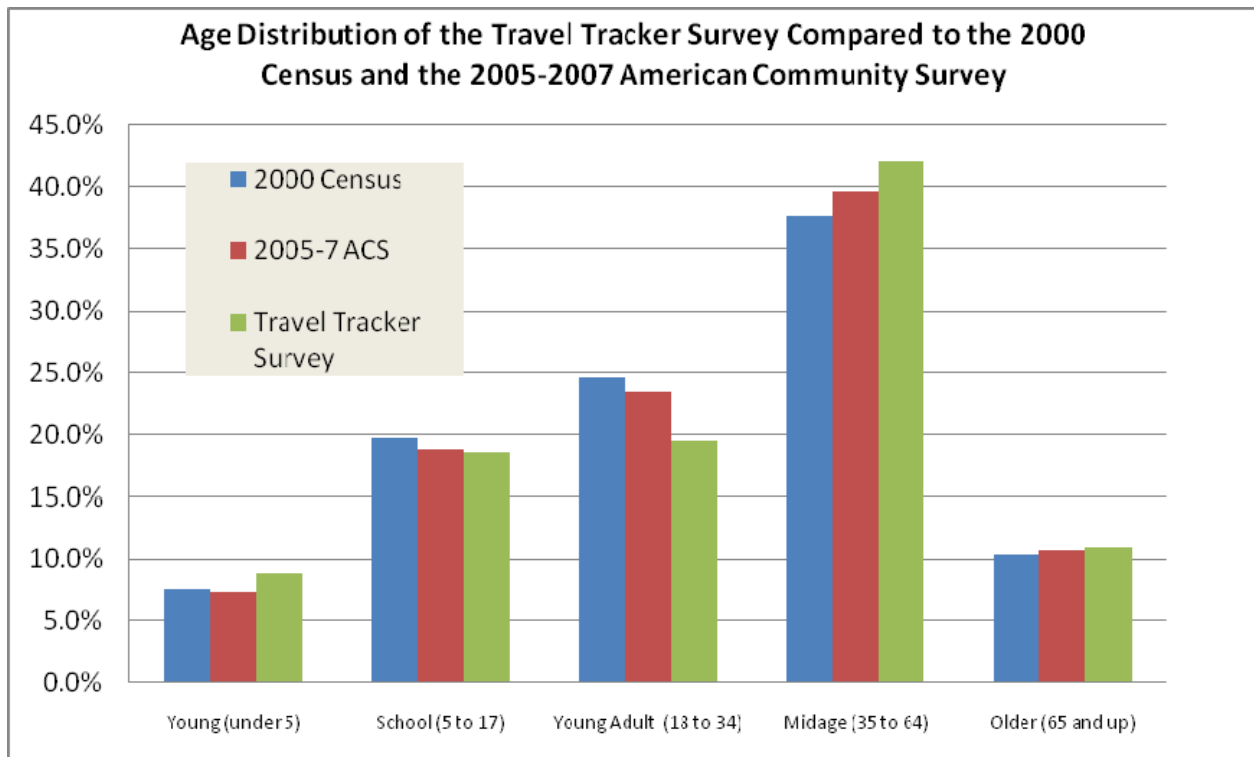
Table 13 Lifecycle Characteristics. Matching Household Survey Data to ACS Control Totals

	ACS Control Totals	ACS Control Percentage	Survey Final Weighted Totals	Survey Final Percentage	Difference: Total Control Final Weights	Raw Survey Data	Raw Survey Percentage
Household with children under age 6	506,018	16.7%	506,485	16.7%	-467	1,173	11.1%
Household with children age 6-18	608,952	20.1%	609,373	20.1%	-421	1,482	14.0%
Single adult under age 35;	154,901	5.1%	154,902	5.1%	-1	217	2.1%
Single adult age 35-64	421,061	13.9%	421,139	13.9%	-78	1,733	16.4%
Single adult age 65 and older	255,711	8.4%	255,748	8.4%	-37	1,440	13.6%
Household with 2 or more adults: oldest under 35	140,701	4.6%	140,548	4.6%	153	294	2.8%
Household with 2 or more adults: oldest age 35-64	606,685	20.0%	606,183	20.0%	502	2,309	21.9%
Household with 2 or more adults: oldest age 65 or older	333,272	11.0%	332,923	11.0%	349	1,904	18.0%
Total	3,027,301		3,027,301			10,552	

The age distribution of the individuals in the weighted survey data is compared to the 2000 census and the 2005-2007 ACS in Figure 2. The age distribution for the survey data is not the same as the census products. There seems to be a higher share of young children compared to older children and there is a higher share of mid-age adults than younger adults.

The households were weighted by the lifecycle variables and so the simple relationship between age and travel, as shown in Table 12, will not necessarily remain the same. It cannot be stated that since there is a higher share of mid-age adults, who typically travel longer total distances, that the survey data will over state total travel. It might be true that the older and younger households, that appear unbalanced since there are too few young adults, are in the same lifecycle phase and so have similar travel patterns. Also, if younger adults with children were under-sampled in the survey, using the lifecycle variable, they were not replaced with younger adults without children, but rather with older adults who had children. This would retain important travel characteristics. Analysis of the weighted survey data will provide additional detail to this issue.

Figure 2 Age Distribution



Section 3 Weighting Verification

Verifying the effects of weighting of the survey data: A comparison of the weighted survey data to the 2000 Census data

If the survey data is weighted correctly for characteristics that influence travel, then the travel in the weighted survey data should reflect the actual travel in the region. The survey data has been weighted based on six categories⁶ to represent the entire population of the region. Specifically, the weighted survey is supposed to reflect the travel patterns of the region.

It is important to estimate how well the weighted survey data fulfills this goal. There are not any current comprehensive travel datasets to compare the survey with, but there is census data that details the mode that people used to travel to work in the year 2000.

The following table compares the share of work trips by mode for the 2000 census and the weighted Chicago Regional Household Travel survey⁷. There are numerous differences between the census data and the survey for the mode share of work trips. In general, the survey data has more transit, walking and bike trips and conversely, there are fewer trips made using auto and trucks.

Table 14 Work trip mode shares for the 2000 Census, 2005-2007 ACS, and the Household Survey (Weighted and Raw data)⁸

Mode to Work	Auto-truck	Bus or trolley bus	Subway or elevated	Railroad	Bicycle	Walked	Taxicab	Total
CTPP 2000	82.65%	5.30%	3.75%	3.65%	0.33%	3.29%	0.34%	99.30%
ACS 2005-2007	82.59%	5.74%	3.43%	3.49%	0.51%	3.09%	0.28%	99.14%
Survey-Weighted	77.17%	6.19%	4.94%	5.00%	1.46%	4.03%	0.09%	98.89%
Survey-Raw	77.31%	5.05%	5.35%	5.64%	1.50%	3.93%	0.12%	98.90%

It should be noted that the 2000 Census and the regional survey do not track the work trip in exactly the same manner. The 2000 census collected information on April 1st from one in seven households on how people usually travel to work, for those who worked in the previous 2 weeks. The census did not track how someone traveled on a specific day.

The ACS collected the same data as the census, but the collection of data was completed throughout the year. Over the three-year period, the ACS surveyed about 2.5% of the households in the region. Compared to the census data, the ACS had a similar amount of auto and transit use, but within the transit category, the ACS had more bus use and less commute by train trips. The ACS had a relatively large increase in bicycle use, but a slightly smaller share of walking trips.

⁶ The household surveys were weighted on household population, the number of workers in the household, the number of vehicles per adult for one person households and also for larger households, the household's race, the household's ethnicity, and the household's lifecycle status

⁷ This is based on how a person describes their general work trip not the mode of each work trip in the survey.

⁸ Totals do not equal 100% due to modes listed as "Other"

The Chicago Regional Household Travel collected information on how people traveled on a specific date and also asked people how they most frequently commute to work. Tracking work trips with two types of questions might have affected how the question was answered. The regional survey also only collected data on households, not group quarters as was done in the census.

A characteristic of travel surveys which is difficult to be controlled for is the potential for people who agree to participate in surveys, to have unusual travel patterns or mode choices. For example, if someone completed the survey because it was good for society, this person might also travel in ways that might be considered good for society. The survey process might attract an unusually large number socially minded participants who travel in environmentally sound ways. It would be very challenging to measure this effect.

The question arising from the different mode share for the census and survey data is how much of the difference in mode share is a result of sampling errors and how much represents a change in how people commute to work. The survey was taken during the beginning of a severe recession and increased fuel prices, which helped transit ridership, reach new highs.

It is important to note that while the use of weights changed the relative share of each work mode, the overall effect was not drastic in a general sense. The auto share for the raw data was 77.31% and the weighted share is 77.17%. In the raw survey data, walking trips accounted for 3.93% of the work trips compared to 4.03% of the weighted trips. There were relatively large percentage shifts between transit modes, but little between transit and the other modes of journey to work travel.

The standard errors involved with survey data will need to be examined in order to determine the margins of error that is associated with these weights. Additional comparisons will need to be made to verify that pedalcyclist, pedestrian and transit work trips have increased as much as the weighted survey data suggests.

Conclusion

The Chicago Regional Household Travel Inventory (Travel Tracker Survey) collected travel information for 23,808 individuals who resided in 10,552 households in the northeastern Illinois region. The surveys have been weighted, based on the characteristics of the 2005-2007 American Community Survey to represent the travel made by the households and the population in the region.

The surveys weights were balanced so that the totals for the following characteristics remained consistent with Census estimates. The household surveys were weighted on household population, the number of workers in the household, the number of vehicles per adult for one person households and also for larger households, the household's race, the household's ethnicity, and the household's lifecycle status

The preliminary results show that the weighted survey data has an increase in the share of transit, pedalcyclist, and walking work trips and a decrease in the share of driving work trips, compared to the 2000 census data. Additional analysis is planned to determine if these trends can be verified.

The survey data can be found at <http://www.cmap.illinois.gov/TravelTrackerData.aspx>