

POPULATION-WEIGHTING FOR MEASURES DISSEMINATED VIA THE TRACKING NETWORK

Background:

Raster datasets (continuous surfaces of regularly sized grids) are often summarized to administrative (e.g., county, census tract) boundaries using zonal statistic methods. Classic zonal statistic methods take a simple average of all grid cell centroids that fall within a polygon boundary. However, this approach does not consider the underlying population distribution of an area and may result in misleading exposure values. Consider counties in the Western U.S. that are typically large in size and may be dominated by arid, desert landscapes. The majority of the population in such a locale may be concentrated in a small area of the county. Yet, in classic zonal statistic methods, equal weight is given to all grid cells that fall within the county boundary. Population-weighting seeks to account for underlying population distributions and calculate exposure values that are more representative of what the majority of the population experiences. Population-weighting is calculated using a weighted average approach.

Data Criteria:

Data eligible for population-weighting must be raster datasets (e.g., temperature, air pollution) and ideally have complete spatial coverage (i.e., no missing data cells).

Methods:

Using population-weighted centroid shapefile data for census tracts from the U.S. Census Bureau ([Centers of Population \[census.gov\]](https://www.census.gov/popest/data/totals/shapefiles.html)), extract values of the raster dataset to these points. Next, calculate a weighted average of the extracted raster values by using the proportion of the total county population that each census tract represents as the weight as shown in equation 1. (1) The sum of the census tract weights in a given county should be equal to 1.

$$\text{Equation 1 Population weighted average} = \sum (W_i * R_i)$$

In equation 1, W_i is the proportion of the total county population that the i th census tract represents and R_i is the raster value extracted to the i th census tract population-weighted centroid.

If there are population-weighted census tract centroids that fall within missing data raster cells (e.g., poor quality cells in satellite-derived data removed due to cloudiness), then the data for that entire county should be set to missing.

Considerations:

Population-weighted census tract centroids are updated by the U.S. Census Bureau each decennial census year (e.g., 2010, 2020, 2030). Depending on the dataset, population-weighting may or may not be applied using population-weighted centroids of the most recent census decade. For any environmental dataset, where possible, the Tracking Program will reprocess to the most current

geographic boundaries. In those cases, historical data will need to be reprocessed using the latest population-weighted census tract centroids as inputs into the formula listed in the methods section. If reprocessing to the most current geographic boundaries is not possible for a given dataset for any reason, then the historical data will not need to be reprocessed using the latest population-weighted census tract centroids. In those cases, the population-weighted census tract centroids would be used for the corresponding decade of data (e.g., the 2010 population-weighted census tract centroids would be used for 2010-2019 data years).

References

- 1) Spangler KR, Weinberger KR, Wellenius GA. Suitability of gridded climate datasets for use in environmental epidemiology. J Expo Sci Environ Epidemiol. 2019 Oct 25;29(6):777-89.



CDC'S ENVIRONMENTAL PUBLIC HEALTH TRACKING PROGRAM

Contact us: trackingsupport@cdc.gov

Visit the Tracking Network today: www.cdc.gov/ephtracking

Follow us on social media:

- X (@CDC_EPHTracking)
- Facebook (facebook.com/CDCEPHTracking)