

# Small Area Demographic Estimates User Notes and Errata

*Preliminary Vintage 20250107*

## Overview

Small Area Demographic Estimates (SADE) data for 2020-2024 are currently available.

These estimates are considered “preliminary” because they use 2010 Census “some other race” data. New data from the 2020 Census are expected to be released later this year and will be incorporated into future SADE releases.

OFM recommends that users limit their analysis to a specific SADE data vintage (e.g., 20250107) because estimates for prior years are updated if new or more accurate information becomes available.

Impacts from the COVID-19 pandemic are apparent in the 2020-2024 data. Many state and private institutions had limited occupancy due to pandemic-related restrictions. Communities with large populations living in group quarters (such as college or university housing, prisons, jails, nursing homes, and most group living situations) were impacted by pandemic regulations. The SADE estimates reflect those changes and users should be aware that the year-to-year variation in the data is greater in the current release than pre-pandemic releases.

## SADE Data Inputs

The SADE models utilizes Census data and OFM; county race estimates, county age and sex estimates, and Small Area Estimates Program (SAEP) data. The SADE model uses an iterative proportional fitting procedure to ensure consistency with these products. The SADE model produces estimates of age, sex, race, and ethnicity by small geographic area and year.

These are the first SADE estimates based on the 2020 Census Privacy-Protected Microdata File (PPMF). The PPMF is a privacy protected person level data file. The file includes simulated person level data with demographic characteristics at the census block level. Previous SADE releases combined several census data tables at various levels of geography to simulate the population characteristics at the census block level. For more information about the PPMF on the [Census Bureau website](#).

## County Race Estimates

Expect changes to race and ethnicity characteristics in the next SADE release. OFM 2020-2024 race and Hispanic origin controls are considered preliminary as they are based on modified race factors from the 2010 Census. The release date for the 2020 Census modified race data is expected to be later this year. OFM will begin using that data when they become available.

2020 Census and OFM 2020-2024 preliminary race and ethnicity data are based on the Federal Office of Management and Budget (OMB) 1997 statistical reporting data standard. On March 28, 2024, OMB announced the revision of the standard through Statistical Policy Directive 15 (SPD 15). As per the directive, Federal agencies are to start using the new standard by March 28, 2029. Information about the SPD 15 revision can be found at <https://spd15revision.gov/>.

## County Age and Sex Estimates

The 2020 Census data exhibited a large amount of “age-heaping”. Age heaping refers to the phenomenon where respondents are more likely to round their age to multiples of 5 rather than provide a specific age. This results in distorted population counts for these ages. For example, the population aged 25 might be larger than expected, while at the same time ages 24 and 26 might be lower than expected.

OFM smoothed the age-heaped single-year age and sex data at the county-level using a Bayesian random walk procedure based on the 2020 Census data that was consistent within 5-year age group totals and controlled that OFM adjusted totals (see [Population Trends – Table 5](#) for more information).

For information about age heaping in the 2020 Census, please see:

<https://www.census.gov/newsroom/blogs/random-samplings/2023/05/age-heaping-2020-census-dhc.html>.

## Small Area Estimate Program (SAEP) Estimates

Because of new census privacy protection measures implemented by the Census Bureau, OFM adjusted census block populations and housing occupancy state to make them logically consistent. Please see the [SAEP User Guide](#) for additional information about the data and methods behind the SAEP estimates, as well as limitations in their use.

Note that OFM makes corrections and adjustments to decennial census populations, and OFM data may not match decennial data. Our adjustments are intended to increase accuracy

and correct issues. Adjustments will manifest differently in different products, but we strive to make our products consistent with the total populations published in OFM's April 1 county and city population estimates. For more information about April 1 estimate adjustments, please see Table 5 in OFM's [Population Trends](#).

## 2000-2010 Intercensal Estimates

The SADE 2000-2010 intercensal estimates have been updated to 2020 geographies. We did not make any changes to the method or input data for the existing estimates.

## 2010-2020 Intercensal Estimates

The SADE 2010-2020 intercensal estimates use 2020 geographies. To create the estimates, we used an interpolation method based on the method used by the Census Bureau. The goal is to adjust the 2010-based postcensal estimates so they line up with the 2020-based postcensal estimates in the year 2020. The method creates a time series that is informed by both the older and the newer postcensal estimates. When comparing the intercensal estimates with the 2010-based postcensal estimates, you will see larger differences in later years than in earlier years.

Intercensal interpolation assumes that postcensal estimate accuracy decreases over time – 2010 is the most accurate year since it is the closest to the 2010 Census, and 2020 is the least accurate year since it is 10 years away from the 2010 Census. SADE assumes that the change in accuracy follows a linear progression, similar to how the Census Bureau interpolates small populations. Using the 2020-based postcensal estimates as a desired endpoint, SADE adjusts the 2010-based postcensal estimates, keeping decreasing accuracy over time in mind. The end result has smaller adjustments closer to 2010 and larger adjustments further from 2010.

Because the intercensal estimates rely on the 2020-based postcensal estimates, any substantive changes to the postcensal estimates method will require a re-release of the intercensals. We expect such a re-release to occur when the SADE postcensal estimates are updated with upcoming 2020 Census race data.

## Changes to the SADE Model

The SADE model used for vintage 20250107 has been updated from previous vintages.

- Changing 2020 Census data inputs. The Census Bureau has released the 2020 Privacy-Protected Microdata File (PPMF), an individual-based dataset for the 2020 Census. The

SADE model has been updated to use this data, allowing estimates to be processed at smaller geographies without relying on larger geography data.

- Changing how age-heaping corrections are applied. The model has been updated to be consistent with single-year age county estimates, rather than 5-year age groups.
- Removed data records for geographies with near-zero populations. This does not affect SADE estimates, but makes the final output files smaller and easier to process for our data users.

## Future Work

OFM will update the estimates with 2020 modified race factors when they are released by the Census Bureau. This will impact intercensal (2010-2019 and postcensal (2020-current year) estimates. We are exploring updates to improve how growing populations are represented throughout the times series, as well as estimating population subsets (such as the group quarters population). We also are planning on adding race classifications from the newly defined OMB SPD-15.