

# **2023 COMMUNITY RESILIENCE ESTIMATES**

## **Detailed Technical Documentation**

Updated January 2025<sup>1</sup>

---

Small Area Estimates Program

Social, Economic, and Housing Statistics Division

U.S. Census Bureau, Department of Commerce

---

<sup>1</sup> 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-FY25-0096).

# Contents

<b>1. Introduction</b>	<b>3</b>
<b>1.1 Overview</b>	<b>3</b>
<b>1.2 Data Sources</b>	<b>3</b>
<b>1.2.1 Decennial Census</b>	<b>3</b>
<b>1.2.2 PEP</b>	<b>3</b>
<b>1.2.3 ACS</b>	<b>3</b>
<b>1.3 Notation</b>	<b>4</b>
<b>2. Auxiliary Population Estimates</b>	<b>6</b>
<b>2.1 Initial 2020 Population Estimates</b>	<b>6</b>
<b>2.2 2023 Population Estimates</b>	<b>9</b>
<b>3.1 Tract Universe</b>	<b>9</b>
<b>3.2 Layer Definitions</b>	<b>10</b>
<b>4. Component of Social Vulnerability Estimation</b>	<b>10</b>
<b>5. Composite Estimators</b>	<b>11</b>
<b>5.1 Direct Estimate Variance</b>	<b>11</b>
<b>5.2 MSE Estimation</b>	<b>12</b>
<b>5.3 Composite Formula</b>	<b>13</b>
<b>5.4 Tract-County Relationship</b>	<b>14</b>
<b>5.5 Processing Notes</b>	<b>16</b>
<b>6. Ranking</b>	<b>17</b>
<b>6.1 County</b>	<b>17</b>
<b>6.2 Tract</b>	<b>17</b>
<b>7. References</b>	<b>17</b>

# **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 2023 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. Ensuing 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 2023 CRE estimates use the Decennial Census Public Law 94-171 (PL94) Redistricting Data summary tables and Demographic Housing Characteristics File (DHC) tables from the 2020 Decennial Census. We treat the DHC and PL94 data as being constants in our modeling, even though data from both releases was injected with noise for disclosure avoidance purposes (United States Census Bureau, 2023).

### **1.2.2 PEP**

The Population Estimates Program 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 2023 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 ACS**

The ACS samples United States households yearly to gather information on housing, population, demographics, health, and finances. We used internal ACS microdata from the 2023 ACS survey to produce ACS estimates for various geographies. Unlike the Decennial and PEP estimates, ACS estimates contain statistical error, in the form of sampling variability, which is estimated using replicate weights.

### 1.3 Notation

CRE creates 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 quarter (GQ) types:

1. correctional facilities
2. juvenile institutions
3. nursing homes
4. institutional GQs, such as in-patient hospice facilities and military treatment facilities.
5. college dormitories
6. military barracks
7. non-institutional GQs, 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. We refer to the bulleted categories as CRE age groups, CRE race/ethnicity categories, and CRE sex categories to distinguish them from the Censal categories listed below.

This document also discusses racial classifications and age groupings available in 2020 Decennial Census data but not listed above. The Censal 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. Censal age groups are varied, depending on the particular DHC or PL94 data being used. 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 *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 GQs, 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, and
- 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 Components of Social Vulnerability

Residents in nursing homes and other types of institutional GQs 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 GQs. 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 Decennial tabulations by age x race/ethnicity x sex for households and the seven GQ types. The only results from the 2020 Decennial Census available in time for use in CRE 2021 were the PL94 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 Decennial data. In May 2023, more detailed demographic data for a variety of geography levels from the 2020 Census became available from the DHC file. Both the PL94 and DHC data were used to create population estimates for CRE 2022. For CRE 2023, we carryover the CRE 2022 methodology and describe our approach in detail over the next two subsections.

### 2.1 Initial 2020 Population Estimates

We want to create 2020 tract-level population estimates broken down by

- four age groups
- five race/ethnicity categories
- two sex categories

and the household/GQ types described in Section 1.3. Our focus is on three groups of DHC tables:

- PCT13A-PCT13I, PCT13
  - Household population broken down by race/ethnicity, age, and sex. Does not have all non-Hispanic racial categories necessary for this project, for instance, does not have a table for Black non-Hispanic.
- PCT18A-PCT18I, PCT18
  - GQ population broken down by race/ethnicity, age, and sex. Does not have all non-Hispanic racial categories and age is restricted to three ranges (0-17, 18-64, and 65 and over).
- PCT12A-PCT12O, PCT12
  - Total population broken down by age, race/ethnicity, and sex. Includes more race/ethnicity categories than the first two groups of tables and has single year age but isn't broken down by household/GQ type.

We discuss how our method works for households; the process for GQs is similar except we impute the 18-44 and 45-64 age groups using PEP data. The PCT13 series of tables includes household population totals for

1. non-Hispanic White

2. Black
3. American Indian and Alaskan Native
4. Some Other Race
5. Two or more races
6. Hispanic

broken down by various age groupings and sex. The published Censal age groups can be combined to form the desired CRE age groups (0-17, 18-44, 45-64, and 65+) so unless otherwise indicated, for the remainder of this section an age group will mean one of 0-17, 18-44, 45-64, or 65+. Notice the Hispanic and non-Hispanic household totals for age group  $age$  and sex category  $sx$  within a given tract  $t$ , denoted by  $PEP\_HH\_HISP_{age,sx,t}$  and  $PEP\_HH\_NHISP_{age,sx,t}$ , can be directly computed from the PCT13 series of tables.

We need to

- impute Hispanic origin for the racial categories in 2.-5.
- estimate totals for non-Hispanic White x Some Other Race, non-Hispanic Black x Some Other Race, and non-Hispanic American Indian and Alaskan Native x Some Other Race
- redistribute totals involving non-Hispanic Some Other Race to the CRE race categories of non-Hispanic White, non-Hispanic Black, non-Hispanic American Indian and Alaskan Native, and non-Hispanic Other, i.e. make the racial categories consistent with PEP categories (U.S. Census Bureau, 2012).

For a race  $rc$  among those listed in 2.-5., let  $TOT_{age,rc,sx,t}$  denote the *total* population in tract  $t$  for age group  $age$  and sex category  $sx$ , and  $TOT\_NH_{age,rc,sx,t}$  be the number of individuals in  $TOT_{age,rc,sx,t}$  who are non-Hispanic. Values of  $TOT_{age,rc,sx,t}$  and  $TOT\_NH_{age,rc,sx,t}$  can be obtained using the PCT12 tables. The proportion of  $TOT_{age,rc,sx,t}$  which is non-Hispanic is given by

$$PROP\_NH_{age,rc,sx,t} = \frac{TOT\_NH_{age,rc,sx,t}}{TOT_{age,rc,sx,t}}.$$

We assume this proportion holds for the *household* populations, i.e. if the household population is denoted by  $HH_{age,rc,sx,t}$  then the non-Hispanic portion of the population is defined as

$$NH\_HH_{age,rc,sx,t} = PROP\_NH_{age,rc,sx,t} HH_{age,rc,sx,t}.$$

For demographics involving White populations, there is no need to use an imputation for Hispanic origin; we can use values from PCT13I to define  $NH\_HH_{age,rc,sx,t}$ . At this point we have Decennial tract household population totals for combinations of age and sex and the following race/ethnicities: Hispanic, non-Hispanic, non-Hispanic White, non-Hispanic Black,

non-Hispanic American Indian and Alaskan Native, non-Hispanic Some Other Race and non-Hispanic Two or More Races.

We want to further decompose the non-Hispanic Two or More Races category to obtain estimates for individuals classifying themselves as exactly two races, one of which is Some Other Race, i.e. we want our estimates to include racial options like non-Hispanic White x Some Other Race (non-Hispanic White and Some Other Race). The PL94 data has multi-racial categories broken down by Hispanic origin for two age groups (18+ or 0-17). Let  $rc$  denote a race from among White, Black, and American Indian and Alaskan Native. Use the PL94 data to compute for non-Hispanic populations

$$PROP\_SOR_{dage,rc,t} = \frac{SOR_{dage,rc,t}}{TWO\_PLUS_{dage,t}}$$

where  $TWO\_PLUS_{dage,t}$  is the total number of individuals in tract  $t$  who identify as 2 or more races and are part of the Decennial age group  $dage$  (18+ or 0-17) and  $SOR_{dage,rc,t}$  represents individuals who are part of  $dage$  and identify as race  $rc$  and Some Other Race. The proportions  $PROP\_SOR_{dage,rc,t}$  allow us to allocate non-Hispanic populations involving Two or More races to non-Hispanic White x Some Other Race, non-Hispanic Black x Some Other Race, and non-Hispanic American Indian and Alaskan Native x Some Other Race. That is, we set

$$TWO\_SOR_{age,rc,sx,t} = PROP\_SOR_{dage,rc,t} TWO\_PLUS_{age,sx,t}$$

with  $TWO\_PLUS_{age,sx,t}$  as the number of individuals of age group  $age$  and sex category  $sx$  who identify as non-Hispanic Two or More Races. Note that when  $age$  is among 18-44, 45-64, or 65+ we use the proportion  $PROP\_SOR_{18+,rc,t}$ .

Finally, we make the race/ethnicity categories consistent with those found in PEP and perform some raking. We define

$$NH\_SOR\_HH_{age,rc,sx,t} = TWO\_SOR_{age,rc,sx,t} + NH\_HH_{age,rc,sx,t}$$

for  $rc = \text{White, Black, and American Indian and Alaskan Native}$ . Then we allocate a portion of the non-Hispanic Some Other Race totals, denoted by  $NH\_HH_{age,SOR,sx,t}$ , to non-Hispanic White, non-Hispanic Black and non-Hispanic American Indian and Alaskan Native using PL94 data and add the results to  $NH\_SOR\_HH_{age,rc,sx,t}$ , obtaining  $PRELIM\_PEP\_HH_{age,rc,sx,t}$ . This gives preliminary household age and sex population totals for non-Hispanic White, non-Hispanic Black, and non-Hispanic American Indian and Alaskan Native.

To define  $PRELIM\_PEP\_HH_{age,rc,sx,t}$  for  $rc = \text{Other}$ , sum  $PRELIM\_PEP\_HH_{age,rc,sx,t}$  for  $rc = \text{White, Black, and American Indian and Alaskan Native}$  and subtract the result from  $PEP\_HH\_NHISP_{age,sx,t}$ , i.e.



$$PRELIM\_PEP\_HH_{age,other,sx,t} = PEP\_HH\_NHISP_{age,sx,t} - \sum_{rc} PRELIM\_PEP\_HH_{age,rc,sx,t}$$

where the  $rc$  in the summation sign indicates summing over  $rc = \text{White, Black, and American Indian and Alaskan Native}$ . We round any negative values of  $PRELIM\_PEP\_HH_{age,rc,sx,t}$  to zero.  $PEP\_HH\_HISP_{age,sx,t}$  are our final Hispanic population totals, while we control the preliminary  $PRELIM\_PEP\_HH_{age,rc,sx,t}$  values to non-Hispanic age and sex totals from the PCT13 tables to produce final estimates. In the remainder of the document, final 2020 population results are denoted by  $POP20$  along with appropriate subscripts to indicate geography, demographic, and household or GQ type.

## 2.2 2023 Population Estimates

We calculate growth rates by demographic and household/GQ type using Vintage 2023 PEP data

$$gr_{c,j,23} = \frac{PEPPOP2023_{c,j}}{PEP2020BASE_{c,j}},$$

and apply the growth rates to the  $POP20$  estimates.<sup>2</sup> Results are summed separately for households and each GQ type at the tract-level to create population estimates  $POP23$ . The  $POP23$  values are 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. For instance, in the case of households, we define

$$rk_c = \frac{PEPPOP2023_c}{\sum_{t \in c} POP23_t}$$

for a given county  $c$  and multiply  $POP23_t$  by  $rk_c$  to obtain a raked estimate  $POPRK23_t$  for each tract  $t$  in  $c$ . The raked values become our final 2023 population estimates. In the remainder of the document, when we refer to  $POP$  or  $POPUNI$ , we are referring to these raked values.

## 3. Estimation Layers

### 3.1 Tract Universe

The 2023 Census geography vintage contains 84,415 tracts which we split into two groups: residential and non-residential. Of the 84,415 total tracts, 83,806 are classified as residential, i.e. the tract's estimated in-universe population is non-zero, and the remaining 609 tracts are

---

<sup>2</sup> A small amount of reconciliation is done, 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.

assumed to have no in-scope residents. In Sections 4-6, unless otherwise noted, by tract we will mean a residential tract.

### 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
  - this value is the weighted (by distance) sum of a tract's population and those tracts near it, scaled so that 100-250 is moderate concentration, and 250+ is high.

Using the two criteria, we define four urbanization strata<sup>3</sup>:

- 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. Define the ratio  $r_{a,j,dg,HH7}$  as the proportion of 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 subset to respondents living in households and appropriate non-institutional GQs. 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.<sup>4</sup> Denote the final

---

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

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

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 the following:

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

Estimation for the remaining GQs 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, we have

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

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.

## 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, see for example (U.S. Census Bureau, 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 that same for all values of  $dg$  within a given tract. To estimate values for  $C$  and  $q$  we fit the following regression for all tracts with a sample size greater than 25,

$$\log(r\_dirvar_{t,dg}) - \log(\hat{R}_{t,dg}(1 - \hat{R}_{t,dg})) = \log C + q \log n_t.$$

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 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	0.113	0.102	0.111	0.100	0.038	0.022
12	0.120	0.109	0.115	0.106	0.037	0.024
3	0.089	0.089	0.079	0.086	0.050	0.023

Source: 2023 Community Resilience Estimates.

## 5.2 MSE 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 & Molina, p. 43-44). If  $T$  is an arbitrary collection of tracts  $T$ ,  $N_T$  in total, then our basic MoM equation for MSE is:

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

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 actual 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 SVgroup 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 SVgroups and tracts.
- we calculate the multiplicative constant of proportionality at an aggregate level, averaging across both SVgroups and tracts.

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

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

The estimated constant, denoted  $F_T$ , was calculated as

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

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  $dg$ :

$$\begin{aligned} wt_{t,dg} &= \frac{r\_gvfvar_{t,dg}}{r\_gvfvar_{t,dg} + MSE_{t,dg}} \\ \tilde{R}_{t,dg} &= wt_{t,dg} \hat{R}_{t,dg} + (1 - wt_{t,dg}) r_{t,dg}. \end{aligned}$$

See (Rao & Molina, 2015p. 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

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

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)<sup>5</sup> for the composite estimates are better than those for the ACS and the indirect estimates. The statistics 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

<sup>5</sup>The 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  $r_{t,dg}$

Table 2: Comparison of Estimates by SV Group

Group	Mean			Median		
	ACS	Indirect	Comp.	ACS	Indirect	Comp.
0	0.336	0.341	0.341	0.330	0.343	0.337
12	0.445	0.447	0.446	0.439	0.441	0.441
3	0.220	0.212	0.213	0.181	0.207	0.196

Source: 2023 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	0.205	0.282	0.170	0.296	0.332	0.218	0.494	0.381	0.290
12	0.188	0.246	0.153	0.239	0.265	0.177	0.322	0.286	0.208
3	0.300	0.410	0.249	0.465	0.470	0.324	0.783	0.522	0.425

Source: 2023 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  $.2$  for 2022)
- variances for all tracts within a given county were equal.

A different approach based on principles from spatial statistics, was developed for CRE 2023, see Cressie (2015) or Smith (2014). Under the new methodology, geographic correlations between tracts within a county are, on average, lower than the previously used values of  $.2$  and  $.4$ , and county-level margins of error are consistently smaller than those seen in CRE 2018-CRE 2022. A detailed description of the new 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:

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

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

$$\overline{Cov}_D = \frac{1}{3N_D} \sum_{dg=0,1,2,3} \sum_{d(t,t')=D} (\hat{R}_{t,dg} - r_{t,dg})(\hat{R}_{t',dg} - r_{t',dg}).$$

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

- $H$  MSE layer if both tracts belong to the  $H$  urbanization category (i.e., UStrat3 & UStrat4),
- $L$  MSE layer if both tracts belong to the  $L$  urbanization category (i.e., UStrat1 & UStrat2),
- $M$  MSE layer if both tracts belong to different urbanization categories,

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

$$\overline{Cov}_{D,urb} = \frac{1}{3N_D} \sum_{dg=0,1,2,3} \sum_{\substack{d(t,t')=D \\ urb_{t,t'}=urb}} (\hat{R}_{t,dg} - r_{t,dg})(\hat{R}_{t',dg} - r_{t',dg}).$$

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

Since distance is a continuous variable it's not practical to perform calculations for specific distances; it's 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 for the covariances  $Cov_{b,urb}$

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

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  $\beta_{urb}$  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

$$\begin{aligned} cov(\tilde{Z}_{t,dg}, \tilde{Z}_{t',dg}) &= Cov(\tilde{R}_{t,dg}, \tilde{R}_{t',dg}) POPUNI_t POPUNI_{t'} \\ &= wt_{t,dg} wt_{t',dg} Cov(\hat{R}_{t,dg}, \hat{R}_{t',dg}) POPUNI_t POPUNI_{t'} \\ &= wt_{t,dg} wt_{t',dg} \sqrt{MSE_{t,dg} MSE_{t',dg}} cor_{t,t'} POPUNI_t POPUNI_{t'}. \end{aligned}$$

For a given county  $c$ , we have

$$\begin{aligned} MSE(\tilde{Z}_{c,dg}) &= \sum_{t \in c} MSE(\tilde{Z}_{t,dg}) \\ &\quad + \sum_{\substack{t, t' \in c \\ t \neq t'}} wt_{t,dg} wt_{t',dg} \sqrt{MSE_{t,dg} MSE_{t',dg}} cor_{t,t'} POPUNI_t POPUNI_{t'} \\ RMSE(\tilde{Z}_{c,dg}) &= \sqrt{MSE(\tilde{Z}_{c,dg})} \\ MOE(\tilde{Z}_{c,dg}) &= RMSE(\tilde{Z}_{c,dg}) * 1.645. \end{aligned}$$

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 2018-CRE 2021, county-level values were used as rate estimates for these tracts, while all count estimates were set to zero. For CRE 2022 and CRE 2023, 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. Rounding typically occurs in several stages with the overall result that point estimates for counts are rounded to whole numbers and point estimates for rates are rounded to two decimal places. Similarly, margins of error for rate estimates are also rounded to two decimal places. Past versions of CRE (CRE 2018, 2021, and 2022) have permitted non-integer values for some, or all, count margins of error, but for CRE 2023, margins of error for counts are rounded to whole numbers.

## 6. Ranking

For 2023 CRE’s “Top Vulnerable Counties and Tracts” tables and forthcoming equity supplement, we rank counties and tracts based on the proportion of individuals within the county or tract who have three or more components of social vulnerability. A rank of one indicates the county/tract has the lowest proportion of individuals with three or more components of social vulnerability and larger rankings indicate higher proportions. Dividing the rankings by the total number of tracts/counties being ranked and multiplying the results by 100 translates the rankings into percentiles. Using bootstrapping principles based on those outlined in Wright, T. Klein, M. and Wieczorek (2014), we obtain lower/upper bounds for the rankings and percentiles.

### 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 each county, from which we obtain upper and lower bounds for the county’s ranking. These rankings and bounds are converted to percentiles, and a subset of counties with the largest percentiles are published.

### 6.2 Tract

Tracts are similar 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  $\Sigma$  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

Cressie, N. (2015). *Statistics for Spatial Data*. John Wiley & Sons Incorporated.

Population Estimates Program, U.S. Census Bureau. (2012). Modified Race Summary File Methodology Statement. Retrieved 4/26/2023 from: <https://www2.census.gov/programs->

[surveys/popest/technical-documentation/methodology/modified-race-summary-file-method/mrsf2010.pdf](https://www.census.gov/surveys/popest/technical-documentation/methodology/modified-race-summary-file-method/mrsf2010.pdf).

Rao, J.N.K. & 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.

Small Area Modeling and Development Branch, U.S. Census Bureau (2024). Small Area Estimation of Health Insurance Coverage in 2010-2022. Retrieved 11/15/2024 from: <https://www2.census.gov/programs-surveys/sahie/technical-documentation/methodology/2008-2022-methods/sahie-tech-2010-to-2022.pdf>.

U.S. Census Bureau. (2023). Disclosure Avoidance and the 2020 Census: How the TopDown Algorithm Works. Retrieved 11/28/2023 from: <https://www2.census.gov/library/publications/decennial/2020/census-briefs/c2020br-04.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.