**Summarized Methodology for**

**Estimates of Local Populations Eligible for Programs (ELPEP)**

**Illinois CCAP Application**

Developed by: NORC at the University of Chicago[[1]](#footnote-2)

Draft date: January 2025

Prepared by Nick Mader (mader-nick1@norc.org)

## Introduction

NORC researchers have developed a new statistical methodology for estimating the number of children whose families are eligible for subsidies through the Child Care Assistance Program (CCAP) in Illinois.[[2]](#footnote-3) In our present approach, we develop estimates at levels of geographic detail that are as small as Census Tracts[[3]](#footnote-4) and aggregate estimates to the level of zip codes[[4]](#footnote-5), counties school districts, and Child Care Resource and Referral (CCR&R) agency region, and by multiple youth age ranges (0-2, 3-5, 0-5, and 6-12).

This document describes an application of the estimation method to Illinois CCAP. The method is flexible and can be adapted to other programs and jurisdictions with attention to aligning program criteria to the information available in the Census data sources.

## Methodology

This section provides a summary explanation of the statistical analysis and assumptions behind the methodology. Our approach uses a combination of statistical methods to produce estimates of children in CCAP eligible families using the R programming language. We use a selection of public Census data releases with complementary strengths that together contain the necessary measures of CCAP eligibility for households with young children, and reflect the near-present, for geographic areas that plausibly reflect local childcare markets relevant to any given household.

The three data sources included in the model are:

* **American Community Survey 1-Year ("ACS1") microdata** – The ACS1 microdata has information about each member of surveyed households, including age, employment and school attendance status, and income. This makes it possible to determine which—and what proportion of—children are in CCAP-eligible households. However, ACS1 data are only as geographically specific as the Census Public Use Microdata Code (PUMA)[[5]](#footnote-6) level, which reflects areas with a minimum of 100,000 individuals. For example, there are 17 PUMAs inside of Chicago, but also 15 PUMAs in Illinois that span multiple counties.
* **American Community Survey 5-Year ("ACS5") aggregate tables** – The ACS5 data release spans five consecutive years of data from the American Community Survey, which allows it to be reported for smaller geographies such as the Census Tract level. However, the ACS5 data are released as aggregate tables—such as the number of children in families with income below the federal poverty line—and not microdata. While CCAP-eligibility status is not reflected, the ACS5 does include a breadth of information useful for distinguishing the general socioeconomic characteristics of smaller geographic areas within Illinois.
* **Current Population Survey ("CPS") microdata** – The CPS follows households across time, including time spans as long as 16 months. These are the same data used to produce US monthly jobs reports but, for our purposes, are valuable for seeing how the economic circumstances of families change over time, including the ability to assess eligibility for CCAP. The CPS surveys a rotating group of households each month, and while each survey has too few households to build a reliable comparison of CCAP-eligibility specific to Illinois, it reflects major economic trends with a reporting lag of only approximately 2 months.

**Table 1. Comparison of Features Among US Census Data Releases**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data Source** | **Level of Information** | **Geography** | **Recency** | **Role in Method** |
| American Community Survey  1-Year (ACS1) | Detailed microdata | PUMA-level » 100,000 individuals  In IL: 15 multi-county PUMAs. Also: 17 in Chicago City. | ~9-21 months lagged[[6]](#footnote-7) | Identifies the income/work circumstances of Illinois families, by PUMA, in the “baseline” year (e.g., 2023) |
| American Community Survey  5-Year (ACS5) | Select aggregates | Tract-level » 4,000 individuals | 5-year aggregate | In conjunction with the ACS1 data, is used to estimate income/work circumstances of Illinois families down to the Census tract level in the “baseline” year (e.g. 2023) |
| Current Population Survey (CPS) | Detailed microdata | National sample | ~1-2 months lagged[[7]](#footnote-8) | Used to estimate counts of children in CCAP-eligible households at the Census tract level, based on how past-year socioeconomic trends have likely shaped recent eligibility dynamics of households. |

We build our estimates model with the following steps:

1. ACS1 data are used to develop a “direct” estimate of the share of children in households eligible for CCAP, in each PUMA.
2. ACS1 and ACS5 data are combined, using statistical regression modeling techniques, to study the association between community characteristics and the share of CCAP-eligible children. These associations are then used to generate predictions of the “model”-based estimate of the share of CCAP-eligible children.
3. The “direct” (step 1) and “model” (step 2) based estimates are combined using a weighted average—based the relative estimated precision of each number—to provide final estimates for the baseline year for tracts in Illinois. To address a natural compression of estimates towards PUMA-level averages, these tract-level estimates are then “fanned out”[[8]](#footnote-9) to match the geographic diversity of each measure.[[9]](#footnote-10)
4. The CPS data is used to build statistical relationships between household socioeconomic status as of the baseline year and CCAP eligibility as of its most recent available data.
5. These CPS-based statistical relationships are applied to the final baseline estimates for Illinois, using the specific estimated socioeconomic status of each Illinois Census tract to estimate CCAP-eligibility as of the most recent month of CPS data.
6. Estimated shares are converted to counts by multiplying the estimates with projected proportions of the overall number of children aged 0-5 and 6-12 forecast using Census decennial data.

Steps (1)-(3) are formally known as “Small Area Estimation (SAE)” statistical methods in the field of academic demography. Steps (4)-(6) represent a novel “now-casting” method developed by NORC researchers. Now-casting is a term adapted from “fore-casting”, meant to imply that the method uses lagged (historical) data to characterize present circumstances.

Recognizing that every estimate yielded from both survey data and statistical methods has a specific amount of uncertainty, we use a range of statistical calculations to capture this combined uncertainty throughout our full method. The amount of uncertainty for a given eligibility estimate is reflected in lower bound and the upper bound estimates. These bounds indicate confidence levels in the precision of each estimate, which is useful for decision-making purposes.

The code used to implement ELPEP—as well as documentation describing its use for new users, and highly detailed technical notes—can be found in open source at [this link](https://github.com/chapinhall/elpep).

**Limitations**

* Precision is naturally reduced when making estimates at more geographically granular level
* Reliance on Census incorporates some known sources of bias, such as documented undercounting of young children
* Some state eligibility rules may not be adequately identified in the Census data, such as families involved in the child welfare system, deployed military families, or youth with documented disabilities.

## Potential Applications

ELPEP has distinct potential relative to other tools in the applied research space. A widely-known tool for estimating eligibility status for means-tested programs is the Urban Institute’s Analysis of Transfers, Taxes, and Income Security (ATTIS) model. ATTIS is designed to capture detailed eligibility criteria for a wide range of programs, and to estimate how patterns of eligibility would change with respect to proposed rules. However, ATTIS reporting is limited to reporting at the PUMA-level at its most geographically granular.

While ELPEP has been developed only for estimating eligibility for child care subsidies, it has several advantages to ATTIS. First and most prominently, ELPEP is a statistical framework that can incorporate the strengths of additional data. As seen in Table 1 above, ELPEP is able to use the granularity of the ACS 5-year data, and economic trends in the CPS to develop estimates that more up-to-date, and more local than ATTIS. Second, ELPEP has to-date been developed [in open source](https://github.com/chapinhall/elpep), offering greater transparency and potential for use and adaptation.

Whereas ATTIS may be best suited as feedback to rule-making and budgeting for states, we believe that ELPEP provides the best evidence for assessing, motivating, and evaluating program local efforts for making public programs accessible. NORC researchers are currently developing automated reports for Illinois that make use of ELPEP estimates to assess both:

1. **Engagement with CCAP among eligible families.** In which areas of the state are children in eligible families more likely to take-up CCAP subsidies—measured by a ratio between the number of CCAP-enrolled and CCAP-eligible and what are the characteristics/contexts of areas that are above/below state averages?
2. **Accessibility/sufficiency of existing provider capacity.** In which areas of the state may there be gaps in the capacity of CCAP-accepting providers to serve potential/likely families?

# Technical Notes on ELPEP Methods

The ELPEP statistical method works in two stages:

1. use of custom **Small Area Estimation** methods to estimate a range of community characteristics related to childcare program eligibility at a small geographic level; and
2. use of statistical methods to bring the first stage estimates up to the “near-present” by capture macroeconomic trends relevant to these smaller geographies.

### 5.1.1 Small Area Estimation

ELPEP’s implementation Small Area Estimation (SAE) uses several modifications to canonical methods to estimate community characteristics down to the Census tract level.[[10]](#footnote-11)

Let be the share of young children in category , e.g. the % of young children with income between 0-100% of the Federal Poverty Level (FPL), 100-200% FPL, etc., in geography , at time . Note that although eligibility of children for public childcare subsidies is typically determined by household status (e.g. presence, labor force participation, and income of parents), our analysis is at the level of children and thus makes use of person–rather than household–sampling weights. This is motivated by the final goal of assessing whether affordable childcare slots are sufficient for the number of children (rather than for the number of local families).

“Direct” estimates are calculated by assuming PUMA-level population shares are, in expectation, representative of tract-level shares:

where is our small geography (tract), and is the big geography (i.e. Public Use Microdata Area, or PUMA).

Within the SAE method, “Model” estimates use PUMA-level data to draw inference using regression model:

where the tract-level estimate is obtained using with tract-level values:

with standard error of estimation of .

Note that the regression equation used above mixes an outcome measured at the larger geographic level and predictors measured at the smaller geographic level . This is an adaptation of canonical SAE methods which would otherwise use a measure of for both parameter estimation and prediction. An outcome of that would represent equivalent geographic basis is not available, and the use of as a substitute invites non-classical measurement error. ELPEP proceeds with this method because it maintained estimation and prediction using the same () measures, and utilizes the full distributional support (i.e. the total variation of the measures across small geographies).

This step produces estimates that capture the “direction” of projection onto local community characteristics, but have compressed variance. To re-establish the appropriate level of variance, ELPEP re-inflates the first estimates using the closest available estimates of variance at the level of as follows:

where is the mean of the estimates, is the standard deviation of those estimates, and is the standard deviation of tract-level estimates for share category .

For categories that:

* correspond to household shares of income-to-poverty ratio, ACS 5-year estimates are used directly to calculate since income-to-poverty ratios are directly represented in ACS5 aggregate tables.
* correspond to a combination of (1) household income-to-poverty rate shares crossed with (2) shares of families that are work eligible (due to all adults in the household working), while that cross-tabbed status is not available in ACS5 tables, each separate measure is. ACS1 PUMS data are used to calculate the correlation between each measure for each PUMA, and these correlations are used within the delta method to calculate the standard deviation of the product of each component of (i.e. income and work eligibility status).

Within the SAE method, the blended estimates (Empirical Best Linear Unbiased Predictor; EBLUP) are a weighted average of the direct and model estimates:

where

Because is not guaranteed in the estimates, ELPEP obtains final estimates and .

### 5.1.2 Now-Casting Methods

The ultimate goal is to predict counts of children who are eligible for program :

where is the target “small” geography, is near-present time, and is the proportion of the count of all young children that are eligible for .

No direct measurements of are available, so ELPEP accounts for both community composition–using SAE results–and recent eligibility dynamics. Thus, ELPEP models binary eligibility status as

with observations from individual child and their observed characteristics which reflect individual and household measures. The parameters can be estimated from analysis of individuals and households in the CPS. However, in the applied exercise of “now”casting for each tract, only tract-level averages exist as analogs to .

ELPEP assumes the following structure of linear expectations with respect to individual eligibility measures to be able to form estimates of using community averages:

The linear expectations form of in the regression above implies that must be estimated using a linear probability model (LPM) rather than a logit or probit. While the use of LPMs are necessary given the community- (rather than individual-)level predictors, in practice, the predictions of program eligibility rates are far enough from 0% (and 100%) that LPMs are reasonable approximations to other non-linear estimation methods.

In practice, is composed of both SAE estimates as well as ACS5 measures that can reasonable be assumed to be representative of “” because they do not change rapidly, or are more accurately observed via ACS5 rather than estimated via SAE. Community measures such as income-to-poverty status, which we presume are highly dynamic given both macro and local economic factors, are sourced via our SAE method. Other measures that we believe are more persistent, such as adult educational attainment, are sourced directly from ACS5 data.

The final estimated counts of young children eligible for a given program are produced by multiplying the estimates of with an estimate of the overall youth population from a simple forecast of Decennial Census counts at the Census tract level.

1. Corresponding author: Nick Mader (mader-nick1@norc.org). [↑](#footnote-ref-2)
2. Work supported by Child Care Policy Research Partnership (CCPRP) Cooperative Agreement (Grant No. 90YE0225), Office of Research, Planning, and Evaluation (OPRE), Administration for Children and Families U.S. Department of Health and Human Services. CCPRP Grant No. 90YE0225 was awarded to Chapin Hall at the University of Chicago and supported foundation ELPEP work from September 2019 – September 2024. [↑](#footnote-ref-3)
3. Census Tracts are constructed by the US Census Bureau as areas that generally include between 1,200 and 8,000 individuals, 4,000 on average. See [this reference](https://www.census.gov/programs-surveys/geography/about/glossary.html#:~:text=Census%20Tract,-Census%20Tracts%20are&text=Census%20tracts%20generally%20have%20a,optimum%20size%20of%204%2C000%20people.) for details. [↑](#footnote-ref-4)
4. Zip codes are calculated as ZIP Code Tabulation Areas (ZCTAs), which are generalized areal representations of Zip Codes that are populated. For example, ZIP Codes for PO Boxes only have no corresponding ZCTAs. In most cases the ZCTA is the same as the ZIP Code. See [this](https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html) reference for details. [↑](#footnote-ref-5)
5. See [this reference](https://www.census.gov/programs-surveys/geography/guidance/geo-areas/pumas.html) for details on Census Public Use Microdata Areas (PUMAs). [↑](#footnote-ref-6)
6. The ACS1 for a given year (e.g. 2021) is typically released by the Census Bureau in the following September (i.e., 09/2022), and is the most recent release until the September of the subsequent release. [↑](#footnote-ref-7)
7. The CPS Basic Monthly is typically released by the Census Bureau in the following month and becomes available on IPUMS shortly after the Census release. [↑](#footnote-ref-8)
8. Specifically, this involves a mean- and order-preserving inflation, where the variance of the resulting estimates equal the variance of the same ACS 5-year measure. Because there is no measure of CCAP eligibility in the ACS 5-year measure, this variance is estimated by using the delta method to calculate the variance of (proportion of households that would meet work eligibility)\*(proportion of children in households that would meet income eligibility) across tracts but within each PUMA, and using a PUMA-level estimate—developed using microdata—of the correlation between these two components of eligibility. [↑](#footnote-ref-9)
9. For example, while some PUMAs may contain many tracts with very similar rates of CCAP eligibility, others may have sharply different subgeographies, especially in more urban or populous areas given how neighborhoods form and differentiate. We perform this operation to ensure that we capture the full contrast between areas within each PUMA that may have distinctly higher and/or lower eligibility rates than the PUMA average. [↑](#footnote-ref-10)
10. See Rao, J. N., & Molina, I. (2015). *Small area estimation*. John Wiley & Sons. [↑](#footnote-ref-11)