



This document addresses two updates to the 2007-08 Travel Tracker dataset: the inclusion of supplemental survey weights developed by the Chicago Metropolitan Agency for Planning (CMAP) and corrected distance traveled values. With the exception of three data elements, all of the data in the Travel Tracker database is consistent with the original information provided by NuStats, who developed and processed the survey. The weights used in the household and person files, and the calculation of the direct distance of each trip have been recalculated for this version of the Travel Tracker database.

Supplemental Travel Tracker Survey Weights

The Travel Tracker Survey collected travel information for 23,808 individuals who resided in 10,552 households in the northeastern Illinois region¹. The survey information was collected in 2007 and 2008 and was done in conjunction with the Northwestern Indiana Regional Planning Commission (NIRPC) household travel survey. While the public-use dataset that is provided includes data from both surveys, this discussion of weighting the survey results only pertains to the data collected for the CMAP area.

Original and Supplemental Survey Weights

The original survey weights developed by the consultant represented data from the 2000 Census; a discussion of this data weighting exercise is included in the Travel Tracker [Final Report](#). The new 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 newer data represents 3,027,301 households and 8,366,434 individuals living in non-group quarters households. This supplemental data weighting analysis is described in this [report](#).

If each household in the region was identical (or the survey was a perfect sub-set of the region), then the surveyed households could each be factored by 287 to represent the 3,027,301 total households in the region. However, the mix of people who participated in the survey does not exactly match the socio-economic profile for the region, thus 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.

In addition to the original weights in the survey dataset developed by the survey consultant, CMAP has calculated supplemental weights for the survey data to represent the travel made by both the households and the population in the region. To accomplish this, two sets of weights

¹ Cook County, DuPage County, Grundy County, Kane County, Kendall County, Lake County, McHenry County, and Will County

were generated. If the number of households is the item of interest, then the “WGTHH” value should be used in the analysis. Conversely, if the population is the unit of analysis then “WGTP” value should be used to weight the trips and people. These supplemental weights were developed to minimize the number of extremely high and low weight values that are included in the original release of the data². CMAP recommends that the original and expanded weights included in the dataset (“wtfin” and “expfin”, respectively) only be used if the user has specific knowledge of their development and appropriate use³.

For the calculation of the supplemental survey weights, the CMAP planning area was divided into 11 sub-regions⁴. 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 reflect under-sampled characteristics. The surveys were weighted to match Census products for the CMAP region; data on both households and individuals were used in the calibration. One piece of information is inconsistent between the Census products (i.e., the household and person files) used in this process: the sum of the people living in the households is not the same as the number of people living in the region. Similarly, the number of households that exist in the person file is not the same as the number of household in the household file.

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 per household,
- the household’s race,
- the household’s ethnicity, and
- the household’s lifecycle status. There are eight lifecycle categories, which are defined by the age of people in the household, the household size and the presence of preschool and older children.

Application and Limits of Survey Weights

Data collection for Travel Tracker occurred over a 15-month period; nearly 2/3rd of the surveys were single day surveys and the remainder were 2-day surveys. The single day surveys were conducted only on weekdays while the 2-day surveys were given throughout the week; however, all 2-day surveys included at least one weekday. The total number of surveys that were collected on Saturdays or Sundays were far fewer than those collected for each of the weekdays.

² The maximum value for the supplemental household weights is 3619, the minimum weight is 36, and the average weight is 286.93. The original weights varied between 0.09 and 46,520.

³ As an example, the 19 households in the survey data which have the highest original weights represent 0.18% of the records, but 29.2% of the weighted households. Each weighted trip that these individual household make represents 1.5% of the CMAP population.

⁴ 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 including portions of Grundy County.

A challenge is determining the proper way to factor these data to represent specific combinations of days. If all of the households in the dataset are represented, then the weighting reflects the composition of the region. Analysis of Monday-Friday travel can be aggregated to represent an “average” weekday; this is a correct use of the weights because all of the households will be included. While the specific travel on each of the weekdays can be calculated, the weighting is not accurate for a single day and this would be an improper use of the weighting values.

The supplemental weights for the survey data are **not** sufficient in design to account for analysis of weekend-only travel. CMAP suggests that users apply the unweighted data for weekend analysis or create a weighting system that matches the weekend travelers’ characteristics to the region’s demographic profile.

Sample Instructions for Applying Survey Weights for Weekday Travel

The following table lists the “DAY” code and actual day of the week for 1-day surveys and for the first day of 2-day surveys. As noted, no 2-day surveys contained only weekend days, thus no 2-day survey started on a Saturday.

| DAY (Code) | Day (Actual) |
|------------|--------------|
| 1 | Monday |
| 2 | Tuesday |
| 3 | Wednesday |
| 4 | Thursday |
| 5 | Friday |
| 6 | N/A |
| 7 | Sunday |

Note: DAY code refers to the first day of travel in a 2-day survey.

For an analysis of weekday activities, the following household weights should be applied:

1. **For single day surveys:** The survey weights should be used with a value equal to “WGTHH”.
2. **For two-day surveys covering two weekdays:** The survey weights should be used with a value equal to $(0.5 * \text{“WGTHH”})$. Using the full weight for both days of travel would over represent the demographics of these households.
3. **For two-day surveys including only one weekday:** The weekend day information should be excluded from the analysis and the weekday data should be used with a weight value equal to “WGTHH”.

Corrected Trip Distance

In the “place_final” table in the dataset, the distance traveled (“DIST”) between the previous location and the location of the current activity has been recalculated due to inconsistencies in previously-reported values. This distance represents the direct distance between two locations (i.e., “as the crow flies” distance and not a street network-based distance) and is measured in miles. The distance calculations are based on the actual location information provided in the

survey responses; the location information provided in the public database has been generalized to Census tracts in order to protect the privacy of the survey respondents. *Note: the public use files contain only anonymized location information, therefore distance calculations based on this geographic information will not match the distance ("DIST") value provided.*

Values for "DIST" were calculated by specifying the location of each point in the NAD1983 State Plane Illinois East FIPS 1201 Feet coordinate system. Each location is defined by two points: the number of feet north of a specific reference point and the number of feet east of the reference point. The Pythagorean theorem was then used to calculate the straight-line distance between locations.