Community Cohort Evaluation Tool

Introduction

This document presents CMAP's Community Cohorts grouping tool for determination of the level of local capacity and technical assistance need for communities in the region. The tool, named Community Cohort Evaluation Tool (CCET), assigns Community Cohorts throughout the CMAP region based on four factors: population, income, tax base per capita, and percent of population located in an economically disconnected or disinvested area (EDA). The Community Cohorts will be assigned on an annual basis every April using this tool and the most current data available.

Data

The factors considered for developing and running the tool are shown in Table 1. Each of the factors should be updated annually and the weights may be altered, if needed, to redefine the relative importance of individual factors.

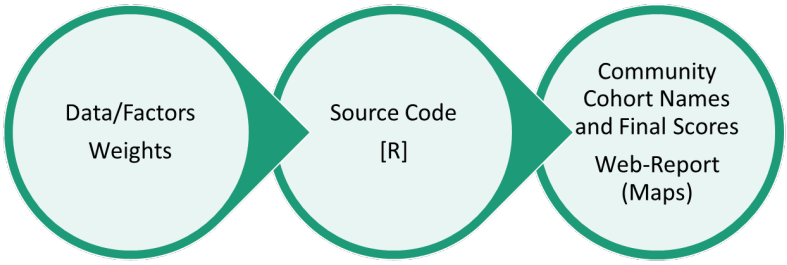
Table 1. Evaluated factors. Factors in bold were used for the final analysis and tool development.

Factor	Data Source	Weight
Population	**U.S. Census Bureau's Population Estimates Program**	**1.0**
Median household income	**U.S. Census Bureau's American Community Survey, via CMAP's Community Data Snapshots**	**1.0**
Total tax base *[total equalized assessed value (EAV) plus total retail sales]*	Illinois Department of Revenue, via CMAP's Community Data Snapshots	1.0
Tax base per capita	**Calculated from population and total tax base factors above**	**1.0**
Percent of population aged 25+ with at least a high school diploma	U.S. Census Bureau's American Community Survey, via CMAP's Community Data Snapshots	1.0
Percent of population located in an economically disconnected or disinvested area (EDA)	**CMAP analysis of internal 2015 Parcel-Based Housing Inventory dataset**	**-1.0**

Methodology

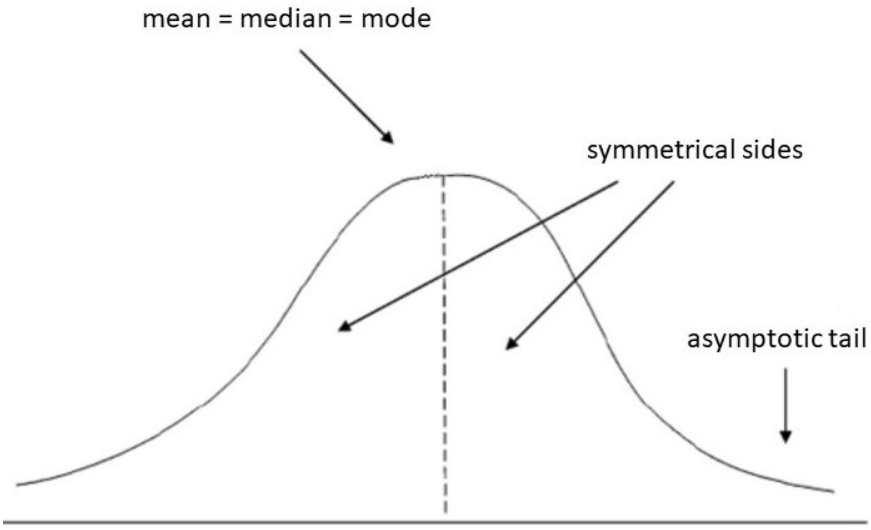
The tool was developed to have three components: input data, source code, and output files (Figure 1). The inputs consist of a single Excel file containing three sheets: the first holds the factors/data for each of the 284 CMAP region communities; the second defines the weights assigned to each factor (see Table 1); and the third defines the score thresholds that are used to assign communities to cohorts (see Table 2). The user can modify this file to change the datasets, weights and cohort thresholds without altering the source code. The source code is the sequence of computational statements written in R. The primary output consists of two CSV file containing the final cohort assignment and associated scores for each municipality and Chicago Community Area. The tool also generates several charts and maps to visualize the results.

Figure 1. Conceptual representation of the CCET for CMAP Community Cohorts designation



- ***Assumption:*** The only assumption is that the datasets (i.e. factors) must be approximately normally distributed. A normal distribution has a bell-shaped density curve that allows a description of a dataset by its mean and standard deviation (Figure 2). The normal distribution curve is symmetrical, centered about its mean/median, and its spread determined by its standard deviation (approximated by z-score from statistics textbooks), allowing to organize the datasets into two categories where one half of values are less than the mean/median and the other half greater than the mean/median.

Figure 2. A conceptual representation of a bell-shaped, normal distribution density curve.



- ***Normal distribution test:*** The input factor distributions were evaluated before inclusion in the tool. Population, total tax base, tax base per capita and median household income have been log-transformed to approximate a normal distribution. Percent of population with HS diploma has a heavily left-skewed normal distribution. Percent of population located in EDAs has a bimodal distribution, with the vast majority of communities having either 0% or 100% of their population in EDAs – this does not remotely resemble a normal distribution. A correlation test was then performed to determine the factors which highly correlated to each other (Figure 3). R2 = 0.7 is considered highly correlated in this analysis. Total tax base and population are strongly correlated (R2 = 0.875); thus, total tax base was not used for the

analysis. Percent of population with a HS diploma was also not used for the analysis as this dataset not only had a heavily skewed distribution but also holds limited importance for the objective of this tool. In contrast, the EDA factor was kept in the analysis despite its non-normal distribution due to the importance of this factor in the region.

- ***Determination of factor-specific scores:*** The decile z-scores of a standard deviation were used to determine five groups at each side of the median value for each of the factor distributions. The analysis used median value instead of mean for the midpoint so that, for each factor, exactly 50% of communities would receive a score of 1-5 and 50% would receive a score of 6-10 (see Figure 4). A score of 1-10 was assigned to the groups with score 1 being the lowest score and assigned to group at the far left side of the bell curve and score 10 assigned to the group at the far right of the bell curve. (The score ordering would be reversed for factors with a negative weight.) The group thresholds were defined as follows:

  Group 1 < median - 1.2816 standard deviations
  Group 2 < median - 0.8416 standard deviations
  Group 3 < median - 0.5244 standard deviations
  Group 4 < median - 0.2533 standard deviations
  Group 5 < median + 0 standard deviations
  Group 6 < median + 0.2533 standard deviations
  Group 7 < median + 0.5244 standard deviations
  Group 8 < median + 0.8416 standard deviations
  Group 9 < median + 1.2816 standard deviations
  Group 10 >= median + 1.2816 standard deviations

  Due to its unique distribution, the percent of population in EDAs factor was grouped differently: Group 1 was assigned if 90-100% of the population lived in EDAs; Group 2 if 80-89% did; Group 3 if 70-79% did; etc. Group 10 (0-9%) contains 181 communities, 155 of which have 0% of their population in EDAs.

- ***Calculation of overall scores:*** The overall score was calculated by multiplying each factor's score by the absolute magnitude of its weight (Table 1), then summing the weighted scores for each community. The overall scores were then dynamically rescaled so that the total range of theoretically achievable scores would be 0-100, regardless of how many factors were included or what their relative weights were.

- ***Conversion of the overall scores into community cohorts:*** The overall scores were converted into community cohorts by applying thresholds of 37/45/65 to the final scores. These thresholds were determined by comparing the newly estimated final scores to the previously used method final scores, maintaining a similar number of municipalities in each cohort as before (Figure 6a-c). The community cohorts determined based on the final scores are shown in Table 2. Any municipality with a population of 1,000,000 or more was also evaluated at the neighborhood level to account for the variability in financial resources among large communities, which otherwise may not be noticed when lumped together.

- ***Calculation of Chicago Community Area (CCA) scores:*** Data were collected at the CCA level. The factor scores derived from the municipal datasets were applied to the CCA data to create comparable scores, and overall scores were calculated in an identical manner (Figure 7). Population was held constant (using the entire city's population) for all CCAs: using CCA-level population instead as a factor would result in a higher score for every part of the city, relative to the city as a whole, despite the CCAs all theoretically having the same access to citywide funding. Retail sales cannot be calculated at the CCA level, so tax base per capita was calculated as a hybrid: *citywide* retail sales per capita plus *localized* equalized assessed value (EAV) per capita. (Localized EAV was estimated from the total property value in each CCA, scaled so that all CCAs summed to the citywide EAV.)

- ***Output:*** The output of the tool constitutes CSV files containing each municipality and CCA's scores and assigned cohorts, as well as a web-based report including maps showing the cohorts' geographic distribution (Figures 8a-b).

Table 2. Community cohort designation with the CCET

Final score	Cohort
0 - 37	Cohort 4 (Very high need)
37.1 - 45	Cohort 3 (High need)
45.1 - 65	Cohort 2 (Moderate need)
65.1 - 100	Cohort 1 (Low and very low need)

Figure 3. Plots and R^2 values showing distribution and inter-correlation of the factors of the CCET for CMAP Community Cohorts designation. From left-to-right: median household income, percent of population with a HS diploma, percent of population in EDAs, log of population, log of total tax base, and log of tax base per capita.

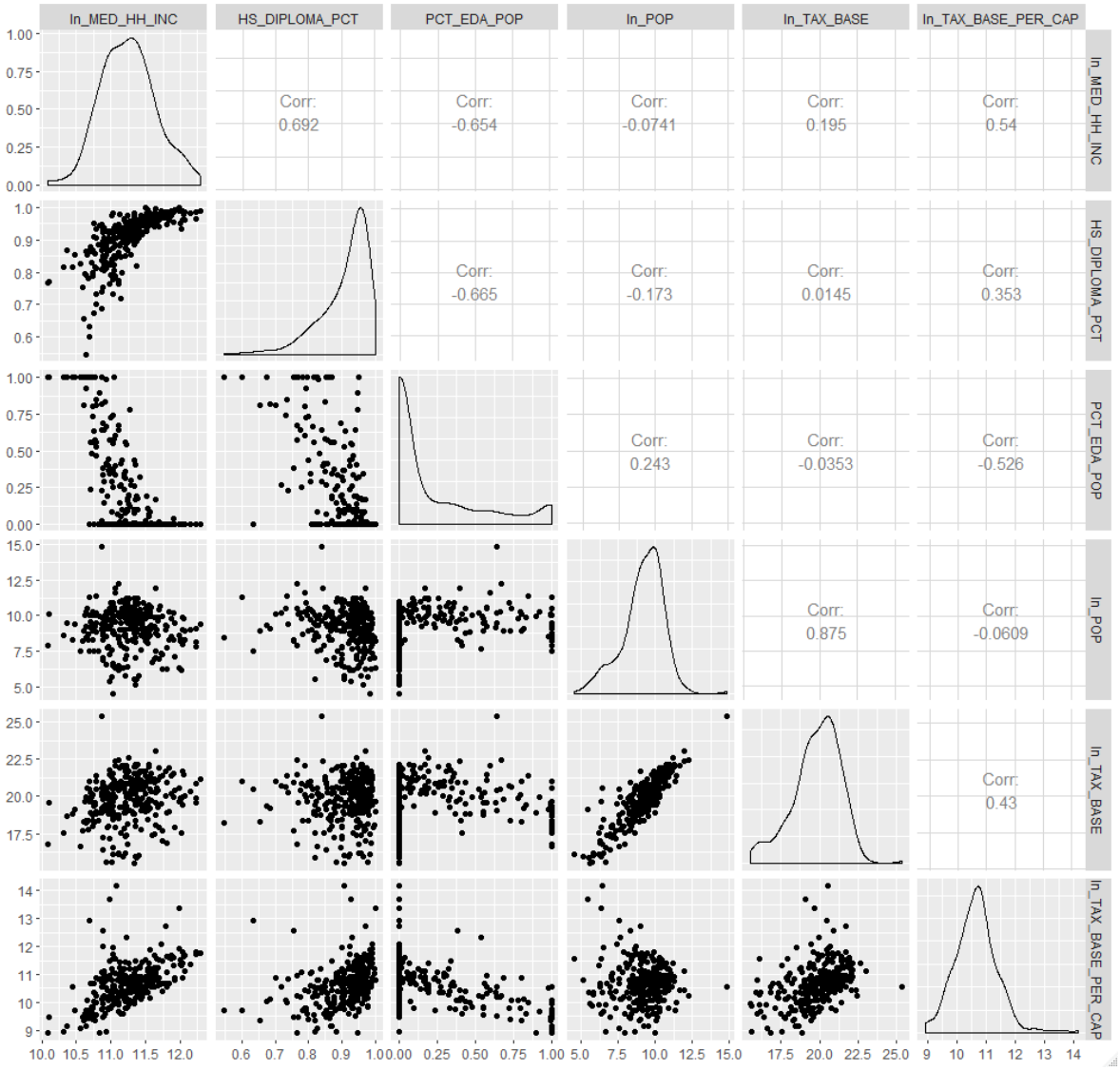


Figure 4. Factor distributions (left) and resultant group/score distributions (right). On each of the factor distributions, the red lines depict the group thresholds. These were based on the decile z-scores of a standard deviation. Median was used instead of mean for the midpoint so that, for each factor, exactly 50% of municipalities would score 1-5 and the other 50% would score 6-10 (see Table 1).

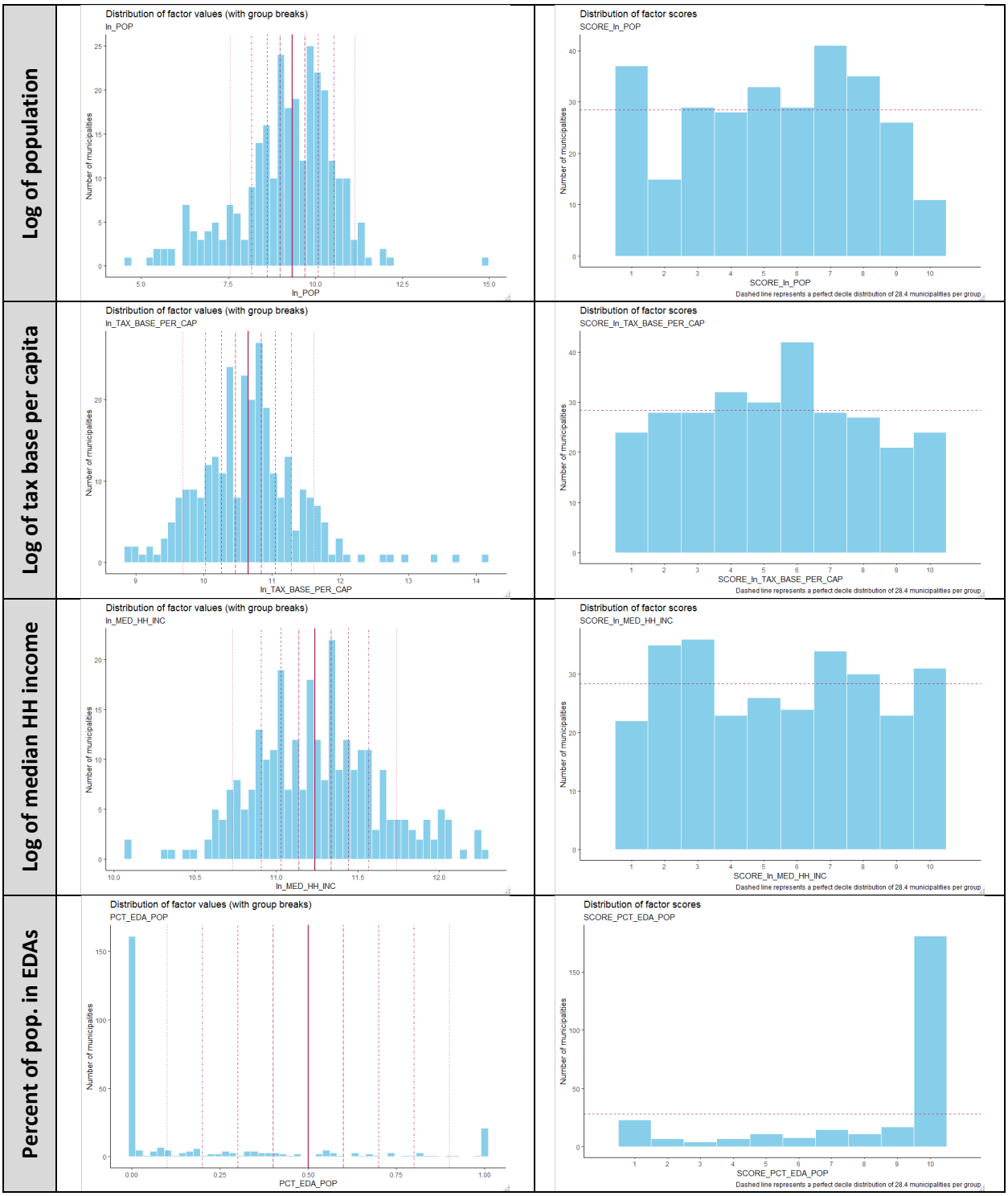


Figure 5. Overall (weighted) scores for municipalities calculated with the CCET

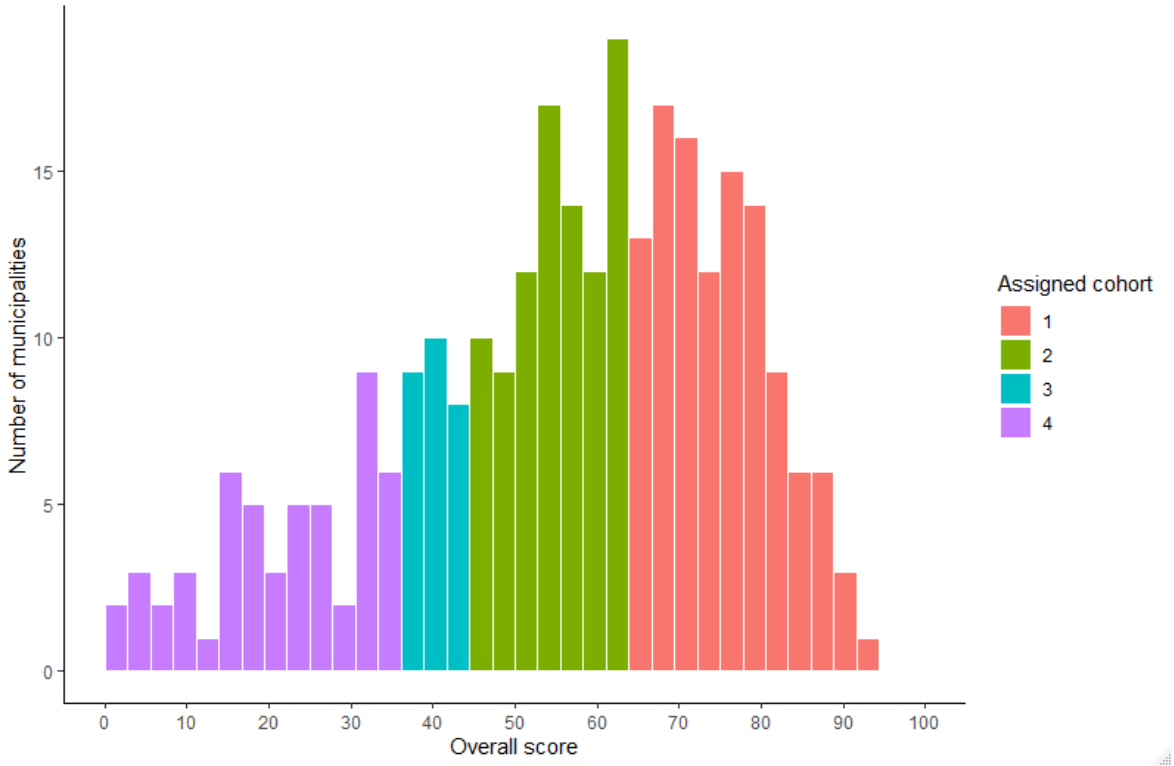


Figure 6a. Comparison of overall scores calculated with the CCET, against the previously assigned scores for each municipality

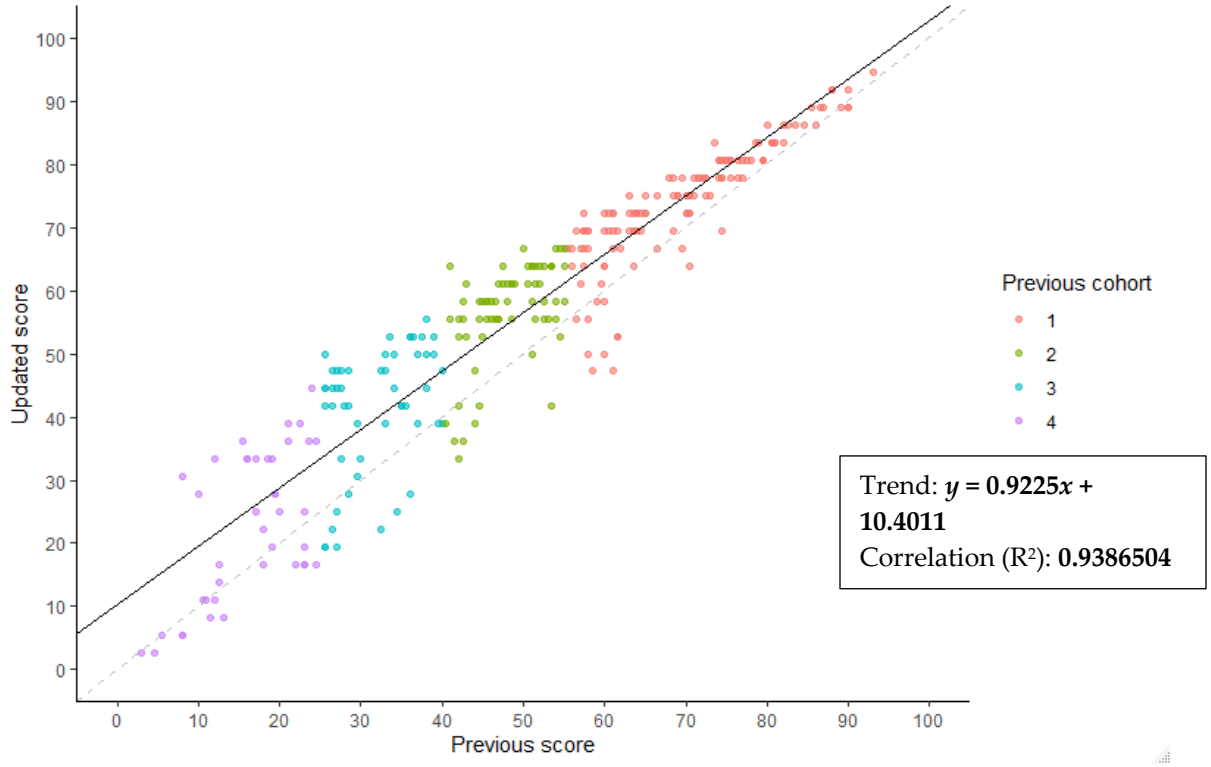


Figure 6b. Comparison of final municipal cohorts assigned by the CCET, against previously assigned scores

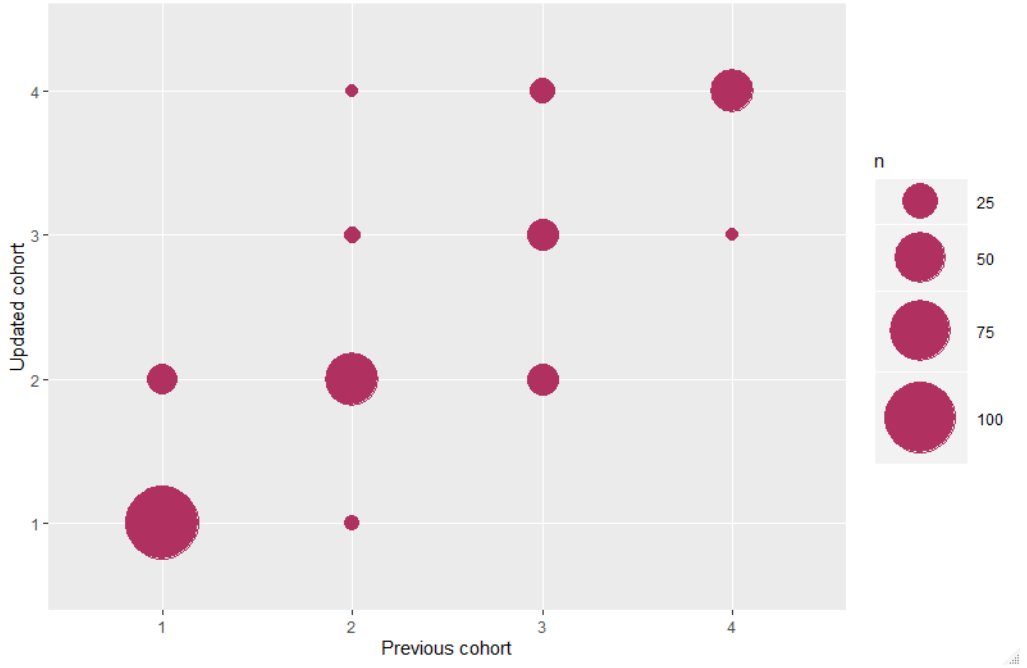


Figure 6c. Comparison of final municipal cohorts assigned by the CCET, against previously assigned scores

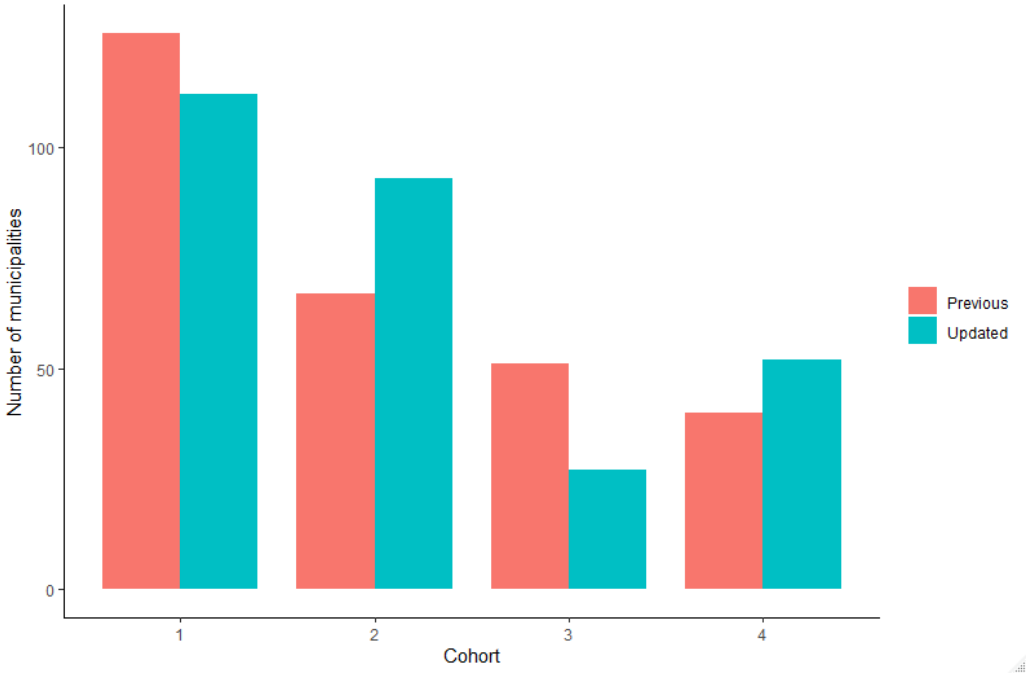


Figure 7. Overall (weighted) scores for CCAs calculated with the CCET

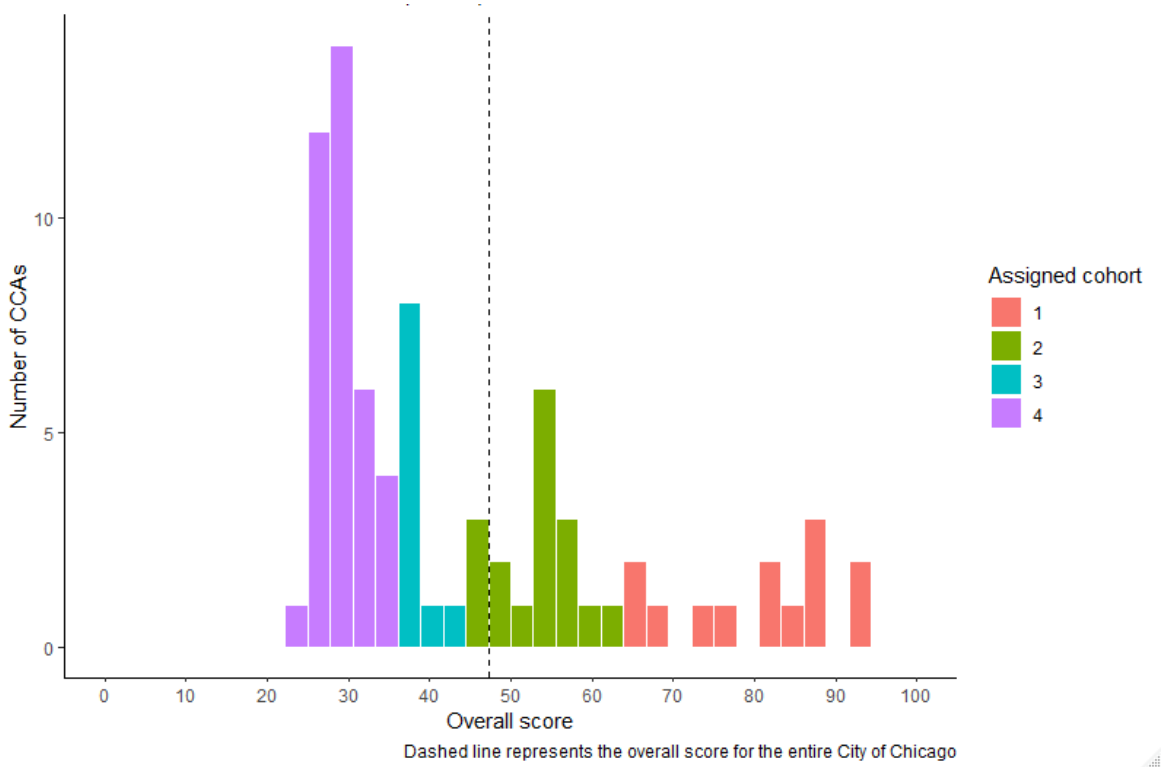


Figure 8a. Map of cohorts assigned to each municipality by the CCET

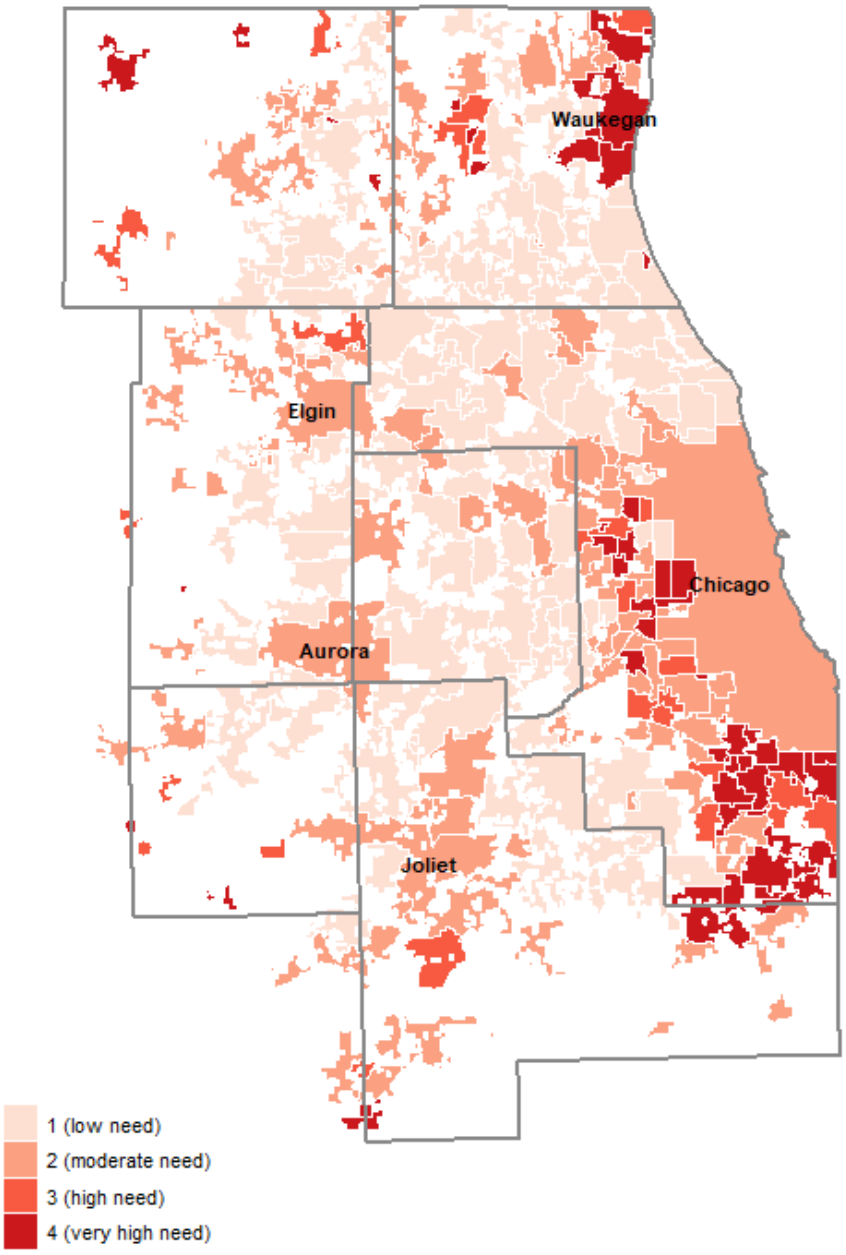


Figure 8b. Map of cohorts assigned to each CCA by the CCET.

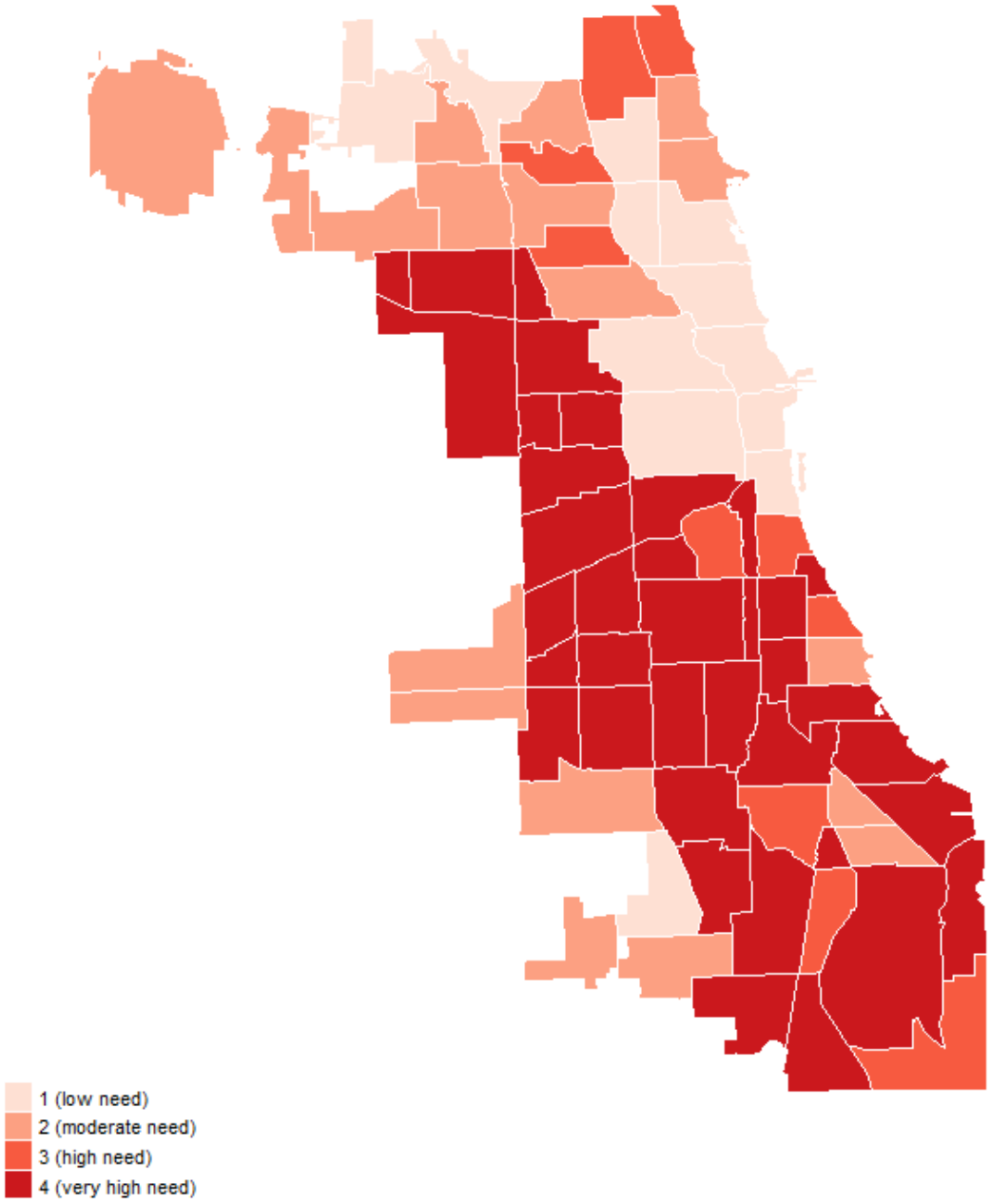
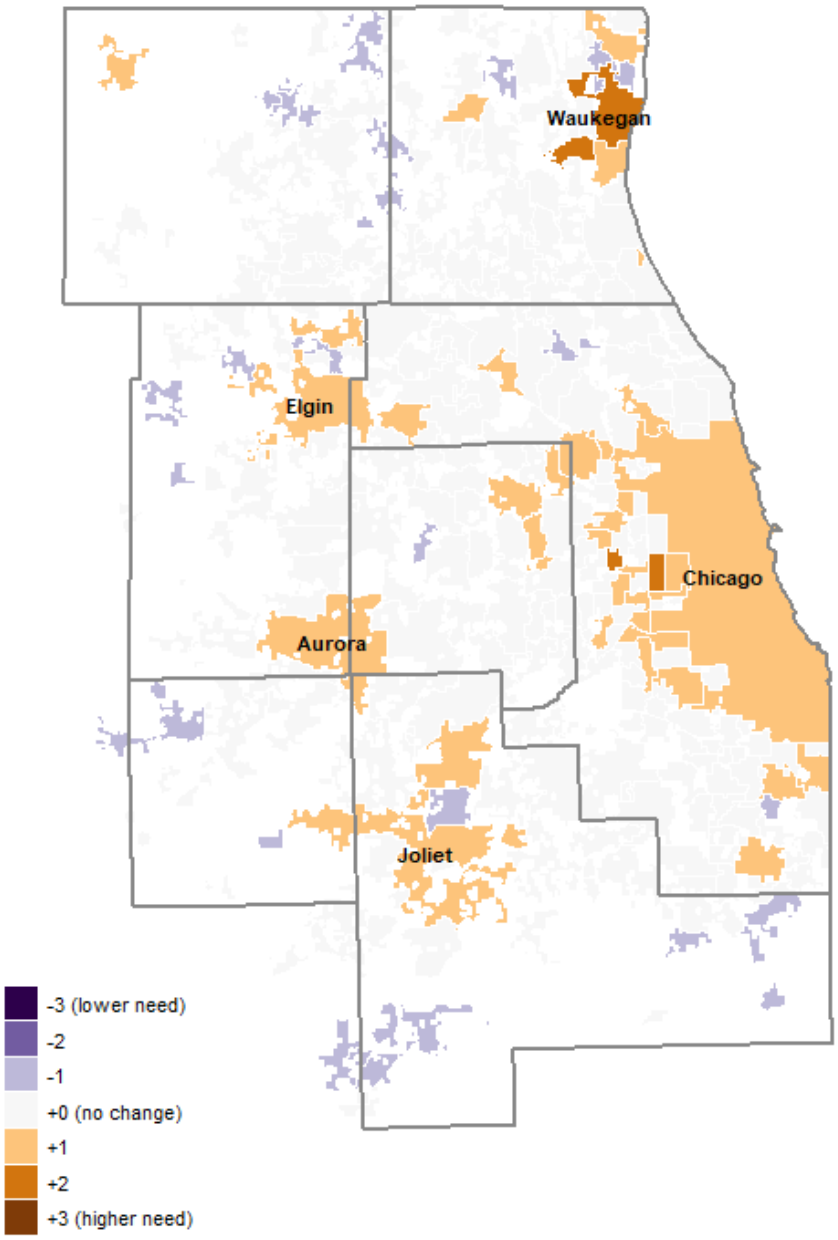


Figure 9. Map comparing final municipal cohorts assigned by the CCET, against previously assigned scores; lower weight of population factor resulted in higher-need cohort assignments for several large municipalities.



Concluding Remarks

The factors shown in Table 1 were used for tool development and cohort analysis. Staff will periodically evaluate the data and the subsequent analysis as new data becomes available or as other appropriate factors are deemed necessary. CMAP intends to update the cohorts with the latest data on April 1 of each year.