

What is the Small Area Demographic Estimates (SADE) model?

The Washington State Office of Financial Management's (OFM) Small Area Demographic Estimates (SADE) model produces a consistent set of small area population data for statewide applications. 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.

The SADE model produces population estimates with "characteristics," a term of art in demography referring to datasets that have population counts subdivided by age, sex, race and ethnicity. The model produces estimates of population at very small geographies; internally the estimates are performed at the block level, but SADE data are aggregated and released only at larger geographies.

What race category system does SADE use?

SADE estimates comply with the U.S. Office of Management and Budget (OMB) 1997 revised standards for the collection, tabulation and presentation of federal data on race and ethnicity. These revised OMB standards identify a minimum of five racial categories: White, Black or African American, American Indian and Alaska Native (AIAN), Asian, and Native Hawaiian and Other Pacific Islander (NHOPI). There is also a multiracial category, which is a tabulated statistic to reflect respondents' selection of two or more races on the questionnaire.

The SADE model uses the Census Bureau's 2010 Modified Race Summary (MARS) file for the benchmark year. The procedure allocates respondents who selected "Some Other Race" in the federal census to one or more of the five OMB 1997 race categories. For more information on MARS please see the following document: <https://www.census.gov/popest/methodology/2013-natstcopr-meth.pdf>.

How are the SADE estimates developed?

SADE estimates are produced using an Iterative Proportional Fitting (IPF) procedure. IPF is a mathematical scaling procedure, which produces the initial estimated population by age, sex and race and also converges the estimates to the constraining totals developed from multiple sources.

To begin, an IPF procedure is executed to allocate census 2010 MARS data at the county level to census blocks. To move one year forward, IPF gradually adjusts the benchmark estimate through repeated calculations to fit the various constraints. These constraints, also known as controls, are annual total population by block from OFM's Small Area Estimates Program (SAEP), OFM's county level population estimates by age and sex and OFM's county level estimates by race and ethnicity. For information about these estimates please visit OFM's website at <http://www.ofm.wa.gov/pop/default.asp>.

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

It is crucial to understand that SADE 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: <http://www.ofm.wa.gov/pop/april1/default.asp>. SADE postcensal estimates are controlled to OFM's official April 1 population estimates by ensuring that they sum to April 1 estimates at the county level.

What are the limitations of using SADE data?

SADE is controlled to SAEP at the block geography for each year, so all the limitations that apply to SAEP estimates apply to SADE as well.

SADE estimates use characteristics proportions from the 2010 Census. These proportions don't reflect aging or migration of the small area population, so SADE data cannot be used to examine trends in small areas.

Because characteristics are calculated for such small areas, there are fractional population counts. These fractional numbers must be summed and rounded to yield more reasonable figures in geographies larger than blocks.

For how small a geography are the SADE estimates valid?

We suggest only using SADE data at geographies more or less on the same scale as census tracts, about 5,000 people per geography, using five year age groupings, six race categories (including two or more races) and two Hispanic categories (Hispanic and non-Hispanic). There are two main reasons for this suggestion:

- (1) The most detailed data with characteristics available for 2010 are the "PCT" files, which report data at the census tract geography using six race categories, single year age groups and sex; any more fine-grained reporting is modeled only and not empirical.
- (2) With sixteen age groupings, two sex categories, two Hispanic categories and six race categories, cell sizes get too small at small geographies to be statistically stable. For example, if a geography has 5,000 people (the normative population for a census tract), with a perfectly even distribution of population among categories, each cell would have $5,000 / (16 \times 2 \times 2 \times 6) = 5,000 / 384 =$ only 13 people in each data cell. At such small numbers, cell data can quickly become unstable. The uneven distribution of population amongst cells exacerbates this problem, for example 80 year old Hispanic NHOPI females compared to 50 year old non-Hispanic white females.

Note that custom geographies need not be the same as census tracts and still meet our recommendations, as long as they have a similar population size or bigger. Analyses can also be performed with less detailed categories, reducing the number of cells and the associated problems.

Why are there decimals, 0.5 of a person for example?

OFM leaves the data in decimals for two reasons. First, we want the results to total to control numbers correctly and rounding would make that impossible. Second, we want to remind users that these data are estimates only and not actual enumerations. See above entry regarding "limitations."

Are small numbers in the estimate “identifiable”?

Some of the cells in the dataset have very small, but nonzero, numbers (for example Asian Hispanic 40-45 males in Adams County); one might think these small numbers allow a user to use SADE data to determine private information such as ethnicity for a single person. However, the estimates are modeled with multiple inputs, none of which are identifiable, so the specific cells don’t reflect “real people”, except statistically. The only input that is based on enumerated people is the 2010 census and the Census Bureau de-identifies small groups. All the other data are de-aggregated from county-based datasets that are not identifiable in the first place.

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

The SADE model produces a consistent set of small area data for use in statewide applications. SADE 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 first.

How can I access SADE data?

SADE estimates can be obtained from OFM’s website at <http://www.ofm.wa.gov/pop/asr/default.asp>.

What kinds of SADE data products are available?

SADE data are available as an annual time series from 2000 onwards for the following geographies:

- State
- County
- Congressional District
- Legislative District
- School District
- Census Tract

There are three tables for each geography representing the total population, the Hispanic population and the non-Hispanic population. The data are tabulated by OMB 1997 race category, sex and five year age group.

What is the update frequency for SADE data products?

SADE data are updated annually in the fall. The intercensal estimates (2001-2009) and the census year data (2000 and 2010) remain the same with each yearly release, but the postcensal estimates (2011 forward) are updated annually. The data are typically made available on our website by the end of the year. This means that current year estimates will not be available until the end of the fourth quarter of that year (e.g., the 2015 estimates will become available in the latter part of 2015).

What if I need estimates for a geography or an age group other than the ones available on the website?

Contact OFM to see about a custom tabulation. If we have time and the tabulation is possible technically, we may be able to provide it. We would likely charge for analyst time to produce the tabulation and we would make it publicly available.

Is there a suggested citation?

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

Washington State Office of Financial Management, Forecasting Division (2015). Small Area Demographic Estimates: Census Tracts [Data file]. Retrieved from http://www.ofm.wa.gov/pop/asr/sade/sade_tract10_t_5y.xlsx.

Is there a liability disclaimer?

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