# Federal Committee on Statistical Methodology (FCSM)
# Equitable Data Toolkit

Version: November, 2023

*Contents:*

- Linkage and Introduction Materials for the FCSM Equitable Data Toolkit
- Race and Ethnicity Data Tools
- Rural Definitions and Measurement
- Analyzing Persistent Poverty Areas Using Federal Data

**Linkage and Introduction Materials for the FCSM Equitable Data Toolkit**

**FCSM Equitable Data Toolkit**
Find resources to help you plan, collect, analyze, and disseminate equity-related data.
As the Equitable Data Toolkit grows and evolves, look for additional topics, resources and tools, and other content that will help you increase available data, improve accuracy of analyses, and ensure ethical and secure data governance.

**About the FCSM Equitable Data Toolkit**
The Federal Committee on Statistical Methodology (FCSM) developed the FCSM Equitable Data Toolkit (Toolkit) to provide federal agencies with tools that support equity analyses with a focus on historically underserved populations. It is intended to support an increase in available data, improve the accuracy of analyses, and ensure ethical and secure data governance to improve the representation of underserved populations in federal data and analyses. The choice of which populations to examine in the Toolkit was influenced by Executive Order 13985 (Advancing Racial Equity and Support for Underserved Communities Through the Federal Government), which stated that:

> "The term 'equity' means the consistent and systematic fair, just, and impartial treatment of all individuals, including individuals who belong to underserved communities that have been denied such treatment, such as Black, Latino, and Indigenous and Native American persons, Asian Americans and Pacific Islanders and other persons of color; members of religious minorities; lesbian, gay, bisexual, transgender, and queer (LGBTQ+) persons; persons with disabilities; persons who live in rural areas; and persons otherwise adversely affected by persistent poverty or inequality."

It is noted that the FCSM provides materials on LGBTQ+ outside of the Toolkit on the webpage Measuring Sexual Orientation and Gender Identity (SOGI) Research Group.

Historically underserved populations are often "hard to measure" in sample surveys and even administrative data systems. In some cases, the population group is small, and its members are difficult to locate and include in probability-based surveys. In addition, obtaining the

information needed to properly classify individuals as members of the population group of interest can be difficult.

Measuring experiences and outcomes for these groups involve a wide range of methodological challenges, including:

- Creating questions that members of the population groups recognize (regardless of the relative size of the subpopulation).
- Devising accurate proxy reporting measures for administrative record systems.
- Developing terms, scales, and definitions that can be useful in variety of contexts and studies.
- Implementing data disclosure protections that still permit public-release of information
- Interpreting statistical results when placing them into a broader context.

The Toolkit provides information and resources for Statistical Officials, Chief Data Officers, agency staff, and practitioners seeking guidance on:

- Refining wording of survey items to operationalize concepts that facilitate identification or self-identification of characteristics in ways that are relevant for equity analysis.
- Overcoming sample-size issues that affect the precision of estimates for population groups that are small in number, geographically diffuse, and possibly difficult to identify or self-identify.
- Linking across data sources to build information including matching strategies and evaluation of linkage success.
- Addressing data-protection issues that spring from a survey participant being a member of a relatively small population group and one of relatively few such members included in the sample; these issues are especially salient when data are examined for subgroups defined by a combination of variables (race/ethnicity, geographic location, etc.) that can result in unacceptably small number of observations in the subgroup
- Issues of definition and measurement involving geography and persistent poverty.

The information and resources provided here are general. Those who use the Toolkit will need to assess the extent to which it is helpful to strive for consistency across agencies and their data collections and when it is helpful to develop an approach tailored for a specific purpose

The Toolkit considers these topics and others in one or more of its three sections:

- Race and Ethnicity Data Tools
- Rural Definitions and Measurement
- Analyzing Persistent Poverty Areas Using Federal Data

**Race and Ethnicity Data Tools**

Recommended Citation:
Irving, Shelley, Shagufta Ahmed, Shelley Burns, Eric Jensen, Gene Koo, Beverly Pratt, and Robert Sivinski. 2023. "Race and Ethnicity Data Tools," *FCSM Equitable Data Toolkit: A Toolkit for Strengthening Federal Data to Analyze Historically Underserved Populations.* Federal Committee on Statistical Methodology. November, 2023.

Table of Contents

# 1. Introduction to Race and Ethnicity Data Tools

The value of detailed race and ethnicity data

Detailed race and ethnicity data are needed to identify small and vulnerable racial and ethnic groups and to understand differences in outcomes of interest (e.g., income and poverty levels, employment and disability status, and birth and mortality rates) across racial and ethnic groups. These data are also relied upon to responsibly enforce civil rights laws. In particular, race and ethnicity data collected by federal agencies through censuses, surveys, administrative forms, etc. are used to monitor equal access in housing, education, employment, and other areas. They are essential for understanding populations that have historically experienced discrimination and differential treatment.

To meet growing demand, federal agencies—both statistical agencies and program agencies[1]—can work to enhance the detail of the race and ethnicity information they collect and produce data that comply with all applicable federal guidelines (e.g., protecting the disclosure of personally identifiable information (PII)) and meet the growing demand of data users. This must be done in the context of maintaining confidentiality standards and data quality, while minimizing data collection costs and respondent burden. Producing accurate and reliable detailed race and ethnicity data can present challenges for federal agencies, including locating and motivating respondents, ensuring sufficient numbers of cases for meaningful analysis, and protecting the confidentiality and privacy of respondents.

## Goals of Race and Ethnicity Tools

The purpose of this section of the FCSM Equitable Data Toolkit is to provide federal agencies with useful tools for supporting equity analyses of racial and ethnic groups, particularly detailed groups beyond the minimum Federal racial and ethnic reporting categories. These tools can help agencies collect more detailed and accurate data; access, analyze and use existing data; and release more granular data while protecting privacy.

While some approaches require significant resources and expertise, there are incremental approaches that every agency can take to improve the representation of underserved populations in their data and analyses.

This report provides tools for improving:

- *Privacy and confidentiality* (e.g., maintaining the confidentiality for data subject who are of small or vulnerable racial and ethnic groups),
- *Data collection* (e.g., locating and oversampling specific groups),

---

[1] "Program agency," for purposes of this Toolkit, refers to an agency or unit, typically within the organization structure of a Federal department, that administers, or helps to administer, a Federal program within which a determination about the rights, benefits, or privileges of individuals, businesses, or institutions is made, including those agencies with regulatory or law enforcement responsibilities.

- *Data analysis and evaluation* (e.g., maximizing the accuracy and reliability of statistics given small numbers of cases), and
- *Data access and dissemination* (e.g., developing strategies for minimizing disclosure risk when releasing information).

Some of the steps above—such as "targeting and oversampling specific groups"—may be especially applicable to a federal agency as it constructs or revises a survey design. Program agencies that conduct surveys or collect data from applicants, clients, or participants in the course of administering a federal program also need to consider the information it solicits on race and ethnicity in light of increased requirements to evaluate program effectiveness in terms of historically underserved populations (e.g., certain racial and ethnic populations).

2. Privacy and Confidentiality

Federal guidelines on race and ethnicity encourage the production of:

> *… as much detailed information on race and ethnicity as possible. However, Federal agencies shall not present data on detailed categories if doing so would compromise data quality or confidentiality standards.*[2]

The following tools can help balance between the need to collect and release detailed, useful data and the legal and ethical obligations to protect privacy and ensure confidentiality.

---

**Privacy concerns individuals.** Privacy refers specifically to individuals and their desire to control whether their personal matters are disclosed or publicized.[3] For example, persons may not want to be seen entering a place that might stigmatize them, such as a mental health clinic at a publicly known address, or to reveal financial information such as outstanding debts or income.

**Confidentiality concerns data.** Confidentiality is a quality or condition accorded to information as an obligation not to disclose that information to an unauthorized party.[4] Many federal surveys collect sensitive information alongside personally identifiable information (PII). To reassure respondents that their information will be protected, surveys often include a pledge to the respondents that their data will be protected and only used for statistical purposes. Confidentiality can only be guaranteed if there is a specific law or regulation to support it.

---

**Federal laws on privacy and confidentiality.** The Privacy Act of 1974 governs the collection, maintenance, use, and dissemination of personally identifiable information (PII) about individuals maintained by federal agencies. No agency shall disclose any record to any person without the prior written consent of the individual (with limited exceptions). A willful disclosure could result in a misdemeanor charge and/or a fine.

The Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2018 protects PII from disclosure when information is collected under a pledge of confidentiality and for

---

[2] U.S. Office of Management and Budget, "Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity," *Federal Register*, 62, no. 210 (October 30, 1997): 58782-58790, https://www.govinfo.gov/content/pkg/FR-1997-10-30/pdf/97-28653.pdf.

[3] Legal Information Institute, "Right to Privacy Definition," Wex Law Dictionary, (accessed June 17, 2022), https://www.law.cornell.edu/wex/right_to_privacy#:~:text=1)%20The%20right%20not%20to,fundamental%20personal%20issues%20and%20decisions.

[4] 44 U.S.C. § 3563.

statistical purposes. A willful disclosure could result in a felony charge with substantial fines and/or a prison sentence.

Some federal agencies may be subject to additional legal requirements. For example, the Education Sciences Reform Act (ESRA) of 2002 requires that:

1. All individually identifiable information about students, their families, and their schools shall remain confidential.
2. Individually identifiable information is immune from the legal process and cannot be used in any judicial proceeding (except in the case of terrorism).

Work with your agency's general council and Senior Agency Official for Privacy to make sure you are in compliance with all relevant laws and policies.

**The Data Protection Toolkit:** The Federal Committee on Statistical Methodology (FCSM) Data Protection Toolkit (DPT) provides easy to use guides and tools to maximize the usefulness of data while protecting privacy and confidentiality.

The goal is to release as much data as possible while ensuring to the greatest extent possible that the identity of the data subjects is protected. Specific statistical disclosure limitation methods are described in detail in the Data Protection Toolkit, including such techniques as cell suppression, data swapping, and partially synthetic data generation. Best practices developed by statistical agencies over the years include the evaluation of the effects of disclosure

## FCSM's Data Protection Toolkit

*Agencies should start with meaningful assessments of the three dimensions of the triple constraint: disclosure risk, access to data, and data accuracy. Agency leadership and staff with statistical expertise should discuss these dimensions, the interrelationships and tradeoffs between them, and develop agency standards that are appropriate based on law and the agency's risk assessment.*

*By leveraging best practice policies, methods, and technologies, agencies can sufficiently mitigate the overall risks while promoting broader access to sufficiently accurate data. This may entail the design and application of statistical disclosure limitation methods best suited to the intended data users' needs; or could involve agencies providing data users with controlled access to the confidential data through one or more tiered access models.*[1]

protection techniques on the reliability of the data (e.g., see Section 4. Evaluating the 2020 Census Data).

Additional Privacy and Confidentiality Resources

- [H.R. 4174 Foundations for Evidence-Based Policymaking Act of 2018](#)
- [National Center for Education Statistics' Statistical Standards](#)
- [Protecting Privacy in Census Bureau Statistics](#)
- [Protecting Privacy with MATH (Collab with the Census)](#)
- [U.S. Office of Management and Budget: Privacy Guidance](#)
- [U.S. Office of Management and Budget: Standards and Guidelines for Statistical Surveys](#)

3. Data Collection

Federal agencies serve the needs of diverse racial and ethnic communities. Many racial and ethnic communities, however, are not adequately represented in federal data.[5] The approaches below may help to address some frequently encountered barriers to participation.

As racial and ethnic populations are heterogenous, the same approach will not work for every audience. A mixture of approaches is preferred to improve engagement by diverse populations. Often, barriers to participation need to be probed before applying a specific approach. For example, many community members do not provide personal data due to a lack of understanding, motivation, or trust.[6]

Data can be collected through surveys or administrative forms, such as applications for permits or benefits. The methods and context of data collection will determine the specific barriers that are present. Agencies should begin by identifying these barriers and then applying the appropriate strategies below.

### Ways to Improve Participation Across Racial and Ethnic Groups

**Identify the population of interest.** Start by being specific about the intended use of the data. Consider possible internal uses of the data, and use public comment and other methods of stakeholder engagement to help identify particular racial and/or ethnic groups that you want to draw inferences about.

**Help respondents see themselves in response options.** Use culturally relevant labels that people relate to, within statutory requirements (as described above). Respect that populations are heterogenous and may require customized approaches. Provide examples that can help respondents match their self-identity to available options and, if possible, allow respondents to select more than one option. Ask people to self-identify their racial and ethnic identity, but do not collect more detailed demographic data than is necessary to avoid adding to fears of surveillance. Offer free-response "fill-in" options if you have the capability to capture and process these responses.

**Collaborate with other agencies, whenever possible.** Share experiences on outreach and engagement through communities of practice. Collaborate with other agencies in data collection to reduce duplication of data and data collection approaches that allow for data

---

[5] Office of Management and Budget, "Study to Identify Methods to Assess Equity: Report to the President," (Washington, DC: Office of Management and Budget, July 20, 2021), https://www.whitehouse.gov/wp-content/uploads/2021/08/OMB-Report-on-E013985-Implementation_508-Compliant-Secure-v1.1.pdf

[6] Sarah Evans, Jenna Levy, Jennifer Miller-Gonzalez, Monica Vines, Anna Sandoval Girón, Gina Walejko, Nancy Bates, and Yazmin García Trejo, "2020 Census Barriers, Attitudes, and Motivators Study (CBAMS) Focus Group Final Report," (Washington, DC: Census Bureau, January 24, 2019), https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/plan/final-analysis/2020-report-cbams-focus-group.html

interoperability. Share networks of trusted partners, organizations, and community leaders to help engage diverse audiences. Your agency's Paperwork Reduction Act staff can help you identify other offices within your agency, and, through the Council of Agency Paperwork Reduction Act (CAPRA), they can connect you with efforts in other Departments.

**Engage communities of interest, including community organizations, advocates, media, respected leaders, and influencers.** Work with local leaders who have the trust of their communities. When resources allow, cultivate relationships with people who can validate and amplify your message across different segments of the community – faith and civic leaders, media that reaches the communities of interest,[7] small business owners, and social media influencers. People who have a high profile with local communities may not be visible to federal agencies. Take the time to look into the background of grassroots contacts when possible. Use trusted community leaders to distribute outreach materials and post them on a publicly accessible website.

Engage communities early to identify barriers to participation. Listen for what might motivate community members to participate or to avoid participation. Ask, for example, how they conceive of demographic concepts such as "race" and "ethnicity" and how they prefer being asked these questions. For example, prior to the 2020 Census, the U.S. Census Bureau conducted a study into barriers, motivators, and attitudes related to participation in the census. While this was an extensive and well-funded effort, agencies can invest in smaller-scale efforts to examine similar issues.

**Front load outreach investments.** It takes time to develop local relationships. It also takes time to raise awareness, understanding, and engagement with data collection efforts. However, an early investment in awareness-building in the form of outreach with community members and leaders who can encourage their communities to participate may be more cost-effective than making individual visits or phone calls to make up for low participation rates. Establish indicators to monitor data collection so you know if you are on track to collect enough data and adjust outreach investments accordingly, rather than waiting until the end of data collection to evaluate response from various communities of interest.

**Communicate the purpose and value of data collection.** Communicate the benefits of participation with those populations so that they better understand the value of engaging. Name specific, positive outcomes that could accrue to the community if individuals provide needed data – answer the question, "What's in it for me/us?" For example, the 2020 Decennial Census offered materials explaining "Why we conduct the decennial census" and "Why we ask each question" during their outreach campaign. Additionally, trusted organizations who provide benefits (e.g., state Supplemental Nutrition Assistance Program (SNAP) benefit offices) may be effective and credible community messengers of the importance of participation in surveys.

---

[7] These may require intentional research into, for example, non-English, digital, and/or social media.

**Co-create materials and outreach plan with key communities if possible.** Build in the time necessary to test an approach with the communities you are targeting to confirm that messages are being communicated effectively. For example, the Centers for Disease Control (CDC) established the Health Message Testing System to pretest materials with audiences and revise based on feedback received.

**When testing key messages and outreach materials, check that materials convey key information, motivate action, allay concerns, and do not inadvertently trigger negative responses.** Use the Paperwork Reduction Act's (PRA) generic clearance process, if appropriate, to gather feedback early and often. When that is not possible, agencies can always test materials with nine or less participants without PRA approval. Online tests can provide fast, low-cost initial validations of your approach with populations that are harder to reach in person or through other means. Try to validate results obtained through testing with trusted community leaders whenever possible for further confirmation on the best approach.

**Use the language of the target community when creating press releases, social media, outreach materials, and public service announcements**. Use different media and channels – print, digital, radio, video – to reach different segments of the community. Employ bilingual interviewers or neighborhood interpreters if budget allows. [See "Transcreate into non-English languages" below.]

**For sample-based collections, oversample areas with higher concentrations of detailed groups of particular interest to your collection purposes.** This will increase the number of respondents of interest included in the sample and will aid in making resulting estimates more reliable. The Census Bureau's Planning Database can be used to identify census tracts and block groups with a high concentration of targeted groups where oversampling, special outreach, and promotion efforts could be considered. It should be noted that there are statistical trade-offs associated with oversampling one particular group. Improved estimates for the oversampled group may come at the cost of less precise estimates for other groups. In addition, results will reflect the characteristics of those group member who reside in the concentrated area and may not reflect well the intra-group diversity that exists across wider areas. For more information about trade-offs and how to make them efficiently, reach out to your agency Statistical Official.

**Offer clear and simple instructions on how to participate and how information will be used.** Explain exactly how the data collection will work and provide clear, step-by-step instructions. Assure potential participants that their responses will be kept private or confidential, if applicable, to the extent permitted by law, and will not result in negative consequences such as criminal prosecution or loss of benefits. Offer specific statutory and/or regulatory protections in outreach materials that may be in place to protect information provided, as was done, for example, with the 2020 Census, the National Health Interview Survey (NHIS), and the Survey of Income and Program Participation (SIPP). If the data will be collected on an administrative form, be clear about whether and how the information will be used for program administration. Talk to your agency counsel and PRA staff to understand what protections and assurances you're

able to offer. Whenever possible, pre-test messages that would best communicate these concerns without triggering additional fears and to ensure that the types of respondents needed understand the confidentiality assurances being shared with them.

**Use multiple modes.** Offering multiple modes of response – phone, in-person interviews, paper and digital forms (e.g., emails, text messages, or other platforms) – can help to reduce barriers and meet people where they are most comfortable.[8]

**Develop a strategy to follow up with non-respondents.** Send reminder letters to respondents who do not participate early in the collection period.[9] For interviewer-administered surveys, send FAQ postcards that convey the importance of participating in advance of data collection to prepare respondents of the upcoming survey. These will also help to communicate the legitimacy of the interviewer who will come knocking on their door. When contacting respondents, cite trusted organizations and leaders who are supporting this effort, when possible.

## Transcreate into non-English languages

Transcreation describes the process by which content is translated from one language to another, while maintaining its intent, style, tone, and context. When adapting outreach messages, survey instruments, and other materials, work with a translator so that the target audience receives not only the intended information, but also style, tone, and context.[10] Use plain language and avoid word-for-word translations that do not resonate with the audience. Ask the translator, "How would you say this?" Be aware of differences in word choices for differing language dialects. Avoid bureaucratic and legalistic language, but also avoid being too familiar.

Validate transcreated materials by testing them with community partners or with focus groups within the intended community. Community partners can help to provide additional context around your materials. In some cases, they may even be able to transcreate materials when there is a lack of resources to do so.

## Alternatives to Data Collection

It may not be feasible to directly collect race and ethnicity data on forms such as benefit applications. The content of forms may be restricted by statute or regulation, voluntary data may have low response rates, and respondents could fear that their responses to demographic

---

[8] Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. "Methods of Data Collection." *Survey Methodology* (2nd ed. Hoboken, NJ, 2009), chap. 5.

[9] Clark, Sandra. "How Effective is a Prenotice Letter in Increasing Self-Response?" (Washington, DC: Census Bureau, May 12, 2016), https://www.census.gov/newsroom/blogs/research-matters/2016/05/how-effective-is-a-prenotice-letter-in-increasing-self-response.html

[10] Jennifer Kim, Jason Kopp, and Marisa Hotchkiss, "Developing Public-Facing Language Products: Guidance From the 2020 Census Language Program," (Washington, DC: Census Bureau, November 2021), https://www2.census.gov/library/publications/decennial/2020/operations/language-product-handbook.pdf

questions will be used to adjudicate applications. In these cases, agencies should consider alternatives to collecting race and ethnicity data directly:

- **Data linkage –** The self-reported race and ethnicity of program participants may exist on other datasets that can be linked to program data. For example, the Census Bureau's Data Linkage Infrastructure collaborates with federal agencies to support high-quality research and evaluation.
- **Modeling or imputation –** Existing information can be used to estimate or impute the race and ethnicity of individuals. For example, information found on administrative forms such as first name, last name, and address can be used to estimate race and ethnicity.[11] These experimental methods are still being tested and developed by several agencies, and must be used with caution to avoid inadvertently introducing bias.

Additional Data Collection Resources

- Translation Services
  - Developing Public-Facing Language Products - Provides guidance from the 2020 Census Language Program
  - Library of Congress's Federal Research Division - Provides translation services for qualitative or quantitative data collection design and implementation.
  - Many marketing firms have translation and transcreation capabilities; see GSA Schedule Special Item Numbers 541910 (Marketing Research and Analysis), 541613 (Marketing Consulting Services), and 541810 (Advertising Services).
- Selected Sources of Race and Ethnicity Data and Statistics
  - American Community Survey - Provides information about jobs and occupations, educational attainment, veterans, whether people own or rent their homes, and other topics, including the race and ethnicity of respondents.
  - Census Race Library - Publications, visualizations, and working papers released by the U.S. Census Bureau related to race and ethnicity.
  - Crime Victimization Survey - Provides information on criminal victimization, in addition to the race and ethnicity of victims and offenders.
  - Data.gov - Source of federal open data. Provides data, tools, and resources to conduct research and design data visualizations.
  - Current Population Survey - The primary source for monthly labor force statistics. Also collects data, including the race and ethnicity of respondents, for other studies to inform on the economic and social well-being of the country.
  - Decennial Census - Provides the number of people living in the United States, in addition to respondents race and ethnicity. Data are used to apportion the number of seats each state has in the U.S. House of Representatives.

---

[11] Ann Haas, Marc N. Elliott, Jacob W. Dembosky, John L. Adams, et al., "Imputation of Race/Ethnicity to Enable Measurement of HEDIS Performance by Race/Ethnicity," *Health Services Research* 54, no. 1 (February 2019): 13-23, https://pubmed.ncbi.nlm.nih.gov/30506674/

- o [Health of the United States, annual report](#) - A report on the health status of the nation. Some tables present data according to race and Hispanic origin.
- o [Labor force characteristics by race and ethnicity](#) - Provides labor force, employment, and unemployment statistics by year by race and ethnicity.
- o [National Assessment of Educational Progress](#) - Provides nationally representative data on what America's students know and can do in various subject areas, in addition to race and ethnicity of students.
- o [National Health Interview Survey](#) - Provides data on a broad range of health topics, in addition to race and ethnicity of respondents.
- o [Survey of Prison Inmates](#) - Provides estimates for the state and sentenced federal prison populations across a variety of domains, including race and ethnicity of offenders.
- o [United States Life Tables by Hispanic Origin](#) - Life tables by Hispanic origin, race for the non-Hispanic white and non-Hispanic black populations, and sex.

4. Data Analysis and Evaluation

This section presents tools to help agencies produce accurate and reliable race and ethnicity statistics. Fortunately, there are proven statistical methods for getting the most value out of limited information and small sample sizes. For help applying these methods, reach out to your agency Statistical Official.

Data Analyses for Small Race and Ethnicity Categories and/or Detailed Groups

**Use existing guidance on the presentation of race and ethnicity data.** For example, OMB offers guidance on the [aggregation and allocation of data on race](#) and [flexibilities and best practices](#) for implementing their race and ethnicity data standards.

**Pool multiple months or years of data together.** This method will help ensure adequate sample size for analysis. The number of pooled cycles or years can be based on target minimum cell sizes or on target maximum uncertainty, or variance. It may be necessary to adjust weights when pooling data[12]. For example, the [American Community Survey and the Puerto Rican Community Survey](#) provide multi-year files to data users.

**Aggregate detailed groups where needed.** If data is collected at very granular levels that will require several years of pooled data to reach publishable sample sizes, work with the data to discover the most granular level possible for publication at any given time.

For example, if data for detailed Black or African American groups are needed but the sample size is too small, then combine Black or African American detailed groups to the high-level identities and/or geographic regions that comprise this category as defined in SPD 15 (e.g., aggregate Jamaican, Haitian, and Bahamian into an Afro Caribbean grouping; or aggregate Nigerian, Liberian, and Ghanaian into a West African grouping). The Census Bureau's [Hispanic Origin and Race Code List](#) (Appendix F beginning on page 193) is an available resource when deciding how to aggregate groups.

**Supplement with other data sources.** Increasingly, federal agencies are [linking administrative records and survey data](#) to create an enhanced data file for analysis. This approach can provide a more detailed picture of the economic and social well-being for racial and ethnic groups. This also allows different federal agencies to help each other to better understand how programs are working and where they could be improved.

Presenting Results for Race and Ethnicity Categories and/or Detailed Groups
Similar attention given to *collecting and analyzing* race and ethnicity information should be given to *presenting* results so that they're safe and useful.

**Present results from detailed race and ethnicity groups with equitable, balanced, and relevant terminology.** To do so, rely on (1) terminology used by specific racial and ethnic

---

[12] You can read about how the American Community Survey combines multiple years of data here: [American Community Survey Multiyear Accuracy of the Data (5-year 2017-2021) (census.gov)](#)

communities and/or (2) terminology of how respondents describe themselves. Use the Census Bureau's Hispanic Origin and Race Code List (Appendix F, which begins on page 193) as a guide.

**Do not use terms like "majority," "minority," "other," and "non-White," and avoid combining specific races and ethnicities into "majority," "minority," "other," and "non-White" groupings.** These terms have several conceptual and practical challenges and have become more complex and contested in recent decades. For more information, see Measuring Racial and Ethnic Diversity for the 2020 Census (beginning with "Measuring Diversity Then and Now"). In cases where the minimum reporting categories cannot be used, agencies may use "all other races" when such a collective description is appropriate.[13]

## Measures of Uncertainty

Especially when dealing with small populations it's important to consider the uncertainty associated with results. The acceptable amount of uncertainty will depend on the particular use and the importance of having accurate and precise estimates.

A **margin of error (MOE)** describes the precision of an estimate at a given level of confidence. The confidence level associated with the MOE indicates the likelihood that the sample estimate is within a certain range (the MOE) of the population value. MOEs are often provided at a 90 or 95 percent confidence level.

**Confidence intervals** are easily calculated using MOEs and are often displayed as an upper or lower bound at a given confidence level for the estimate.[14] As with MOEs, it is common to produce confidence intervals with a confidence level of 90 or 95 percent. Confidence intervals or MOEs are excellent tools for communicating uncertainty to a non-technical audience. The larger the MOE or confidence interval for a particular estimate, the more caution is required when using the estimate. The size of a CI may be used a criterion for determining whether to report an estimate to the public—a practice that is adopted by the National Center for Health Statistics.[15]

Measures of statistical uncertainty can also be used to produce **coefficients of variation (CV)**, which are measures of how close the observed data points are to their mean.[16] The CV is also called the **relative standard error**. It is calculated by dividing the standard error of an estimate by its mean and is usually expressed as a percentage. According to the U.S. Census Bureau's

---

[13] U.S. Office of Management and Budget, "Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity," *Federal Register*, 62, no. 210 (October 30, 1997): 58782-58790, https://www.govinfo.gov/content/pkg/FR-1997-10-30/pdf/97-28653.pdf.

[14] Technically, a 90 percent confidence interval means that 90 percent of samples drawn the same way as the working sample would include the true estimate within the estimated confidence interval. For a more in-depth definition of confidence intervals, see the Census Bureau's Basic Explanation of Confidence Intervals.

[15] Parker JD, Talih M, Malec DJ, et al. National Center for Health Statistics data presentation standards for proportions. National Center for Health Statistics. Vital Health Stat 2(175). 2017; Parker JD, Talih M, Irimata KE, Zhang G, Branum AM, Davis D, et al. National Center for Health Statistics data presentation standards for rates and counts. National Center for Health Statistics. Vital Health Stat 2(200). 2023

[16] U.S. Census Bureau. "U.S. Census Bureau Statistical Quality Standards." (Washington, DC: U.S Census Bureau, July 2013) pg. 154, https://www.census.gov/content/dam/Census/about/about-the-bureau/policies_and_notices/quality/statistical-quality-standards/Quality_Standards.pdf

Statistical Quality Standards, serious data quality issues related to sampling error occur when the estimated CVs for the majority of the key estimates are larger than 30 percent.[17]

Sensitivity Analysis

A sensitivity analysis is a method used to evaluate how the results change when the inputs or assumptions used in the analysis are varied. It can be a useful tool when modeling with either survey data or administrative records. Sensitivity analyses can be particularly useful for administrative records, where it's difficult to estimate the uncertainty in the results because the respondents aren't randomly sampled.

For example, birth records *do not list* the race and ethnicity of the child *but do list* this information for the child's mother and father. An analysis of births by race and ethnicity will need to assign these characteristics to the child using the mother's and/or father's characteristics. A possible sensitivity analysis could be to change how race and ethnicity are assigned to the child and then evaluate how this impacts the results for an outcome variable. Ideally, a known outcome variable would be used to calibrate the methods to best approach the correct value.

## Suitability of Detailed Race and Ethnicity Group Data for Publication

When preparing materials for publication, it is important to use disclosure avoidance procedures to protect the confidentiality of respondents, especially for small and vulnerable groups. These procedures may range from relatively simple methods such as suppression (i.e., redacting cells with small, unweighted counts) and rounding of cell counts (e.g., to the nearest thousand) to more complex methods such as differential privacy. The Data Protection Toolkit provides a good source information and resources on methods to protect data.

Measures of statistical uncertainty (described above) can be used to determine whether estimates are suitable for release. When releasing estimates, provide the accompanying measures of statistical uncertainty. Measures of uncertainty, accompanied by explanation and interpretation provided by subject matter experts, help data users understand the reliability and limitations of the estimate.[18]

Additional Analysis and Evaluation Resources

---

[17] The 30 percent CV means that the ratio of the standard deviation to the mean is 0.3, which is the extent of the variability relative to the mean. U.S. Census Bureau. "U.S. Census Bureau Statistical Quality Standards." (Washington, DC: U.S Census Bureau, July 2013) pg. 114, https://www.census.gov/content/dam/Census/about/about-the-bureau/policies_and_notices/quality/statistical-quality-standards/Quality_Standards.pdf. Data users should use discretion when computing the CV (or relative standard error) for proportions. The CV for an estimated proportion of 0.05 is much larger than the CV for its complement (0.95), which could be used to convey the same information. For example, if the proportion of the population having characteristic X is 0.05, then the proportion of the population without characteristics X is 0.95. The CV for the proportion having characteristic X will be larger than the proportion without characteristic X. Jennifer D. Parker et al., "National Center for Health Statistics Data Presentation Standards for Proportions," *Vital and Health Statistics*, 2, no. 175 (National Center for Health Statistics, November 1, 2018), https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf

[18] See, for example, the NCES Condition of Education report which informs readers when the coefficient of variation for an estimate is so large that it threatens its validity.

- Examples of Data Quality Standards
  - [Bureau of Justice Statistics](#)
  - [U.S. Census Bureau](#)
  - [Economic Research Service](#)
  - [National Center for Education Statistics](#)
  - [National Center for Health Statistics](#)
- Resources on Data Disaggregation
  - [Annie E. Casey's By the Numbers: A Race for Results Case Study](#)
  - [Urban Institute's Combining Racial Groups in Data Analysis Can Mask Important Differences in Communities](#)
  - [Urban Institute's Considerations for Ensuring Data Aggregation Is as Inclusive as Possible](#)

5. Data Access and Dissemination

This section provides guidance on the release and dissemination of race and ethnicity data. Federal agencies have an obligation to release data files and a range of data products to meet data users' needs in a timely manner. Additionally, they should provide information necessary to ensure that users interpret data accurately. However, it is critical that agencies adhere to disclosure laws and guidelines to keep respondent information confidential.

Agencies should establish a dissemination plan and communicate the plan to the general public to promote transparency and build public trust.[19] The dissemination plan should provide timely access to all users and information to the public about the agencies' dissemination policies and procedures, including those related to any planned or unanticipated data revisions. For example:

1. Develop a schedule and determine the mode for the release (e.g., printed hardcopy, PDF file, or HTML format) of information products;
2. Inform the general public, as well as targeted audiences; and
3. Ensure timely access to data and data products to all users.

Tiered Data Access

Tiered data access is a strategy for disseminating data that includes multiple versions of the same data in order to help the agency monitor and control the risk of disclosure. Less sensitive or lower risk versions might be released publicly (public use data, or PU), while more detailed and sensitive versions may require a signed data use agreement, or oversight from a Disclosure Review Board.

Public use (PU) data is released by agencies in multiple formats. These can include PU files that have undergone additional disclosure mitigation procedures to allow public release (e.g., collapsing detailed race and ethnicity groups to mitigate disclosure risk and suppressing data), online data tools (e.g., National Center for Education Statistics DataLab) that include procedures to protect data generated, or release of tabular data either in reports or separately. Agencies must evaluate their audiences and their data needs to determine which PU data products best suit their users' needs.

Some data users require access to more sensitive data that cannot be made available through PU data releases. This might include information on small race or ethnic categories that are collapsed or suppressed on PU data files. Restricted use (RU) access is a process that allows federal agencies to make restricted data available to approved users in a secure data environment that minimizes risk of disclosure. Restricted access to the least modified federal statistical data has been through restricted-use data centers (i.e., The Federal Statistical

---

[19] For example, in line with OMB Statistical Policy Directive 4, the National Center for Education Statistics' (NCES) 2012 Statistical Standards include standards and guidelines for the dissemination of and access to data.

Research Data Centers (FSRDCs) managed by the Census Bureau[20]) or through licensing researchers as CIPSEA agents (i.e., National Center for Education Statistics, National Center for Health Statistics).[21] Movement toward remote RU access is underway (i.e., Coleridge initiative at the National Center for Education Statistics). Contact your agency's statistical official for updates on the standard application process (SAP) and remote access options.

Since data must be protected by law, providing a wide range of data products can allow access by the widest audience.

Additional Access and Dissemination Resources

- Statistical Policy Directive No. 4: Release and Dissemination of Statistical Products Produced by Federal Statistical Agencies
- Selected Data Dissemination Plans
  - Bureau of Economic Analysis
  - Bureau of Justice Statistics (Section VI)
  - National Center for Education Statistics (Chapter 7)
  - National Center for Health Statistics (Data Release Policies)
  - U.S. Census Bureau (Section F)

---

[20] Restricted data "are available to qualified researchers with approved projects at secure Federal Statistical Research Data Centers (RDCs). There are currently 31 open Federal Statistical Research Data Center (RDC) locations. The RDCs partner with over 50 research organizations including universities, non-profit research institutions, and government agencies.

[21] Any results the researcher proposes to release must still go through the disclosure avoidance protocols.

6. Collecting Detailed Data While Complying with OMB's Race and Ethnicity Standards for Federal Data

The U.S. Office of Management and Budget's (OMB) Statistical Policy Directive No. 15: [Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity](#) (SPD 15) establishes a uniform framework for the collection of race and ethnicity data that all federal agencies must use when collecting race and ethnicity information.[22] Sometimes cited as a barrier to collection of more detailed race and ethnic categories, it actually encourages agencies to collect more detailed racial and ethnic categories so long as they roll up to the following minimum categories.

The two required minimum categories for ethnicity,[23] and their corresponding definitions, are:

1. **Hispanic or Latino:** A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race.
2. **Not Hispanic or Latino**

Note, people of Hispanic or Latino origin may be of any race, per SPD 15.

The five required minimum categories for data on race,[24] and their following definitions are:

1. **American Indian or Alaska Native:** A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment.
2. **Asian:** A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam.
3. **Black or African American:** A person having origins in any of the black racial groups of Africa.
4. **Native Hawaiian or Other Pacific Islander:** A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
5. **White:** A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

Note, people may report multiple races.

**SPD 15 permits the collection of more detailed racial and ethnic groups provided that any additional information can be aggregated into the minimum categories.** It is important for

---

[22] The reporting guidelines can be found at [OMB Bulletin No. 00-02 - Guidance on Aggregation and Allocation of Data on Race for Use in Civil Rights Monitoring and Enforcement](#) and [Provisional Guidance on the Implementation of the 1997 Standards for Federal Data on Race and Ethnicity.](#)

[23] Multiple responses to the ethnicity question are not permitted.

[24] SPD 15 does not permit an "Other" race category. The only exception is the decennial census, including the American Community Survey, which is [required by law](#) to include a "Some Other Race" response category.

some uses of data to have race and ethnicity information disaggregated beyond – or more detailed than – the minimum categories provided in SPD 15. For example, there are a wide variety of detailed groups that fall under the Asian category (e.g., Chinese, Asian Indian) and Hispanic or Latino category (e.g., Mexican, Cuban). In addition to the 2020 Census, many surveys collect more detailed information about race and ethnicity that can be aggregated to the minimum racial and ethnic categories in SPD 15 (e.g., American Community Survey (ACS), National Health Interview Survey (NHIS); Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011); and Consumer Expenditures Survey (CE)).

SPD 15 is currently under review by OMB and may be revised in the future. You can learn more about the review process here: SPD15revision.gov.

**BLANK PAGE TO SEPARATE SECTIONS OF THE TOOLKIT**

**Rural Definitions and Measures Tools**

Recommended Citation: Cromartie, John and Michael Ratcliffe. 2023. "Rural Definitions and Measures Tools," *FCSM Equitable Data Toolkit: A Toolkit for Strengthening Federal Data to Analyze Historically Underserved Populations*. Federal Committee on Statistical Methodology. November, 2023.

Table of Contents

1. Introduction

The term "rural" leads a double life. On the one hand, it is used to describe broadly shared images of open countryside, farms, crossroad communities, and small towns at some distance from a big city. If asked, most Americans would agree generally on the physical attributes that define rural: small populations, low-density settlement, and remoteness. On the other hand, the term applies to statistical definitions used to delineate a precise boundary between rural and urban territory for the purpose of generating statistics and insights about rural people and places. These boundaries also are used to determine eligibility for rural-based programs and to allocate billions of dollars of federal funding. The effective and equitable distribution of these funds depends, in large part, on selecting an appropriate definition of rural.

Drawing a precise line between urban and rural requires answering two questions:

- At what population threshold do rural places become urban?
- Where along the urban periphery do suburbs give way to rural territory?

Answers vary tremendously. Most population thresholds dividing rural from urban places range from 2,500 to 50,000. Methods to designate an urban periphery include narrow definitions based on municipal boundaries and much broader ones based on county-level commuting zones. Different choices lead to dramatic differences in the resulting rural populations, in terms of overall size, geographic distribution, and socio-economic characteristics. This leads to debates among experts on what constitutes rural, a variety of rural definitions currently in use across Federal agencies, and confusion and frustration among program leaders and policy makers looking for a straightforward solution to a tricky classification issue.

This section of the FCSM Equitable Data Toolkit is designed as a brief guide to rural definitions and measures, and supplements information available from the Economic Research Service (https://www.ers.usda.gov/topics/rural-economy-population/rural-classifications/what-is-rural.aspx) and the Census Bureau (www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html). It begins with an explanation of why different definitions exist and a synthesis of key similarities and differences among them. This is followed by a detailed description of the primary sources for rural definitions. These descriptions and comparisons are designed to enhance understanding surrounding rural definitions and enable better decision-making in choosing definitions that best fit the purpose of specific applications.

Federal definitions generally follow three basic approaches to defining rural and urban areas:
- Place-based: Urban and rural areas are based on boundaries for places (generally municipalities), with rural places defined as those having populations below a specified threshold.

- Settlement structure: Urban and rural areas are defined based on measures of the built environment and settlement patterns, such as population density, housing density, and land uses.
- Functional relationships: Urban and rural areas defined based on measures of social and economic interaction, such as commuting, or measures of distance from or proximity to an urban center of specified size.

Different approaches are adopted for different reasons, including:
- Analytical purposes: Some questions may require information about settlement and development patterns (a settlement-structure approach) while others may focus on socio-economic relationships between communities and across urban and rural territory (a functional-relationship approach).
- Statutory versus statistical requirements: Some questions may require the use of rural definitions created for statutory programs that may differ from rural designations defined for statistical purposes. Statutory programs may require a definition that references governmental units, such as counties or municipalities.
- Geographic scale and data access: Some questions may require access to data that are only available at select geographic scales. For example, the abundance of county-level demographic economic, health, and other data often the abundance of demographic, economic, health, and other data for counties often leads to use of a county-based definition of rural (a place-based approach).
- Ease of use and implementation: density-based definitions that utilize small geographic areas (census blocks, census tracts) as building blocks generally require use of a Geographic Information System (GIS) for efficient and effective mapping and analysis.

Choosing a rural definition begins by considering the underlying concept that guides the construction of a specific definition and deciding whether that perspective fits the purpose of the application.
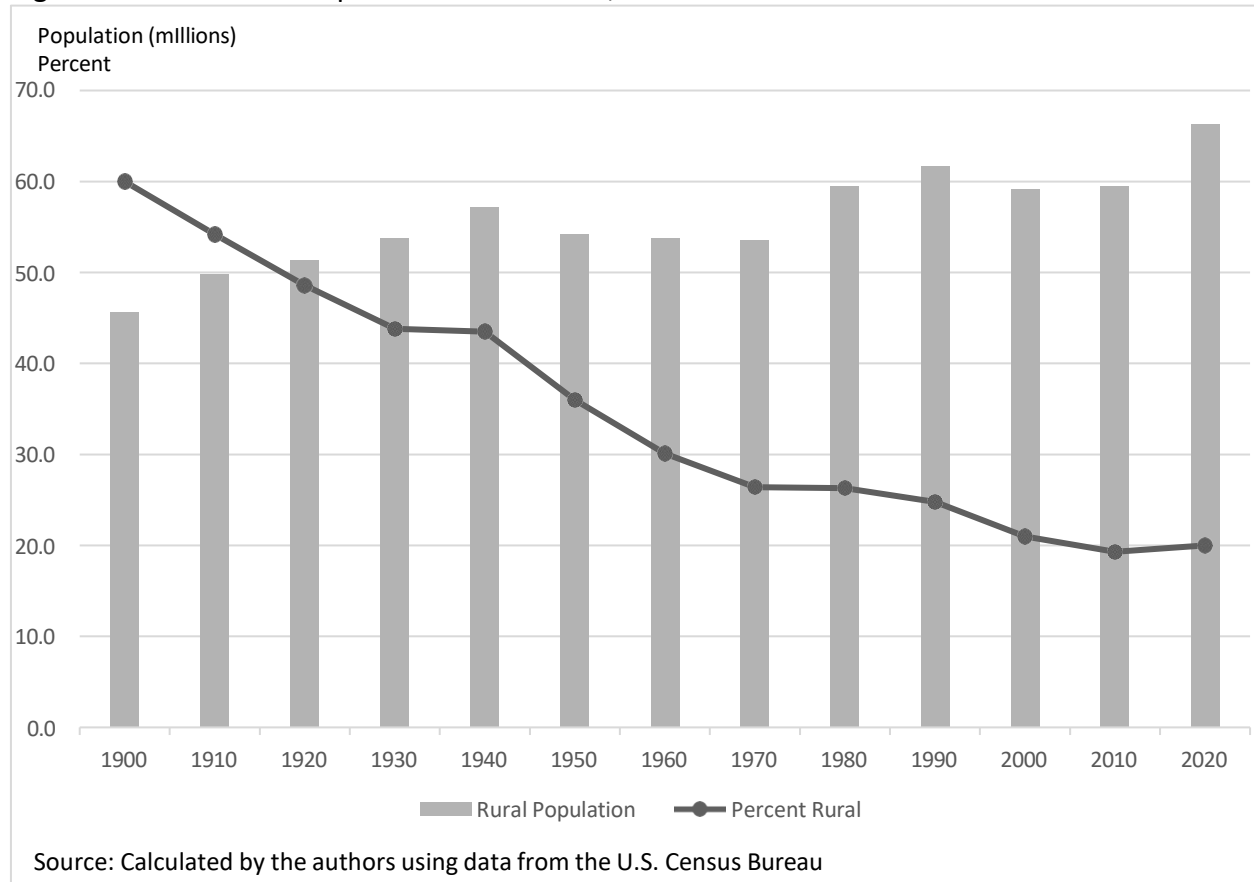
2. The Challenge of Classifications

The classification of people and territory as rural poses a variety of challenges to researchers, policy makers, and program managers throughout the federal system and beyond. Most Americans share a common image of rural—open countryside and small towns at some distance from large urban centers—but disagree on where and how to draw the line between rural and urban. Similar conundrums exist for any measurement term that imposes a dichotomy on a continuum (e.g., the income level that serves as a poverty line, or the age at which a person enters old age). Confusion over rural definitions stems from the unavoidable task of drawing a categorical line through a broad continuum. Different solutions to this problem lead to different configurations of rural and urban space.

There is no single, official definition of rural or rurality in use by federal agencies. Individual agency and individual program definitions vary. As a result, areas of the U.S. that are considered rural encompass a wide variety of settlements and landscapes, from low-density housing subdivisions on the fringes of large urban areas, to small towns in predominantly agricultural areas, to frontier-like areas located long distances from urban centers. The individuals and communities in these diverse settings have different needs, face different challenges, and interact differently with other rural communities and urban centers. To classify them all under a single category, or to utilize a single definition of rural, masks their various characteristics, interactions, and needs. Different purposes for seeking to identify, count, understand, and serve the rural population can call for different definitions of rural.

Two factors contribute to multiple rural definitions. First, no consensus exists on a key question: At what population size does a place shift from rural to urban? Around 1910, the U.S. Bureau of the Census (Census Bureau) defined as urban any incorporated place with 2,500 or more people; all other territory, including incorporated places of less than 2,500 people, were defined as rural. While other elements of this classification were modified over time to reflect changing realities, this basic population threshold remained unchanged until 2022 when the Census Bureau announced that for Census Bureau publications and analysis starting with the 2020 Census the minimum threshold for urban would be either 2,000 housing units or 5,000 people. Arguments have been made for moving to a higher threshold because the average size of places is so much larger today and the country has switched from predominantly rural to majority urban (figure 1). Transportation and communication advances helped reorganize economic and social activities around larger towns and cities. Places of 2,500 people typically do not provide the same levels of employment, goods, and services that existed in many of those places in 1910. However, the choice of an appropriate population threshold delimiting rural and urban places has never been resolved and limited research exists to aid in choosing such a cut-off. Rural development programs within USDA employ various thresholds for defining a place as rural or urban, anywhere from 2,500 to 50,000 people. For example, the USDA's Business and Industry Loan Guarantees Program uses a 50,000-person threshold when identifying qualifying rural places. On the other hand, the USDA's Water and Waste Disposal

Loan and Grant Program identifies places of less than 10,000 persons as rural (see links in table 2 for details on the programs and selection of rural definitions in them).

Figure 1. Census Rural Population and Percent, 1900-2020

Population (mIllions)
Percent



Source: Calculated by the authors using data from the U.S. Census Bureau

Second, rural definitions typically are based on what is not urban, with boundary lines drawn around each urban entity, marking where urban ends and rural begins along the urban periphery. This "urban-centric" approach generally resulted from operational considerations in which it was easier to compile and manage data for a discrete set of central places that could form the starting points for defining densely settled urban areas. Today, many U.S. residents live in suburbs or exurbs that intentionally combine rural and urban elements. A complex, interstitial zone of bedroom communities, office parks, retail corridors, residential subdivisions and individual house lots characterizes the urban periphery around almost any sizeable U.S. city. A narrow definition of the urban boundary risks defining a large segment of suburban population as rural. A broad definition inevitably classifies some rural residents as urban. No definition divides these populations in a way that could satisfy all users.

Despite the variety of definitions of rural (and urban) that are in use across federal agencies and programs, we can identify four basic components of a definition that track with the discussion of basic approaches followed in federal definitions noted at the opening:

- *settlement size*
- *population, housing, or development density*
- *proximity to or isolation from larger urban centers*
- *the broader social and economic context*

The first two components relate to structural aspects of settlement; that is, the nature of the built environment and settlement patterns. The second two relate to the functional aspects of settlement—how different communities are connected socially and economically. These two ways of looking at settlement patterns and community interaction—structural and functional—form the foundations of different definitions of rural and should be considered when deciding on a definition for a specific purpose, whether research, policy analysis, or program implementation. The Census Bureau's Urban-Rural Classification provides a good example of a structural approach to defining urban and rural. Examples of functional definitions include the Core-Based Statistical Areas (CBSAs) provided by the U.S. Office of Management and Budget (OMB) and the Rural-Urban Commuting Areas (RUCAs) provided by the Economic Research Service (ERS). The National Center for Education Statistics' (NCES) Locale Codes incorporate both the structural and functional aspects of settlement by referencing settlement size as well as proximity to larger urban centers. Each of these classifications is discussed in more detail below.

These four components to measuring and defining urban and rural areas intersect in several ways with the three approaches discuss earlier. For example, an administrative-based approach could specify that municipalities below a minimum population size are considered rural. Likewise, municipalities with populations at or above the specified minimum threshold would be considered urban, regardless of the overall density of development. Definition of urban areas based on continuity of densely populated or densely developed territory represents a structural approach and assumes that social and economic interactions occur between communities within the larger area. Functional approaches to defining urban and rural areas often start a densely developed core area of specified population size and include additional areas of varying settlement size and density based on measures of social and economic interaction. Measures of proximity or isolation, often based on distance or travel time, can act as proxies for more explicit measures of functional relationships.

Across legislation and federal agencies, population thresholds dividing rural from urban places have ranged from 2,500 to 50,000. Methods to designate an urban periphery include narrow definitions based on census blocks, census tracts, or municipal boundaries and much broader definitions based on counties. The geographic "building block" (i.e., census block, census tract, place, county) used in definitions may have been chosen for a variety of reasons, including interest in a high degree of spatial precision (which census blocks offer); presence of a local government capable of applying for and receiving program funds (i.e., municipalities, counties); ease of use/smaller numbers of geographic units with which to work (i.e., census tracts, counties); or the availability of a wide variety of statistical data for use in analysis and decision

making (i.e., counties). Different choices lead to dramatic differences in the resulting rural populations, in terms of overall size, geographic distribution, and socio-economic characteristics.

3.  Examining Definitions in a Settlement Concentration-Social and Economic Relationships Framework

We begin by acknowledging the multi-dimensionality of rural definitions—the fact that they combine size, density, and distance criteria in very different ways. Degrees of rurality (or urbanness) can be viewed in terms of both levels of concentration, which refers to the density of the population, and levels of social and economic ties, which captures the size of the area within which residents are commuting from home to work (figure 2). For example, intensely developed urban cores of major cities are focal points for commuting and other economic and social ties within large metropolitan regions. They also tend to be surrounded by suburbs and lower density exurban communities that are linked to the urban center. At the other end of the spectrum are medium- and small-sized towns that are densely developed (though not as much as a large urban center) and are beyond the commuting zone of a larger urban center. These smaller places may serve as an economic center and focal point for sparsely populated surrounding areas. Recognizing and accounting for these differences are important. A low-density area on the outskirts of a large urban center in a major metropolitan area exists in a different context than a densely settled small town located at a great distance from any other population center or a sparsely populated area isolated from urban centers of various sizes. A clearer summary of definitional issues can be achieved, and better choices made, by recognizing there can never be a one-size-fits-all definition of rural. For researchers and policy makers alike, the appropriate choice of a rural definition depends first and foremost on the purpose of the enterprise.

Figure 2. Concentration and Metropolitanness are Two Dimensions for Understanding Rurality
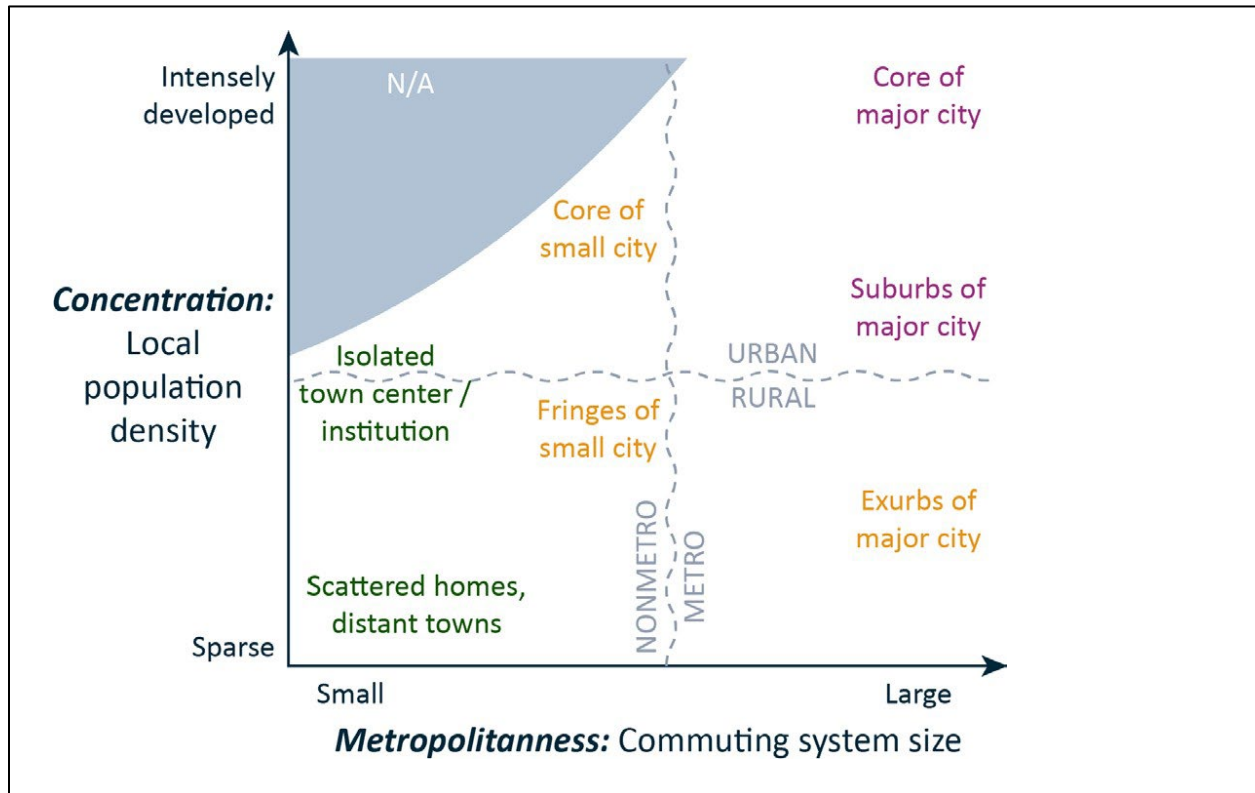


Source: Schroeder and Pacas (2021)

Choosing a rural definition begins by considering 1) what you want to accomplish and which populations and communities you want to include in your research, analysis, or program and 2) the data to which you have access. For instance, tracking land-use change or studying the effect of urbanization on farmland prices may benefit from a geographically detailed definition of rural that distinguishes built-up territory from surrounding, less-developed land. Mapping health care accessibility and analyzing its effect on rural well-being may need to incorporate distance measures into the rural definition, along with distinguishing more isolated and less isolated settings in relation to an urban center that supports health clinics, medical services providers, and hospitals.

In any application involving empirical research, practical aspects of data availability will play a major role and may preclude choices that otherwise would be desirable. County-based definitions dominate rural demographic and economic research because of the wider availability of detailed statistical data at the county level. As a result, researchers have treated nonmetropolitan counties as rural even though, based on the Census Bureau's classification, some of those counties contain urban areas and many metropolitan counties contain rural territory. In fact, based on the Census Bureau's definition, over half of the rural population resides within metropolitan counties. Counties can be quite large in land area, especially in the West, and can include a wide variety of land uses. Many metropolitan counties in the western half of the United States contain vast areas of sparse, if any, settlement. In such counties, large

proportions of the population are clustered in densely developed urban centers, while a large proportion of the land area is decidedly rural in nature, with some small communities separated from urban centers by large distances. San Bernardino County, California and Mohave and Coconino Counties, Arizona offer good examples of metropolitan counties that contain large expanses of sparsely settled or uninhabited territory.

For policymaking and economic development applications, the choice of a desirable rural definition may not be as limited by data considerations. Choices of a rural definition may vary depending on program goals. A program providing housing assistance may choose to target more isolated or economically distressed rural settings compared with programs designed to stimulate business starts and job creation. For example, the need for access to municipal water and sewage disposal systems or transit systems may lead to the selection of a higher population threshold since larger places are more likely to contain the infrastructure needed to promote and sustain economic development. Rural communities lacking access to health services may not be the same ones dealing with limited telecommunications infrastructure. The infrastructure needed to support health clinics serving largely rural populations led to the definition of rural for the Rural Health Clinics Program (administered by the Center for Medicare and Medicaid Services) as any area outside a Census Bureau urban area of 50,000 or more persons (see link in table 2 for more program details). The need to encourage and support expansion of telecommunications infrastructure into sparsely populated areas might argue for a lower population threshold and use of a maximum population density criterion.

4. Classifications

The federal government currently uses over two dozen rural definitions. A study undertaken to identify key similarities and differences among these definitions revealed that many of these definitions are grounded in four geographic sources, each of which defines a particular set of urban entities (Cromartie and Bucholtz 2008):

1. Census Places: The Census Bureau's list of incorporated and unincorporated places.
2. Urban Areas: Rural is the residual territory not included in a Census Bureau-defined urban area.
3. Core-Based Statistical Areas (CBSAs): Defined by the Office of Management and Budget (OMB) using urban area population and commuting data applied at the county-level, and divided into metropolitan and micropolitan statistical areas (metro and micro areas). Territory not included in either a metropolitan or a micropolitan statistical area is labeled "outside CBSAs."
4. Rural-Urban Commuting Areas (RUCAs): Defined by ERS using urban area population and commuting data applied at the census tract-level and at the ZIP Code-level.

The NCES Locale Codes represents a fifth source for consideration, offering a multilevel classification that incorporates both the structural aspect of settlement by referencing the Census Bureau's urban and rural areas as well as principal cities of CBSAs and the functional aspect by measuring distance from Census Bureau urban areas of varying size.

We label these "sources" because they do not necessarily describe just one definition, as each can be manipulated to derive different definitions. Some definitions combine elements from more than one source. For example, both CBSAs and RUCAs utilize the Census Bureau's urban areas as cores or centers to which commuting is measured. The National Center for Health Statistics' Urban-Rural Classification ([Data Access - Urban Rural Classification Scheme for Counties (cdc.gov)](https://www.cdc.gov)) provides a good example of a multi-category county-based classification that uses OMB's CBSA classification as its basis. This six-level classification distinguishes counties as follows:

- Large Central Metro: central counties of metropolitan statistical areas of one million or more population.
- Large Fringe Metro: outlying counties of metropolitan statistical areas of one million or more population.
- Medium Metro: counties in metropolitan statistical areas of 250,000 to 999,999 population.
- Small Metro: counties in metropolitan statistical areas of less than 250,000 population.
- Micropolitan: counties in micropolitan statistical areas.
- Non-core: counties outside metropolitan and micropolitan statistical areas.

Choosing a definition appropriate for a particular purpose begins with understanding the features that differentiate these four sources. One fundamental difference is the underlying

concept of urban built into each source. Urban areas may be conceptualized as administrative units defined by the legal boundaries of municipalities, as structural entities delineated by high population or housing unit density and infrastructure, or as economic-functional units (such as labor market areas defined by commuting patterns). Other differences among the four sources include the geographical building blocks used to construct them (e.g., counties vs. census tracts), the criteria employed (e.g., the number of people or housing units per square mile in an urban area), and the population ranges covered by each source (e.g., micropolitan areas range in size from 10,000 to 50,000).

Table 1.  Rural Definitions and Their Key Features

| Classification | Underlying Concept | Geographic Building Block | Criteria for Setting Urban-Rural Boundary | Population Threshold Between Urban and Rural Places |
|---|---|---|---|---|
| U.S. Census Bureau's List of Places | Administrative: rural areas defined in relation to legal or locally recognized place boundaries | Municipalities and census designated places (CDPs) | Corporate boundaries or locally defined unincorporated place boundaries | Varies: can be set at any level |
| U.S. Census Bureau's Urban Areas | Land use: rural areas defined in relation to high-density, built-up areas | Census blocks | Population density (prior to 2020)<br><br>Housing unit density (2020) | 2,500 (prior to 2020)<br><br>5,000 (2020) or 2,000 housing units |
| Office of Management and Budget's Core-Based Statistical Areas (CBSAs) | Economic:  rural areas defined in relation to labor market areas | Counties | Population density and commuting | Usually 50,000; can be adjusted upward or down to 10,000 |
| Economic Research Service's Rural-Urban Commuting Areas (RUCAs) | Economic:  rural areas defined in relation to labor market areas | Census tracts or ZIP Code areas | Population density and commuting | Usually 50,000; can be adjusted upward or down to 2,500 |
| National Center for Education Statistics' Locales | Land use and proximity: rural areas defined relative to high-density areas | Census blocks and distance buffers | Population/housing density (post-2020) and distance | 5,000 (2020) or 2,000 housing units |

Source:  authors' analysis

*Census Bureau's list of incorporated and unincorporated places*
The Census Bureau maintains a list of places (https://www.census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html) and publishes decennial census and American

Community Survey (ACS) data and a limited set of intercensal population estimates and Economic Census data for places (accessed through data.census.gov). The majority of places are incorporated entities with legally prescribed boundaries. The list also includes a set of census designated places (CDPs), representing locally recognized, unincorporated communities. CDP boundaries are determined with local input. Incorporated places and CDPs are mutually exclusive and do not overlap.

Rural is not defined officially using places. Instead, in several policy contexts (e.g., designating rural health clinics), rural consists of open countryside and places below a selected population threshold. Several rural development programs within USDA follow such a strategy, including the Single-Family Housing Program and the Community Facilities Loan and Grant Program. Although the programs share a common strategy, they do use different thresholds. The underlying concept is that a place-based approach provides another structural definition to the extent that municipal boundaries and CDPs accurately distinguish more and less densely settled territory. However, the underlying concept also includes the administrative functions of municipalities, which may provide advantages in the administration of economic development programs.

Advantages:
- geographically detailed
- decennial census data available along with intercensal estimates
- easy to understand; consists of geography to which the public easily relates
- municipalities are active political jurisdictions that provide critical services relevant to rural development and are sometimes the targeted entity for federal funding

Disadvantages:
- not statistically consistent in delineating high-density areas
- municipal incorporation and annexation laws differ by State; those differences and other factors contribute to variation in whether and by how much places expand over time to incorporate their suburbs
- actual population growth may or may not be reflected in boundary changes
- the CDP concept encompasses a variety of unincorporated communities, ranging from quasi-municipal special districts (such as, the unincorporated towns in Nevada), to places with the same range of social and economic activities as incorporated places, but lacking local government, to communities with unique characteristics and needs for data (such as colonias in the border regions of Texas, New Mexico, Arizona, and California).

*Census Bureau's urban areas*
https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html

The Census Bureau delineates and publishes data for the only federal classification system that uses the term urban in an official, statistical capacity. The classification itself defines rural by default (what is not urban) and the Census Bureau produces statistics for rural areas based on

this definition. In this scheme, rural areas encompass a variety of settlements, including low-density subdivisions (often on the fringes of urban areas), settlements with fewer than 2,000 housing units or 5,000 residents (fewer than 2,500 residents prior to 2020) that are not part of a densely developed urban area, and open countryside. Prior to 2020, the Census Bureau identified two types of densely settled urban areas: urban clusters (with populations of at least 2,500 and less than 50,000) or urbanized areas (if the population is 50,000 or higher). For the delineation of areas based on the 2020 Census, the Census Bureau increased the minimum population threshold for an urban area to 5,000 and added a minimum housing unit threshold of 2,000 as an alternative to qualification based on population. It is a structural definition only, one that distinguishes less densely settled areas from more densely settled areas and smaller from larger settlements. Urban areas include most territory that would be considered suburban but does not incorporate a functional perspective to reflect the broader economic sphere of influence of urban areas.

This version of rural is the most geographically detailed. Census blocks are used as the primary geographical building block. Urban area delineation begins by aggregating census blocks with densities of at least 425 housing units per square mile (1,000 persons per square mile prior to 2020) to form an initial core. Additional blocks are then added based on a lower density threshold of 200 housing units per square mile (500 persons per square mile prior to 2020). Additional measures are employed to account for nonresidential urban land uses. The Census Bureau's adoption of housing unit density-based criteria provides a nuanced reflection of the structural footprint of human settlement, moving closer to a definition based on the built environment rather than distribution of residential population.

Although there is an established population threshold that is defined as rural for statistical purposes, it is possible to establish alternative levels by combining smaller urban areas with the area already defined as rural. For instance, to move the threshold to 10,000, simply label as rural those urban areas with populations less than 10,000.

Advantages:
- geographic detail
- statistical consistency of estimates over time—the basic concept of a densely settled area since 1910. The 2,500-person threshold was in place from 1910 through 2010 (the higher 5,000-person threshold was adopted in 2022 for use with 2020 Census data)
- detailed decennial census statistics are available for urban areas and rural components of census geography (tracts, counties, etc.)
- annual data are available from the American Community Survey with limited geographical detail; more detail available with ACS five-year averages

Disadvantages:
- not part of the Census Bureau's annual population estimates program

- not included in the Census Bureau's Economic Census data tabulations and no annual economic data such as those available for counties in the Census Bureau's County Business Patterns or available from the Bureau of Labor Statistics and elsewhere
- does not follow municipal boundaries, so not a geography familiar to the public
- though the basic concept has not changed, criteria have been adjusted over the years, somewhat hampering analysis of urbanization trends over time

*OMB's Core Based Statistical Areas (CBSAs)*
https://www.census.gov/programs-surveys/metro-micro.html.

OMB defines broad economic-functional regions known as CBSAs. They are defined in terms of their economic function and represent labor market areas that extend well beyond the built-up urban core. There are two types of CBSAs. Metropolitan statistical areas (metro areas) are defined as (1) central (or core) counties with one or more Census Bureau urban areas with 50,000 people or more, and (2) outlying counties that are economically integrated with the core counties. Economic integration is measured by the share of employed population that commutes to core counties to work or the share of workers coming from core counties (reverse commuting). Micropolitan statistical areas are defined using the same criteria as metro areas, except the core county (or counties) contains a Census Bureau urban area of at least 10,000 population.

Nonmetropolitan (nonmetro) counties are outside the boundaries of metro areas and have been widely adopted as a version of rural in many research and policy contexts. Micropolitan statistical areas, first introduced in 2000, subdivide previously undifferentiated nonmetro territory into two distinct types of counties—micro and noncore, thus providing a window on the diversity found in nonmetro America.

It is important to note that OMB provides explicit guidance that the CBSA classification is not an urban/rural classification. CBSAs contain a mix of urban and rural territory (based on the Census Bureau's urban and rural classification).

Advantages
- composed of familiar geographic units; counties are typically active political jurisdictions, have stable borders, and usually have programmatic importance at the Federal and State level
- in addition to decennial census data, estimates of county population, employment, and income are available annually
- Economic Census, County Business Patterns, and other economic data are available for counties
- the Current Population Survey's individual and household characteristics, such as age, race, education, migration, and poverty status, are estimated annually for metro and nonmetro areas by State

- most studies of rural conditions and trends, and rural research generally, refer to conditions in nonmetro areas

Disadvantages
- counties are too large, especially in western States, to adequately capture the growing complexity of settlement patterns and labor market areas; this guarantees that, along with populations living in open countryside and small towns that are economically tied to cities (as reflected in their commuting levels), metro areas include people who are legitimately rural from both a structural and economic perspective
- the underlying concept and 50,000-person threshold have remained the same since they were first defined in 1950, but criteria have changed considerably with almost every decennial update, hampering the study of trends over time

*USDA's Rural-Urban Commuting Areas (RUCAs)*
https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx

RUCA codes are based on the same theoretical concepts used by OMB. Population density, urbanization, and daily commuting identify urban cores and adjacent territory that is economically integrated with those cores. Rural can be defined in several ways but consists of open countryside and small towns outside the economic influence of larger cities. OMB's metro, micro, and nonmetro terminology is adopted to highlight the underlying conceptual connectedness between the two classification systems. The use of census tracts instead of counties as building blocks for RUCA codes provides a different and more detailed geographic pattern of settlement classification.

The classification contains 10 primary and 30 secondary codes. Few, if any, research or policy applications need the full set of codes. Rather, the system allows for the selective combination of codes to meet varying definitional needs. Metropolitan cores (code 1) are defined as census tract equivalents of urbanized areas. Micropolitan and small-town cores (codes 4 and 7, respectively) are tract equivalents of smaller urban areas (known as urban clusters until 2022). High commuting (codes 2, 5, and 8) means that the largest commuting share was at least 30 percent to a metropolitan, micropolitan, or small-town core. Low commuting (codes 3, 6, and 9) refers to cases where the single largest flow is to a core but is less than 30 percent. The last of the general classification codes (10) identifies rural tracts where the primary flow is local or to another rural tract.

The primary RUCA codes are further subdivided to identify areas where settlement classifications overlap, based on the size and direction of the secondary, or second largest, commuting flow. For example, 10.1, 10.2, and 10.3 identify rural tracts for which the primary commuting share is local, but more than 30 percent also commute to a metropolitan, micropolitan, or small-town core, respectively.

Advantages

- provides an alternative to OMB's county-based system for situations where more detailed geographic delineation is needed
- allows for the identification of economic functional areas surrounding small towns, those between 2,500 and 10,000 population; counties are too large to adequately delineate these small labor market areas
- identifies rapidly growing, potentially urbanizing zones within current nonmetro territory; conversely, identifies territory within metro counties that are outside the economic influence of the metro core
- permits stricter or looser delimitation of metropolitan, micropolitan, and small town commuting areas

Disadvantages
- census tracts are less familiar geographic entities, compared with counties or municipalities
- no intercensal data
- unstable boundaries; census tracts in many States are routinely redefined (split or merged) between decennial censuses in response to population change and the need to comply with specified population size criteria
- depends on measuring tract-to-tract commuting using data from the American Community Survey, which can include very small flows with high margins of error

*National Center for Education Statistics' Locale Codes*
https://nces.ed.gov/programs/edge/Geographic/LocaleBoundaries

The National Center for Education Statistics Locale Code typology extends the Census Bureau urban area framework and classifies all U.S. territory into an annually updated 12-category continuum. The framework initially classifies areas into four core types – City, Suburban, Town, and Rural – with each type further classified into three subtypes. City classifications are based on principal city designations determined for CBSAs (limited to the portion of Principal City boundaries contained within a Census urban area with a population of 50,000 or more); Suburban classifications are based on Census urban areas with a population of 50,000 or more; Town classifications are based on Census urban areas with a population less than 50,000; and Rural classifications are based on territory located outside of Census urban areas.

City and Suburban locales are further classified as Large, Midsize, and Small based on population size. Town and Rural locales are further classified as Fringe, Distant, and Remote based on proximity to urban areas of different sizes. This classification is an extension of the Census Bureau's standard urban definition and can be collapsed into the standard Census urban/rural dichotomy or extended into 12 detailed classifications to address a variety of analytic needs.

Advantages:
- Consistent with Census Bureau urban areas and collapsible into urban/rural dichotomy

- Provides additional granularity for Rural-Fringe, Rural-Distant, and Rural-Remote areas based on varying proximity to large and small urban areas
- Rural classifications are not constrained by county or tract boundaries
- Locale boundaries are easily accessible as data web services and static geospatial data files
- Education conditions for public schools, private schools, postsecondary schools, and school districts are available by locale
- Easily explorable through dedicated reference tools (https://nces.ed.gov/programs/maped/LocaleLookup/)

Disadvantages:
- Social/economic data are limited to core types and cannot be disaggregated for City/Suburban without custom tabulation
- Does not follow county or municipal boundaries except for Principal Cities of CBSAs
- Newer and less familiar to data users than CBSAs and RUCAs

A study comparing these sources using 2000 Census data showed major differences in overall population size and population characteristics (Cromartie and Bucholz, 2008). The biggest differences can be attributed to the underlying concept of rural as administrative, structural, or economic-functional, which strongly affects the placement of suburban and exurban populations.

For additional information about these classifications, see https://www.ers.usda.gov/data-products/rural-definitions/.

For example, an administrative definition in which municipalities of a specified population size or more are considered urban (for example, 25,000 persons) would result in designation of adjacent densely populated unincorporated territory as rural. In other words, a densely settled subdivision adjacent to, but outside, the boundaries of an urban city or town, would be classified as rural even though the settlement pattern is similar to that of neighborhoods within the city limits.

The inconsistencies created by use of a definition focused on administrative units are solved by use of a definition based on population density, housing density, or other criteria that measure the settlement structure without reference to municipal boundaries. The Census Bureau's urban areas are structural entities defined at a high degree of spatial resolution, which provides for a highly detailed (but also complicated) demarcation between urban/suburban and rural territory. These structural entities, however, do not provide a measure of the economic relationships that might exist between the entity and the surrounding rural territory, or even the economic relationships between multiple urban centers. For that level of insight, the economic-functional approach is needed.

Rural definitions based on an economic-functional perspective generate larger socioeconomic contrasts between rural and urban populations. This result emerges because a higher proportion of the population living in lower density suburban and exurban areas as well as some rural settings is counted as urban. The choice of a population threshold separating rural from urban entities affects the total size of the rural population for definitions based on census places, but does not significantly affect population size or characteristics for definitions based on the other three sources.

Table 2 provides examples of rural definitions used to administer federal programs targeting rural populations. Although many of these examples use a 50,000-person threshold to distinguish rural and urban places, the use of different definitional sources (e.g., places versus urban areas) means sometimes large differences in the targeted populations.

Table 2. Examples of Rural Definitions Used by Federal Agencies for Program Eligibility or Statistics

| Department/Agency/Program or Data Product | Purpose/Website | Source of Rural Definition | Population threshold |
|---|---|---|---|
| Department of Agriculture, Water & Waste Disposal Loan & Grant Program | Water and Waste loans and grants eligibility | Census Bureau Places | Less than 10,000 |
| Department of Agriculture, Business & Industry Loan Guarantees | Business and Industry loans eligibility | Combination of Census Bureau Places and Census Bureau Urban Areas | Less than 50,000 |
| Department of Agriculture, Economic Research Service, State Fact Sheets | Statistics for rural and urban areas by State | OMB Core Based Statistical Areas | Less than 50,000 |
| Department of Education, Rural Education Achievement Program | Rural school district grant eligibility | NCES Locales and Census Bureau Urban Areas | 5,000 |
| Department of Health and Human Services, National Center for Health Statistics natality and mortality data; monitoring of health of urban and rural residents | Statistics for rural and urban areas by county | OMB Core Based Statistical Areas | Less than 50,000 |
| Department of Health and Human Services, Centers for Medicare & Medicaid Services, Rural Health Clinics | Rural Health Clinics certification | Census Bureau Urban Areas | Less than 50,000 |

| Department of Health and Human Services, Health Resources and Services Administration, Federal Office of Rural Health Policy | Rural health funding eligibility | ERS Rural-Urban Commuting Areas (RUCAs) | 50,000 |
|---|---|---|---|
| Department of Transportation Metropolitan Planning Organization (MPO) Program | MPO designation; transportation planning and funding | Census Bureau Urban Areas | 50,000 or more |
| Department of Transportation, Transportation Infrastructure Finance and Innovation Act (TIFIA) | TIFIA Rural Project Initiative eligibility | Census Bureau Urban Areas | 150,000 |
| Department of Veteran Affairs, Veterans Health Administration, Office of Rural Health | Statistics for rural veterans | ERS Rural-Urban Commuting Areas (RUCAs) | 50,000 |

5. Practical Considerations and Pitfalls in Choosing a Rural Definition

The Census Bureau's classification of urban and rural territory is widely accepted as the baseline definition for statistical purposes. It is a geographically detailed delineation of the Nation's built-up territory that forms the basis for both OMB's delineation of CBSAs, the ERS RUCA codes, and the NCES Locale Codes. As mentioned previously, it is the only federal classification system that uses both "urban" and "rural" as terminology in data products (ERS RUCA codes and NCES Locales use the terms "rural areas" and "rural," but employ other terminology, such as "metropolitan area core," "city," "suburb," or "small town" for other categories). Researchers and other users commonly adopt the term rural to describe definitions based on the other three geographical sources. Research reports analyzing conditions and trends in nonmetropolitan counties often include stipulations such as: "In this report, the terms 'nonmetro' and 'rural' are used interchangeably." Legislation creating rural-based economic development programs typically include language such as "The term 'rural' is here defined as…" Confusion regarding the use of rural definitions comes in part from this ubiquitous practice of applying the term rural interchangeably across vastly different classification schemes. In the policy arena, the practice often leads to simply replacing the proper term (such as census places or nonmetro areas) with rural without adequate explanation.

Over the years, congressional acts creating rural-based programs have applied different rural definitions to establish eligibility. For example, a 2002 proposal to amend the Consolidated Farm and Rural Development Act included this language:

> *Except as otherwise provided in this paragraph, the terms 'rural' and 'rural area' mean any area other than—*
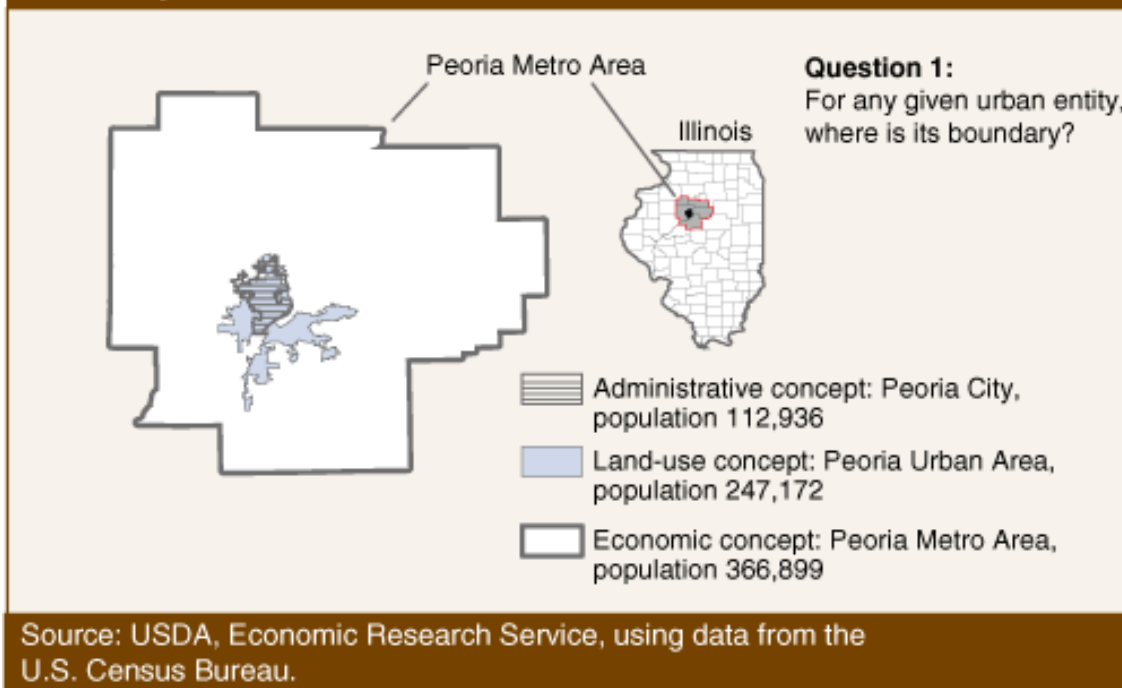>
> (i)  *a city or town that has a population of greater than 50,000 inhabitants; and*
> (ii)  *the urbanized area contiguous and adjacent to such a city or town.*

The rest of the paragraph described alternative definitions to be applied to specified programs. Many of these rural definitions are written into legislative or regulatory language with major financial implications for rural communities and families throughout the country.

All of these sources define what is urban. That is, they delineate boundaries for, and identify by name, individual entities centered on areas of population concentration, but do so in vastly different ways. Thus, someone looking for statistics on Peoria, Illinois may choose among three versions, ranging in population from just over 100,000 (Peoria city) to just under 367,000 (Peoria metro area).

Figure 3. Three Ways to Define Peoria

**Three ways to define Peoria**

Peoria Metro Area

Illinois

**Question 1:**
For any given urban entity, where is its boundary?

Administrative concept: Peoria City, population 112,936

Land-use concept: Peoria Urban Area, population 247,172

Economic concept: Peoria Metro Area, population 366,899

Source: USDA, Economic Research Service, using data from the U.S. Census Bureau.

Source: Cromartie and Bucholtz, 2008, https://www.ers.usda.gov/amber-waves/2008/june/defining-the-rural-in-rural-america/

Rural definitions typically constitute territory lying outside these entities. For instance, the Census Bureau defines rural as all territory lying outside urban areas. Rural researchers, depending heavily on county-level data, often classify as rural all counties lying outside metro areas. Other rural definitions include, in addition to the undefined territory, some of the defined entities small enough in population to be considered rural. For example, the USDA rural housing program determines eligibility, in part, as residence outside places or in places with fewer than 25,000 people.

The urban focus of the geographic sources discussed earlier makes it more economical, in almost all cases, to describe rural in terms of the territory that is not included, such as the above descriptions of rural as territory outside urban areas or outside metro areas. Part of the confusion faced by users of rural definitions comes from having to mentally "flip" definitions to focus on what is not described. On maps, it often means focusing on the white areas rather than the highlighted portions that naturally draw the eye. The descriptions of the five geographic sources—their underlying concepts, geographical building blocks, criteria, and population thresholds—often take an urban perspective. However, good decision-making in choosing an appropriate rural definition requires an understanding of the key characteristics of urban entities and how they, in turn, determine the characteristics of rural definitions derived from them.

Nonmetro territory has long served as a proxy for rural in research applications based on county-level data. So ubiquitous is the county as a unit of analysis for demographic and other

social science research, and so extensive is the adoption of metro and nonmetro as a fundamental rural-urban divide, that nonmetro has become a standard definition of rural for many research purposes. So far, the availability of micro areas (first defined in 2000) has not resulted in any shift away from this standard practice of defining rural as nonmetro. Most researchers view micro areas as a useful means to subdivide previously undifferentiated nonmetro territory, thus providing an important window on the economic and social diversity found in nonmetro America.

Counties are familiar geographic units, typically active political jurisdictions with relatively stable borders and usually of programmatic importance at the Federal and State level. In addition to decennial census data, estimates of county population, employment, and income are available annually. In the Census Bureau's Current Population Survey, individual and household characteristics (e.g., age, race, education, migration, employment, income, poverty status) are estimated annually for nonmetro areas by State. Most ERS studies of rural conditions and trends, and rural research generally, refer to conditions in nonmetro areas.

6. Conclusions

Rural researchers and policy makers constantly face a complex and vexing geographic puzzle that defies simple answers. Delineating a line between rural and urban America has always been problematic, and the complexity of today's settlement system now makes futile any search for a one-size-fits-all solution. A better strategy is to recognize that urban and rural are multi-dimensional concepts incorporating size, density, distance, and other perspectives. The choice of a rural definition should be based first and foremost on the purpose of the enterprise. It is both a challenge and an opportunity that the range of definitional options available today is quite broad.

The five types of urban entities (as defined by the four geographic sources in Table 1) and the range of thresholds available within each source permit flexibility in tailoring definitions to suit a given application. Choices for research activities are sometimes limited by data requirements. Ideally, researchers understand and report the implications of any definitional choice: Who is included and who is left out? What is being masked by using large geographical building blocks? How does the urban-rural geography vary by State? Policy makers have the opportunity to craft eligibility rules that best fit particular programs. Careful consideration of rural definitions has the potential to improve the overall efficiency of economic development programs by better targeting the intended beneficiaries.

## 7. References

Berube, Alan, Audrey Singer, Jill H. Wilson, William H. Frey. 2006. *Finding Exurbia: America's Fast-Growing Communities at the Metropolitan Fringe*, The Brookings Institution, Living Cities Census Series.

Cromartie, John, Shawn Bucholtz. 2008. "Defining the 'Rural' in Rural America," *Amber Waves,* June 1, 2008. https://www.ers.usda.gov/amber-waves/2008/june/defining-the-rural-in-rural-america

Galster, George, Royce Hanson, Michael R. Ratcliffe, Harold Wolman, Stephen Coleman, Jason Freihage. 2001. "Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept," *Housing Policy Debate* 12(4): 681-717.

Hart, Gary L., Eric H. Larson, Denise M. Lishner. 2005. "Rural Definitions for Health Policy and Research," *American Journal of Public Health* 95(7):1149-1155.

Heimlich, Ralph E., William D. Anderson. 2001. *Development at the Urban Fringe and Beyond: Impacts on Agriculture and Rural Land*, Economic Research Service, U.S. Department of Agriculture, Agricultural Economic Report No. 803.

Hewitt, Maria. 1989. *Defining "Rural" Areas: Impact on Health Care Policy and Research,* U.S. Office of Technology Assessment Staff Paper.

Isserman, Andrew M. 2005. "In the National Interest: Defining Rural and Urban Correctly in Research and Public Policy," *International Regional science Review* 28(4):465-499.

Isserman, Andrew M. 2007. "Getting State Rural Policy Right: Definitions, Growth and Program Eligibility," *Journal of Regional Analysis and Policy* 37(1):72-79.

Morrill, Richard L., John B. Cromartie, Gary L. Hart. 1999. "Metropolitan, Urban, and Rural Commuting Areas: Toward a Better Depiction of the U.S. Settlement System," *Urban Geography* 20:727-748.

Ratcliffe, Michael R. 2015. "A Century of Defining a Changing Landscape: The Census Bureau's Urban and Rural Classification, 1910 to 2010. https://www2.census.gov/geo/pdfs/reference/ua/Century_of_Defining_Urban.pdf.

Schroeder, Jonathan P. and José D. Pacas. 2021. "Across the Rural-Urban Universe: Two Continuous Indices of Urbanization in for U.S. Census Microdata," *Spatial Demography* 9:131-154.

Slifkin, R.T., R. Randolph, T.C. Ricketts. 2004. "The Changing Metropolitan Definition Process and Rural America." *Journal of Rural Health* 20: 1-16.

Waldorf, Brigitte S. 2007. *What is Rural and What is Urban in Indiana?* Purdue University, Purdue Center for Regional Development, Report No. R-4.

**BLANK PAGE TO SEPARATE SECTIONS OF THE TOOLKIT**

Analyzing Persistent Poverty Areas Using Federal Data

Recommended Citation:

Farrigan, Tracey and Jessica Crowe. 2023. "Analyzing Persistent Poverty Areas Using Federal Data," *FCSM Equitable Data Toolkit: A Toolkit for Strengthening Federal Data to Analyze Historically Underserved Populations.* Federal Committee on Statistical Methodology. November, 2023.

Table of Contents

1. Introduction: The value of persistent poverty area indicators

The Executive Order on [Advancing Racial Equity and Support for Underserved Communities Through the Federal Government](#) provides guidance to federal agencies to allocate resources with the aim of advancing fairness and opportunity for historically underserved communities, marginalized populations, and others adversely affected by persistent poverty and inequality. A data-based component of that effort involves the ability of federal agencies to produce and employ persistent poverty area indicators and related statistics, such as racial and ethnic diversity, employment and disability status, and educational attainment.

Federal agencies are building on a legacy of past efforts as they strive to better understand and address issues of equitable access. Persistent poverty indicators have been relied upon to target, implement, and monitor federal grants and programs designed to support educational and employment opportunities, health care services and healthy food access, transit service and community facilities improvements, housing assistance and land development loans, fiscal health and administrative capacity of local governments, energy savings and climate change resilience, and aid to underserved groups (see Appendix Tables 1 and 2 for links to related legislation and federal agency examples).

## Goals of this section of the FCSM Equitable Data Toolkit

Poverty is a long-standing and fundamental measure of economic well-being and inequality in the United States. Its measurement occurs initially at the levels of families and individuals. The emergence of spatial poverty measures, which characterize the extent and nature of poverty for geographic areas, reflects the recognition that the geographic concentration of poverty has its own socioeconomic dynamic (see for example [Understanding Neighborhood Effects of Concentrated Poverty](#)).  The concept of "persistent poverty" introduces a temporal dimension. It reflects the recognition that the persistent concentration of poverty in a geographic area over multiple decades also has its own socioeconomic dynamic, which differs from that associated with concentrated poverty that is intermittent or that exists for just a short time.

Though well established in federal program design, persistent poverty area measurement in the federal government is not uniform in its methodology or application. The purpose of this section of the FCSM Equitable Data Toolkit is to provide federal agencies with information on existing persistent poverty area indicators and on the underlying constructs of persistent poverty area measurement. With this information, federal agencies will know of their options when seeking a persistent poverty area measure that can meet their research or programmatic needs. For consistency with other agencies, an agency may elect to adopt an existing measure.

Discussion of tools to measure persistent poverty areas consists of three main sections:
1) A conceptual and methodological history of persistent poverty's emergence in federal policy.
2) Guidance on navigating the existing landscape of federal persistent poverty area indicators, their methodologies and uses, and alternative concepts and measures.

3) Discussion and guidance to help users make decisions about and locate resources for generating their own persistent poverty area indicator and related demographic, social, and economic statistics.

2. The persistent poverty concept and its emergence in federal policy

## What is persistent poverty?

In most countries around the world, the term persistent poverty refers to a form of chronic poverty that is defined by a person or household with relatively low income (e.g., below 60 percent of median area disposable income) for several consecutive years (see for example, Persistent Poverty in the UK and EU). Given that it is a relative measure, this definition of persistent poverty doesn't imply abject despair or even a necessarily low standard of living, but rather is understood as a measure of being at risk of poverty. Persistent poverty typically has a very different definition and meaning when used in reference to economic well-being in the United States.

The term persistent poverty in the United States generally refers to a spatial concept of poverty defined by long-standing geographic concentrations of the poor. It is commonly defined by a high rate of poverty (usually 20 percent or more) in a given geographic area over a number of consecutive decades (most often three or four, as indicated in the data years and sources column in Appendix Table 1 and in details available from links in Appendix Table 2). As an indicator of spatial well-being, such a measure effectively captures the interwovenness of localized private sector disinvestment, deficiency of community resources, and limited economic opportunities.[25] The long-term entrenchment of these conditions is often characterized by a lack of multiple baseline necessities for area residents, such as access to health care facilities, grocery stores that offer affordable and nutritious food, an adequate housing market, a sufficient educational system, jobs that pay a living wage, and essential public services. Likewise, relative material deprivation (the inability to consume goods and activities that are the norm in society) may be prevalent given lack of access to things like public transportation, parks and recreation, and civic services.

The geographic concentration of poverty can exacerbate the income poverty of individual residents by limiting the availability of services and employment prospects.[26] In conjunction, persistent poverty areas (PPAs) tend to have disproportionate numbers of people with characteristics that make them prone to disadvantage, such as low educational attainment and chronic health issues.[27] They also tend to have higher than average proportions of underserved racial and ethnic populations and other groups that historically have had trouble gaining access to economic opportunities. However, while PPAs often share similar challenges and

---

[25] This and the following statements in this section are supported by an extensive body of scholarly literature on concentrated poverty and neighborhood effects. See for example: Kuhn, 2005; Jargowsky, 2013 Meade, 2014.
[26] This phenomenon is often referred to as 'double poverty exposure' and the outcome 'poverty amplification' or 'compound deprivation.' The research on the impacts of double exposure to poverty are mixed. It is difficult to tease out cause and effect given the circularity and complexity of locational poverty and individual poverty. Yet there are ample correlates to support that the impacts can be significant, particularly for children that have been exposed for the duration of their developmental years (for a seminal summary of this work see Brooks-Gunn et. al, 1997). However, double exposure effects aren't limited to children; for an example of this research for the adult population see Ludwig, et. al, 2012.
[27] USDA Economic Research Service, 1995. Understanding Rural America, Agricultural Information Bulletin #710.

characteristics, they are not socially, culturally, economically, and environmentally homogeneous. PPAs represent a complex form of poverty that manifests across unique contexts.

## Meeting earlier demand for poverty area measurement

Persistent poverty area measurement is rooted in federal research and policy initiatives dating back to the early 1960s, namely the Johnson Administration's Great Society programs, which sought to address issues such as inequities in education and access to medical care, as well as racial discrimination and poverty. With respect to the latter, the President famously declared a War on Poverty (WOP) in 1964.[28] While poverty was the focus, there was no universal concept or measure of poverty to serve as a target or by which to determine success. Federal researchers and other social scientists "were enlisted to help define and measure poverty, to plan programs, and later to evaluate them and measure the progress achieved."[29]

Poverty definition and measurement subsequently followed two distinct paths. The better-known path is the development of the official poverty measure for families and individuals. In 1965, the Office of Economic Opportunity (OEO)[30] developed a working definition of poverty constructed from a basic income needs approach for determining the poverty status of families and for counting the poor among them.[31] Four years later the Office of Management and Budget (OMB) issued a memorandum establishing an 'official' poverty measure (OPM) and assigned the Census Bureau the task of collecting and reporting poverty statistics. Widespread use of the OPM did not come about until nearly a decade later when OMB issued a statistical policy directive (#14, May 1978) specifying the OPM as the definition of poverty to be used by all executive departments and establishments for statistical purposes.[32]

⬧ Resource: The history of the official poverty measure (census.gov)

The lesser-known path is that of poverty area measurement, which was a growing field of study for federal researchers and rural development analysts during the years leading up to the WOP (see for example, ERS Legacy of Poverty Area Measurement). Federal spatial initiatives focusing on regional development as a means of poverty area alleviation grew extensively following the late 1950s with an awareness that poverty remained high for certain places and subpopulations while the nation as a whole prospered. This concern fueled many of President Johnson's Great Society initiatives and subsequent need for spatial information on poverty conditions, which led to the commissioning of poverty area research. As part of the focus on urban renewal, the OEO charged the Census Bureau with the study of urban poverty areas. The study of rural poverty

---

[28] For a full discussion of the War on Poverty history, policies, and impacts see: Haveman et al., 2015. The War on Poverty: measurement, trends, and policy. Journal of Policy Analysis and Management, 1-46.

[29] Sawhill, I., 1988.Poverty in the U.S.: Why is it so persistent? *Journal of Economic Literature*, 26(3): 1073-1119.

[30] The Office of Economic Opportunity (OEO) was established in 1964 as an independent agency and was responsible for administering most of the War on Poverty programs.

[31] For more information, see The development of the Orshansky poverty thresholds and their subsequent history as the U.S. Official Poverty Measure.

[32] For information on how this directive was implemented see the related Census Bureau page.

areas was tasked to the USDA Economic Research Service (ERS) as part of the work of the President's National Advisory Commission on Rural Poverty (Rural Commission).

There was no single established method to identify poverty areas in the mid-1960s. However, earlier in the decade, ERS researchers developed poverty area methodologies for identifying the extent and persistence of poverty in rural areas.[33] They consisted of relative composite indices, which aggregate multiple variables into a single number that can be used to determine an area's position from the lowest to highest levels of economic well-being in society. For instance, if an area was positioned in the lowest quartile of index scores then it was categorized as a poor area. Such indices serve as a means to capture the complexity and persistence of poverty by highlighting additional deprivations experienced by area residents. They may include measures of income, population age structure, housing conditions, employment status, and educational attainment. ERS built upon this work with its contributions to the seminal Rural Commission report 'The People Left Behind' (1967) and research volume "Rural Poverty in the United States" (1968). In those publications, rural poverty areas were defined at the county-level.[34] ERS researchers continued to use an index approach for county-level poverty area measurement up until the early 1980s.

The Census Bureau published a technical report 'Characteristics of Families Residing in Poverty Areas' (1966). It outlined an index approach similar to that used by ERS, but with variable selection based on relevancy to the urban context. The spatial scale for this work was census tracts within metropolitan areas. A series of more in-depth reports on metropolitan areas followed 1967-1972.[35] Over that period, the Census Bureau transitioned from the use of a relative composite index to an approach based on a single variable – an area was defined as poor if it had an OPM poverty rate of 20 percent or more. The poverty rate cut-off was selected after Census Bureau research showed that on average previously designated metro poverty areas had an OPM poverty rate of 20 percent or more while the average for their non-poor counterparts was below 20 percent. Since then, the Census Bureau has consistently published poverty area reports for the entire country using the same approach,[36] which is often referred to as an absolute measure. That is, an area's poverty status is not determined by its economic position relative to society as with the former index approach, but rather it is solely based on a pre-defined level of disadvantage as measured by an absolute poverty rate cutoff.

---

[33] ERS developed similar indices for low-income and low levels of living agricultural and rural areas. This work is referenced in many early 1960s reports, including:  USDA Economic Research Service, Agricultural Information Bulletin #234 by Inman, 1960 and Agricultural Economic Reports:#63 by Bird, 1964 and #79 by Cowhig, 1965.

[34] At that time, anything outside of a Census defined urban area was considered rural. Rural census tract geography did not exist (urban census tracts were defined for select metropolitan areas in combination with urban area geography). Therefore, county-level geography was chosen and 'rural' was defined as the balance of the county population and land area that was not urban.

[35] See for example Trends in social and economic conditions in metropolitan areas (1969). And Trends in social and economic conditions in metropolitan and nonmetropolitan areas (1970).

[36] See for example: Changes in poverty rates and poverty areas over time: 2005 to 2019, which compares county poverty rates spanning three consecutive time-periods covering fifteen years and Census tracts defined as poverty areas based on a 20 percent cutoff.

## Evolution of definitions of persistent poverty areas

USDA rural poverty area researchers at ERS and within the Rural Development mission area have historically focused on measuring poverty at the regional or county level and have emphasized the duration of high poverty as an important indicator of spatial distress. The Census Bureau and other federal entities with an interest in urban issues have leaned toward measures based on other geographical units while often considering current rather than an extended period of time. A number of factors have influenced these differences in measures across federal agencies, such as the geography of economic regions, spatial location of administrative and other governmental bodies, and issues of special interest to a particular agency or program. However, the timing of the publication of nationwide poverty estimates and the spatial scale at which rural and urban data were produced were driving forces behind poverty area measurement decisions up until at least 1990.[37] Spatial poverty measures for rural areas, such as ERS's seminal persistent poverty classification discussed below, were developed at a time when the county was the lowest geographic unit for nationwide coverage.

In 1985, ERS published the first formal classification of persistent poverty areas ever released by a federal agency, based on the characteristics of current (1979) OMB designated nonmetropolitan counties.[38] Persistent poverty counties were defined as those in the lowest quintile of per capita income among all U.S. counties in each of the years 1950, 1960, 1970, and 1980.[39] Per capita income was chosen as a comparable measure of economic well-being in absence of an official poverty measure. Considering the policy context over the same timeframe and the characteristics of persistent poverty counties compared to others, ERS researchers noted that:

> Persistent poverty counties are among those affected disproportionally by various federal and state programs directed against poverty. However, such programs over the past three decades have not been enough to move people in these counties into the mainstream of economic activities.[40]

---

[37] The architecture of Census geography that we have today was in its infancy in the 1960s. The Census Bureau delineated urbanized areas 'to provide a better separation of urban and rural population in the vicinity of larger cities.' Basically, rural consisted of anything that was not urban. For highlights on the various measures of rurality, see the Rural Definitions and Measures Tools section of the Equitable Data Toolkit. For a full description of current Census geography see Guidance for geography users. For further explanation of urban and rural population designations in the 1960s see 1960 Census supplementary report on the population of urbanized areas (1961).

[38] For discussion of the original ERS persistent poverty county type, see the USDA Economic Research Service, Rural Development Research Report, RDRR #49, The Diverse Social and Economic Structure of Nonmetropolitan America, 1985. Reports contributing to the 1985 persistent poverty county definition included: USDA Economic Statistical Cooperative Service, Rural Development Research Report, RDRR #12, *Persistent Low-income Counties in Nonmetro America*, 1979 and USDA Economic Research Service, unpublished staff report, *A Decade of Change in Persistent Low-income Counties*, 1981.

[39] This reflected Decennial Census income years (1949, 1959, 1969, and 1979) with the earlier years pre-dating the OPM.

[40] USDA ERS Rural Development Research Report, The Diverse Social and Economic Structure of Nonmetropolitan America, RDRR #49, 1985, page 15.

In 1994, ERS published an expanded and revised version of the 1979 county classification (commonly the ERS Typology).[41] The effort reflected the need to be consistent with observed changes in the economy and society as well as federal statistical reporting practices. This included a shift in the persistent poverty methodology to an absolute measure based on an OPM poverty rate cutoff (similar to the earlier shift made by the Census Bureau). Sensitivity analysis was conducted to determine a useful cutoff for nonmetro counties. Persistent poverty status became defined by a county poverty rate of 20 percent or higher in each of the Decennial Census years 1960, 1970, 1980, and 1990. ERS has continued to use the 1994 methodology, updating the years to cover a sliding three-decade span (baseline plus three evaluation periods).[42]

    ⬇ Resource: ERS County Typology Codes

Analysis of the 1994 persistent poverty counties found that the distinct regions of persistent poverty of decades prior (Central Appalachia, the Black Belt, the Mississippi Delta, the Southwest borderlands, and Native American lands), and their racial and ethnic disparities, remained prominent. This would be true of subsequent updates to ERS's persistent poverty county list – the number of counties declined somewhat over time, but the geography and demography changed little.[43]

These findings are interpretable. Economic growth in the 1950s and 1960s and the series of federal efforts from the Great Society programs changed the landscape of poverty dramatically.[44] Poverty rates fell nationwide during the 1960s and early 1970s then leveled out through the 1980s. By 1990, poverty rates had improved almost everywhere, and the national poverty rate had fallen well below the 1960 rate of 22 percent to 13 percent. The most dramatic reductions in poverty took place between 1960 and 1970 with an average county poverty rate decrease of nearly forty percent nationwide. Counties with poverty rates above 20 percent in 1960 experienced the largest decennial poverty rate decreases, yet more than one-third remained above the 20 percent cutoff in 1970.[45]

---

[41] For more information see the USDA Economic Research Service, rural development research report, RDRR-89, The revised county typology: an overview, 1994.

[42] ERS updates the 1994 persistent poverty list every decade, maintaining the methodology that spans three decades by using four decennial years. In 2005, the Census Bureau changed the income and poverty data collection from the Decennial Census to the American Community Survey (ACS). Since then, nationwide OPM poverty statistics have been reported on a rolling 5-year basis (see Section 3). The 5-year ACS period that most corresponded to what would be the next decennial census income year (2009) was 2007-2011. The subsequent ERS persistent poverty county update included 1980, 1990, and 2000 decennial census data and 2007-2011 ACS 5-year estimates. The next update will take place following the release of the 2017-2021 ACS 5-year estimates.

[43] For a discussion of the correlation between persistently high regional poverty and race and ethnicity, see Beale, 2004, Anatomy of nonmetro high poverty areas: common in plight, distinctive in nature, Amber Waves, USDA Economic Research Service.

[44] Islam, T., J. Minier, and J. Ziliak, 2015, On persistent poverty in a rick country, Southern Economic Journal, 81(3): 653-678.

[45] This is demonstrated by a 2017 update to the 1967 ERS poverty area map as it appears in the Rural Commission report. Research by T. Farrigan, B. Weber, and A. Glasmeier presented at the Rural Poverty Research Institute conference: Rural poverty, 50 years after the People Left Behind - a research conference, looking backward and forward, March 2018.

🔸 Illustration: High poverty counties over time [interactive](interactive) and [static](static) maps

Since 1970, aside from minor changes coinciding with macroeconomic conditions, poverty rates have remained stable for the majority of counties. There have been exceptions where, for some counties, poverty rates have continued to drop significantly since 1970. However, the opposite trend of rising poverty rates has also occurred. For example, in many counties with high concentrations of Native Americans, poverty rates have gone from a level that is considered high (20 percent or more) to one that is generally considered to be extreme (40 percent or more).[46]

🔸 Illustration: Spatial concentration of Native American and Alaska Native poverty using a [racial and ethnic typology](racial and ethnic typology) of high poverty counties

---

[46] Research suggests that Native Americans residing on tribal lands, such as residents of the nine reservations in South Dakota, have not had the same success as others in accessing federal resources. For example, according to ERS analysis (Farrigan 2022) the poverty rate for Jackson County, SD, which contains part of the Pine Ridge Reservation, increased by more than 70 percent between 1960 (decennial Census poverty rate 26.4%) and 2019 (2015-2019 5-year ACS poverty rate 45.5%). The Department of Housing and Urban Development notes that about one-third of reservation homes lack electricity, adequate plumbing, and running water. In their [Fiscal Year 2017 Congressional Justifications report](Fiscal Year 2017 Congressional Justifications report) it is emphasized that 'lack of housing and infrastructure in Indian country is severe and widespread.'

3. The federal landscape of persistent poverty area indicators

## Persistent poverty area indicators are embedded in recent federal policy

Since the 1990's, other federal agencies have widely adopted ERS's measure of persistent poverty, or some variation of it. A measure akin to ERS's appeared in federal legislation in the American Recovery and Reinvestment Act of 2009 (ARRA, P.L. 111-5). The ARRA addressed how USDA was to allocate appropriated funds to three rural development programs. The legislation required USDA to allocate at least 10 percent of funds to persistent poverty counties, which the ARRA identified using a poverty rate of 20 percent or more for each Decennial Census year from 1980 to 2000; the definition in ARRA used one time period fewer than ERS's definition but referred to the time period as a span of 30 years. The ARRA provision became known as the 10-20-30 provision. Since ARRA, the 10-20-30 provision has been applied to other federal programs outside of rural development and updated to include more current data.

The Consolidated Appropriations Act of 2021 (CAA-2021, P.L. 116-260) updated and redefined the 10-20-30 provision definition of persistent poverty counties. The CAA-2021 also expanded the focus on poverty areas to include high poverty census tracts, identified by a poverty rate of 20 percent or more for a single time period.[47] Altogether, CAA-2021 includes multiple definitions of persistent poverty in conjunction with the Act's provisions to various federal agencies and initiatives (see Appendix table 2 for links to specific policies).

A feature of these definitions is that they combine geographic levels of counties and census tracts. The poverty concept used for census tracts in CAA-2021 is high poverty measured by a 5-year average (from the American Community Survey) in contrast to the multiple decades that are used for defining persistent poverty at the county level. There is growing demand for persistent poverty measurement at the census tract level. In 1990, the Census Bureau first assigned census tract geography to the entire nation and collected tract-level income and poverty data for all tracts (for more information see Tracts and Block Numbering history at Census.gov). With the additional years of nationwide tract-level data now available, persistent poverty measures for census tracts can be constructed and have begun to be adopted by federal agencies.[48] A challenge with defining the census tract rather than the county as the geographical unit is that the tract level can involve greater methodological complexity, as discussed in Section III below.

Illustration: Comparison of persistent poverty counties and census tracts

## Diversity of persistent poverty area indicator uses by federal agencies

Persistently poor areas are generally defined by a high proportion of residents with incomes below the federal poverty level over multiple decades (see for example, ERS Poverty Area

---

[47] This is akin to many of Census Bureau's post 1970 publications on poverty areas. For example, Census Bureau reports published in 1995, 2005, 2011, 2014, and 2020 analyzed high and concentrated poverty census tracts.

[48] The National Institute of Health, National Cancer Institute recently developed and implemented a census tract measure of persistent poverty (using an updated version of ERS's county methodology) for grants and program use. See Cancer control research in persistent poverty areas.

[Measures](#) definitions and others in Appendix tables 1 and 2). By this definition, persistent poverty for an area is both systemic and enduring. Persistent poverty reaches beyond economic well-being to also encompass social, demographic, political, cultural, and environmental outcomes of interest. As such, persistent poverty indicators produced by ERS or others are relied upon to target, implement, and monitor federal initiatives aimed at addressing a wide range of issues. They are used by grants and programs designed to assist with educational and employment opportunities, health care services and healthy food access, transit service and community facilities improvements, housing assistance and land development loans, fiscal health and administrative capacity of local governments, energy savings and climate change resilience, and aid to underserved groups.

Appendix table 2 provides a summary of federal agencies that use persistent poverty area indicators in their programs and granting initiatives and, where available, links to their definitions and data resources. The summary is not exhaustive but illustrative of the diverse programs across the federal government. As application of persistent poverty indicators has become more common across the federal government, it has been accompanied by the development of tools that can assist stakeholders with determining their persistent poverty status and with obtaining corresponding socioeconomic and demographic information.

- Resource: [ERS Poverty Area Measures data product](#)

## Alternative concepts and indicators

There are a number of other spatial indicators of economic well-being used in federal program implementation and research. In some cases, federal programs establish multiple eligibility criteria using several single-dimension indicators such as income, poverty, health, education, or housing quality.[49]  They may also include geographic and demographic metrics to reflect known disparities, such as rural/urban designations and race/ethnicity. Others aim to capture multiple dimensions of areawide economic hardship and material deprivation in a single indicator, such as an index (see for example, the Appalachian Regional Commission's index-based system for [classifying economic distress in Appalachian counties](#)).

- Resource: [Link to Rural Definitions and Measures Tools](#)
- Resource: [Link to Race and Ethnicity Tools](#)

These indicators are typically used as relative measures similar to early persistent poverty indicators as previously described, while contemporary persistent poverty indicators use a poverty rate cutoff constructed from the official poverty measure, which is an absolute measure. They are conceptually different in that the former is based on the comparative economic status or standard of living in society as a whole, whereas the latter is based on a threshold meant to reflect a minimum acceptable level of economic well-being.

The relative multidimensional approaches used in the federal government today closely resemble those used in the past. Most of these originated in the 1960s, stemming in large part

---

[49] For an example of variable selection from the American Community Survey see [A multidimensional poverty measure using the American Community Survey](#).

from the Public Works and Economic Development Act of 1965 (PWED) and have since changed little if at all. The PWED states that to be eligible for assistance a project must be located in an area that meets one or more of the following criteria:

- Low per capita income – the area has a per capita income of 80 percent or less of the national average.
- Unemployment rate above national average – the area has an unemployment rate that is, for the most recent 24-month period for which data are available, at least 1 percent greater than the national average unemployment rate.
- Unemployment or economic adjustment problems – the area is in an area that has experienced or is about to experience a special need arising from actual or threatened severe unemployment or economic adjustment problems resulting from severe short-term or long-term changes in economic conditions.

The PWED is directly referenced by multiple federal programs while other programs use a similar design though not directly referential. Examples are listed in Appendix table 3.

4. Methodological considerations for persistent poverty area measurement

Developing a measure of persistent poverty or understanding an existing measure better for programmatic or research purposes involves several key considerations. They can be summarized in a conceptual framework that is introduced here using four D's: Data, Duration, Depth, and Decisions. With a focus on the Official Poverty Measure, which is prominent in current federal methodologies, the discussion of the 4-D Framework for Persistent Poverty Measurement examines each factor in turn. The issue of spatial scale, which is embedded in each of the four D's, is discussed as well.

## Data – available sources

Ultimately, any methodological considerations depend on the nature, strengths, and limitations of the data that are available. Federal data resources that can help meet the need for multiple years of spatial poverty statistics include Decennial Census data, the American Community Survey 1- and 5-year estimates (ACS), and the Small Area Income and Poverty Estimates (SAIPE).

*Decennial Census.* The Decennial Census is a national survey that dates back to 1790, but its content has changed over time. Between 1970 and 2000 the Decennial Census collected certain demographic and housing information from the entire population using what was called the short form. A subset of the population – about one in six households – answered a second questionnaire, called the long form, that collected more detailed information including data on income that were used for measurement and analysis of poverty.[50] The long form was eliminated following the 2000 Census with the advent of the American Community Survey, which posed those questions and others on an ongoing basis instead of once each decade.
- Resource: [The Decennial Census of Housing and Population Data](#)

A benefit of the Decennial Census is that it has an extensive selection of geographies and the availability of corresponding demographic, socioeconomic, and housing characteristics. However, it meant that county level poverty estimates based on the Decennial Census were only available every ten years from 1960 (poverty measures were added post Census) through 2000. Similar data at the census tract level are also available for the entire country for 1990 and 2000.
- Resource: [Decennial Census Geographies](#)
- Resource: [Decennial Census Data sets](#)
- Resource: [Historical County Level Poverty Estimates Tool](#)

*American Community Survey (ACS).* In 2005, the American Community Survey replaced the Decennial Census long form. The ACS is an annual, nationwide survey with a sample size of about 3.5 million addresses across the 50 states and Puerto Rico. One of the main purposes of ACS is to help Congress determine funding and policies for a wide variety of federal programs. To do so, the ACS includes a diverse set of social characteristics (e.g., disability, educational

---

[50] The Census also collected sample data similar to the long form in 1940 and 1950, but collection of all information was done through a single form instead of two.

attainment, language spoken, and veteran status), economic characteristics (e.g., employment status, health insurance, income, and earnings), housing characteristics (e.g., computer and internet use, monthly owner costs, rent, year structure was built), and demographic characteristics (e.g., age, sex, race, Hispanic origin, and relationship to householder). It also includes a rich set of geographies: nation, states, congressional districts, counties, places, census tracts, and other localities.

    ✦ Resource: [Geography in the American Community Survey](#)

ACS is used to obtain one-year estimates for select geographies (that are sufficiently large to support statistical estimates based on a single year of data) and five-year estimates (which pool data across 5-years to generate period averages) for all geographic areas down to the census tract and block-group levels. The Census Bureau recommends that a comparison of five-year estimates over time for a given geographical unit be limited to five-year periods that do not overlap (see [How should users compare 5-year estimates?](#) at Census.gov); a comparison of overlapping five-year periods would include one or more years of the same data, which would make interpretation of the comparison problematic. The release of the 2015-2019 estimates represented the first time that three consecutive non-overlapping five-year periods were available, thereby offering trend data for most Census geographies for a combined 15-year period. All ACS data products can be found on Census' digital data platform.

    ✦ Resource: [American Community Survey, Multiyear Accuracy of the Data](#)
    ✦ Resource: [Census data digital platform](#)

The choice of using ACS or Decennial Census data can influence the measurement of persistent poverty areas. The two data sources use different timeframes for measuring income and the population bases differ somewhat. The Census Bureau offers a summary of these differences and guidance about making comparisons across the two data sources for temporal analysis. Census also provides to ACS users a series of special topic handbooks, which include handbooks targeted for federal agencies and for researchers.

    ✦ Resource: [Differences between the ACS and Decennial Census](#)
    ✦ Resource: [Handbooks for ACS Data Users](#)

*Small Area Income and Poverty Estimates (SAIPE).* The Small Area Income and Poverty Estimates program of the Census Bureau offers annual estimates of income and poverty statistics for all states, counties, and school districts. Its main objective is to provide poverty and income estimates for the administration of federal programs and the allocation of federal funds to local jurisdictions. The estimates are available for 1993 then annually from 1995 to present. The details of SAIPE methodology differ year to year. In general, income and poverty for states and counties are modeled estimates derived from a combination of Census population data and poverty inputs from surveys, specifically the Current Population Survey up until 2004 and ACS onwards, and administrative records. Therefore, they are not direct counts from enumerations or administrative records, nor are they direct estimates from sample surveys.

    ✦ Resource: [Small Area Income and Poverty Estimates Program](#)
    ✦ Resource: [Income and Poverty Interactive Data Tool](#)

The main appeal of SAIPE is that it provides single-year estimates that are updated annually; in contrast, the ACS provides estimates based on five-year period averages (for most substate geographies). New five-year ACS estimates are released by the Census Bureau annually, but four of the five years will be the same as in the previous year's five-year period average. SAIPE estimates also generally have lower variance than ACS estimates. Census data user guidance notes that for counties and school districts, particularly those with populations below 65,000, SAIPE provides the most accurate subnational single-year estimates of poverty. Typically, SAIPE estimates are most useful when single year poverty estimates for all ages, ages 5-17, or less than age 18 for US counties or for ages 5-17 at the school district level are desired. SAIPE is the only complete source for these estimated domains. The Census Bureau provides guidance for data users that desire poverty estimates for other subgroups characteristics and geographies.

- Resource: Which data source to use for poverty
- Resource: Differences between available surveys/programs for poverty

Limitations of SAIPE include substantially fewer geographic scales than the Decennial Census or ACS and the lack of supplemental variable selection that exists with the other data sources. However, the limitation that may be most impactful when considering whether to use SAIPE is technical. The modeling methodology results in numerous cautions about the use of the estimates, which have implications for persistent poverty area measurement. For instance, correlations amongst the estimates should be taken into account to provide a more accurate test for significant year-to-year changes. Some data years are more concerning than others, for example, when considering transition years from the use of CPS to ACS in the models. SAIPE technical details, cautions, and guidance are available on the Census Bureau website for those with at least moderate statistical expertise.

- Resource: General cautions about comparing estimates
- Resource: Guidance for making year-to-year comparisons of SAIPE estimates

A summary of data options is provided below to assist with making data year/source selections.
- ⟹ County level
  - o For years 1990 or earlier, Decennial Census.
  - o From 1995 to 2005, annual SAIPE and 2000 Decennial Census.
  - o From 2005 to present, annual SAIPE and ACS 5-year period estimates, beginning with 2005-2009 (use of non-overlapping 5-year periods recommended).
- ⟹ Census tract level
  - o For years 1980 or earlier, tract geography does not exist for the entire nation, but what does exist is available through the Decennial Census. Additional census tract geographies and poverty estimates derived from and available through various data sources are also an option (use with caution). This issue and alternative data sources are discussed below, under the spatial scale heading.
  - o From 1990 to 2000, Decennial Census.

- o From 2005 to present, ACS 5-year period estimates, beginning with 2005-2009 (use of non-overlapping 5-year periods and consideration of margins of error recommended, also discussed later with spatial scale).

## Duration – appropriate timeframe

Federal data users often seek definitive answers about the appropriate timeframe for persistent poverty area measurement. Yet, the temporal scale for any poverty area measurement is context specific, depending on the poverty phenomenon of interest and purpose.

*Research.* ERS's 1985 persistently low-income county type and subsequent persistent poverty county type (beginning in 1994) differ by the specific indicator of well-being and the use of a relative or absolute approach, but they both use four data points spanning 30 years. The decision for this approach was driven by the availability of data and by the research context. Specifically, ERS researchers sought to examine spatial trends in poverty over as long a period as possible, with a particular interest in examining the distributional impacts of 1950s and 1960s economic prosperity trends, the War on Poverty initiatives, and related issues of interest to ERS such as agrarian technological change and trends in migration. More generally, there was an intention to provide researchers and federal stakeholders with tools to inform and evaluate federal policies and programs. At the time, the sole data source available for long-term analysis with sub-state geography was the Decennial Census.[51]

⇒ The original motivation for the measurement of persistent poverty was to examine the endurance and distribution of high poverty rates over as long a time as possible. From this specific research perspective, it could be argued that the methodology might be to maintain or extend rather than shorten the persistent poverty timeframe used by ERS.

*Policy objectives.* Poverty area measurement as applied in the federal context has been tailored to specific policy objectives. This is demonstrated by the 10-20-30 provision of the ARRA and the CAA-2021, which were discussed in Section II. The persistent poverty area methodologies defined in those legislative acts are very similar to ERS's. The main difference is that they use three data points instead of four. A result of using fewer data points is that more areas can meet the persistent poverty criteria because an area would not have to exhibit high poverty for as many consecutive points in time.

⇒ Decisions about the number of data points to include in the persistent poverty area definition can be used to expand or contract inclusivity. Broad policy objectives meant to reach a large contingency of places in need might consider fewer data points.

---

[51] The Current Population Survey Annual Social and Economic Supplement (CPS ASEC) was/is another poverty data option. It is one of the oldest, largest, and most recognized surveys in the U.S. and serves as the data source for the Official Poverty Measure and the Supplemental Poverty Measure. However, the geography in the unrestricted public use files is limited. The Census Bureau recommends that it is best used for national and state-level (3-year averages recommended) analysis.

Also similar to ERS's definition is that the persistent poverty measures of the Acts are described as spanning thirty years. Yet some use a shorter time span, while others use a potentially longer time span. For instance, the rural development definition that appears in the 10-20-30 provision and CAA-2021 (table 1,) uses three data points including 1990, 2000, and 2007-11. They are approximately equally spaced apart, by about 10 years, considering the transition from Decennial Census data to ACS 5-year estimates. The time span between the first and the last data point is 20 years. In comparison, the public works definition of the CAA-2021 (table 1) also uses three data points including 1990 and 2000. The third data point is stated as the most recent Small Area Income and Poverty Estimates. The most recent SAIPE to date is 2021, meaning that the public works definition uses three data points, unequal distances apart, spanning more than thirty years. Given that SAIPE is updated annually, the time span for this definition will increase annually.

A potential result of annual updates is that the persistent poverty status of some areas, particularly those with poverty rates nearest to 20 percent, will fluctuate. Poverty rates can change rapidly from year to year due to cyclical changes in the macroeconomy, causing short-term economic difficulty or improvement. Research has shown that this is especially problematic for rural manufacturing and natural resource-based economies that rely heavily on one relatively unstable industry.

⇒ Cyclical economic trends, which last a little more than five years on average, should be taken-into-account when evaluating change in PPA status to ensure that the change captured represents permanence rather than fluctuation.

*Program needs.* Persistent poverty area indicators are often constructed to reflect the specific needs of federal programs, hence the diversity shown in Table 2. Many adopted persistent poverty area indicators before any definition had appeared in federal legislation, while others developed their own following the 10-20-30 provision of the ARRA, even though their program areas did not fall under rural development. The lack of uniformity of definitions found in CAA-2021 (Table 1) is likely influenced by the need to conform to the definitions already embedded in specific program areas. Since the 10-20-30 provision was first introduced, which targets persistent poverty counties, there has been widespread recognition that counties are not the appropriate unit of geography for all situations. Communities with entrenched concentrations of poverty can fail to meet program eligibility because they exist within counties that do not meet the criteria for persistent poverty status. This led to a search of more nuanced targeting mechanisms that can identify the diverse array of persistently poor communities across the nation. One result has been the adoption of census tract level poverty area measures in place of or in addition to county level persistent poverty area measures.

All of the census tract level poverty area measures that appear in legislation to date are defined by one data period. Very few program agencies have developed and implemented persistent poverty area census tract indicators. The lack of availability of comparable tract level data over the long-run and methodological complexities, as discussed in Section II and elsewhere in this

report, can create hindrances. The use of one data period, however, can be problematic if the program goal is to address long-term economic difficulties.

The nature of economic difficulty associated with the persistent poverty phenomenon is systemic. There are fundamental structural differences between persistently poor areas and their counterparts. Allocating aid using one data period (single-year or a multi-year average) or multiple data points in the short-run may provide a misleading picture of long-term economic well-being. Research has found that a timeframe of less than five years results in estimates that do not represent well the phenomenon of persistent poverty (as measured using four time periods of data spanning thirty years). This research stems from the concentrated neighborhood poverty and resource economics literature. For instance, the characteristics most associated with persistent poverty are not especially prevalent in chronic or short-term concentrated poverty areas.[52]

⇒ The poverty area indicator chosen to target aid ought to vary depending on the nature of economic difficulty that the program is meant to address. A timeframe equal to or spanning beyond thirty years is feasible and useful when there is interest in areas with historical legacies of poverty-related conditions.

## Depth – poverty rate cutoff

As discussed in Section 1, fifty years ago the Census Bureau first used a 20 percent cut-off for the OPM poverty rate to measure depth of area poverty. The decision was based on correlation between such a measure and the metropolitan area poverty status derived from a multi-dimensional index measure of economic well-being. About 20 years later, ERS was considering its own study of poverty rate cutoffs relative to nonmetro counties and decided to use a 20 percent cutoff as well.

More recently, the contemporary relevance of the 20 percent cutoff was tested as part of a '50 years later' exploratory update to the ERS poverty area research that appeared in the seminal 1967 People Left Behind report on rural poverty. Using the 1967 relative index methodology and data from the 2012-2016 American Community Survey five-year estimates, ERS researchers found that the corresponding OPM poverty rate for nonmetro and metro poor counties was about 22 percent and 18 percent, respectively. This finding suggests that when considering the conceptual and methodological groundwork for persistent poverty area measurement, a 20 percent OPM cutoff is appropriate to the combined metro/nonmetro persistent poverty county designation today.

⇒ The 20 percent poverty rate cut-off is widely adopted for high and persistent poverty area measurement; it is regarded to be relevant to rural (nonmetro) and urban (metro) areas.

Over the same timeframe (approximately 1970 to 2020) research by some academic researchers also suggest that 20 percent is the critical poverty rate cutoff at which residents

---

[52] Gans, H., 2010. Two American problems: concentrated poverty, a critical analysis. *Challenge*: 53(3) 82-96.

begin to experience the impacts of area-wide poverty.[53] Those impacts appear to be greater at even higher poverty rates, raising the question of whether a "critical impact" point might exist above 20 percent. These effects tend to plateau or slow significantly beyond a 40 percent poverty rate, which is commonly referred to as the threshold of extreme poverty.

⇒ The 40 percent poverty rate cut-off, or extreme poverty area indicator, has been used widely in academic research. It has not received as much attention by federal researchers, nor has it been adopted widely for use with federal programs, but it has been growing in popularity as of late.

It is possible that a poverty rate just above the 20 percent threshold over multiple decades may be more damaging to areawide well-being and thus have greater potential to impact residents than a poverty rate nearing the extreme threshold of 40 percent for just a few years. The long-term erosion of government financial resources in the face of a limited residential tax base is one such scenario.[54] Given these considerations, the 20 percent poverty rate seems a reasonable and defensible threshold for identifying high spatial poverty.

⇒ The 20 percent poverty rate cut-off is typically applied in federal policy and research as >= 20.0 percent. Different rounding options are often discussed but to date have failed to impact standard practice.

## Spatial scale – counties and census tracts

Historically, persistent poverty areas have been defined at the county level. Even so, persistent poverty can be measured at any spatial scale for which appropriate data are available. Demand for sub-county measures has grown in recent years, bringing increased attention to census tract data (often used as a proxy for neighborhoods). As previously noted, these data have been available for the entire nation since 1990, initially based on the decennial Census followed by the ACS (as five-year estimates).

There are special considerations with tract level data, including how boundaries of census tracts change over time much more than do boundaries of counties. Many statistical areas (like census tracts and block groups) are updated once per decade to reflect the most recent Decennial Census. Census tract geography can change dramatically from one decade to the next making temporal comparisons difficult. In order to maintain the greatest geographic coverage when constructing a persistent poverty area indicator, the geographic normalization of tract level data over time should be considered. There is no set methodology for normalizing data, but tutorials for doing so exist as do options to use open access data or to purchase proprietary normalized census tract datasets from private vendors.

🕂 Resource: Updates to census tract boundaries and how to compare them decade to decade

---

[53] See for example, Galster and Booza, 2010 The mechanisms of neighborhood effects: theory, evidence, and policy implications.  And Galster et.al, 2006 The social costs of concentrated poverty.

[54] A broad literature supports that the conditions found in persistent poverty and economically distressed areas make them less attractive to private sector investment, thereby discouraging private revitalization efforts and further decreasing the local government tax base.

County geography also changes periodically, sometimes with the addition of new counties or the splitting of old ones or the annexation of counties or county equivalents. While the change in county geography is less problematic than it is with census tracts, attention must be paid to this issue when constructing a persistent poverty area indicator, particularly in the case of Alaska where the county-equivalent geography has changed in every decade since statehood. The Census Bureau provides a list of these geographic changes by decade beginning with 1970, though note that there is no authorized recommendation for comparing county-level updates as is true for census tracts.

Another issue with census tracts, and to a lesser degree with counties, is the error of the estimates, often measured in terms of margin of error (when using survey data). Poverty estimates for smaller geographies have higher margins of error. If there is a desire to subset the population into smaller groups, such as by race, then the magnitude of error increases and the estimates can be highly unstable. Data error estimation and interpretation for Decennial Census sample data (long form) requires some degree of statistical expertise. Conversely, the Census Bureau provides calculated margins of error for all ACS estimates (all geographies) and guidance on how to use them, making it amenable to the beginner. Similarly, the Census Bureau also publishes SAIPE estimate (counties and school districts) confidence intervals.

Measures of uncertainty should be used when available. One common practice is to develop an index of reliability from the MOE's. The index can be translated into a scale, such as low, moderate, and high reliability. This information can be used to inform decisions about whether poverty estimates for select geographies are reliable enough to report and analyze. There are several different versions of reliability indices used in the federal government, but for spatial analysis the most popular is that developed by ESRI (an international GIS software and applications supplier).

Another common practice is to use the MOE's to estimate upper and lower bounds of the estimate (lower = estimate – MOE; upper = estimate + MOE). This provides information on the potential range of the estimate, within the margin of error. There is no standard practice on how to use this information, but when developing poverty area measures one option is to use it to determine if the MOE impacts the high poverty status of a given area. For instance, if the lower bound, the estimate, and the upper bound yield different poverty status outcomes (e.g., using a 20.0 percent cutoff) then caution should be considered when using that estimate for poverty area analysis and with interpreting findings. A less common practice is the opposite; to define an area as high poverty if *any* of the values (lower bound, estimate, upper bound) are 20 percent or higher. This is not recommended. The potential for false positives is high, particularly for census tracts. In general, when possible other measures of well-being, contextual information, and alternative data sources should be used to validate findings.

- Resource: [Using ACS estimates and MOEs and additional resources](#)
- Resource: [ESRI importance of margins of errors and mapping](#)
- Resource: [Calculating margins of error the ACS way](#)
- Resource: [Suppressing unreliable observations and transparency of reliability](#)

## Decisions – putting it all together

There are numerous decision factors when producing a persistent poverty area indicator. The primary factors,[55] as just described, should be considered in conjunction with the motivation and purposes of the end user. At first glance, it might seem that a user could select any one of the existing federal persistent poverty area methodologies, which for the most part do not vary substantially. However, even slight variation in methodology can change the make-up of the areas identified as persistently poor. This phenomenon can impact how well the policy or program is identifying its target population. For the counties and communities in need, difference in definitions and measurement can make the difference between the area being eligible or ineligible for federal funding. At the least, attention should be paid to data limitations and to the motivation and conceptualization of a particular measure. The decision-making exercise and ERS example provided below may help with this process.

### *Decision-making exercise.*

A. Questions / answers that may be of help include, but are not limited to:
- Is the purpose of the persistent poverty indicator for research, meeting broad policy objectives, meeting specific program need, or a combination?
- Is the target population the most historically impoverished areas, chronically poor areas, newly or temporarily poor areas, or a combination?
- Is the interest only in persistent poverty, as traditionally measured using a high poverty rate cutoff (20 percent or more) or is persistent extreme poverty (40 percent or more) also of interest?
- Is a ten-year period between data points acceptable or desirable? Is there a need to consider an alternative (shorter, longer, or varying)?
- How important is it to have the most current, annual, single-year poverty rates?
- Is direct access and comparability of other socioeconomic, demographic, and/or housing variables important?
- Is the ability to determine reliability of the estimates a priority? If yes, what degree of difficulty in doing so is acceptable (novice, intermediate, or expert)?
- Is there a need for a census tract measure of high or persistent poverty, in addition to or instead of a county measure of persistence?

---

[55] The factors presented are those that are most critical to resultant persistent poverty area geography and counts. They are also the most representative of ongoing definitional debates and where there are differences in existing legislative language. However, there are additional discrete factors to consider, such as rounding decisions with respect to the poverty rate cutoff. This is less of an issue in terms of the impact on the persistent poverty area count. And as of the writing of this report this factor is uncontested in legislation – the language consistently references a poverty rate cutoff of 20 percent or more, which infers 20.0 percent or higher (not 19.9 percent).

B. Once the answers to these and other questions (relevant to motivation, data limitations, concept) are established, as a next step, consider:

✓ Reviewing the existing persistent poverty area definitions and poverty area measures in tables 1, 2, and 3, as well as ERS's definition, to determine their potential for adoption.

✓ If none of the existing measures are acceptable, keeping your answers in mind, revisit the data, duration, depth, and spatial scale discussions of Section III to be reminded of the various aspects of the decision process and what to consider in making decisions.

C. Also, consider consulting with other federal program agencies about the resultant persistent poverty area definition (adopted from existing definitions or uniquely designed) and their experiences with the same/similar/unique persistent poverty area indicators.

*Example decision-making exercise related to ERS's definition:*
A. By answering the exercise questions, it was determined that in order to meet ERS's primary research and secondary federal agency support needs, there are two fundamental (temporal) issues for determining persistent poverty area status and change in persistent poverty area status.

1) Poverty should be measured over the long run to capture structural poverty rather than cyclical poverty, using a timeframe adequate to reflect extent of economic difficulty.
2) Economic cycles should be taken into consideration when evaluating change, using a time span adequate for capturing permanent improvement or lack thereof.

B. Upon review of definitional options that might address the two issues, ERS's decision was to continue to use the county-level methodology that was established in 1994:

✓ Use a timeframe that spans thirty years, with a consistent 10-year time span between poverty indicator data points (baseline plus three evaluation periods) and updates.

✓ Generate a comparable census tract level persistent poverty area indicator, allowing for within county analysis of persistent poverty.

C. Federal stakeholder consultation: Agencies with programs aimed at addressing persistent poverty directly or aimed at addressing various issues associated with long-term economic difficulty (e.g., access to healthcare) have reported that ERS's approach works well. It captures areas consistent with program concerns and observed conditions. It limits cyclical variation in program eligibility status and provides a sufficient timeframe for program impact evaluation. A change to an update every five years (using non-overlapping 5-year ACS) or every year (using concurrent 5-year ACS or SAIPE) instead of ten would diminish the usefulness of the indicator.

## 5. Appendix

Table 1. Consolidated Appropriations Act of 2021 poverty area definitions

| Policy objective / program area | Geographic scale | Concept | Indicator of well-being | Data years included and sources |
|---|---|---|---|---|
| 1. Rural development | county | persistent poverty | poverty rate of 20 percent or more | 1990 and 2000 Decennial Censuses and the 2007-2011 American Community Survey 5-year estimates |
| 2. Public works | county | persistent poverty | poverty rate of 20 percent or more | 1990 and 2000 Decennial Censuses and the most recent Small Area Income and Poverty Estimates |
| 3. Comprehensive environmental response, compensation, and liability | county | persistent poverty | poverty rate of 20 percent or more | 1990 and 2000 Decennial Censuses and the most recent Small Area Income and Poverty Estimates |
| 4. Community development financial institutions | a. census tract | high poverty | poverty rate of 20 percent or more | 2011-2015 American Community Survey 5-year estimates |
| | b. census tract | high poverty | poverty rate of 20 percent or more | 2010 Island Areas Decennial Census |
| | c. county | persistent poverty | poverty rate of 20 percent or more | 1990 and 2000 Decennial Censuses and the 2011-2015 American Community Survey 5-year estimates |
| | d. county | persistent poverty | poverty rate of 20 percent or more | 1990, 2000, and 2010 Island Areas Decennial Census or equivalent data of the Bureau of the Census |
| 5. Transportation infrastructure | a. county | persistent poverty | poverty rate of 20 percent or more | 1990 and 2000 Decennial Censuses and the most recent Small Area Income and Poverty Estimates |
| | b. census tract | undefined | poverty rate of 20 percent or more | 2014-2018 American Community Survey 5-year estimates |

| 6. Covid-19 pandemic response and recovery (emergency capital investment program) | a. communities | low- and moderate-income | unspecified | unspecified |
| | b. communities | underserved | unspecified | unspecified |
| | c. counties | persistent poverty | unspecified | unspecified |

Table 2. Summary of federal agencies that use persistent poverty area indicators

| Program, policy, or grant name | Administering department/agency | Issue(s) addressed | PPA spatial scale | Website |
|---|---|---|---|---|
| Areas of Persistent Poverty Program | Department of Transportation, Federal Transit Administration | Support planning, engineering, and financing to improve transit services in areas of long-term economic distress | Counties and census tracts | DOT FTA Areas of Persistent Poverty Program |
| Bank Enterprise Award Program | Department of Treasury, Community Development Financial Institutions Fund | Provides formula-based grants to successful applicants for increasing Qualified Activities | Counties | DOT CDFI Bank Enterprise Award Program |
| Community Development Financial Institutions Program | Department of Treasury, Community Development Financial Institutions Fund | The CDFI Program offers both Financial Assistance and Technical Assistance awards to CDFIs | Counties | DOT CDFI Community Development Financial Institutions Program |
| Native American CDFI Assistance Program | Department of Treasury, Community Development Financial Institutions Fund | Financial Assistance and technical assistance awards are made to Native CDFIs. | Counties | DOT CDFI Native American CDFI Assistance Program |
| Distressed Cities and Persistent Poverty Technical Assistance Program | Department of Housing and Urban Development | Improve fiscal health and build administrative capacity of relatively small units of general local government (UGLGs or local governments) and their nonprofit partners in places experiencing persistent poverty and economic distress. | Census tracts | HUD Distressed Cities and Persistent Poverty Technical Assistance Program |
| Rural Community Development Grants | Department of Health and Human Services Administration for Children and Families' Office of Communities Services | RCD grants support training and technical assistance for creating and maintaining safe and affordable water and wastewater systems in the nation's lowest income rural communities, including | Counties | HHS ACF Rural Community Development Grants |

| | | tribal areas, many of which have populations at or below 2,500 individuals. | | |
|---|---|---|---|---|
| Expanding Cancer Control in Persistent Poverty Areas | Department of Health and Human Services, National Cancer Institute | Provide resources to support highly collaborative, multi-disciplinary Program Projects (P01s) that focus on the development and conduct of cancer control research in low-income and/or underserved populations living in persistent poverty (PP) areas. | Counties and census tracts | NIH NCI Expanding Cancer Control in Persistent Poverty Areas; includes link to census tract persistent poverty data |
| Cancer control grants; Persistent Poverty Initiative; Cancer Control Research in Persistent Poverty Areas | Department of Health and Human Services, National Cancer Institute | Cancer control in designated NCI cancer centers and cancer control research to understand the causes and distribution of cancer in populations, support the development and delivery of effective interventions and monitor and explain cancer trends. | Counties | NIH NCI Cancer Control Research in Persistent Poverty Areas |
| Rural Community Facilities Program; Community Facilities Grant | U.S. Department of Agriculture Rural Development | Provides affordable funding to develop essential community facilities in rural areas. | Counties | USDA RD Community Facilities Program |
| Rural Business Program Account; Rural Business Development Grants | U.S. Department of Agriculture Rural Development | Provides technical assistance and training for small rural businesses. | Counties | USDA RD Rural Business Development Grants |

Table 2. Summary of federal agencies that use persistent poverty area indicators (continued)

| Program, policy, or grant name | Administering department/agency | Issue(s) addressed | PPA spatial scale | Website |
|---|---|---|---|---|
| Rural Business Program; Rural Economic Development Loans Program | U.S. Department of Agriculture Rural Development | Provides funding for rural projects through local utility organizations. | Counties | USDA RD Rural Economic Development Loan and Grant Program |
| Rural Business Program; Rural Cooperative Development Grants | U.S. Department of Agriculture Rural Development | Improves the economic condition of rural areas by helping individuals and businesses start, expand or improve rural cooperatives and other mutually-owned businesses through Cooperative | Counties | USDA RD Cooperative Development Grant Program |
| Water and Waste Disposal Program Account; Water and Waste Disposal Loan and Grant Program | U.S. Department of Agriculture Rural Development | Provides funding for clean and reliable drinking water systems, sanitary sewage disposal, sanitary solid waste disposal, and storm water drainage to households and businesses in eligible rural areas. | Counties | USDA RD Water and Waste Disposal Loan and Grant Program |
| Rural Electrification and Telecommunications Loans Program | U.S. Department of Agriculture Rural Development | Provides financing for the construction, maintenance, improvement and expansion of telephone service and broadband in rural areas. | Counties | USDA RD Telecommunications Program |
| Distance Learning and Telemedicine and Broadband Program | U.S. Department of Agriculture Rural Development | Helps rural communities use the unique capabilities of telecommunications to connect to each other and to the world, overcoming the effects of remoteness and low population density. | Counties | USDA RD Distance Learning and Telemedicine Program |

| | | | | |
|---|---|---|---|---|
| Distance Learning and Telemedicine and Broadband Program; Delta Health Care Services Grant | U.S. Department of Agriculture Rural Development | Provides financial assistance to address the continued unmet health needs in the Delta Region. | Counties | USDA RD Delta Health Care Services Grant |
| Rural Housing Insurance Fund Program; Direct Single Family Housing Loans | U.S. Department of Agriculture Rural Development | Assists low- and very-low-income applicants obtain decent, safe and sanitary housing in eligible rural areas by providing payment assistance to increase an applicant's repayment ability. | Counties | USDA RD Single Family Housing Direct Home Loans |
| Rural Housing Insurance Fund Program; Single Family Housing Repair Loans | U.S. Department of Agriculture Rural Development | Provides loans to very-low-income homeowners to repair, improve or modernize their homes or grants to elderly very-low-income homeowners to remove health and safety hazards. | Counties | USDA RD Single Family Housing Repair Loans |
| Rural Housing Insurance Fund Program; Rural Housing Site Loans | U.S. Department of Agriculture Rural Development | Provides loans to acquire and develop sites for low- or moderate-income families, with no restriction as to the method of construction | Counties | USDA RD Rural Housing Site Loans |
| Rural Housing Insurance Fund Program; Self-Help Housing Land Development Loans | U.S. Department of Agriculture Rural Development | Provides loans are to acquire and develop sites only for housing to be constructed by the Self-Help method. | Counties | USDA RD Self Help Housing Land Development Loans |
| Rural Housing Insurance Fund Program; Mutual Self Help Housing Grants | U.S. Department of Agriculture Rural Development | Provides grants to qualified organizations to help them carry out local self-help housing construction projects. | Counties | USDA RD Mutual Self Help Housing Grants |
| Rural Housing Assistance Grants; Rural Housing Preservation | U.S. Department of Agriculture Rural Development | Provides grants to sponsoring organizations for the repair or rehabilitation of housing owned or occupied by low- and very-low-income rural citizens. | Counties | USDA RD Housing Preservation Grants |

Table 3. Federal programs that use poverty area measures similar to the Public Works and Economic Development Act (1965)

| Administering department/agency | PPA spatial scale | Website |
|---|---|---|
| Department of Commerce, Economic Development Administration (EDA) | Region | EDA Economic distress levels |
| Department of Transportation, Federal Aviation Administration (FAA) | Community | Economically distressed areas special rule |
| Appalachian Regional Commission | County and community | ARC distressed counties and areas |