# SMALL AREA ESTIMATE PROGRAM
## User Guide

*Contributors: Nate Chase*
*Mike Mohrman*

December 2023

This page intentionally left blank.

## Contents

# What is the Small Area Estimate Program?

The Washington State Office of Financial Management's (OFM) Small Area Estimates Program (SAEP) provides a consistent set of small area population and housing data for statewide applications.  For the purposes of the SAEP, small areas are defined as geographic areas below the level of the state that may or may not be coincident with the boundaries of U.S. Census Bureau geographic entities.  SAEP estimates are generated for select census areas, 'special areas' required by statute, plus other areas of statewide significance.  Special areas include Highway Urban Areas, Public Transportation Benefit Areas, and the Thermal Electric Generating Facility Area.

# How are the SAEP estimates developed?

The SAEP estimates include both intercensal and postcensal estimates.  The data for years 2010-2019 are considered intercensal estimates whereas the data for years 2021-present are considered postcensal estimates.  The two estimate series differ in their development and revision cycles but both series are based on 2020 census blocks. The methodology used to develop the SAEP intercensal and postcensal estimates data are quite different, however.

### Intercensal estimates (2010-2019)

The SAEP intercensal estimates (2010-2020) take into account changes in census blocks and geographic boundaries between the 2010 and 2020 censuses. A spatial allocation procedure assigns 2010 census population and housing units to 2020 census blocks. SAEP intercensal estimates are then produced using the SAEP postcensal estimates and an interpolation method that assumes error is directly proportional to time elapsed between censuses. The SAEP intercensal estimates are controlled to county intercensal estimates to provide consistency between published intercensal estimates series. The county intercensal estimates can be found on OFM's official historical estimates webpage.

### Postcensal estimates (2021-present)

The SAEP postcensal series takes OFM April 1 city and county population and housing estimates and distributes these to census tabulation blocks based on information from one or more sources including: residential building permits, assessor records, postal delivery statistics, and federal census data.  The SAEP uses whichever data are judged to be the most representative of local housing change for a given jurisdiction each year.

The entire postcensal series is updated each time the SAEP data process is run, typically in the fall of each calendar year.  This allows the system to take advantage of revisions to input data sets and updates to geographic boundaries.  Because the underlying data may change at each revision, we urge SAEP data users to use only the most recent SAEP data series in their work.

## How are SAEP postcensal estimates related to OFM's official April 1 city and county population estimates?

Please note that SAEP estimates are NOT the official state population estimates used for revenue distribution and program administration for cities and counties.  Users interested in city and county estimates should visit the state's official April 1 Population Estimates Program website for more details.

Although city and county population totals from the state's official population estimates program are used as control numbers when building internal SAEP data tables, the two programs differ in how population in annexed areas is treated.  Therefore, population estimates from the SAEP and the state's official April 1 Population Estimates Program are unlikely to match below the county level.

## How are SAEP intercensal estimates different from OFM's official April 1 intercensal estimates for the state, counties, and cities?

SAEP intercensal estimates match April 1 intercensal estimates for the state and counties but differ for cities. If you are looking for city totals, use the Official April 1 intercensal estimates.

## Can you provide more detail about how SAEP postcensal estimates are developed?

Yes.  The SAEP makes use of the most current data available each time the process is run.  Some of the input data sets change over time.  New data will replace older data and the SAEP estimates will reflect those changes.  These changes impact not only the estimate for the current year but can also impact estimates for prior years as well.  Because of this, we recommend that end users utilize the most recent version of SAEP data at the time of their analysis.  Please do not combine data from different SAEP releases.

The estimate process is separated into three distinct parts: group quarter population estimation, housing unit estimation, and household population estimation.  The estimates are not rounded to whole numbers to remind users that these are estimates, not census counts.  By no means do the decimal places imply precision.  The process is run year by year, with each new estimate building off the prior year's estimate.

The three primary data processing steps use some form of proportional allocation and iterative proportional fitting (raking) to distribute control data (OFM official city and county population estimates) to census blocks.  The block data represent intermediate data—the population and housing values are always aggregated to larger geographic areas.  The block-level estimates are not released as a standalone data product.

### Data considerations

The base data used in the SAEP estimate process comes from OFM's adjusted Census 2020 and from OFM's annual April 1 Population Estimates Program for cities and counties.  OFM estimates of population and housing are used as controls throughout the estimation process.

City boundaries may change over time due to annexation/deannexation activity.  Cities may also incorporate or disincorporate.  In Washington, every city annexing or deannexing is required to conduct a door-to-door census of population and housing in the affected area.  Annexation censuses are reviewed and certified by OFM.  Any changes to population, housing, and the group quarter population resulting from boundary changes are accounted for in OFM's city and county control numbers.

The SAEP uses the concept of 'tabulation areas' to approximate city limits and the unincorporated portions of counties.  Tabulation areas consist of aggregations of census blocks that approximate each jurisdictional boundary.  Every census block in the state is assigned to a tabulation area that represents an incorporated city or the unincorporated portion of a county.  SAEP tabulation areas are developed using a multistep GIS model as outlined below:

- Census 2020 blocks are joined with the city limit boundary file provided by the Washington State Department of Transportation that most closely approximates the conditions as of April 1.
- Everywhere the city limit file splits a census block, the system calculates the percentage of each block's area that falls inside or outside of a city.
- Each census block is assigned to the tabulation area based on the jurisdiction that covers the greatest proportion of the block's area.
- The process is run annually in order to capture boundary changes over time.

Tabulation areas effectively serve as control areas.  They allow each jurisdiction to be treated as a standalone entity on an annualized basis.

3

## OFM adjusted 2020 Census block data

OFM adjusts census block data after each decennial census for SAEP data processing. Corrections are made to the locations of housing and group quarters facilities, as well as to the counts of household and group quarters populations where special situations exist. In 2020 more adjustments were necessary than in the past. With the 2020 data, the Census Bureau introduced a new privacy protection system called the Disclosure Avoidance System (DAS), which adds statistical noise to the counts of occupied housing units, population, and population characteristics to protect respondent privacy. For SAEP the noise infusion is particularly problematic because the noise created illogical and implausible situations in the data. Examples include census blocks with occupied housing units but no household population blocks with household population but not occupied housing units, and numerous blocks with improbable occupancy rates and household sizes. Since the SAEP model follows the housing unit method, OFM adjusted the block data to remove some of these logically impossible situations.

In the 2020 data, the Census did not add noise to the number of total housing units, so OFM re-distributed household population and occupied housing based on total housing units in each block and data from adjacent blocks. An average of 11 adjacent blocks were grouped together to help with this smoothing. During this process OFM maintained consistency with published Census Bureau data as much as possible, but there are differences in smaller geographies and non-census geographies. To help users understand our adjustments, SAEP products now contain tabulations of the 2020 Census data as published as well as estimates based on OFM's adjusted 2020 data.

## Group quarters population allocation

The SAEP group quarters population estimate is based on data from a variety of sources. Individual cities report group quarter populations by facility as part of OFM's annual population and housing survey. State agencies such as the Department of Corrections and Department of Social and Health Services and military facilities also report populations to OFM for the facilities they manage.

For the SAEP, each facility is geocoded to a census block. Population change is calculated for each block by subtracting the census year administrative population count from the current year administrative population count. The population change is then added to the Census 2020 group quarter population count by block. The group quarters population of any individual census block is not allowed to fall below zero. Any population change that cannot be allocated based on a facility's geocoded location is proportionally allocated to all census blocks that contribute to the jurisdiction's total group quarter population. It is important to note that in a few cases (certain colleges, military bases, etc.), OFM receives a total count for the entire institution but the actual 2020 census population counts are spread across multiple census blocks. Changes in population for these facilities are distributed to census blocks based on the

proportions of populations counted in the 2020 census.  In rare cases, the jurisdiction as a whole may show an overall loss of group quarters population.  In such cases, the population is ultimately subtracted from the household population.

### Housing unit allocation

The allocation of housing units to small areas is a data driven process.  A master dataset consisting of new units and demolitions by structure type (single family, multi-family, mobile homes, and specials) by year is created from an internal database used for OFM's April 1 Population Estimates Program.  Housing units are allocated to census blocks within each SAEP tabulation area by year according to one of three data sources: 1) geocoded housing units, 2) postal delivery statistics, and 3) block-level Census 2020 housing unit counts.  Note that geocoded housing units consist of residential building permits and/or assessor records that have been parcel matched and/or address matched using a GIS.

The first step in the process is to determine which of the three data sources should be used to distribute housing units to blocks that comprise a tabulation area.  Geocoded housing unit information and postal delivery data are not available for every part of the state.  The actual distribution method that is used is dependent upon which data sources are available for the jurisdiction and which is considered to be the most accurate.

Geocoded housing data are usually considered the most accurate indicator of housing change and are given priority consideration in the housing unit allocation process.  If the number of housing starts in the SAEP tabulation area is within 60% of the control count then the units are distributed to census blocks based on the spatial distribution of the geocoded housing.  If there is no reasonable match between the number of geocoded housing starts and control count of new housing units, then the data may be distributed based on the change in postal delivery data.

Postal delivery statistics represent our second choice for allocating housing units to tabulation areas.  Postal delivery data are available by carrier route area and are therefore less geographically precise than new housing data.  Delivery area boundaries can change markedly over time since they are developed for ease of mail delivery, not statistical analysis.  Carrier route areas do not cover the entire state nor do carrier routes exist in areas where there is no mail delivery (i.e., areas with very low populations or areas served exclusively by post office boxes).  To mitigate the effects of these boundary problems, the carrier route data are allocated to census blocks.  Census block boundaries are updated every 10 years—providing a stable geographic area over the course of the estimate cycle.  Each tabulation area is tested to see how well the allocated postal data matches the decennial census count.  If a jurisdiction fails the test (i.e., the allocated postal data does not closely match 2020 census data), then block-level information from the 2020 census is used to distribute housing unit change.

Although not ideal, we assume that the pattern of new housing growth follows the pattern of existing housing in areas where geocoded housing or postal delivery data prove unworkable. In these cases, new units are allocated to blocks based on the proportion of each block's Census 2020 housing unit count to the total housing unit count for the tabulation area. This procedure generally applies to small towns in rural areas.

While geocoded housing data can be used to distribute new housing with reasonable success, precise information about the location of demolitions is often lacking. Consequently, demolitions are usually allocated based on postal change or, less frequently, according to federal census data.

After the initial housing unit allocation process is complete, any differences between the control totals and the tabulation area totals are raked over the census blocks that make up each tabulation area based on the distribution of all housing within the tabulation area. The result of this process is a table of estimated housing units by census block by year.

## Population allocation

The distribution of household population is based on the distribution of housing units from the housing unit allocation process. The household population estimates themselves are based on the following formula:

Estimated household population = total housing units * occupancy rate * persons per household

Housing units added during the current estimate year, referred to as new units, are processed separately from units that existed prior to the estimate year because new units may be located in census blocks which did not have any housing at the time the 2020 census was conducted. The population in new units is estimated using a jurisdiction's occupancy rate and household size ratio (average persons per household).

The population in existing housing (i.e., units which are not new this estimate year) in a tabulation area is estimated using adjusted Census 2020 occupancy rates and household size information that is adjusted annually to mirror trends in OFM's city and county estimate series.

The new household population estimates and existing household population estimates are combined into an initial population estimate for the tabulation area. Any difference between the control total and the tabulation area total is raked over the census blocks comprising the tabulation area based on the distribution of all housing within the tabulation area.

The total population for the jurisdiction is then calculated by adding the household and the group quarter population estimates together.

# How does the SAEP account for annexation and boundary changes?

To account for jurisdictional boundary changes, the SAEP uses the concept of tabulation areas. SAEP tabulation areas are groups of census blocks that approximate the boundary of each jurisdiction as of April 1st each year.  The tabulation areas are designed to be as consistent as possible with the boundaries used for OFM's official April 1 city and county estimates.

# What are the differences between the census area, special area, and other area estimates?

The SAEP data products are logically organized into three sets: 1) census area estimates, 2) special area estimates, and 3) 'other' area estimates.  Census area estimates are made by aggregating census block data into larger geographic units based on common geographic IDs. Special area and 'other' area estimates have boundaries that are not coincident with census block boundaries.  Estimates for these areas are developed using area interpolation techniques. Special area estimates are distinguished from 'other' area estimates because special area estimates are either required by law or support a specific law.

# How are estimates developed for special and other areas?

The SAEP utilizes area-weighted and parcel-point weighted areal interpolation techniques to develop estimates for small areas that do not follow census block boundaries.  This strategy is intended to take advantage of more geographically precise information where it is available.

### Area-weighted areal interpolation

Area-weighted areal interpolation is a method used to allocate data from one set of areal units, the source zone, to another set of units, the target zone.  The source zone data are proportioned to the target zone based on the amount geometric overlap between the two zones. This method assumes that the data being proportioned are uniformly distributed throughout the source zone.

Area-weighted areal interpolation is defined as:

$$P_t = \sum_{r=1}^{q} \frac{A_{ts\,r} P_{s\,r}}{A_{s\,r}}$$

where $P_t$ is the estimated target zone population total, $q$ is the number of source zones that intersect the target zone, $P_s$ is the population of the $r$th intersecting source zone, $A_s$ is the area of the $r$th source zone, and $A_{ts}$ is the area of intersection between the $r$th source zone and the target zone (Fisher & Langford, 1996).  In the simplified example illustrated in Figure 1, 41.67% of the source zone's population is estimated to be in the target zone because 41.67% of the source zones area overlaps with the target zone.



**Source Zone**
Total area = 36 sq. mi.
Poupulation = 100

**Target Zone**
Overlapping Area = 15 sq. mi.
Weight = (15 / 36) * 100
Population estimate= 41.67
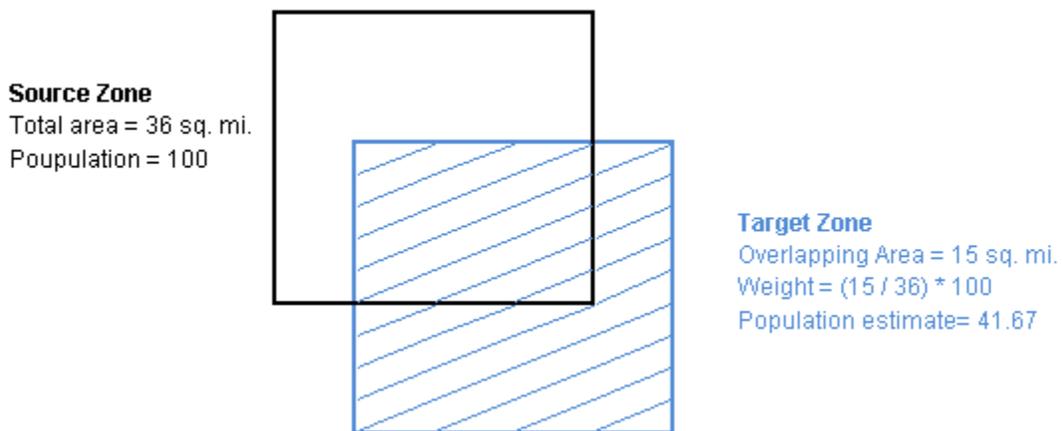
Figure 1. Area-weighted areal interpolation

**Parcel point-weighted areal interpolation**

Parcel point-weighted areal interpolation uses tax parcel centroids as weights to allocate data between zones rather than geographic area.  This method assumes that the location of tax parcels adequately reflects the spatial distribution of population and housing in the source and target zones.

Parcel point-weighted areal interpolation is defined as:

$$P_t = \sum_{r=1}^{q} \frac{D_{ts\,r} P_{s\,r}}{D_{s\,r}}$$

where $P_t$ is the estimated target zone population total, $q$ is the number of source zones that intersect the target zone, $P_s$ is the population of the $r$th intersecting source zone, $D_s$ is the count of all parcel points in the area of the $r$th source zone, and $D_{ts}$ is the count of all parcels in the area of intersection between the $r$th source zone and the target zone. In the simplified example illustrated in Figure 2, 66.67% of the population in the source zone is estimated to be in the target zone because 66.67% of the parcel center points in the source areas are also in the target zone.
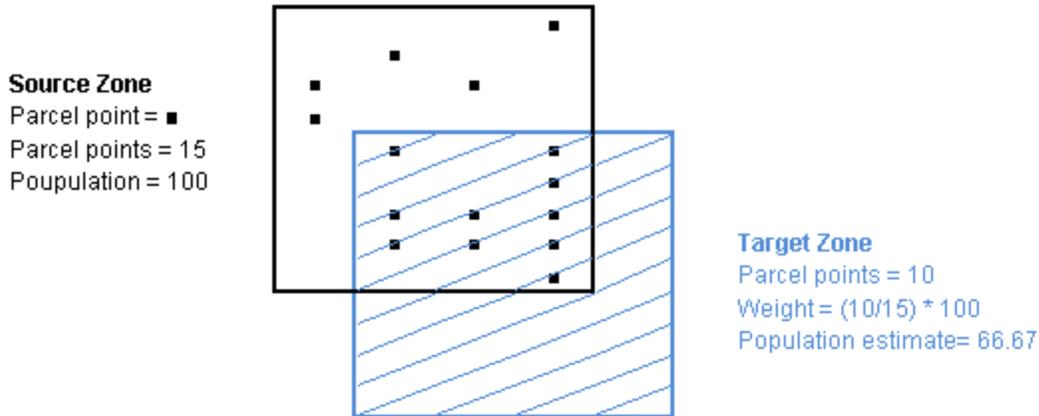


Figure 2. Parcel point-weighted areal interpolation

We believe that tax parcel centroids provide a reasonable approximation of settlement patterns and thus represent an improvement over the uniform density assumption. The SAEP makes use of parcel point-weighted areal interpolation in areas where parcel data are available. Area-weighted areal interpolation is used when parcel data are lacking.

## What are the limitations of using SAEP data?

Users need to be aware of certain limitations in the use of SAEP data. SAEP estimates are subject to errors arising from census count error, problems associated with incomplete or inaccurate source data, as well as modeling error. If users wish to calculate occupancy rates and household size ratios from SAEP data, they do so at their own risk. SAEP estimates are based on adjusted decennial census data and controlled to OFM's official April 1 intercensal and postcensal estimates. Counts of total housing and occupied housing are distributed spatially and raked to meet the official OFM totals. In some cases, our methodology may exaggerate declines in occupancy rates and erroneously increase person per household ratios

over time.  OFM would like to further caution users that trends in population and housing from the SAEP should not be used to develop long-range forecasts for small areas.  OFM recommends that long-term growth expectations for small areas be consistent with long-range county growth projections.

Estimates are not comparable across SAEP data releases due to boundary, data, and methodology changes.  It is recommended that users limit comparisons over time to a given data release (e.g., comparing the 2022 population to the 2020 population using the 2022 data release is perfectly acceptable but we do not recommend comparing values between SAEP data releases).  Also, please be aware that estimates for the current year are not as reliable as estimates for same year in the subsequent release because more spatially precise geocoded building permits or assessor data replace less precise postal data when they become available.

## Should I seek out other sources of regional and small area data?

The SAEP produces a consistent set of small area data for use in statewide applications.  SAEP estimates are not meant to replace regional council of government and other local estimate programs that are integrated with regional transportation and land use planning.  County and regional planning agencies are able to make localized adjustments to their data that are difficult to implement on a statewide basis.  If your analysis area is completely within the jurisdiction of your local planning agency we suggest you contact them for your data needs.

- [Benton-Franklin Council of Governments](#) (Benton, Franklin, and Walla Walla counties)
- [Cowlitz-Wahkiakum Council of Governments](#)
- [Grays Harbor Council of Governments](#)
- [Puget Sound Regional Council](#) (King, Kitsap, Pierce, and Snohomish counties)
- [Spokane Regional Transportation Council](#)
- [Thurston Regional Planning Council](#)
- [Whatcom Council of Governments](#)

## How can I access SAEP data?

The SAEP data products can be accessed from OFM's Small Area Estimate Program website at (http://www.ofm.wa.gov/pop/smallarea/) and OFM's Special Areas website at (http://www.ofm.wa.gov/pop/special/)

## What kinds of SAEP data products are available?

The primary SAEP data products are tables and GIS shapefiles.  SAEP data products are summarized by geography.  The smallest level of geography available for public release is the census block group.

Each table contains information on the following demographic measures: total population, household population, group quarters population, total housing units, and occupied housing units.  For each demographic measure, the following variables are included: Census 2020 values, OFM adjusted 2020 values, SAEP postcensal estimates from 2021-present, change since 2020, and percent change since 2020.  Also available on the SAEP website are GIS metadata and reference maps for specific geographies.

## What is the update frequency for SAEP data products?

SAEP data are updated annually in the fall.  The intercensal estimates are essentially static.  The postcensal estimates are updated annually each fall.  The data are typically made available on our website in early October.  This means that current year estimates will not be available until the beginning of the fourth quarter (e.g., 2022 estimates will become available in the latter part of 2022).

## Is there a suggested citation?

No.  Use whatever citation style your organization, field, or discipline uses but please try and reference the SAEP data layer and the specific SAEP release.  Here's an example of an appropriate citation in APA style:

Washington State Office of Financial Management, Forecasting Division (2022). Small Area Estimate Program: Census Block Groups [Data file]. Retrieved from http://www.ofm.wa.gov/ pop/smallarea/data/xlsx/saep_blkgrp.xlsx.

## Is there a liability disclaimer?

Yes.  The disclaimer shown below applies to many OFM data products including SAEP data products.

*By using these data the user agrees that the Washington State Office of Financial Management shall not be liable for any activity involving these data with regard to lost profits or savings or any other consequential damages; or the fitness for use of the data for a particular purpose; or the installation of the data, its use, or the results obtained.*

# References

Bureau, US Census. "Disclosure Avoidance for the 2020 Census: An Introduction." Census.gov, www.census.gov/library/publications/2021/decennial/2020-census-disclosure-avoidance-handbook.html. Accessed 17 Oct. 2022.

Fisher P. & Langford M. (1996). Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation by dasymetric mapping. *Professional Geographer, 48*(3), 299-309.

Reese-Cassal, K. (2007, March). <u>An evaluation of land parcel-weighted areal interpolation in small area estimates</u>. Paper presented at the 2007 Population Association of America Conference, New York City, NY.