

2024 COMMUNITY RESILIENCE ESTIMATES

Detailed Technical Documentation

Updated January 29, 2026¹

Small Area Estimates Program

Social, Economic, and Housing Statistics Division

U.S. Census Bureau, Department of Commerce

¹ The U.S. Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product (Data Management System (DMS) number: P-7517412; Disclosure Review Board (DRB) approval number: CBDRB-FY26-036).

Contents

1. Introduction	3
1.1 Overview	3
1.2 Data Sources.....	3
1.2.1 Decennial Census.....	3
1.2.2 Population Estimates Program	3
1.2.3 American Community Survey	4
1.3 Notation	4
1.4 Assumptions for the Components of Social Vulnerability	5
2. Auxiliary Population Estimates	6
2.1 Initial 2020 Population Estimates	7
2.2 2024 Population Estimates	8
3. Estimation Layers	9
3.1 Tract Universe	9
3.2 Layer Definitions.....	9
4. Component of Social Vulnerability Estimation	9
5. Composite Estimators	11
5.1 Direct Estimate Variance	11
5.2 Mean Squared Error Estimation.....	12
5.3 Composite Formula	13
5.4 Tract-County Relationship.....	14
5.5 Processing Notes	17
6. Ranking.....	18
6.1 County	18
6.2 Tract.....	19
7. References	20

1. Introduction

1.1 Overview

The general framework of the Community Resilience Estimates (CRE) follows a standard area-level approach from small-area estimation: a weighted average is taken of a direct estimator and an indirect estimator to produce a composite estimator; see Rao and Molina (2015) for more information on small-area estimation methods. The composite estimator is less volatile than either of the two original estimators. In the case of CRE, the direct estimates are 2024 American Community Survey (ACS) estimates for the number of individuals within a given tract who possess zero, one or two, or three or more components of social vulnerability (SV). The indirect or synthetic estimates, are developed from applying modeled proportions to auxiliary population data. The following sections will look at each step in more detail.

1.2 Data Sources

There are three sources of data for CRE: the 2020 Decennial Census, the Population Estimates Program (PEP) of the United States Census Bureau, and ACS 1-year estimates.

1.2.1 Decennial Census

Every ten years, the United States Census Bureau conducts the Decennial Census. The intent of the Decennial Census is to enumerate the United States resident population for the purposes of apportionment in the United States House of Representatives. Information collected during the Census includes demographic data such as age, race, and sex. The 2024 CRE estimates use published data from the Privacy-Protected Microdata File (PPMF) and Modified Age & Race Census (MARC) file. The creation of both files relied on disclosure avoidance techniques to protect respondent privacy. The CRE, however, ignores the corresponding uncertainty created by such measures (e.g., we treat the MARC totals as constants in our modelling).

1.2.2 Population Estimates Program

The Population Estimates Program (PEP) of the United States Census Bureau produces postcensal resident population estimates for geographies within the United States and Puerto Rico that update the size and demographic composition of the population since the last decennial census. Population estimates are published annually for households and group quarters (GQ) and are broken down by age, race, sex, and Hispanic origin. The 2024 CRE estimates use internal data which is also broken down by household/GQ type. CRE methodology assumes county-level PEP estimates are measured without statistical error.

1.2.3 American Community Survey

The American Community Survey (ACS) samples United States households yearly to gather information on housing, population, demographics, health, and finances. We used internal ACS microdata from the 2024 ACS survey to produce ACS estimates for various geographies. Unlike the Decennial data and PEP estimates, ACS estimates contain statistical error, in the form of sampling variability, which is estimated using replicate weights.

1.3 Notation

We use unpublished tract-level population estimates for combinations of:

- four age groups
 - 0-17,18-44,45-64,65+
- five race/ethnicity categories
 - White non-Hispanic
 - Black non-Hispanic
 - American Indian/Alaska Native non-Hispanic
 - Other non-Hispanic
 - Hispanic
- two sex categories

broken down by households and seven group quarters (GQ) types:

1. correctional facilities
2. juvenile institutions
3. nursing homes
4. institutional GQ, such as in-patient hospice facilities and military treatment facilities
5. college dormitories
6. military barracks
7. non-institutional GQ, such as group homes.

Only GQ types 3, 4, 6, and 7 are in-universe for CRE. All possible combinations of the four age groups and five race/ethnicity categories are denoted by *AR* and similarly *ARS* indicates all possible combinations of the age groups, race/ethnicity categories, and two sex categories.

This document also discusses racial classifications and age groupings available in 2020 Decennial Census data but not listed above. The Census race groups include White, Black, Asian, Native Hawaiian and Pacific Islander, American Indian and Alaskan Native, Some Other Race, and combinations of two, three, four, five, or six races. Throughout, we refer to a combination of age, race/ethnicity, and sex as a demographic combination, or simply, a

demographic. Often a demographic will be indicated using subscripts such as age, h, rc, sx for a specific age group age , Hispanic origin h , race category rc , and sex category sx or denoted more abstractly using the single letter j .

For a given ACS respondent y , w_y will denote the respondent's ACS weight, and for a demographic category j , social vulnerability component k , and geographic area g , let:

- $ACS_{j,k}$ represent the collection of all ACS respondents with a demographic combination j and social vulnerability component k .
- $ACS_{g,j,k}$ be the subset of respondents from $ACS_{j,k}$ with residence in g .
- $ACSPop_{g,j}$ denote the ACS population estimate for demographic combination j and geography g .

An area's total in-universe population, i.e. households and all in-scope GQ, will be denoted by $POPUNI_{g,j}$ and $POPUNI_g$. Similar descriptive names are used for other population counts, such as 2020 Decennial counts and PEP estimates.

We use $dg \in \{0,12,3\}$ to differentiate between different SV groups where:

- 0 denotes the zero SV group
- 12 denotes the one or two SV group
- 3 denotes the three or more SV group

For instance, $ACS_{g,j,dg}$ is the subset of ACS respondents in demographic j and geography g in the dg SV group.

1.4 Assumptions for the Components of Social Vulnerability

Residents in nursing homes and other types of institutional GQ are only eligible for a subset of the ten components of social vulnerability:

- crowding
- low income to poverty ratio (IPR)
- communication barrier
- age 65 and over
- presence of a disability

We assume the residents do not possess any of the remaining five social vulnerability components. The crowding social vulnerability component is assumed to hold for all residents in these GQ. For military barracks we estimate only one social vulnerability component, the

crowding component, i.e. we assume all residents in military barracks have this social vulnerability component.

2. Auxiliary Population Estimates

In the 2018 (experimental) and 2019 versions of CRE, the starting points of our population estimates were tract-level 2010 Census tabulations by age x race/ethnicity x sex for households and the seven GQ types. The only results from the 2020 Census available in time for use in CRE 2021 were the Decennial Census Public Law 94-171 (PL94) Redistricting Data summary tables. The PL94 summary tables contain population totals for different age x race/ethnicity combinations down to the block level but lack a great deal of information previously used for the CRE auxiliary population. The CRE 2021 methodology imputed this missing information using 2010 Census data. In May 2023, more detailed demographic data for a variety of geography levels from the 2020 Census became available from the Demographic Housing Characteristics File (DHC) file. Both the PL94 and DHC data were used to create population estimates for CRE 2022 and CRE 2023.

For CRE 2024, we use the PPMF and MARC files. The PPMF release has separate person-level and housing-unit level files. The files are synthetic data which result from applying a disclosure avoidance algorithm to 2020 Decennial responses to ensure respondent confidentiality. Earlier releases of 2020 Census data, such as the DHC and PL94 files, were based on the PPMF. The files include geographic information for each person or housing unit down to the Census block level. Of particular importance for the CRE, the person-level file contains demographic information, e.g. age or sex, and whether the individual resides in a household or GQ. The MARC file contains county-level Decennial tabulations broken down by age, sex, and modified race for both those residing in households and group quarters. Modified race refers to reassigning the racial information of Decennial respondents so that it aligns with the racial categories used to create PEP's population estimates. More information about the creation, uses, and contents of the MARC and PPMF can be found in U.S. Census Bureau (2024a) and U.S. Census Bureau (2024b).

Our strategy for creating population estimates can be broken into three major steps. First, we use the PPMF person-level data to create tract-level counts by age, race/ethnicity, and sex for households and each of the seven GQ types. Second, we control our tract-level estimates to the MARC counts. Finally, we use the Vintage 2024 PEP estimates to update the Decennial tract counts. Briefly, we describe the first and second steps in Subsection 2.1 and focus on the third step in Subsection 2.2.

2.1 Initial 2020 Population Estimates

We start by aggregating the PPMF data up to the tract level for cross classifications of our specified age groups, sex, and the following Census race/ethnicity categories: Hispanic, Non-Hispanic White, Non-Hispanic White and Some Other Race, Non-Hispanic Black, Non-Hispanic Black and Some Other Race, non-Hispanic American Indian and Alaskan Native, Non-Hispanic American Indian and Alaskan Native and Some Other Race, and Some Other Race alone. Note that Census race categories include combinations of Some Other Race while PEP categories do not, so at this point we need to reconcile the Census race categories with the PEP categories. Racial categories involving Some Other Race in combination with a valid PEP racial category, i.e. everything but Some Other Race alone, are reclassified using the non-Some Other Race category; afterward the demographic breakouts are recalculated. For example, instances of ‘Some Other Race and White’ become ‘White’. The only unsupported category remaining after this re-aggregation is Some Other Race alone, which we distribute according to a tract-level household or state-level GQ distribution of valid PEP race/ethnicity categories by age and then reaggregate. For a given tract t and demographic j denote the results of this final aggregation as $HH_CRE_{t,j}$ for households and $GQ1_CRE_{t,j}, \dots, GQ7_CRE_{t,j}$ for each of the seven GQ types.

When summed to the county level, the household and GQ totals do not necessarily match the county totals in the MARC file. Therefore, we rake, or control, our totals to county MARC values. Since we also want a tract’s overall population for households and each of the GQ types to remain unchanged, we alternately rake to the MARC totals, original tract-level household counts, and GQ counts.

The raking process is described below for households; raking for GQ is similar. Let $HH_MARC_{c,j}$ be the household MARC count for county c and demographic j and $HH_CRE_{t,j}$ be as above. Then the raking proceeds in the following steps.

1. Set $hh_marc_rk_{c,j} = \frac{HH_MARC_{c,j}}{\sum_{t \in c} HH_CRE_{t,j}}$
2. Redefine $HH_CRE_{t,j} = hh_marc_rk_{c,j} HH_CRE_{t,j}$
3. Control back to the original household total population HH_TOT_t , which can be obtained directly from the PPMF or PL94 and DHC tables, giving:

$$hh_tr_rk_{t,j} = \frac{HH_TOT_t}{\sum_j HH_CRE_{t,j}}$$

4. Redefine once more setting $HH_CRE_{t,j} = hh_tr_rk_{t,j} HH_CRE_{t,j}$

Steps 1-4 are iterated numerous times to ensure all raking factors between successive iterations differ in absolute value by no more than .01. Use $HH_POP20_{t,j}$ to denote the post-raking household count for tract t and demographic j . Similarly, allow $GQ1_POP20_{t,j}$, ..., $GQ7_POP20_{t,j}$ to represent the corresponding GQ counts.

2.2 2024 Population Estimates

We calculate growth rates by demographic and household/GQ type using Vintage 2024 PEP data. For instance, if c is a county and j a demographic, and $HH_PEPPOP24_{c,j}$ is the Vintage 2024 PEP household pop estimate for 2024 and $HH_PEP20BASE_{c,j}$ the corresponding base pop estimate, set:

$$HH_gr_{c,j,24} = \frac{HH_PEPPOP24_{c,j}}{HH_PEP20BASE_{c,j}}$$

Apply the growth rate to $HH_POP20_{t,j}$ for each tract t in c , obtaining $HH_POP24_{t,j}$. Estimates for $GQ1_POP24_{t,j}$, ..., and $GQ7_POP24_{t,j}$ are handled similarly to households.² Results are summed separately for households and each GQ type at the tract level to create population estimates which are then adjusted, or raked, so that their sum is equal to PEP totals at the county level for households and state level for each of the GQ types. In the case of households, for example, we define:

$$rk_c = \frac{\sum_j HH_PEPPOP24_{c,j}}{\sum_{t \in c} \sum_j HH_POP24_{t,j}}$$

for a given county c and multiply $HH_POP24_{t,j}$ by rk_c to obtain a raked estimate $HH_POPRK24_{t,j}$ for each tract t in c and demographic j . The raked values become our final 2024 population estimates. In the remainder of the document, when we refer to POP or $POPUNI$, we are referencing these raked values.

² A small amount of reconciliation is done for both households and GQ, e.g. we account for demographic groups added since the Decennial Census to the PEP data by splitting a county demographic total among constituent tracts.

3. Estimation Layers

3.1 Tract Universe

The 2024 Census geography vintage contains 84,415 tracts which we split into two groups: residential and non-residential. 83,808 tracts are classified as residential, i.e. the tract's estimated in-universe population is non-zero. The remaining 607 tracts are assumed to have no in-scope residents. In Sections 4-6, any reference to tracts will be for residential, unless otherwise noted.

3.2 Layer Definitions

Only residential tracts are assigned to a layers, and construction of the a layers is based on two criteria:

- tract-level population density
- concentration of urban population (UCI) within a given tract and those near it. UCI is the distance-weighted sum of a tract's population and those tracts near it, scaled so that a UCI of 100-250 is a moderate concentration and over 250 is high.

Using the two criteria, we define four urbanization strata³:

- UStrat1 consisting of tracts with UCI < 100
- UStrat2 consisting of tracts with UCI 100-250 and high-density proportion less than 0.5
- UStrat3 consisting of tracts with UCI 100-250 and high-density proportion greater than or equal to 0.5
- UStrat4 consisting of tracts with UCI > 250.

The a layers are the urbanization strata broken down by Census Division. Since there are nine Census Divisions and four urbanization strata there are 36 total a layers. Note that a given a layer can cross state and county boundaries and are not necessarily contiguous.

4. Component of Social Vulnerability Estimation

We estimate SV group membership for households and the seventh GQ type together and separate from the remaining GQ types. We define the ratio $r_{a,j,dg,HH7}$ as the proportion of

³ The East South Central Division has slightly lowered thresholds, otherwise the populations for some of the estimation layers would be unacceptably small.

individuals within an a layer a and demographic $j \in AR$ residing in a household or appropriate GQ type with zero, one or two, or three or more components of social vulnerability:

$$r_{a,j,dg,HH7} = \frac{\sum_{y \in ACS_{a,j,dg,HH7}} w_y}{\sum_{y \in ACS_{a,j,HH7}} w_y}$$

We use the subscript $HH7$ for the ACS data to indicate it is a subset to respondents living in households and appropriate non-institutional GQ. The proportion $r_{a,j,dg,HH7}$ is broken up into a marginal and conditional probability using Bayes' Theorem. The marginal is modelled in a linear regression within each a layer and then the joint values are reconstructed.⁴ We denote the final modelled value of $r_{a,j,dg,HH7}$ as $\tilde{r}_{a,j,dg,HH7}$. For a given tract t within an a layer a , and $j \in AR$ we estimate the number of individuals in group dg as follows:

$$\hat{Y}_{t,j,dg,HH7} = \tilde{r}_{a,j,dg,HH7} * POP_{t,j,HH7}$$

Estimation for the remaining GQ is similar (we estimate GQ types 3 and 4 together and GQ type 6 by itself) and yields values $\hat{Y}_{t,j,dg,34}$ and $\hat{Y}_{t,j,dg,6}$. Summing over demographic groups:

$$\begin{aligned} \hat{Y}_{t,0} &= \sum_{j \in AR} (\hat{Y}_{t,j,0,HH7} + \hat{Y}_{t,j,0,34} + \hat{Y}_{t,j,0,6}) & \hat{R}_{t,0} &= \frac{\hat{Y}_{t,0}}{POPUNI_t} \\ \hat{Y}_{t,12} &= \sum_{j \in AR} (\hat{Y}_{t,j,12,HH7} + \hat{Y}_{t,j,12,34} + \hat{Y}_{t,j,12,6}) & \hat{R}_{t,12} &= \frac{\hat{Y}_{t,12}}{POPUNI_t} \\ \hat{Y}_{t,3} &= \sum_{j \in AR} (\hat{Y}_{t,j,3,HH7} + \hat{Y}_{t,j,3,34} + \hat{Y}_{t,j,3,6}) & \hat{R}_{t,3} &= \frac{\hat{Y}_{t,3}}{POPUNI_t} \end{aligned}$$

We let $r_{t,0}$, $r_{t,12}$, and $r_{t,3}$ denote the ACS estimates for the proportion of individuals within tract t who have zero components of social vulnerability, one or two components of social vulnerability, and three or more components of social vulnerability, respectively.

⁴ The modelling helps to account for differences between ACS population totals and has a limited impact on the final results.

5. Composite Estimators

5.1 Direct Estimate Variance

Estimates for the variance of $r_{t,dg}$ with $dg = 0, 12$, or 3 can be calculated directly from the ACS microdata using replicate weights. To improve the reliability of these variance estimates, we smoothed them using a generalized variance function (GVF). The CRE GVF is analogous to GVFs that have been used in other contexts (Bauder et al., 2024), and its exact form is given by:

$$r_gvfvar_{t,dg} = C\hat{R}_{t,dg}(1 - \hat{R}_{t,dg})n_t^q$$

where n_t is the unweighted ACS sample size of tract t and C and q are parameters to be estimated. Note that n_t varies by tract but not values of dg , i.e. n_t is the same for all values of dg within a given tract. Values for C and q are estimated via least squares by fitting the following model:

$$\log(r_dirvar_{t,dg}) - \log(\hat{R}_{t,dg}(1 - \hat{R}_{t,dg})) = \beta_0 + \beta_1 \log n_t + \epsilon_t$$

We used the estimates for q and C to calculate a GVF variance for all tracts, even those with a sample size of 25 or smaller. Table 1 compares the GVF standard errors with the original standard errors from the ACS for the three different SV groups. The distributions of GVF standard errors have means and medians similar to those of the ACS, but the standard error of the distributions tends to be smaller.

Table 1: GVF and ACS Standard Errors by SV Groups

Group	Mean		Median		Standard Error	
	ACS	GVF	ACS	GVF	ACS	GVF
0	0.114	0.103	0.112	0.100	0.039	0.022
12	0.121	0.110	0.116	0.107	0.038	0.024
3	0.089	0.089	0.080	0.086	0.052	0.023

Source: 2024 Community Resilience Estimates.

5.2 Mean Squared Error Estimation

To estimate the mean squared error (MSE) of our synthetic estimators $\hat{R}_{t,dg}$, we used a method-of-moments (MoM) approach, roughly following the derivation outlined in (Rao and Molina, 2015, p. 43-44). If T is an arbitrary collection of tracts T , N_T in total, then our basic MoM equation for MSE is:

$$MSE_{T,dg} = \frac{1}{N_T} \left(\sum_{t \in T} (\hat{R}_{t,dg} - r_{t,dg})^2 \right) - \frac{1}{N_T} \sum_{t \in T} r_{t,dg}^2$$

A MoM variance estimator can be negative, and the choice of T , i.e. the layer of aggregation, must be made with care. Additionally, for a fixed tract t , the number of individuals with zero, one or two, and three or more social vulnerability components have a multinomial relationship and this places additional constraints on $MSE_{T,dg}$. For the time, being we'll ignore the selection of T , but note that to address the other concerns:

- We assume the MSE for a given SV group proportion estimate, $\hat{R}_{t,dg}$, is proportional to $\hat{R}_{t,dg}(1 - \hat{R}_{t,dg})$, and that the constant of proportionality can be assumed stable over a wide range of SV groups and tracts.
- We calculate the multiplicative constant of proportionality at an aggregate level, averaging across both SV groups and tracts.

Under this strategy, we modified the previous equation to average across SV groups and tracts:

$$MSE_T = \frac{1}{3N_T} \sum_{dg=0,12,3} \left(\sum_{t \in T} (\hat{R}_{t,dg} - r_{t,dg})^2 \right) - \frac{1}{3N_T} \sum_{dg=0,12,3} \sum_{t \in T} r_{t,dg}^2$$

The estimated constant, denoted F_T , was calculated as:

$$PmP_T = \frac{1}{3N_T} \sum_{dg=0,12,3} \sum_{t \in T} \hat{R}_{t,dg}(1 - \hat{R}_{t,dg})$$

$$F_T = \frac{MSE_T}{PmP_T}$$

so that for an individual tract t in T and each dg , $MSE_{t,dg} = F_T \hat{R}_{t,dg} (1 - \hat{R}_{t,dg})$. Returning to the choice of T , we created two national urbanization layers: UStrat1 combined with UStrat2 (L) and UStrat3 combined with UStrat4 (H). These layers were chosen empirically and result in estimates which are stable.

5.3 Composite Formula

Our final estimator is a shrinkage, or composite, estimator which is a weighted average of $\hat{R}_{t,dg}$ and $r_{t,dg}$ for each tract t and value of dg :

$$w_{t,dg} = \frac{r_{gvfvar_{t,dg}}}{r_{gvfvar_{t,dg}} + MSE_{t,dg}}$$

$$\tilde{R}_{t,dg} = w_{t,dg} \hat{R}_{t,dg} + (1 - w_{t,dg}) r_{t,dg}$$

See (Rao and Molina, 2015, p. 57-58) for a derivation of the weight formula. For relatively large values of $r_{gvfvar_{t,dg}}$, the synthetic estimator is weighted more heavily while for smaller values of $r_{gvfvar_{t,dg}}$ the opposite is true.

We quantify the uncertainty of our estimates $\tilde{R}_{t,dg}$ using:

$$\begin{aligned} MSE(\tilde{R}_{t,dg}) &= w_{t,dg} MSE_{t,dg} & MSE(\tilde{Z}_{t,dg}) &= MSE(\tilde{R}_{t,dg}) POPUNI_t^2 \\ RMSE(\tilde{R}_{t,dg}) &= \sqrt{MSE(\tilde{R}_{t,dg})} & RMSE(\tilde{Z}_{t,dg}) &= \sqrt{MSE(\tilde{R}_{t,dg})} POPUNI_t \\ MOE(\tilde{R}_{t,dg}) &= RMSE(\tilde{R}_{t,dg}) * 1.645 & MOE(\tilde{Z}_{t,dg}) &= MOE(\tilde{R}_{t,dg}) POPUNI_t \end{aligned}$$

where $\tilde{Z}_{t,dg} = \tilde{R}_{t,dg} POPUNI_t$. As seen in Table 2 and Table 3, the direct, indirect, and composite estimates are all very similar but the coefficients of variation (CVs)⁵ for the composite estimates are better than those for the ACS and the indirect estimates. The statistics

⁵The coefficient of variation is defined as an estimate's standard error divided by its point estimate, for instance, $\frac{\sqrt{r_{gvfvar_{t,dg}}}}{r_{t,dg}}$ for $r_{t,dg}$

in Table 2 are calculated over all tracts with ACS sample and those in Table 3 are over tracts where the ACS estimate is positive.

Table 2: Comparison of Estimates by SV Group

Group	Mean			Median		
	ACS	Indirect	Comp.	ACS	Indirect	Comp.
0	0.333	0.338	0.338	0.327	0.341	0.335
12	0.449	0.452	0.450	0.443	0.446	0.445
3	0.218	0.210	0.211	0.179	0.205	0.194

Source: 2024 Community Resilience Estimates.

Table 3: Comparison of CVs by SV Group

Group	P25			Median			P75		
	ACS	Indirect	Comp.	ACS	Indirect	Comp.	ACS	Indirect	Comp.
0	0.208	0.282	0.172	0.298	0.334	0.219	0.499	0.385	0.292
12	0.188	0.243	0.152	0.238	0.262	0.176	0.320	0.284	0.206
3	0.303	0.413	0.251	0.471	0.470	0.327	0.792	0.524	0.427

Source: 2024 Community Resilience Estimates.

5.4 Tract-County Relationship

Prior to CRE 2023, county-level aggregates were produced using two simplifying assumptions:

- all tracts within a given county had a fixed geographic correlation coefficient ($\rho = .4$ for 2018-2021 and $\rho = .2$ for 2022), and
- variances for all tracts within a given county were equal.

A different approach based on principles from spatial statistics, see Cressie (2015) or Smith (2014) for more information, was developed and used for CRE 2023-2024. Under this methodology, geographic correlations between tracts within a county are, on average, lower than the previously used ρ values of .2 and .4. County-level margins of error are consistently smaller than those seen in the CRE for 2018-2019 and 2021-2022. A detailed description of the spatial procedure for estimating county variances is below.

Given two tracts t and t' with $t \neq t'$, we want to estimate the covariance between $\tilde{Z}_{t,dg}$ and $\tilde{Z}_{t',dg}$. Since $POPUNI_t$ is considered constant, it's enough to measure the covariance between $\tilde{R}_{t,dg}$ and $\tilde{R}_{t',dg}$. Notice, we can expand $Cov(\tilde{R}_{t,dg}, \tilde{R}_{t',dg})$ into four terms:

$$\begin{aligned} Cov(\tilde{R}_{t,dg}, \tilde{R}_{t',dg}) &= wt_{t,dg}wt_{t',dg}Cov(\hat{R}_{t,dg}, \hat{R}_{t',dg}) \\ &+ (1 - wt_{t,dg})(1 - wt_{t',dg})Cov(r_{t,dg}, r_{t',dg}) \\ &+ wt_{t,dg}(1 - wt_{t',dg})Cov(\hat{R}_{t,dg}, r_{t',dg}) \\ &+ wt_{t',dg}(1 - wt_{t,dg})Cov(\hat{R}_{t',dg}, r_{t,dg}) \end{aligned}$$

We assume the covariance between ACS estimates for a given SV group is zero, i.e. the covariance is sufficiently small, compared to the other terms in this sum, as to be negligible. We also assume the covariance between the direct and indirect estimates is zero, i.e.

$Cov(\hat{R}_{t,dg}, r_{t',dg}) = Cov(\hat{R}_{t',dg}, r_{t,dg}) = 0$. Therefore, the covariance between $\tilde{R}_{t,dg}$ and $\tilde{R}_{t',dg}$ is driven by the covariance between the indirect estimates $\hat{R}_{t,dg}$ and $\hat{R}_{t',dg}$.

For a given distance $D > 0$, a MoM estimator for the indirect estimates' covariance is:

$$\overline{Cov}_{D,dg} = \frac{1}{N_D} \sum_{d(t,t')=D} (\hat{R}_{t,dg} - r_{t,dg})(\hat{R}_{t',dg} - r_{t',dg})$$

where $d(t, t')$ is the distance between t and t' , and N_D is the number of ordered tract pairs with distance equal to D . Similar to the MSE estimation, it's possible to average over the different groups to obtain:

$$\overline{Cov}_D = \frac{1}{3N_D} \sum_{dg=0,12,3} \sum_{d(t,t')=D} (\hat{R}_{t,dg} - r_{t,dg})(\hat{R}_{t',dg} - r_{t',dg})$$

Given tracts t and t' , we define $urb_{t,t'}$ as follows:

- H MSE layer if both tracts belong to the H urbanization category UStrat3 or UStrat4,
- L MSE layer if both tracts belong to the L urbanization category UStrat1 or UStrat2,
- M MSE layer if both tracts belong to different urbanization categories.

Refer to Section 5.2 for a description of UStrat1, UStrat2, UStrat3, and UStrat4. Then we can further decompose \overline{Cov}_D , allowing:

$$\overline{Cov}_{D,urb} = \frac{1}{3N_D} \sum_{dg=0,12,3} \sum_{\substack{d(t,t')=D \\ urb_{t,t'}=urb}} (\hat{R}_{t,dg} - r_{t,dg})(\hat{R}_{t',dg} - r_{t',dg})$$

where $urb \in \{H, L, M\}$.

Since distance is a continuous variable, it is not practical to perform calculations for specific distances; it is more convenient to consider distances within a given range. So instead of considering pairs of tracts where $d(t, t') = D$, we assign tract pairs to bins based on distance. The bins, which were chosen empirically, are by:

1. tenths of a mile for values less than one mile
2. one-mile increments for values of at least one mile and less than fifty miles
3. ten-mile increments for values of at least fifty miles and less than one hundred miles
4. increments of one hundred miles for values at least one hundred miles.

Notice for each bin b we have three different empirical covariances denoted by $Cov_{b,urb}$ with $urb \in \{H, L, M\}$. These empirical covariances, however, can be noisy so we smooth them. Specifically, we assume the following functional form:

$$Cov_{b,urb} = \frac{V_{urb}}{\left(1 + \frac{D_b^2}{k_{urb}^2}\right)^{\beta_{urb}}}$$

where the value of V_{urb} is the layer MSE when urb is H or L and $\sqrt{MSE_H MSE_L}$ when $urb = M$ and D_b represents the average distance for tracts in bin b (Cressie, 2015). The parameters β_{urb} and k_{urb} are estimated using non-linear least squares and the resulting estimates are denoted by $cov_{t,t'}$ for tracts t and t' in a distance bin b with $urb_{t,t'} = urb$. The covariances are translated to correlations $cor_{t,t'}$ so that we can estimate:

$$cov(\tilde{Z}_{t,dg}, \tilde{Z}_{t',dg}) = Cov(\tilde{R}_{t,dg}, \tilde{R}_{t',dg}) POPUNI_t POPUNI_{t'}$$

$$\begin{aligned}
&= wt_{t,dg}wt_{t',dg}Cov(\hat{R}_{t,dg},\hat{R}_{t',dg})POPUNI_tPOPUNI_{t'} \\
&= wt_{t,dg}wt_{t',dg}\sqrt{MSE_{t,dg}MSE_{t',dg}cor_{t,t'}}POPUNI_tPOPUNI_{t'}
\end{aligned}$$

For a given county c , we have:

$$\begin{aligned}
MSE(\tilde{Z}_{c,dg}) &= \sum_{t \in c} MSE(\tilde{Z}_{t,dg}) + \\
&\quad \sum_{\substack{t,t' \in c \\ t \neq t'}} wt_{t,dg}wt_{t',dg}\sqrt{MSE_{t,dg}MSE_{t',dg}cor_{t,t'}}POPUNI_tPOPUNI_{t'} \\
RMSE(\tilde{Z}_{c,dg}) &= \sqrt{MSE(\tilde{Z}_{c,dg})}
\end{aligned}$$

$$MOE(\tilde{Z}_{c,dg}) = RMSE(\tilde{Z}_{c,dg}) * 1.645$$

Similar to the case for tracts, we divide $MSE(\tilde{Z}_{c,dg})$, $RMSE(\tilde{Z}_{c,dg})$, and $MOE(\tilde{Z}_{c,dg})$ by the appropriate power of $POPUNI_c$ to obtain $MSE(\tilde{R}_{c,dg})$, $RMSE(\tilde{R}_{c,dg})$, and $MOE(\tilde{R}_{c,dg})$. For higher geographies, we assume mean squared errors are summable, i.e. county estimates are independent, and state estimates are independent.

5.5 Processing Notes

Non-residential tracts are tracts that have an estimated population of zero. In CRE for 2018-2019 and 2021, county-level values were used as rate estimates for these tracts, while all count estimates were set to zero. In CRE for 2022-2024, in addition to setting count estimates to zero, non-residential tracts have their rate estimates set to zero as well.

All CRE model estimates are rounded prior to publication. Final results are rounded to whole numbers for count estimates and two decimal places for rate estimates. Margins of error for rate estimates are handled similarly to the rate point estimates. Past versions of CRE (CRE 2018, 2021, and 2022) have permitted non-integer values for some or all count margins of error, but in CRE for 2023-2024, margins of error for counts are rounded to whole numbers.

6. Ranking

County and tract estimates for the three or more SV group are ranked across the nation and within states by various criteria such as weather hazard data from the National Risk Index, see (Federal Emergency Management Agency, 2023) for more details about the hazard data. Using bootstrapping principles based on those outlined in Wright et al. (2014), we obtain 90% lower/upper confidence interval bounds for the rankings.^{6,7} To express rankings as percentiles, linear scaling is applied to ensure values lie within the interval [0.01, 99.99]. For an area i , let $rk_{i,cr}$, $rk_lb_{i,cr}$, and $rk_ub_{i,cr}$ denote its ranking, lower bound, and upper bound with respect to some criteria cr . The corresponding percentiles are given by:

$$pct3_{i,cr} = m \frac{rk_{i,cr}}{M} + b$$

$$pct3_lb_{i,cr} = m \frac{rk_lb_{i,cr}}{M} + b$$

$$pct3_ub_{i,cr} = m \frac{rk_ub_{i,cr}}{M} + b$$

where M is the total number of areas being ranked and $m = 99.98$ and $b = .01$ are constants fixed across all areas. Note that the county and tract-level estimates being ranked are densely concentrated within a small interval [0, 100], so small changes in the point estimates can cause large shifts in the rankings. Additionally, there is greater separation between areas as you approach the maximum or minimum values of the distribution, which typically means less uncertainty in the ranking for areas with more extreme point estimates. That is, the ranking margins of error may be smaller for areas near the tails of the distribution and larger for those near the middle.

6.1 County

For a county c and $1 \leq i \leq 100,000$, we draw a random variable $\tilde{R}_{c,3,i} \sim N(\tilde{R}_{c,3}, MSE(\tilde{R}_{c,3}))$. For each i , we rank the counties according to $\tilde{R}_{c,3,i}$. This gives us 100,000 different rankings for

⁶ Because we assume normality, a simulation may produce a value that falls outside the interval [0, 100]. We adjust any negative simulated values to equal zero and values over 100 to equal 100 before calculating the bounds for the rankings. However, this can inflate the lower bound or deflate the upper bound particularly for cases in the tails of the distribution. In rare cases, this can cause an area's ranking to lie outside its confidence interval. When this occurs, we set $pct3_lb_{i,cr} = pct3_{i,cr}$ or $pct3_ub_{i,cr} = pct3_{i,cr}$.

⁷ Since multiple areas can have negative simulations that were set to zero during a given iteration, or draw, multiple ties can occur.

each county, from which we obtain upper and lower bounds for the county's ranking. These rankings and bounds are converted to percentiles.

6.2 Tract

Tracts are processed similarly to counties, except we also model the correlation between tracts of a given county. For each county c and $1 \leq i \leq 100,000$, we draw a vector of rates $\vec{r}_{c,i} \sim MVN(\vec{r}_c, \Sigma)$ where \vec{r}_c is a vector with each component a value $\tilde{R}_{t,3}$ for a tract t within c and Σ is a covariance matrix constructed from the values for $MSE(\tilde{R}_{t,3})$ for all tracts within c and an assumed between-tract correlation. We use one correlation for all tract pairs in a given county and allow the correlation to vary by county. Just as with the counties, we develop rankings (percentiles) for each i , from which we obtain upper and lower bounds for each tract's rank (percentile).

7. References

- Bauder, M., Luery, D., and Szelepka, S. (2024). Small Area Estimation of Health Insurance Coverage in 2010-2022. Social, Economic, and Housing Statistics Division, U.S. Census Bureau. Retrieved 11/15/2024 from: <https://www2.census.gov/programs-surveys/sahie/technical-documentation/methodology/2008-2022-methods/sahie-tech-2010-to-2022.pdf>.
- Cressie, N. (2015). *Statistics for Spatial Data*. John Wiley & Sons Incorporated.
- Federal Emergency Management Agency (2023). National Risk Index. Retrieved 12/4/2025 from: <https://hazards.fema.gov/nri/>.
- Rao, J.N.K. and Molina, I. (2015). *Small Area Estimation* (Second Edition). John Wiley & Sons Incorporated.
- Smith, T.E. (2014). *Notebook on Spatial Data Analysis* (online). Retrieved October 10, 2024, from: <https://www.seas.upenn.edu/~tesmith/NOTEBOOK/index.html> October 9.
- U.S. Census Bureau. (2024a). 2020 Modified Age & Race Census File Methodology Statement. Retrieved 9/30/2025 from: <https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/modified-race-summary-file-method/marc2020-imprace-us.pdf>
- U.S. Census Bureau. (2024b). Privacy Protected Microdata File (PPMF). Retrieved 9/30/2025 from: <https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/privacy-protected-microdata-file/2020census-privacy-protected-microdata-file.pdf>
- Wright, T. Klein, M. and Wieczorek, J. (2014). Ranking Populations Based on Sample Survey Data. *Research Report Statistics # 2014-07*. Center for Statistical Research and Methodology, U.S. Census Bureau.