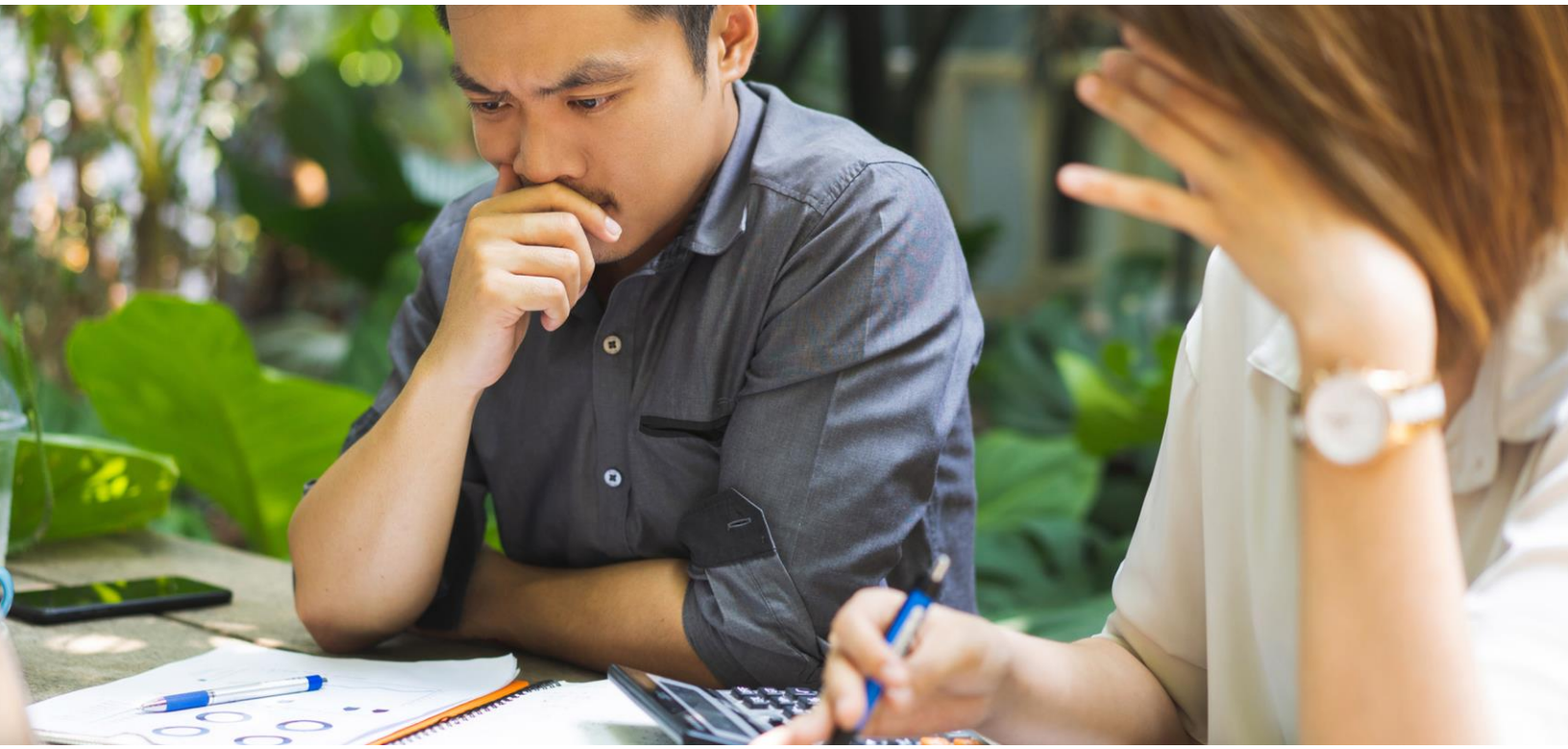

RESEARCH REPORT

In Search of "Good" Rural Data

Measuring Rural Prosperity

Corianne Payton Scally
URBAN INSTITUTE

Eric Burnstein
URBAN INSTITUTE

Matthew Gerken
URBAN INSTITUTE

with Evelyn Immonen
HOUSING ASSISTANCE COUNCIL

April 2020

URBAN INSTITUTE

Community Strategies Group
THE ASPEN INSTITUTE

HAC

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

ABOUT THE HOUSING ASSISTANCE COUNCIL

The Housing Assistance Council, founded in 1971, is a nonprofit organization that supports affordable housing efforts in rural areas of the United States. HAC provides technical housing services, financial products, policy assistance, trainings, and research and information services. HAC is an equal opportunity lender.

ABOUT THE ASPEN INSTITUTE COMMUNITY STRATEGIES GROUP

Since 1985, the Aspen Institute Community Strategies Group has helped connect, inspire, and equip local leaders as they build more prosperous regions and advance those living on the economic margins. More than 75 percent of CSG's work in those years has focused on and in rural America. CSG has worked with rural doers from nearly every state, both developing strategy on the ground and convening them to learn from one another. In turn, CSG has learned from the people doing the best work of building and rebuilding strong, inclusive rural communities and economies.

Contents

Acknowledgments

This report was funded by the Housing Assistance Council, with support from the Ford Foundation and created in partnership with the Aspen Institute Community Strategies Group. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

Many thanks go to David Lipsetz, Evelyn Immonen, and Lance George at the Housing Assistance Council for guidance and resources to complete this work. We also thank Katharine Ferguson and Clifford Deaton at the Aspen Institute Community Strategies Group for their thoughtful input at every stage. Immonen and Deaton completed a series of practitioner interviews for this report. We would also like to thank our colleague Benny Docter for his assistance with the researcher interviews. Graham MacDonald on our Technology and Data Science team was instrumental in understanding key datasets and helping us get access to some proprietary datasets for analysis. Kathy Pettit also provided key insights and direction.

We would like to thank the following rural researchers and practitioners we interviewed for sharing their wisdom, experience, and frustrations with accessing good rural data (titles and affiliations current as of December 2019):

- Bill Bishop, Cofounder and Contributing Editor, Daily Yonder
- Dr. John Cromartie, Geographer, US Department of Agriculture, Economic Research Service
- Dr. Matthew Fannin, Professor, Louisiana State University College of Agriculture
- Jason Gray, Senior Fellow for Research and Policy, NC Rural Center
- Travis Green, Vice President, Solutions Consultant, LOCUS Impact Investing
- Mark Haggerty, Economic Geographer, Headwaters Economics
- Eric Hangen, Senior Research Fellow, Center for Impact Finance, Carsey School of Public Policy, University of New Hampshire
- Dr. Sarah Low, Associate Professor of Regional Economics, Fred V. Heinkel Chair in Agriculture, College of Food, Agriculture, and Natural Resources, University of Missouri

- Graham MacDonald, Chief Data Scientist, Urban Institute
- Donald Macke, Senior Vice President of LOCUS Impact Investing
- Tim Marema, Editor, Daily Yonder
- Dr. Marybeth J. Mattingly, Assistant Vice President, Community Development Research and Communications, Regional and Community Outreach, Federal Reserve Bank of Boston
- Dr. David McGranahan, Senior Economist, US Department of Agriculture, Economic Research Service
- John Molinaro, President and CEO, Appalachian Partnership Inc.
- Dr. John Pender, Senior Economist, US Department of Agriculture, Economic Research Service
- Kathy Pettit, Principal Research Associate, Metropolitan Housing and Communities Policy Center, Urban Institute
- Sandy Ratliff, Assistant Vice President, Community Impact Advisor, Virginia Community Capital
- Duane Yoder, President, Garrett County Community Action

Executive Summary

Economic and demographic data drive research, policy development, distribution of government resources, and private investment decisions. But many of the datasets that policymakers, practitioners, and researchers rely on to understand and guide resources to rural communities fall short in representing rural realities. Given the increasing attention to rural areas in public policy and popular discourse, along with notable trends disadvantaging rural places—persistent poverty and global economic shifts—this search for "good" rural data is timely.

This report explores opportunities for using data to more accurately measure and understand prosperity in rural areas, particularly those aspects of prosperity that fall within built, financial, human, and political community capitals. The community capitals framework is used throughout this report to help frame and define prosperity. A joint project, conducted by the Urban Institute in collaboration with the Housing Assistance Council and the Aspen Institute Community Strategies Group, our research is based on a series of interviews with researchers and practitioners who have expertise working with rural data, as well as a scan of 22 datasets—both commonly used and more emergent, innovative sources. The report includes recommendations for data owners, policymakers, practitioners, and researchers for how to find and use the best data possible and promote equitable data collection and release practices that better capture rural realities across all geographies.

Why Better Rural Data Matters

Rural policymakers, practitioners, and researchers need better rural data to strengthen practice in rural communities, improve policy that affects rural areas, and help change the high-level narrative that frames our discussions and understandings of rural places. Better data will

- enable more accurate, useful, and impactful research on and in rural communities;
- support local practitioners in making more informed decisions about how to adopt strategies and evolve programs to increase rural prosperity;
- improve local governments' ability to make decisions and set policy—including operations and long-term planning;
- guide private investments to expand access to credit for individuals and businesses; and

- counter the national narrative on rural America, which tends to homogenize rural places and people and promote stereotypes.

How Data Fall Short for Rural Communities

Most available data for measuring prosperity do not accurately capture rural realities and have well-known quality challenges.

The indicators of prosperity in rural places are distinct and require different measures from urban ones. We heard how measures around entrepreneurship, agricultural employment, retirement communities, social capital, government capacity, and volunteerism, for example, are particularly crucial for understanding rural places, yet they are often poorly covered by existing public data sources.

Differing definitions of "rural" places also complicate analysis. In some cases, "rural" is defined as simply residual, or what's left over, once urban or metropolitan areas are defined. This can result in more prosperous rural communities—those surrounded by urban communities and those experiencing population and economic growth on the urban fringes—being categorized as part of metropolitan areas, leaving behind only the most distant and struggling communities to count as rural. Other definitions consider how rural places relate to other places in ways such as density, travel time, concentrated economic activity, commuting patterns, and character. Although more granular and often available at smaller geographic levels, most definitions still lack the precision needed for local analysis and decisionmaking. The use of multiple, competing definitions makes painting a clear picture of rural places difficult and complicates comparisons.

Small populations make rural data collection and reporting challenging. The American Community Survey (ACS) is commonly used as the only available source for comprehensive data across geographies on topics such as demographics, housing, and income. Yet, when the ACS is based on a sample of an already small-population geography, multiple years of data are needed to get enough responses in many rural communities, affecting the timeliness of data, and the numbers reported may have high margins of error. Data collection is challenging because of poor communications infrastructure, including broadband. In other datasets, long reporting periods and small sample sizes make numbers appear "sticky," meaning that a measure will appear to not change for a long period of time and then jump or drop dramatically when in reality the change was more gradual. Finding good data covering tribal lands is even more difficult: definitions of tribal lands vary, land boundaries are different from dataset geographies, and data sources are even fewer.

Finding the "Good" in Existing Datasets

To gain insights into the challenges, strengths, and emerging opportunities for using data to better understand rural communities, we interviewed rural data experts and collected information on the datasets they recommended. We scanned 22 datasets spanning a range of prosperity measures across the four community capitals covered in this report—built, financial, human, and political—including business markets, consumer finances, education and social mobility, employment and earnings, government capacity, health, real estate, and transportation. The datasets and community capitals we selected were drawn from expert recommendations as well as the Urban Institute's expertise. The datasets were summarized based on the smallest level of geography covered, data collection and reporting characteristics and practices, and data-quality issues. We compared datasets within similar categories on relative strengths and weaknesses. We did a deeper dive on five of them to test for missing or incomplete data on rural areas. Key findings from the datasets we scanned include the following:

- **Many rich, public datasets on employment and financial well-being do not work for rural places** because they are available only at the county level or larger geographies.
- **Data for small populations are helpful, but privacy is a concern.** There is a range of data available at smaller geographies, such as zip codes, census tracts, and census blocks. To ensure that data cannot be traced back to a specific person or organization, data may be suppressed or not reported if response numbers are too small, or random "noise" may be added that changes the values slightly. These privacy measures mean that data for some rural places may be missing entirely or that some data were changed in a manner that may obscure their true values.
- **Out of all 22 datasets scanned, data on education and social mobility perform best for rural communities**, based on the depth and quality of data available at small levels of geography. These data are often reported at an individual level, such as with Internal Revenue Service Statistics of Income data, or are administrative data reported by institutions, such as the number of students enrolled in schools and graduation rates. However, some data may be missing for some rural places and populations.
- **Proprietary datasets hold promise, but access is tough.** These data may include individual-level data or transaction-based data, but they may be expensive to purchase, come with restrictions on their use and publication, and may still have missing or withheld data.
- **Data aggregators are only as good as their underlying data.** Data aggregators can be a powerful tool for bringing together measures that offer a more nuanced view of a community.

However, if an aggregator pulls together data from multiple datasets to create a new one, it is limited by any data issues within the underlying datasets. Aggregating "bad" data together do not make them "good" or more reliable. Instead, errors can be compounded. For example, a number of available data aggregators rely on ACS data, which is prone to significant error for small geographies.

Moving toward Better Rural Data

Creative use of existing datasets can provide new avenues for improving the accuracy, depth, and nuance of rural-focused data analysis. Administrative data collected by public or private program managers as they administer specific programs can be a rich, untapped resource if used carefully and with attention to the validity of the underlying data. Some state and local governments have begun integrating their internal administrative databases to better align services and gain a more complete picture of populations served, although this effort remains nascent. Linking datasets—particularly administrative data with demographic datasets, such as connecting Supplemental Nutrition Assistance Program data to local demographic breakdowns—can help make new connections for policymaking and practice.

Emerging data sources offer modest potential for finding new measures of prosperity to describe rural realities. These online, social media, and app-based tools can scrape data from sources such as restaurant reviews and frequencies as a measure of economic growth or health status updates to track outbreaks of illness, but they require consistent internet access and use to be valid measures. Government agencies such as the US Department of Agriculture are exploring how to use data in new ways to better serve their mission, including engaging app developers to use their public data. Users are also producing easy-to-navigate catalogs of available data that are downloadable and ready to analyze. Individual-level proprietary data also hold promise for greater accuracy in small geographic areas. Finally, new surveys, scales, and typologies are emerging to provide different rural data than are currently available (e.g., new surveys and interviews asking new questions) and to categorize data in new ways that highlight similarities and differences across rural places more distinctly and consistently.

Next Steps on the Path to "Good" Rural Data

Rural communities deserve data that represent their on-the-ground realities. This will help change the sweeping narrative of rural decline, illuminate nuanced opportunities and challenges to strengthen rural

practices, and allow for more data-driven policymaking that advances rural prosperity. More research is needed to explore the concept of rural prosperity and prompt data owners and users to make progress on generating, using, analyzing, and interpreting better rural data.

There is more work to do to conceptualize the idea of prosperity and find data to measure those aspects not explored in this report. Although the community capitals framework helps identify community resources, it does not identify specific measures of prosperity that should be tracked. The development of better measures around work readiness, individual health, individual ownership of land and other assets, and seasonal employment and earnings, for example, could provide a clearer picture of rural realities. Additionally, expanded community measures can capture additional capacities important for advancing community well-being, including institutional capacity, economic diversity, and anchor institutions.

We suggest a series of next steps for those who collect and own data and those who use it to set policies, work in rural communities, and conduct research to build evidence for informing rural policies and practice.

Data owners—both public and private—can consider the following:

- Increase the rural sample size in survey research by asking more people to complete a survey or ensuring that surveys can be completed in multiple formats and not just online.
- Reexamine practices around data suppression and noise to identify possible improvements in the publicly collected data made available to rural places to help small-population communities track progress and plan for the future.
- Create partnerships between owners of proprietary data and governments and researchers to increase access to rural-specific data for policymaking and research.
- Try new strategies for increasing rural participation in important surveys like the ACS, particularly on tribal lands and other hard-to-count places, including providing safe, secure internet in community spaces; providing on-site translators; and mobilizing community leaders as survey advocates (Brumfield 2018).[1]

Data users can ensure that they are assisting with high-quality rural data collection; partnering on rigorous, influential analyses; and leveraging analyses of "good" data to make informed decisions for rural policy and practice. They can also all work to change the narrative of rural communities as new, better rural data and analyses become available. Those data users engaged in policymaking should

- ensure that they are accessing the best data on prosperity possible, using this report as a starting point;
- seek partnerships with administrative and proprietary data owners to access the best community-level data possible;
- support improvements to rural data collection and dissemination, such as expanding rural samples, providing reliable internet locations, and engaging community volunteers; and
- partner with researchers who understand the nuances of existing rural data, the possibilities for accessing emergent data, and the proper analytic techniques to provide the strongest evidence needed for informed policymaking.

Rural practitioners who use data for making local decisions can

- advocate to data owners for better data on the communities they serve;
- encourage rural residents to participate in important data collection activities to improve the quality and coverage of data for rural communities;
- use this report to understand the opportunities and challenges with commonly used datasets and make sure that they are using the best accessible data;
- partner with rural researchers to help navigate data with which they are unfamiliar and analyze it appropriately, including academic, nonprofit, and public research institutions that are natural partners for rural data work, including many interviewed for and referenced by this report;
- explore ways to work with local and state agencies that may have administrative data they can provide after ensuring that they are protecting people's privacy; and
- collect their own data, if necessary, to track measures of prosperity that are important for their work.

Researchers focused on rural places also need to

- think carefully about the rural definitions and datasets they use to ensure that they are reflecting the most accurate rural realities possible;

- conduct policy-relevant research, guided by rural practitioners and the real needs of rural communities for better data and analyses on prosperity to inform policymaking and practice;
- partner with rural practitioners and rural communities to provide the analytic talent needed to make sense of rural data that will in turn provide better-informed policy decisions;
- work to include rural areas in their research, if they do not regularly, and use high-quality datasets with the most granular geographic coverage possible to do so;
- seek innovative ways of collecting, linking, and analyzing data for rural places, including leveraging administrative data and tapping into new data sources; and
- exercise caution when doing rural research to ensure that they are doing research "with" communities instead of "on" them, especially where there is a history of misuse and well-placed mistrust.

In Search of "Good" Rural Data

Economic and demographic data are crucial for understanding the realities of rural America, and they frequently drive investment in rural communities. But many of the datasets that researchers, the media, policymakers, and others rely on pose specific challenges for rural communities, where small populations across sometimes large geographic areas make high-quality data collection and reporting challenging. Sampling in these contexts is more prone to error, and individual privacy is more at risk. Inadequate or misleading data can distort policies, discourage private investment, and limit the information rural practitioners and local officials need to make well-informed decisions.

Rural America is considerably diverse. Rural areas are in all 50 states, covering vast plains, mountainsides, coastline, reservations, forests, and farmlands. Although the term "rural" often evokes images of agricultural communities in the Midwest and deep South, it is equally applicable across industries and geographies, from New England and the mid-Atlantic coast to the Pacific Northwest and the Sun Belt. Data show that rural America is home to an increasingly ethnically and culturally diverse population (Lichter 2012). Yet researchers and policymakers have frequently painted these areas with the same broad brush: poor, uneducated, white communities experiencing economic failure, population loss, addiction, and general hopelessness. Although poverty, economic struggles, and opioid misuse are very real in many communities—rural, suburban, and urban—they tell an incomplete and misleading story of rural America, which can influence important decisions around government policies and funding, as well as private investment in rural places.

This report explores challenges and opportunities for using data to more accurately measure and understand prosperity in rural areas. Our understanding of prosperity is loosely guided by the emerging asset-based, "wealth-creation" approach to community economic development, championed by Kretzmann and McKnight (1993), as well as Flora and Flora's community capitals framework (Beaulieu 2014). This approach focuses on generating and retaining a range of capitals within the community, reinvesting that wealth for future productivity, and improving the quality of life for community residents, rather than viewing only growth and jobs as the primary measures of success. The community capitals framework identifies seven types of community-level "resources that can be invested or tapped for the purpose of promoting the long-term well-being of communities" (Beaulieu 2014). These span the natural and built environments, human capital and financial capital, and measures of a community's cultural, social, and political capital.

Although we intentionally did not predefine prosperity for this exploratory study, most of the datasets recommended to us by rural experts and practitioners generally fell within built, financial, human, and political capital, defined in box 1 below. Because we did not exclusively ask for community-wide measures of prosperity, we also received suggestions for measuring prosperity for people living in rural places, particularly in the categories of human and financial capital. While natural capital was also suggested during a couple of interviews, it was beyond the scope of this project.

BOX 1

Defining Community Capitals Discussed in This Report

- Built capital: human-made infrastructure that supports society, including roads and utilities, housing, and commercial and health care facilities
- Financial capital: access to financial resources and institutions necessary for economic mobility and growth
- Human capital: how educated, skilled, and healthy the workforce is, as well as access to additional education and training resources
- Political capital: capacity to influence distribution of resources within the community

Source: Adapted from Beaulieu 2014.

This explorative study was conducted by researchers from the Urban Institute in collaboration with the Housing Assistance Council and the Aspen Institute Community Strategies Group. The work included a series of interviews with researchers and practitioners who have expertise working in or with rural communities. Based on these, we collected and reviewed existing datasets for assessing various economic and social aspects of prosperity in rural communities—both commonly used sources and newer, emerging ones. We also documented promising data collection and integration practices that may surface new data on rural realities in the future.

Over the next pages, we discuss the following topics:

- why better rural data matters for strengthening practice, improving policy, and changing the narrative around rural communities
- the challenges of capturing accurate and consistent data for rural communities
- the findings from our scan of 22 datasets for understanding rural prosperity

- new and emerging data sources, methods, and innovations
- next steps in conceptualizing rural prosperity and working together with data owners and users to improve and mobilize data for improved policies, practices, and narratives about rural America

Why Better Rural Data Matters

Policymakers, researchers, businesses, thought leaders, and rural practitioners depend on data to shape markets, influence behavior, and make crucial decisions. The more available and more accurate data are, the more likely decisions will lead to expected outcomes. Rural places face an uphill climb, however, in accessing "good" data that is collected consistently and is of high enough quality for granular analysis to guide these decisions.

Better rural data means improving outcomes. At the local level, better data could support local decisionmaking and policy development, including immediate operations and long-term planning. On the state and national levels, it would allow government agencies to more accurately distribute funding and other supports to the communities where it would be most effective. Private institutions and investors could also have better data on which to base their decisions, be that expanding a business or providing credit in an underserved area. Finally, in the area of public discourse, improved data would provide researchers and public commentators with a more accurate and nuanced view of the diversity of places, people, and experiences in rural communities across the country.

Strengthen Rural Practice

Several experts we spoke with observed that much rural practice in government, nonprofit, or other service delivery is based on experience and instinct, without the benefit of much data. Although these rural practitioners have extensive local knowledge, they have limited resources for data collection and analysis. Urban governments frequently have people on staff with expertise in data and analytics, but this is uncommon in small, rural governments. Big cities also have more resources to fund data-driven initiatives because they have larger tax bases and, in many areas, greater legal authority to tax than most rural communities do. They also benefit from higher-accuracy economic and demographic data from public sources because they have larger and denser populations.

For most rural places, I would say by and large they have ground truth and intuition but are flying blind. They are not using data to ground what they do. There are lots of communities who are not using data. The last time they did a study was 15 years ago.
—Rural practitioner

A lack of accurate, longitudinal, and current data both restricts the tools for planning and limits a community's vision for its future. As one expert we spoke with noted, decrying the lack of reliable local data, "you are what you measure; let's measure what we want to become." Increased access to good data would provide policymakers with a clear baseline of their community capitals, as well as areas of need, and allow them to set identifiable and quantifiable goals. In addition, good data would allow local leaders to leverage existing community capitals to market their communities to potential investors, target funds and supports toward businesses in well-suited sectors, and invest in addressing underdeveloped resources.

Improve Rural Policy

Better rural data could help policymakers make better decisions and develop public and private-sector policies that are more responsive to the needs of rural places. Rural policymakers rely on existing demographic and economic information to make decisions that influence rural prosperity through the distribution of public funds. More than 300 federal programs use US Census Bureau data, including the American Community Survey (ACS), to appropriate and distribute around $900 billion in funds to state and local governments. About 60 programs are explicitly rural focused and distribute around $30 billion (Brumfield 2018). These funds are distributed as the following:

- grants, such as the US Department of Transportation's Formula Grants for Rural Areas
- direct or guaranteed loans, such as US Department of Agriculture (USDA) housing, business, and community facility loans
- direct payments to individuals or institutions, such as USDA Rural Rental Assistance payments (US Census Bureau 2017)

However, there can be significant challenges to using ACS data to accurately describe rural realities, including small sample sizes prone to error and undercounting in rural communities (box 2).

Yet these data on population, housing, income, and other community characteristics are used to determine eligibility criteria, allocation formulas, and even interest rates on loans (Brumfield 2018).

BOX 2

Challenges with Using the American Community Survey to Describe Rural Communities

The American Community Survey (ACS) is a key source of free, publicly available demographic and economic data for US researchers and policymakers. The Census Bureau regularly administers the ACS to an evolving sample of the country's population. The bureau publishes survey results in two forms. The one-year datasets provide "snapshot" data from the sample responses collected in a single year to provide data on a regular, timely basis. The five-year datasets combine the survey responses gathered over five years, which means a higher count of responses from each place. It also means that the five-year estimates are generally more accurate than the one-year estimates.

Although the ACS provides reasonably reliable data for large populations via both types of estimates, its accuracy suffers for smaller-population areas, especially for subpopulations (e.g., racial or ethnic groups). Because of these issues, the Census Bureau publishes one-year estimates only for geographic areas with populations of 65,000 or more. Even for larger-population areas, many one-year estimates are suppressed because of small sample sizes. Five-year estimates are available for more detailed geographic levels, down to the block group (which usually contains 600 to 3,000 people). However, margins of error for smaller geographic units remain high.

In addition to the accuracy challenges inherent to the ACS's sampling structure, rural communities are often undersampled because of high costs and logistical challenges in administering the survey. The cost of in-person collection becomes expensive and arduous as travel times between individual homes and communities that data collectors must visit in the sampled regions increase. The Census Bureau depends on mail, phone, and the internet to gather information from target populations for the ACS. (The 2020 Decennial Census will be the first to be primarily conducted over the internet.) However, according to the Federal Communications Commission's *2018 Broadband Deployment* report, more than 10 percent of residents in rural counties lack access to either fixed broadband (25 Mbps download/3 Mbps upload service) or high-speed mobile internet (5 Mbps download/1 Mbps upload service). These numbers do not take into account places where the internet is slow, unreliable, or available but unaffordable.

Sources: Amanda Gold and Yipeng Su, "Rural Communities Aren't Immune from a Census Undercount. Here's How They Can Prepare for 2020," *Urban Wire* (blog), Urban Institute, October 31, 2019; US Federal Communications Commission, *2018 Broadband Deployment Report* (Washington, DC: FCC, 2018); Kathleen Miller, "The American Community Survey and Rural Data Analysis" (Iowa City, IA: Rural Policy Research Institute, 2012); US Census Bureau, *Understanding and Using American Community Survey Data: What Users of Data for Rural Areas Need to Know* (Washington, DC: Census Bureau, 2019).

Through our conversations with rural experts, we heard how private institutions such as banks also use public data to determine their investment policies: where to invest in businesses, open bank branches, provide home mortgage loans, and make broader community investments through grants or loans. Because of inadequate data for smaller-population areas and a lack of alternative measures of prosperity and capacity, many banking institutions screen them out as unsuitable for investment because they fail to meet thresholds for population size and industry-specific data needed to ascertain risk. This can lead to business loans being declined when the data mischaracterize the viability of the community and local economy within which the business would be located.

Numbers available for industry, the algorithm models used by banks, they rely on datasets from the Bureau of Labor Statistics and other data sources for markets—and conclude that you can't have a viable shop because of the population, so the bank denies credit.
—Rural practitioner

Banks and other lending institutions could improve the quality of their lending practices by moving away from algorithm-based lending that disqualifies rural investments based on inadequate, incomplete data rather than comprehensive human review, but this is unlikely to change. More often than not, it is community development financial institutions that step in to fill this gap by providing access to credit in rural communities and elsewhere based on personal community knowledge and relationships. However, community development financial institutions are not evenly distributed, and many rural communities still depend on mainstream banking for investment loans (Theodos and Hangen 2017). Improving rural data quality is one component of prompting large banks to more accurately assess loan applications and investment opportunities in rural areas.

Change the Narrative

Researchers and other thought leaders looking at trends across the country use available national demographic and economic datasets to build narratives around the state of rural America, often looking at measures of population, poverty, and industry growth and decline. These analyses, and the stories told with them, can draw national attention and influence how stakeholders and the public view rural communities. They can also frame how we discuss rural policies and decisionmaking. Ensuring that

these influencers are accessing accurate data and using these data appropriately can help expand the national narrative on rural America in a way that counters tendencies to homogenize rural places and people and promote stereotypes unsupported by data.

Recent media such as Paul Krugman's *New York Times* piece "Getting Real About Rural America: Nobody Knows How to Reverse the Heartland's Decline"[2] and *CityLab*'s "Most of America's Rural Areas Are Doomed to Decline"[3] by David Swenson reinforce the concept of rural decline and despair with no acknowledgment of the diversity and value of rural economies and cultures. Krugman and Swenson both used lengthy time frames and single indicators to make broad generalizations about rural decline, ignoring the differences between rural communities and the diversity of experiences of people living in rural areas.

My problem is more with perception, sympathy, and understanding [of rural communities] than with anything having to do with data. And the data gets weirdly tied in….[One media source] identified anywhere with a population under 25,000 as rural, so...[suburbs are] counted as rural, so there's a lot of bungling on the data. There's a desire to paint rural as something other than what we are.
—Rural practitioner

More accurate data can help challenge and change the misleading narrative of rural decline. Having high-quality data available for geographies smaller than counties would allow users a higher-resolution view of the diversity of rural communities and contexts and bring differences between and within regions into greater relief. Likewise, more nuanced approaches to differentiating rural communities, especially those on the edge of or inside metropolitan areas, can capture rural communities that are often excluded from analysis. Ultimately, having better data can inspire new analyses and result in clearer, more comprehensive stories on rural America.

How Data Fall Short for Rural Communities

Datasets commonly used to measure prosperity do not accurately capture rural realities. This inadequacy has multiple causes. First, datasets designed to measure prosperity may miss measures that

are more important to rural areas than to their suburban and urban counterparts. Second, varying definitions of "rural" complicate analysis. Third, data collection challenges are exacerbated in areas with low population density. Fourth, data describing prosperity on Native lands are even more problematic when it comes to accuracy, completeness, and aligning with appropriate geographies.

Rural Prosperity Looks Different, Requiring Unique Measures

Economies, natural amenities, social structures, and government infrastructure and capacity can differ greatly in rural areas compared with other parts of the country. This means that different types of measures of community capitals and local capacity than are commonly available may be needed to accurately capture rural realities. Rural data experts we interviewed highlighted the following differences:

- **Entrepreneurship.** Commonly used measures for financial and human capital—employment rates, numbers of jobs, educational attainment, and others—may obscure important aspects of rural economies, especially those that are not based on services or manufacturing. By some counts, nonmetropolitan areas outpace their metropolitan counterparts in self-employment, used as a measure of entrepreneurship.[4] Rural entrepreneurship is of increasing interest to policymakers, practitioners, and researchers, with government and philanthropic programs such as the Appalachian Regional Commission, the W.K. Kellogg Foundation, the Ewing Marion Kauffman Foundation, and the Small Business Administration incentivizing and supporting entrepreneurship and small business development as a path toward economic development (Goetz et al. 2010). However, entrepreneurship trends are not visible in all data sources, so researchers must often rely on indirect measures (such as small-business creation or self-employment). Existing measures also cannot differentiate between those who become entrepreneurs by choice and those who do so by necessity because of a lack of other employment options—a crucial detail needed to correctly assess local economic conditions and set policies for the future (Goetz et al. 2010).
- **Agricultural employment.** Agricultural economies are also at risk of inaccurate representation in standard data sources. These rural communities may show relatively low levels of salaried employment, and wage rates and even investments may look depressed compared with other regions. However, much local wealth may be held in land or agricultural products. Likewise, employment measures may or may not capture seasonal workers, which may skew data for

areas with large leisure or recreation-based economies and some types of manufacturing as well as agriculture.

- **Retirement communities.** The prevalence of high-income retirees in rural communities that have desirable natural amenities (e.g., mountains, lakes, or mild weather) and low costs of living may distort economic measures by making data points such as average household income unreliable measures of a region's wealth. Experts we spoke with recommended using indicators of how cost-burdened people are in rural communities such as housing affordability, measures that show educational outcomes for youth rather than the community as a whole, and measures of financial and other vulnerability such as health insurance rates.

For rural prosperity…community outcomes are a little different. These are things like school readiness scores community-wide or housing affordability, such as how many households are paying more than 30 percent [of their incomes in housing costs], and the rate of health insurance [coverage].
—Rural practitioner

- **Social capital.** Where there are fewer people and institutions, informal networks and associations can play more prominent roles. However, measures commonly used for understanding how to support and expand economic and social gains may not capture this social capital. Conventional indicators of capacity for economic development (such as the number and density of employment services, public facilities, business support networks such as chambers of commerce, business improvement districts, business incubators, and others) may be lower for rural communities. But these data points do not tell the whole story of what is happening in rural places.
- **Government capacity.** The political capital of rural governments can be low. Some communities may share service agreements among governments and social or economic development services. Arrangements such as these are essential to how a region works but complicate measuring individual community capacity. Alternatively, measuring local capacity may underestimate the capacity gained from regional supports.

- **Volunteerism.** In rural communities, a small number of people may play many roles. Volunteerism is often used as a measure of social capital, but in rural areas, high rates of participation in civic organizations or religious communities may more accurately be a measure of potential burnout.

Things like participation in nonprofits can overestimate the resilience in rural communities. [Data show] high per-capita nonprofit participation, but we know there's lots of burnout. It can be a liability to have the same people doing all the different roles in civil society. The purpose of the data is sometimes misused here to overstate rural resilience.
—Rural data expert

Inconsistent Definitions of "Rural" Complicate Interpretations

No single, official definition or designation exists to help determine whether a place is rural. The differences between the extremes of major cities and sparsely populated, remote regions may be clear. But how to classify the areas in between may not be, and boundaries are difficult to set. Our expert sources said that determining whether a place is rural has become increasingly difficult as suburban communities and "commuting sheds" have spread into areas that were previously undeveloped. How we define rural can affect not only how many people are included, but also what sorts of challenges rural people and places face.

Because determining what counts as rural can be done in several ways, basic measures such as the number of people who live in rural communities in the United States can vary greatly. For example, the Census Bureau estimated in 2017 that about 60 million Americans lived in rural areas[5] (using a definition based on population density), whereas a 2018 report by Kenneth Johnson at the University of New Hampshire's Carsey School of Public Policy estimated just 46 million (counting those living outside of counties included in metropolitan areas). Johnson's estimate was similar to the 46 million estimate that USDA's Economic Research Service (ERS) published the same year (Johnson 2018; USDA 2018).

RURAL AS RESIDUAL

Experts we spoke with observed that rural areas are frequently defined as whatever land remains after urban areas have been defined. For example, a metropolitan statistical area (MSA) is a geographic

designation created by the US Office of Management and Budget and used by the Census Bureau to identify urbanized areas at the county level, as used by Johnson and ERS above. To calculate whether a county should be included in a core city's MSA, the Census Bureau assesses the percentage of the county's population that commutes to the core city, as well as other characteristics such as whether the county's population density is above a minimum threshold.[6] As a result, many researchers use all areas not in an MSA as a proxy for rural communities.

Rural community data is in many contexts a residual, what's left. This is how [the OMB] defines metropolitan statistical areas. They start with a city center. Then they will start to add counties in based on commuting patterns. It's all a very urban-centric method of defining rurality. Hospitals can be made or killed on the basis of that fairly arbitrary process.
—Rural data expert

Defining rural in contrast to urban in this fashion creates two substantial problems for researchers and policymakers. The first problem is that MSAs include many areas that most residents and researchers would identify as rural based on their sparse settlement patterns or economic structures. Additionally, many counties, especially those on the outer edge of urban areas, are not evenly developed. Areas closer to the core city may be developed and depend more on the urban core, while outer areas remain sparse. In other cases, urban or suburban development may follow a highway or major road, leaving areas away from the main corridor primarily rural. Experts we spoke with observed that rural areas inside MSAs are often the most prosperous and populated rural areas. This means excluding rural areas inside MSAs from analyses may artificially depress economic measures for rural areas nationwide.

The second problem is that when rural communities are economically successful, they often attract investment and workers looking for employment opportunities, leading to increases in building development and population. Some communities grow enough to join an MSA. As they do, they are no longer included in rural analyses. From 2003 to 2013, the number of counties not included in an MSA declined by 81, from 2,066 to 1,985.[7] By passively excluding counties that have grown, longitudinal analyses of rural areas further bias trends toward economic and demographic decline.

RURAL AS RELATIONAL

To add more nuance to how "rural" is defined, researchers and policy analysts have developed several alternative methods. ERS researchers have developed two classification schemes, the rural-urban commuting area (RUCA) codes and the frontier and remote area (FAR) codes.[8] The RUCA codes use census tracts instead of counties to categorize areas based on population density, urbanization, and daily commuting. Rather than creating a single category for urban and another for rural, RUCA codes provide a range of categories that allow the user to determine which categories to include in a more nuanced way. Likewise, FAR codes use zip code level geography to provide gradations for "sparsely-settled, remote areas of the U.S." based on their distance to the closest urban area.

Our data scan uncovered several other publicly available tools for defining "rural" and "urban" produced by the Consumer Financial Protection Bureau, ERS, the Housing Assistance Council, and USDA Rural Development. Table 1 provides a summary of each, including a description of intended use and the smallest geographic level available.

TABLE 1

Rural Definition Tools

Tool	Description	Source	Smallest geographic level
2019 rural and under-served counties and website tool	This list helps creditors determine whether a property is in a rural or underserved area for purposes of applying certain regulatory provisions related to mortgage loans. Rural is defined for the purposes of this list as in a county that is outside of an MSA or within a census block defined as rural based on the Census Bureau's density-based definition.	Consumer Financial Protection Bureau	County
Frontier and remote area codes	Created by ERS based on travel time by car and highway to cities of various sizes, used to classify frontier areas as an alternative to using commuting data.	USDA Economic Research Service (ERS)	Zip code
Rural-urban commuting area codes	An ERS product since 2000 in collaboration with the Department of Health and Human Services based on population density, urbanization, and daily commuting.	ERS	Census tract
Typologies of rural	The codes classify all counties according to six mutually exclusive categories of economic dependence (farming-dependent, mining-dependent, manufacturing-dependent, federal/state government-dependent, recreation, and nonspecialized) and six overlapping categories of policy-relevant themes (low education, low employment, persistent poverty, persistent child poverty, population loss, and retirement destination).	ERS	County
Urban influence codes	Distinguishes metropolitan counties by population size of their metro area and nonmetropolitan counties by size of the largest city or town and proximity to metro and micropolitan areas. This scheme allows researchers to break county data into finer residential groups, beyond metro and nonmetro, particularly for the analysis of trends in nonmetro areas related to population density and metro influence.	ERS	County
Rural and small town tract designation	Created to provide a subcounty designation of small-town areas incorporating housing density, tract-level commuting, and a rural character measure. Categorizes tracts into rural, small town, exurban, outer suburban, inner suburban, and urban.	Housing Assistance Council	Census tract
Areas rural in character	A category created by the 2008 farm bill and used by USDA in discussion of rural development. Areas rural in character are "pockets of rurality [that] still exist within a municipality that otherwise would be excluded because of being adjacent and contiguous to a city or town greater than 50,000." The ultimate designation of such an area is determined by the undersecretary for rural development but is typically 40 miles or less across, with a population of less than 10,000.	USDA Rural Develop-ment	Custom

Rural Data Collection Is Difficult, and Release Cycles Are Tricky

Several aspects of rural communities can create challenges for collecting and publishing accurate data, including the geographic size of the community, the population size and distribution of population within the community, whether the community aligns with politically or statistically defined boundaries, and the level of geographic isolation. Release cycles can also lag substantially behind the data collection, and when released, data can appear to jump dramatically because of a small underlying change. Experts spoke in detail on the following problems:

- The small number of responses that low-population communities typically have to surveys can result in high margins of error. Many demographic and economic datasets, including the ACS, report data collected by surveying a sample of respondents. In high-population areas, the large number of responses means these data are frequently reflective of the reality on the ground. However, in low-population areas, these data reflect fewer responses, making their true value more difficult to measure because of high margins of error. This means that even though the data give a single-point estimate for each survey answer based on the responses, the range within which the true value falls can be so large that the point estimate is effectively meaningless. It can also make comparison across geographies difficult. A detailed example of this is in box 3.
- Rural areas with populations that are not uniformly distributed can skew data results. Sparsely populated counties with a single population center, for example, might appear as uniformly rural in data, but in reality, people living in the core town may have a different economic reality than those living in the countryside.
- Isolated areas with less-developed communications infrastructure may suffer from undercounting in data collection activities. This infrastructure includes mail delivery services, in-person access for survey and interview purposes, and access to broadband or mobile internet for online data collection.
- Delays in public data releases can lead to gaps between when data were collected and when analysis is done, as an interviewee mentioned was the case for Internal Revenue Service (IRS) data.
- Public data can also be both "sticky" and volatile because of reporting periods and small sample sizes, meaning that a measure will appear to not change for a long time and then jump or drop dramatically when in reality the change was more gradual. For instance, in areas with a small number of employers, the closing of a business can appear devastating to a community. Also,

when a lot of time passes between the reporting of data, employment numbers may appear volatile when in fact growth has actually been steady.

BOX 3

How Margin of Error Affects Demographic Data: Comparing Los Angeles and Yuba Counties

In politically defined geographies such as counties and incorporated municipalities, the degree of accuracy of sample-based data is dependent on the size of the population in question and survey response rates. As an example, consider the share of African Americans with a four-year college degree (or higher) in Los Angeles County, the nation's largest county by population, and Yuba County, which is in California's Central Valley, according to the 2013–17 ACS five-year estimates.

	County population	% of African Americans with 4-year degree	Margin of error (% pts)	Actual range (%)
Los Angeles County	10,105,722	24.9	0.5	24.5–25.4
Yuba County	74,644	28.5	9.2	19.3–37.7

The point estimates (column 3) show that the achievement rate for a four-year college degree or higher in Yuba County is more than 3.5 percentage points higher than it is in Los Angeles County. However, because Los Angeles County is much larger than Yuba County, the margin of error for Los Angeles County is smaller, which means that the estimate is more reliable. The higher margin of error for Yuba County (+/-9.2 percentage points) means the true proportion could be well above or below the estimate for Los Angeles County. This lack of accuracy is due both to Yuba County's smaller population and its smaller number and proportion of African Americans (2,484 African Americans with a margin of error of 255, making up 3.3 percent of the county population, with a margin of error of 0.3 percent). Because of this, we cannot say which county has a higher proportion of African Americans with college degrees.

Native American Lands Face Additional Challenges

Assembling an accurate picture of Native American reservations via data is even more difficult than doing so for rural areas overall because of the collection costs, geographic challenges, history of betrayal by the American government, and distrust of government efforts. Existing data typically capture only fragmented characteristics of the diverse areas present in what is commonly referred to as Indian Country.

Several factors make collecting and publishing reliable data about tribal lands challenging:

- **High levels of remoteness and diversity across communities.** Collecting data on Native American lands can take longer than in other areas because of high levels of geographic remoteness and diversity across communities. As of 2019, there were 326 separate designated tribal lands belonging to 577 federally recognized tribes spread across rural, suburban, and urban areas.[9] This variety in character makes generalized trends within these communities difficult or impossible to define.
- **Historical distrust.** Hundreds of years of discrimination and betrayals by the US government have led many Native American communities to distrust federal initiatives, including survey efforts. That can result in low participation rates among residents of reservations (Schmidt 2011). Additionally, collecting household-level data can be difficult because residents of reservations with economic challenges move frequently to follow employment opportunities on and off reservations and because of crowding of households, as relatives live under one roof on an informal or temporary basis (Housing Assistance Council 2010). As a result, many tribal governments conduct their own surveys when other datasets are inadequate (National Congress of American Indians 2018).
- **Multiple definitions of tribal lands.** Federal reservations are the most clearly identified geography for Native American lands, but several other designations exist, including Oklahoma statistical areas, Alaska Native villages, Hawaiian homelands, and some off-reservation trust lands. Some data incorporate all of these into their definitions, while others include only some.
- **Available data do not match boundaries.** Many datasets do not include tribal geographies. Because reservations and other tribal lands do not typically match statistical or political boundaries such as census tracts, zip codes, or counties, researchers must frequently compile data using the closest available geographic unit available and then weight data based on the proportion of the geography that falls within tribal lands. For researchers seeking to aggregate across Native American lands, the diversity of territories in size, population, and economic characteristics creates challenges. For those looking at individual communities, small sample sizes mean that data must be aggregated over several years and "blurred" to preserve confidentiality, usually with high margins of error and high variability.
- **"Checkerboarding."** The mixed and complicated legal framework for land and property ownership creates further challenges for determining the geography of tribal lands. Land on federal reservations set aside for tribal nations or tribal individuals is held "in trust" by the federal government and often cannot change ownership outside the tribe. However, most reservations also have "fee simple" land held privately by people who historically have mostly

not been Native Americans (Housing Assistance Council 2010). Public information about where trust land ends and fee simple land begins can be difficult or impossible to find.

- **Census and ACS undercount.** Data indicate a significant undercount in the decennial census and ACS survey in tribal lands. Although the Census 2010 undercount outside of reservations was not statistically significant, the undercount of Native Americans living on reservations was statistically significant at 4.9 percent.[10] This indicates a potential for higher levels of inaccuracy in the data available for tribal lands, in addition to the issues that rural areas in general present for researchers because they have smaller populations to survey.
- **Lack of other data sources.** Few datasets aside from the ACS can provide data for Native American areas. Researchers interested in assessing which data are available for reservation land must estimate with county-level or census tract–level data weighted to the proportion of the county in tribal lands.

Better data for Native American lands would help provide a more nuanced, accurate view of Native Americans. Inconsistent definitions can warp data and lower comparability, resulting in competing or inconclusive evidence upon which to base policy and practice. An example of the difficulty of measuring Native American populations in a large, sparsely populated western county is in box 4.

BOX 4

Population Distribution and Data Accuracy: The Case of Coconino County, Arizona

The distribution of people within an area can make demographic and economic measures less reliable. This unreliability is particularly pronounced in many rural counties, especially those in the western US, many of which cover large land areas. Coconino County, Arizona, for example, is 18,661 square miles, approximately twice the size of New Hampshire. Even with a relatively small population (138,639), the county has significant demographic diversity and a large Native American population.

Although countywide statistics from 2013–17 ACS five-year estimates paint a picture of a region with a large Native American population, many of whom are living in poverty, more granular data for the county seat, the city of Flagstaff, leave a different impression. The data show a smaller share of residents who identify as Native American and a smaller share of Native American residents whose incomes are below the federal poverty level. However, the high margin of error for poverty among Native Americans in Flagstaff indicates that the rate could be as high as it is in the county overall.

	Population for whom poverty status is determined	Share living in poverty	Share identifying as American Indian	Share of American Indians living in poverty
Coconino County	126,937 (+/- 1,187)	21.0% (+/- 1.2 % pts)	31.1% (+/- 0.3 % pts)	30.0% (+/- 2.1 % pts)
Flagstaff City[a]	58,607 (+/- 1,202)	21.7% (+/- 2.3 % pts)	12.8% (+/- 1.4 % pts)	24.5% (+/- 6.7 % pts)

More accurate data could improve decisions on where to distribute funds and how to provide services to have the greatest impact on the prosperity of Native Americans across the county.

[a] Although the city of Flagstaff is an incorporated political entity, 38 percent of the Flagstaff metropolitan statistical area land is part of the Navajo Nation, which is tribal land. The data here do not include people living on tribal lands.

A Summary of Well-Known Challenging Datasets

Many go-to public datasets that provide information on demographics, housing costs and housing finance, poverty, financial well-being, and safety net program participation have well-documented challenges in accurately describing rural realities. These problems stem from how the data are collected or providing data at too large of a geography to capture regional variation and small community realities. By familiarizing ourselves with the limitations of commonly used datasets, we can identify shortcomings that could be addressed by looking to other data sources or new approaches to analysis (table 2).

TABLE 2

Datasets and Their Limitations for Research in Rural Areas

Data collection type	Dataset	Description	Limitations
Survey, annual	American Community Survey	National survey conducted annually across a sample of the population across demographics, including statistics on income, education, ancestry, disability status, language spoken, race, sex, housing type, location and quality, family composition, health insurance coverage, poverty, and other measures.	▪ Geographies with less than 65,000 people require pooled data over multiple years to get a large enough sample to be included. ▪ When reported, measures may still have margins of error so large that they are statistically meaningless.
Transaction records	Home Mortgage Disclosure Act dataset	Reported at the transaction level, this dataset covers the number, type, and location of home mortgage applications and demographic information about applicants.	▪ It covers only banking institutions subject to the Community Reinvestment Act. Small mortgage lending institutions and those with limited activity in nonmetropolitan areas are exempt from reporting. However, Home Mortgage Disclosure Act coverage of lending in the majority of rural areas is adequate to good. The number of transactions in rural communities is dwindling, leaving these areas with incomplete or nonexistent data.
Survey, panel	Panel Study of Income Dynamics	A national, longitudinal panel survey of families records where they lived, when they entered the panel, and where they moved.	▪ Rural areas in high-population or highly rural states may have an inadequate number of addresses included in rural regions to yield state-level estimates.
	Survey of Income and Program Participation	A national, longitudinal panel survey of households collects integrated information on "tax, transfer, and other government and private policies."	
Hybrid	Small-Area Income and Poverty Estimates	Annual estimates of income and poverty statistics for all school districts, counties, and states. Measures are modeled estimates based on combinations of survey data, population estimates, and administrative records.	▪ Its estimates rely on statistical models subject to error, and the estimates available are not diverse, including only poverty rates and median household income.

Sources: Moises Loza (executive director, Housing Assistance Council), unpublished letter to US Department of Commerce, re: "proposed information collection; the American Community Survey," May 9, 2011; Joseph Belden (deputy executive director, Housing Assistance Council), unpublished letter to Board of Governors of Federal Reserve System, re: "Home Mortgage Disclosure Act (HMDA); Notice of Hearings: RIN 7100-AD51," August 20, 2010; US Department of Health and Human Services, "Data on Health and Well-Being of American Indians, Alaska Natives, and Other Native Americans. Small Area Income and Poverty Estimates (SAIPE)," December 1, 2006.

Assessing Available Data Sources

In this section, we review the datasets we scanned for this report and identify which are better at describing rural realities and which fall short.

What We Scanned and Why

To better understand the limitations and opportunities of rural data, we scanned various public, administrative, and proprietary datasets and assessed how well they reflect rural realities. We took suggestions from our interviewees, as well as data experts within the Urban Institute, to develop a list of possible datasets to review. We then narrowed the list with our research partners to 22 datasets loosely connected to several community capitals associated with rural prosperity—built, financial, human, and political. The selected datasets reflect a range of geographies, from as small as the census block level to as large as the state and national level. Our goal was to include both datasets that are commonly used in rural research and those with promise.

We pulled a high-level summary of each dataset, and to assess their utility in rural areas, we collected the following information from publicly available codebooks and technical documentation:

- **Geography.** We identified the smallest geographic level available.
- **Data collection and reporting.** We identified underlying data sources, the population the dataset describes, the frequency of data collection, differences in data collection by geography, and data limitations that affect rural applications, such as populations excluded from certain estimates. We also identified data release practices such as suppressing data or incorporating noise that may make rural estimates less complete and accurate.
- **Data quality.** We noted the availability of data, data comparability over time, accuracy of data at rural geographies and variation in that data accuracy by geographic level, and methods used to address missing values.

We grouped these datasets by the category of data collected and within each category assessed the relative strengths and weaknesses of the datasets included. We summarize key takeaways below.

For a subset of five datasets, we also conducted a "deep dive" analysis by downloading and testing the actual data. Through these deep dives, we identified missing data for key measures across a selected set of rural areas (two census tracts, a small municipality, two counties, and a zip code). We also compared measures in the five datasets with similar measures in the ACS, which is the most

comprehensive source of national data on rural communities despite issues noted earlier. We provide key takeaways from this analysis below and our detailed analysis in the appendix.

Data That Experts Use

We asked researchers to identify the datasets they use in their day-to-day research to help us identify the data they consider most necessary and/or valuable for reflecting rural realities to inform policy and practice. The researchers mentioned using datasets from the following sources the most frequently:

- ACS
- Bureau of Labor Statistics
- Bureau of Economic Analysis
- USDA

Also mentioned were the Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES), as well as Zillow data, US Postal Service property vacancy data, and Black Knight/CoreLogic data that collects assessed property values, ownership, and transactions. Data from a credit bureau were recommended for describing consumer finances, and the National Establishment Time Series (NETS) was described as an additional source for information on employers. The IRS was noted as a source for information on consumer finances and businesses, and the Picture of Subsidized Households dataset from the US Department of Housing and Urban Development was mentioned as a good source for housing data.

Finding the "Good" in Existing Datasets

We scanned 22 datasets on businesses, consumer finances, education, employment and earnings, government, health, real estate, and infrastructure to identify those that are the best sources we could find on rural communities and those that should be used with more caution (table 3).

TABLE 3

Scanned Datasets

Type of capital and measure	Dataset	Source	Description	Smallest geographic level
Financial capital: business markets	MarketTrends	CoreLogic	Data on home sales, foreclosure filings, and mortgage performance	Zip code
	National Center for Charitable Statistics Core data*	Internal Revenue Service Statistics of Income	Data on nonprofit organizations	Zip code
Financial capital: consumer finances	Credit bureau sample	A major credit bureau	Credit data on consumers	Zip code
	US Financial Health Pulse*	Financial Health Network	Snapshot of financial health of Americans	Rural/urban
	Individual Income Tax Statistics zip code data	Internal Revenue Service	Information on individual income tax returns	Zip code
	Mastercard	Mastercard	Anonymized and aggregated consumer transaction data	Census block
	Survey of Consumer Finances	Federal Reserve	Triennial survey on family balance sheets, pensions, and income	MSA/non-MSA
Human capital: education and economic mobility	Education Data Explorer	Urban Institute	Compiled pre-K–12th grade and higher data	School
	Opportunity Insights*	Opportunity Insights	Social mobility and other life outcomes	Census tract
	Panel Study of Income Dynamics (PSID)	University of Michigan Institute for Social Research	Longitudinal household survey on topics including intergenerational income and education	Other
Financial capital: employment and earnings	LEHD Origin-Destination Employment Statistics (LODES)	Census Bureau	Geographic patterns and characteristics of workers	Census block
	National Establishment Time Series (NETS)	Walls & Associates	Detailed information on establishments	Zip code
	Quarterly Census of Employment and Wages (QCEW)	Bureau of Labor Statistics	Data on establishments, employment, and wages	County

Type of capital and measure	Dataset	Source	Description	Smallest geographic level
	Quarterly Workforce Indicators (QWI)	Census Bureau	Labor market statistics by industry, worker, and employer characteristics	County
	Regional Economic Information System (REIS)	Bureau of Economic Analysis	Data on employment, personal income, and GDP	Census block
Political capital: government capacity	Economic development dataset	International City/County Management Association	Survey of local governments on economic development activities	Municipality
	Census of Governments*	Census Bureau	Scope and nature of US state and local government	Municipality
Human capital: health	National Vital Statistics System	Centers for Disease Control and Prevention	Data on vital events	City
Built capital: real estate	Property data	CoreLogic	Data on land and property characteristics, property ownership, property sales, and more	Property
	US Postal Service vacancy data	US Department of Housing and Urban Development	Data on vacant addresses and "no-stat" addresses, or addresses not included in carrier service delivery	Census tract
	Zillow Rent Index (ZRI), Zillow Home Value Index (ZHVI), and Zillow Transaction and Assessment Database (ZTRAX)*	Zillow	ZRI and ZHVI: estimated market rent and home value ZTRAX: real estate transaction data and data on property-level characteristics	ZRI and ZHVI: neighborhood ZTRAX: property
Built capital: infrastructure	Local Area Transportation Characteristics for Households (LATCH) survey	Bureau of Transportation Statistics	Estimates of household and vehicle trips and miles traveled	Census tract

Notes: An asterisk indicates a dataset included in our deep-dive analysis. Neighborhoods for the ZRI and ZHVI refer to specific neighborhoods, like Northeast Dallas, Texas, or Harlem, New York.

EDUCATION AND ECONOMIC MOBILITY DATA PERFORM BEST FOR RURAL COMMUNITIES

The datasets we considered that cover educational attainment and economic mobility tended to represent rural communities the best, given the datasets' depth and quality at small geographic levels, as they typically include data from all schools, states, and counties. They cover a wide range of measures useful in assessing educational attainment and economic mobility, in part because two of them (Opportunity Insights and the Urban Institute's Education Data Explorer) pull from several datasets. Opportunity Insights—which pulls from the decennial census, ACS, and federal income tax returns—includes measures of household income, poverty, school district standardized test scores, job growth, and income and racial segregation. The dataset includes the results of a deep analysis that mapped a sample of 20.5 million Americans born between 1978 and 1983 to the communities they grew up in to measure outcomes across census tracts. The Education Data Explorer—which uses data from the Common Core of Data, the Civil Rights Data Collection, the Small Area Income and Poverty Estimates, the Integrated Postsecondary Education Data System, the College Scorecard, and EDFacts—includes data on enrollment, completion, admission, and other school characteristics. The third dataset, the Panel Study of Income Dynamics (PSID), is the longest-running longitudinal household survey in the world and includes information on employment, income, wealth, health, and education.

Compared with the datasets we scanned in other categories, these datasets tended to have the fewest or least consequential data collection or data suppression practices in terms of their effect on rural geographies. They do have notable limitations, however. Estimates for subpopulations, such as household income by gender or by race, in the Opportunity Insights dataset are not always available for rural areas. That may be because the number of people on which the estimates are based is too small or because a particular subpopulation simply is not represented at a particular geographic level. In addition, a few datasets that populate the Education Data Explorer exclude some groups from data collection, such as schools that do not receive federal funding. This exclusion makes the Education Data Explorer less comprehensive. Estimates that pull from the ACS, including several measures available in Opportunity Insights, can also be problematic because of the high margins of error for smaller-population areas that were discussed previously. Additionally, the public use files for the PSID include only broad geographic areas like region or state of residence; only the restricted files allow for linking to smaller geographic levels.

SOME RICH EMPLOYMENT AND FINANCIAL WELL-BEING DATASETS HAVE LIMITATIONS

As noted earlier, assessing rural realities can be difficult when using datasets in which the smallest geographic level is the county or datasets that have just one rural versus urban categorization. Several datasets we scanned were available only at the county or larger geographic level (table 4). The

Quarterly Census of Employment and Wages (QCEW) and Quarterly Workforce Indicators (QWI) provide useful information on the number of business establishments, monthly employment, and employee earnings, but county-level estimates complicate understanding employment in rural areas. The other two datasets, the Survey of Consumer Finances and the US Financial Health Pulse, are rich in how they are structured as surveys and thus offer measures not available in other datasets, such as consumer financial health and asset information. However, the only geographic designation that allows for rural comparisons in both datasets is a set of categories that the data owners call "urbanicity." In the Survey of Consumer Finances, urbanicity is simply whether the geography is inside an MSA or not. In the US Financial Health Pulse data, urbanicity can be rural, urban, or mixed, but how the data owners categorize a place remains unclear. These categories make leveraging these rich data in a way that is useful for rural research or policymaking difficult.

It is also important to note that any dataset that lists data limited to employers included in unemployment benefit programs will have significant gaps in large parts of Indian Country because of different systems and jurisdiction in those areas. Quarterly Workforce Indicators, for example, are drawn from state unemployment compensation system earning records, which exclude self-employment, federal employment, seasonal employment, some nonprofit employment, and most tribal government employment.

TABLE 4

Financial Well-Being Datasets with Broad Geographic Focus

Dataset	Level on which rural is defined
Quarterly Census of Employment and Wages (QCEW)	County
Quarterly Workforce Indicators (QWI)	County
Survey of Consumer Finances	"Urbanicity" (MSA or non-MSA)
US Financial Health Pulse	"Urbanicity" (rural, urban, or mixed)

Source: Urban Institute dataset scan.

DATA FOR SMALL POPULATIONS ARE HELPFUL, BUT PRIVACY IS A CONCERN

Among the datasets we examined were some that included smaller geographic levels, such as zip codes or census blocks (table 5). Data availability at smaller geographic levels improves our ability to provide additional local nuance that would be blurred if only larger geographic levels were available. For example, National Center for Charitable Statistics Core data, Opportunity Insights, the Regional Economic Information System, and the Zillow Rent Index and Zillow Home Value Index all provide data describing local characteristics across different dimensions, including information on the nonprofit

sector, social mobility, educational attainment, and housing affordability. Other datasets—like the IRS's Individual Income Tax Statistics, LODES, the Local Area Transportation Characteristics for Households (LATCH) survey, and US Postal Service vacancy data—provide aggregate information on groups of people or properties at small geographic levels, detailing household income, worker characteristics, means of transportation, and the extent of vacancy across properties. If these data were available only aggregated across multiple rural communities or within a county that contained suburban or urban communities as well, characteristics of the individual rural community would be overshadowed by those of its larger-population neighbors.

The availability of data at smaller geographic levels improves our capacity to examine rural realities, but data suppression to preserve confidentiality complicates rural research efforts. When a sample size used to construct an estimate falls below a set threshold, data are suppressed, or not reported, and an additional amount of random "noise" may be added to estimates to further protect the identities of individuals or entities whose data were included. Given small population sizes, some subpopulations may not even be represented at smaller geographic levels across rural areas, and data estimation techniques such as imputation may be used to replace missing data with substituted estimates. For example, not every population subgroup and income level is represented in a census tract in the Opportunity Insights dataset. If, for instance, a census tract has no parents at the 25th income percentile, the researchers use parents at other nearby values (a higher one or a lower one) to impute the estimates. Although such estimation techniques are a useful alternative to excluding unrepresented groups across small geographic levels, the accuracy of those techniques is a concern.

Assessing how well proprietary datasets with smaller geographic levels represent rural areas is difficult; public documentation of data limitations or data suppression is often limited. From public documentation alone, we could not assess whether any data suppression or other data confidentiality actions occur for the CoreLogic MarketTrends dataset, the CoreLogic property data, Mastercard, and NETS.

TABLE 5

Datasets That Include Small Geographic Levels

Dataset	Geographic level	Data suppression	Other data confidentiality actions
CoreLogic MarketTrends	Zip code	Unknown	Unknown
CoreLogic property data	Property	Unknown	Unknown
Credit bureau sample	Zip code	Unknown	Unknown
National Center for Charitable Statistics Core data	Zip code	Nonprofits with under $25,000 in revenue are excluded	None
Individual Income Tax Statistics	Zip code	Zip codes with fewer than 100 returns are aggregated at the state level	None
LEHD Origin-Destination Employment Statistics (LODES)	Census block	Some data suppression for small geographic units	Worksites are geocoded and then "fuzzed" or masked within a given distance of the actual location to obscure employers' identities
Local Area Transportation Characteristics for Households (LATCH) survey	Census tract	None observed	None observed
Mastercard	Census block	Unknown	Unknown
National Establishment Time Series (NETS)	Zip code	Unknown	Unknown
National Vital Statistics System	City	Varies by measure (e.g., data representing fewer than 10 deaths are suppressed in the mortality data)	None
Opportunity Insights	Census tract	Do not publish estimates based on 20 or fewer people	Add noise to all estimates to protect privacy
Regional Economic Information System (REIS)	Census block	None observed	None observed
US Postal Service vacancy data	Census tract	None observed	None observed
Zillow Rent Index (ZRI) and Zillow Home Value Index (ZHVI)	Neighborhood	Quality control rules lead to county-level suppression	None
Zillow Transaction and Assessment Database (ZTRAX)	Property	None observed	None observed

Source: Urban Institute dataset scan.

Note: Neighborhoods for the ZRI and ZHVI refer to specific neighborhoods like Northeast Dallas, Texas, or Harlem, New York.

PROPRIETARY DATASETS HOLD PROMISE, BUT ACCESS IS TOUGH

Proprietary datasets tend to be better sources of data given that the data are both available at smaller geographic levels and often not aggregated. For example, both CoreLogic and Zillow provide data at the property level, with address information available at small levels of geography, allowing for extremely localized research. The data allow for summarizing property or land classification, average lot size, average tax levied on properties, property use classification, real estate loan types, average loan amount, and more. Because the data are organized by property, one could also focus in on properties that meet certain characteristics such as property type (e.g., mobile home) or lot size. Credit bureau data include credit data on individual consumers, which excludes all identifying information as well as most demographic information, and Mastercard provides anonymized and aggregated transaction data to understand consumer behavior.

Although they hold promise, proprietary datasets do have notable challenges around their use. Proprietary datasets can be expensive to purchase, and even if they are provided for free to the public, the process of requesting access can be lengthy. Once proprietary datasets are purchased or accessed, they can also be difficult to use. Zillow, for example, organizes its ZTRAX dataset separately by transaction and assessment files and organizes those files by state. The files can be quite large given the large number of properties or parcels they represent, posing analysis challenges, especially if analysis over time is desired.

Like other datasets, proprietary datasets are also affected by data collection challenges and data suppression. For Mastercard data, a minimum number of retailers must be contained at a geographic level for certain credit scores to be presented. The publicly available ZTRAX data dictionary notes when specific fields may have high missing rates because of the frequency with which those fields are reported across counties, and other public documentation states that county-by-county differences in data reporting contribute to missing data.[11]

Another challenge when using proprietary datasets is that public documentation detailing data collection or data suppression practices is often limited. When such documentation is available, it may not provide the level of detail necessary to understand rural limitations without having access to the data. This increases the risk of purchasing data, or going through a lengthy data acquisition process, without knowing whether the data are usable for the desired analysis.

DATA AGGREGATORS ARE ONLY AS GOOD AS THEIR UNDERLYING DATA

Datasets that pull together data from other datasets known to have high margins of error at small geographic levels, such as the ACS, should be used with caution for rural research. These aggregators

show data from multiple sources simultaneously and may create indices or rankings using data from multiple sources (table 6). Before using outputs from data aggregators, users should understand the opportunities and limitations of each underlying dataset and pay attention to the warning signs, including high margins of error, data suppression or noise, and missing or unreported data. This is especially the case when aggregators use data for geographic areas smaller than counties.

TABLE 6

Common Data Aggregators

Aggregator	Source	Data type	Smallest geographic level
Atlas of Rural and Small-Town America	USDA Economic Research Service	Provides statistics on socioeconomic characteristics, including demographic data, economic data, county classifications, data on household income and poverty, and data on veterans	County
Census Business Builder	Census Bureau	Provides demographic and economic data	Census tract
County Health Rankings and Roadmaps	Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute	Provides data on vital health factors, including obesity, smoking, unemployment, high school graduation, access to healthy food, income inequality, air and water quality, and teen births	County
Esri	Esri	Provides demographic, psychographic, and socioeconomic data, including data on population, households, age, ethnicity, consumer spending, traffic, crime, and housing	Census block, depending on specific measure and source
PolicyMap	PolicyMap	Provides data covering topics such as demographics, income and spending, housing, the economy, education, health, quality of life, and lending activity	Varies. Smallest geographic level available across all datasets is the block group.
Prosperity Now Scorecard	Prosperity Now	Provides data on family financial health spanning five issue areas: financial assets and income, businesses and jobs, homeownership and housing, health care, and education	County
Rural Opportunity Map	Center on Rural Innovation	Provides interactive maps detailing a diverse set of indicators, including broadband internet coverage, higher education, New Markets Tax Credit and Opportunity Zone tracts, direct federal funding and federal agency spending, health, employment, and education. More measures are currently being tested.	Rural areas, which includes places with less than 50,000 people and excludes some places directly adjacent to an urban area. New updates, currently being tested, will allow for multiple definitions of "rural," including FAR and RUCA codes.
Veterans Data Central	Housing Assistance Council	Provides data on the social, economic, and housing characteristics of veterans in the US using ACS data	County
Rural Data Portal	Housing Assistance Council	Social, economic, and housing characteristics of US communities using data from Housing Assistance Council tabulations of the 2010 Census of Population and Housing, the American Community Survey, and Home Mortgage Disclosure Act data	County

Aggregating "bad" data does not make it "good." Instead, errors can be compounded. This is particularly relevant to rural communities that have more error-prone data. For example, several data aggregators pull from ACS estimates and may treat certain variables across geographic levels as missing in aggregator results when the underlying data have high margins of error.

Moving toward Better Rural Data

Researchers and other data users are continuously working to identify techniques to expand what the data we have can tell us and to improve our ability to understand our communities in deep and nuanced ways. Records kept by government agencies, nonprofits, and other service providers on the people they serve and the services they offer, often referred to as administrative data, offer new ways of looking at a community or population. When linked together or combined with other data sources, administrative data can enrich more general demographic and economic data.

Leveraging Administrative Program Data

Public or private program managers collect administrative data as they administer specific programs, tracking a plethora of individual and community characteristics required for enrolling in programs and monitoring program eligibility and participation over time. Programs can range from public education, to continuing education and job training, to transportation, to social benefit programs that address health, food, and housing. These program data could stand in where other data fail, resulting in more accurate, alternative measures of rural prosperity to inform rural practice, policymaking, and research by providing new types of information and overcoming inaccuracies in other data collection practices. Administrative data may also expand on less reliable survey data, be combined with other data to provide more nuanced analyses, or be integrated into a single system to track participation and outcomes across multiple programs as they relate to improving rural prosperity.

COMBINING ADMINISTRATIVE DATA

Several experts we spoke with said they used administrative data sources on the federal and state level with more standard economic and demographic indicators to better understand the distribution of social assistance programs and, therefore, rural poverty. State databases on the use of the Supplemental Nutrition Assistance Program (SNAP), for example, can be connected to population counts and demographic breakdowns to show what share of people are accessing food assistance at

levels more granular than ACS poverty estimates. In addition to providing an alternative measure of poverty, SNAP participation may also provide insight into areas of food insecurity and instability.

Rural researchers are also seeking new ways to link administrative data to other public and proprietary data sources to provide clearer evidence to drive rural practice and policy and expand rural prosperity. Linked datasets can help researchers seeking more nuanced data around economic development strategies. One example of work in this area is the linking of administrative broadband internet data to National Establishment Time Series data on small-business growth to determine the impact that high-quality internet access can have on economic development prospects. This could provide evidence on where broadband access is more crucial for rural business development. Other examples include linking Small Area Income and Poverty Estimates with the Quarterly Census of Employment and Wages to provide more detailed information about the connection between poverty and employment by industry in rural communities. Understanding these connections can suggest industries more associated with improving rural individual and community financial capital as areas for future investment.

We're getting into trying to link program and administrative data to other data sources. We have people here working with food/nutrition service data, linking it to SNAP and ACS data. —Rural data expert

Combining data holds promise, but it comes with challenges. One expert noted that connecting a record accurately to a geographic area in statistical or mapping software can be difficult: "Place of residence versus place of work is a tricky thing to pull apart. They often get conflated." In addition, combining datasets does not always remove scale issues. Many datasets are limited to the county level and cannot provide more granular insight even if combined with smaller-scale data. Finally, combining datasets does not reduce the limitations that exist in some of them, such as the ACS.

Combining datasets is a complex task that requires specific skills. Most datasets do not fit together easily, especially if they were created by different organizations. Increased collaboration among agencies that produce datasets would greatly improve their accessibility, especially for rural communities that may not have the staff or tools to execute combinations themselves.

It really would be great if there were more collaboration between foundations and agencies who put this data together. There should be a one-stop shop for all the data. Point and click. Always the most up to date. Where lay audiences can use it.
—Rural practitioner

IMPLEMENTING INTEGRATED DATA SYSTEMS

Some state and local governments have begun integrating their administrative databases across the programs they operate to improve their practices by better aligning their service delivery. Although these efforts have primarily been for internal purposes of transparency, monitoring, and cross-department data sharing, these integrated data systems (IDS) can generate increasingly comprehensive datasets for other purposes. For example, IDS can be used to analyze associated outcomes for program participants to suggest where policies and related programs can be transformed to promote increased rural prosperity.

Some leaders in IDS implementation are emerging. Allegheny County, Pennsylvania, has developed a county data warehouse that consolidates human services data in areas such as behavioral health, child welfare, homelessness, and aging and makes them available to the public. These data could allow analysts to identify areas with a concentrated need for assistance to help build individual and community capital, for example, without resorting to proxies such as income. Other governmental bodies that have developed IDS are the State of Washington and the South Carolina Office of Research and Statistics. Education and child-focused services have also led in this area; for example, the Juvenile Welfare Board of Pinellas County, Florida, has used IDS to seek solutions to chronic student absenteeism, and the Policy and Services Research Data Center at the University of South Florida has developed an IDS incorporating data from Medicaid, Emergency Medicine, and state services in substance abuse, mental health, and child welfare (Derian 2016).

It depends on the state and how well-linked they are, but a well-linked dataset across agencies, you can learn a lot—e.g., tax returns linked to education linked to childhood health. Rural survey data will always be disadvantaged because sample size is too small to really have any power.
—Rural data expert

Investment from public or private sources could contribute to significant expansion of usable combined datasets in IDS. One group, Actionable Intelligence for Social Policy at the University of Pennsylvania, has been working on creating a network of governments working on IDS and supporting evidence-based collaboration since 2008. However, this initiative remains small and would benefit from incorporating public datasets.

New and Emerging Data Sources

Our conversations with experts and practitioners uncovered several new sources for data that are beginning to supplement the go-to sources discussed in this report. These range from new web-based data like social media activity, to new government initiatives to leverage public data in new ways, to private data sources that provide access to individual-level data unavailable through public sources, and new efforts to collect and arrange data via surveys, typologies, and scales. Although these sources are promising, the full set of benefits and drawbacks is yet to be seen.

ONLINE AND APP-BASED SOURCES

Some data are available on the internet but not in an easily downloadable format. Applications are available to help scrape data from these sites when it is allowed (e.g., the data are available for public use). Social media can also be a rich source of user-generated data. One interviewee noted some approaches that could be used for rural research, including tracking restaurant listings and reviews on platforms such as Yelp as an indicator of economic growth or tracking posts on social sites such as Twitter for health status updates (e.g., to track disease outbreaks in real time). These sources require reliable internet access and a broad pool of users to be effective, however.

GOVERNMENT DATA INNOVATION

Government agencies are exploring how to use data in new ways to better serve their missions. Since 2016, the Census Bureau's Opportunity Project has provided seed funding for tech companies to leverage public data in new ways through app-based development. This has included several rural-focused ones, centered on finding housing, displaying indices of opportunity, and comparing similar communities across education, jobs, and economic indicators.[12] Also, the chief technology officer of the US Department of Health and Human Services runs the agency's Idea Lab to promote data innovation internally and externally through partnerships with startups, entrepreneurs, and health-focused businesses.[13]

Even in state and local governments where datasets are not yet integrated, efforts to improve transparency and data access have been expanding. All 50 states and the District of Columbia have open data portals.[14] Many municipal and county governments have such portals as well, although they are concentrated in urban areas. The Urban Institute has released a data catalog that compiles data from local governments and states "created, enhanced, cleaned, or otherwise added value to by Urban's staff" and other datasets related to health, income, and local capacity.[15]

USE OF INDIVIDUAL-LEVEL DATA FROM PROPRIETARY SOURCES

One way to overcome limitations in existing rural data is to use individual-level measures such as credit scores as more accurate measures or supplemental measures of rural prosperity. These individual measures can be available in proprietary data. These pay-to-access datasets scrub their data of identifying characteristics so that researchers and rural stakeholders can access the otherwise private data, but proprietary data come with challenges. First, unlike free, public datasets, these datasets often require a fee that may be prohibitive to some local governments, researchers, and practitioners. Second, many rural communities may not have the technical capacity or ability to pay to process and analyze these data. This includes having the necessary computer hardware and software to run the data and the staff members with the skills to execute the analyses and interpret the outputs (although partnerships with institutions of higher education can help meet this need). Third, even though proprietary data offer more granular information than is available through most public sources, providers may suppress individual-level data for smaller-population areas to protect privacy. Some proprietary data owners offer data only down to the zip code or census tract, for example, while others simply do not report data for geographic areas below a certain population threshold. To complicate matters, many proprietary data providers do not disclose how complete their datasets are for a specific geographic level before purchase.

NEW SURVEYS, SCALES, AND TYPOLOGIES

An emerging set of new resources provides opportunities for insights into rural communities (table 7). These include surveys, which collect data based on questions asked to a sample of individuals in the study area of interest; scales, which assess levels of prosperity based on a combined set of measures; and typologies, which group similar communities together based on common features to better understand them as a whole. Among the new resources are

- rural-focused national surveys that provide more nuanced data on rural realities across important concepts of rural prosperity,

- survey datasets that incorporate qualitative interview data yielding rich descriptions of individual experiences in rural communities,
- ways of grouping similar rural communities to point to cross-cutting approaches to advancing prosperity within and across them, and
- nontraditional scales that redefine prosperity as thriving across various measures not captured in the community capitals framework.

Although not necessarily focused on representing the exact realities of every small rural community in the country, these new data sources and approaches are geared toward providing a more nuanced narrative and tools for more thoughtful rural policy and practice.

TABLE 7

Emerging Tools for Understanding Prosperity in Rural Communities

Title	Type	Description	Source(s)
Life in Rural America Poll	Survey	A telephone poll of adults selected to be nationally representative of rural communities on economic and health issues. The poll has been conducted twice, once in summer 2018 and once in winter 2019, with samples of 1,300 and 1,405, respectively. Although raw data are not available online, summaries can be found at the program website.	Harvard University and NPR
American Voices Project	Interviews	Still in the collection phase, this project seeks to interview people from across the US to learn details of their day-to-day welfare and activities. Rather than focusing on a specific area of research, the project seeks to "learn what's going well, what needs to be improved, and how we might make our neighborhoods and country work better." The protocol will be administered to 5,000 households across 200 communities and incorporate income and employment data, as well as psychological impact data.	Stanford and Princeton Universities
American Communities Project	Typology	Using data sources such as County Health Rankings and Roadmaps, Gallup, and Simmons Consumer Research, the American Communities Project identified nine rural community types: (1) African American South, (2) Aging Farmlands, (3) Evangelical Hubs, (4) Graying America, (5) Hispanic Centers, (6) LDS Enclaves, (7) Native American Lands, (8) Rural Middle America, and (9) Working Class Country. An overview of the major findings on demographics, economics, infrastructure (physical and virtual), daily life, health care, and well-being are available at the project website.	American Communities Project, George Washington University
Crisis-to-Thriving Scale	Scale	Intended as a way to assess the life conditions of individuals or families, the crisis-to-thriving scale tracks indicators across a range of categories, including food and nutrition, housing, child care, transportation, and financial management. In each area, families can be identified as in crisis, vulnerable, safe, stable, or thriving. These measures can provide insight into community health as well.	Garrett County Community Action Committee

Next Steps on the Path to "Good" Rural Data

Rural communities deserve "good" data that represent their on-the-ground realities. Better data would improve rural research and analyses and enable practitioners and policymakers to make evidence-informed decisions to advance rural prosperity. They would also promote an expanded national narrative on rural communities, one of diversity and divergence across dynamic communities, rather than a story of homogenous decline. Without better data, rural communities will remain ill-equipped to understand their past, including the causes and consequences of failures in policy and practice, and to plan and implement their futures with the benefit of solid evidence.

Understanding the limitations of existing data and the promise and potential of new data sources and collection techniques is a step in the right direction. However, more work is needed to explore the concept of rural prosperity and prompt data owners and users to make progress on generating, using, analyzing, and interpreting better rural data. We suggest a series of next steps to strengthen practice, improve policymaking, and change the narrative surrounding rural communities.

Reconceptualizing Individual and Community "Prosperity"

More work is needed to conceptualize what we mean by "prosperity" and to find the data to measure aspects of prosperity not explored in this report. The community capitals framework helps identify resources that a community can mobilize for growth and well-being but falls short in identifying individual measures of prosperity. Additional measures may need to be incorporated to both understand rural realities and identify ways to influence them through new policies or practice. The following are some possible measures suggested by rural experts:

- **Work readiness.** Measures such as individual educational achievement level could provide more nuanced data on the skills of the local workforce.
- **Individual health.** Measures of individual health risk factors and incidences of chronic diseases and health conditions could give clearer indications of needed health-related interventions.
- **Owned assets.** Measures of individual ownership of land and other assets as alternative revenue sources could provide a more accurate picture of assets and wealth than income or earnings.
- **Seasonal employment.** Measures of individual employment or earnings that take into account seasonal work over the course of a year avoid the over- or underestimates of employment and income endemic to data that rely on a point-in-time measure or rate.

Measures of community capitals provide insight into local resources—built, financial, human, political, and others—but additional measures can capture capacity needed to leverage these resources to advance community prosperity. Some suggestions include

- **institutional capacity,** which captures whether local government, nonprofit organizations, and private institutions have what it takes to implement policy agendas, economic development initiatives, or physical infrastructure developments, including revenue sources, funding flows, and tax base;
- **economic diversity,** which measures the diversity of employers and employment opportunities to assess the vulnerability of local employment and wealth to economic shocks; and
- **the presence and strength of anchor institutions,** including hospitals, universities, and community colleges, as well as other major long-term employers.

A Path Forward for Data Owners and Users

Innovations in data collection and technology are improving the availability and quality of rural data, but there is still a long way to go. Those who collect and own data have some important decisions to make about how to improve the quality and use of their data for advancing rural prosperity. There are also implications for data users: policymakers who use data to make evidenced-based decisions for rural places, practitioners who need data to improve their work, and researchers who analyze data on rural communities. Although the following suggestions are not an exhaustive list of possible actions, we believe they are important steps toward improving access to "good" rural data.

WHAT DATA COLLECTORS AND OWNERS CAN DO

Public and private data collectors and owners should gather the best rural data possible and share it with rural communities in such a way that it can be analyzed and used to make decisions to advance individual and community prosperity. Data owners can consider doing the following:

- Increasing the size of the rural sample in survey research. This may require asking more people to complete a survey to yield an adequate response rate and reduce margins of error or ensuring that surveys can be completed via paper or in-person interview instead of online.
- Trying new strategies for increasing rural participation in important surveys like the ACS, particularly on tribal lands and other difficult-to-count places (O'Hare 2017). This could include providing safe, secure internet in community spaces such as libraries, schools, and employment

centers; providing on-site translators; and mobilizing community leaders as advocates to encourage more responses (Brumfield 2018).[16]

- Reexamining practices around data suppression and noise as they relate to rural communities. Valuable local data should be made available to rural policymakers and practitioners in the communities from which they have been collected, as well as to their research staff and/or partners, in as complete a form as possible to help track progress and plan for the future.
- For owners of proprietary data, partnering with governments and researchers to increase access to rural-specific data for the purpose of policymaking and research.

RECOMMENDATIONS FOR DATA USERS

Data users can ensure that they are assisting with high-quality rural data collection; partnering on rigorous, influential analyses; and leveraging analyses of "good" data to make informed decisions for rural policy and practice. They can also work to change the narrative surrounding rural communities as more, new, and better rural data and analyses become available.

Policymakers

- Ensure that you are accessing the best data on prosperity possible. This might mean using different data sources for rural individuals and communities from those traditionally used or finding better proxies for measures of prosperity that are uniquely important to rural communities. It may also mean looking more closely at margins of error in survey estimates and determining whether other data sources may be helpful or necessary in corroborating survey estimates.
- Seek partnerships with data owners to access the best community-level data possible. Data-sharing agreements between local governments and data owners may also help small communities access data necessary for informed decisions while maintaining privacy standards. Assisting rural communities in paying for data access may be necessary.
- Support improved rural data collection and dissemination, such as expanding rural samples, providing reliable internet locations, and engaging community volunteers.
- Partner with researchers who understand the nuances of existing rural data, the possibilities for accessing emergent data, and the proper analytic techniques to provide the strongest evidence needed for informed policymaking.

Practitioners

- Advocate for better data for the communities you serve.
- Encourage rural residents to participate in important data collection activities to improve the quality and coverage of data for rural communities.
- Use this report and its dataset summaries and examples to understand the opportunities and challenges with commonly used datasets and make sure you are using the best data accessible to you.
- Partner with rural researchers to help navigate data with which you are unfamiliar and analyze it appropriately. Skilled rural researchers can be found in academic institutions (e.g., land grant universities), nonprofit research organizations (e.g., the partners on this report), and public agencies (e.g., USDA's Economic Research Service), and they are natural partners for rural data work.
- Look for ways to work with local and state agencies that may have administrative data they can provide to you after ensuring that individuals' privacy is protected.
- In the absence of "good" data elsewhere, collect your own data to track measures of prosperity important for your work.

Researchers

- Conduct policy-relevant research guided by the needs of rural communities to inform policymaking and practice.
- Partner with rural communities to provide the analytical talent needed to make sense of rural data and inform policy decisions.
- For those less familiar with rural research, work to include rural areas in your research by searching for quality datasets with the most granular geographic coverage possible. This may include using proxies for measures that are commonly used but are not "good" enough to use for rural analysis, as detailed in this report.
- Seek innovative ways of collecting, linking, and analyzing data for rural places, including leveraging administrative data and tapping into new sources. This may mean working with state and local agencies and proprietary data owners to obtain data on communities of interest.

Appendix. Supplemental Deep-Dive Analysis

Codebooks and technical documentation provide useful information on how well datasets capture rural realities. For a subset of five datasets, we went beyond this information and conducted a "deep dive" to better understand rural representation by downloading data and testing its quality across different rural areas. We chose the following sample of datasets, representing a mix of community capitals and associated measures.

- **Opportunity Insights**, which provides information on demographic characteristics, earnings, educational attainment and quality, employment, health, and poverty. The dataset includes a wide range of geographic levels, from the census tract to the state and nation.
- The Census Bureau's **Census of Governments** contains data on the nation's state, county, municipal, and overlay governments, such as the number of full- and part-time employees.
- The **National Center for Charitable Statistics (NCCS) Core data**, populated by IRS Form 990 data, provide information such as revenue, number of employees, and fundraising activities for nonprofit organizations categorized as 501(c)3 through 9. The data include individual records that can be aggregated to the zip code level and above.
- The **US Financial Health Pulse** provides information on consumer financial health, including state of residence and "urbanicity," which identifies respondents as living in an urban, rural, or mixed area.
- The **Zillow Transaction and Assessment Database** (ZTRAX) contains information on property transactions, such as deed transfers or mortgages, as well as geographic information and characteristics of individual properties.

For the deep dives, we divided the work into two components. First, we summarized data completeness for key indicators across rural geographies to assess how rural geographies are reflected in the data. Second, we compared these key indicators to similar ones from the ACS to understand how the estimates relate to one another and to determine the scanned dataset's value for understanding rural realities. We compare with the ACS because it is the most comprehensive source of national data on rural communities despite known higher margins of error at the smallest geographic levels and unavailability of data for smaller-population areas.

We sampled six geographies to use to compare the deep-dive datasets with the ACS. We chose these geographies to reflect various sizes and contexts, including RUCA codes for census tracts, rural-urban continuum codes (RUCC) for counties, and diverse rural regions across the US. The six geographies are as follows:

- **A rural census tract in a metropolitan county**. We chose a census tract (with FIPS code 06055-2017), designated as rural by its RUCA code assignment, in Napa County, California, which is metropolitan according to its RUCC designation. The census tract had a population of 5,852, based on 2013–17 ACS five-year estimates, and a land area of almost 70 square miles.
- **A rural census tract in a nonmetropolitan county.** We selected a census tract in Iowa (with FIPS code 19069-3603), designated as rural by its RUCA code assignment, in Franklin County, Iowa, which is nonmetropolitan, according to its RUCC designation. The census tract had a population of 2,413, based on 2013–17 ACS five-year estimates, and a land area of almost 300 square miles.
- **A zip code in a rural county.** We selected a zip code (32066) in Lafayette County, Florida, a rural county, according to its RUCC designation. The zip code had a population of 7,488, based on 2013–17 ACS five-year estimates.
- **A city in a rural county.** We selected Bottineau, North Dakota, a city with a population of 2,300, based on 2013–17 ACS five-year estimates. Bottineau is in Bottineau County, which is rural, according to its RUCC designation. The city itself is just over 1 square mile.
- **A county with an established tourism industry.** We selected Hardy County, West Virginia, which is nonmetropolitan, according to its RUCC designation. Hardy County has a population of 13,812, based on 2013–17 ACS five-year estimates.
- **A rural persistent-poverty county.** We selected Choctaw County, Mississippi, a persistent-poverty county[17] identified as rural by its RUCC designation. Based on 2013–17 ACS five-year estimates, it has a population of 8,360.

Extent of Missing Data across Rural Areas

The extent of missing data across the deep-dive datasets varies significantly, not only by category of data but by the level of geography. Factors that influence the underlying data, such as reporting thresholds or survey sampling, similarly influence the extent of data unavailability.

Opportunity Insights

We examined two data aggregations available for public download from Opportunity Insights and assessed how well various indicators within each were represented across available rural geographies, as well as by the six rural geographies tested. Overall, the extent of missing estimates is generally low for rural geographies, except for important subpopulations at both the census-tract and county levels. Health indicators, in particular, tend to have the most missing estimates for rural areas.

The first data aggregation, "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility," presents estimates on household income from low-income households, as well as estimates on various neighborhood characteristics for census tracts and counties. We analyzed data for rural census tracts identified through RUCA codes and rural counties according to RUCC designations.

We found that estimates on incarceration and household income for children of incarcerated parents were generally missing for less than 6 percent of rural census tracts, and the extent of missing data for those estimates for rural counties was lower, generally missing for less than 4 percent of rural counties. Subpopulation estimates, such as mean household income for gender and racial subgroups, were missing across rural geographies at much higher rates, generally for as high as 90 percent of rural census tracts and 85 percent of rural counties. For the four rural census tracts and counties in table A.1, we calculated the share of all subpopulation estimates that were missing and found that about a quarter were missing for the rural Napa County census tract and Choctaw County and more than half were missing for the rural Franklin County census tract and for Hardy County. Neighborhood characteristics were missing much less frequently, generally below 10 percent for rural census tracts across estimates and much lower for rural counties, often close to 0 percent. Although neighborhood characteristics data and aggregate estimates seem to be available for most rural geographic levels, rural researchers should be aware that estimates for particular subpopulations are often unavailable at small geographic levels when using these data.

The second data aggregation, "The Association between Income and Life Expectancy in the US, 2001–14," reports on county characteristics, mostly related to health outcomes but also including information on income and racial segregation, unemployment, poverty, social capital, population density, crime, government expenditures, and educational attainment. It also provides county-level life expectancy estimates by gender and income quartile. Our analyses of the data across all rural counties and our two test counties revealed the following:

- Data absence across rural counties varies significantly by measure, from as low as 0 or close to 0 percent for population, labor force, and segregation indices, but as high as more than 70 percent for health variables on smoking and obesity.
- County characteristics data were generally available for the two rural counties we tested, except for smoking and obesity data for Choctaw County. Rural representation was much lower in the county-level life expectancy data file. Only seven rural counties are in that dataset, and neither Choctaw County nor Hardy County is among them.

Census of Governments

The Census of Governments is based on a government units survey that collects data from all state and local governments across the US every five years. The 2017 survey had a response rate of 85.4 percent. Among respondents, there were no missing data based on the size of government or the population it serves. It is unknown what share of survey nonrespondents were in rural areas and, thus, how representative the survey results are of all rural-serving local governments. However, there may be a nonsampling error because of the mechanics of the survey itself—including nonresponse bias, incorrect reporting, misclassification, and administrative errors—that may be higher for rural communities, where minimal government structures may mean a lower response rate to surveys.

Because the unit of measurement for the Census of Governments is governments, any geographic level that does not align with a government unit—such as a census tract, zip code, or unincorporated area—is not reflected in the dataset. All our test cities and counties—Bottineau, North Dakota; Hardy County, West Virginia; and Choctaw County, Mississippi—are included in the survey results.

The National Center for Charitable Statistics

Any nonprofit organization that files a tax return in a given year is included in that year's NCCS Core data file. Nonprofits that file late would be represented in data of the year filed, not the tax year. Also, before 2010, the minimum threshold for an organization required to file a Form 990-EZ was $25,000 in gross receipts; in 2010, that minimum was raised to $50,000. This is likely to leave nonprofits in small communities, which likely have smaller budgets than those serving larger-population areas, underrepresented in this dataset. Organizations that fall below this threshold and choose not to complete the Form 990-EZ are not represented in the data. If they choose to file a Form 990-EZ or 990, that information will at times be reflected in the NCCS data, but in other instances, the IRS will enter "0" for every financial field, indicating that the organization falls below the threshold. Unfortunately, for

these "0 filers," it is not generally possible to distinguish between organizations that authentically have $0 in all financial fields and those coded as such by the IRS for falling below the threshold. Unfortunately, there is no way to know how many organizations are missing from the data. However, in 2015 Core data (the latest available), 2,293 of 147,772 records have $0 in gross receipts.

US Financial Pulse Survey

Because the US Financial Health Pulse is a survey, missing data are information that should have been collected but was not. We summarized the extent of missing data across the survey questions and found low missing rates across responses, generally under 1 percent.[18] When the data were missing, they were less likely to be missing for respondents in rural areas than for respondents in mixed or urban areas.

Zillow Property and Assessment Data

The publicly available ZTRAX data dictionary notes when specific fields in the assessment and transaction files are not recorded consistently or other reasons why there may be missing data across fields. For example, differences in how counties report the data contribute to the extent of missing data across geographies.

For the two rural counties in our deep-dive analysis, Hardy County, West Virginia, and Choctaw County, Mississippi, we assessed the extent of missing data for a set of measures.[19] Property or land classifications and land use classifications had high missing rates for those two counties. Other property characteristics, such as lot size and tax amount levied on the property, had low missing rates below 10 percent for Hardy County, whereas the tax amount levied on properties was missing for all records in Choctaw County.

Comparison with the American Community Survey

To see how well our selected datasets compare with similar measures from other data sources, we summarize five-year estimates from the 2013–17 ACS for the six rural places we tested (table A.1). Selected measures span demographic characteristics, earnings, educational attainment, employment, housing, and poverty.

TABLE A.1

Profile from American Community Survey Five-Year Estimates for Selected Geographic Areas

	Rural census tract in California	Rural census tract in Iowa	Zip code in rural Florida county	Bottineau, North Dakota	Hardy County, West Virginia	Choctaw County, Mississippi
Demographics						
Total population	*5,852*	*2,413*	*7,488*	*2,300*	*13,812*	*8,360*
Percentage of women	51%	48%	40%	51%	50%	51%
Median age	37	41	37	42	45	44
Percentage of minorities	38%	9%	35%	14%	9%	32%
Earnings						
Median household income (2017 dollars)	$81,842	$47,955	$41,089	$52,560	$42,573	$34,542
Educational attainment						
At least a high school education	89%	92%	71%	91%	81%	81%
Employment						
Unemployment rate	6%	6%	16%	6%	5%	10%
Housing						
Total housing units	*2,216*	*1,209*	*2,405*	*1,121*	*8,217*	*4,186*
Owner occupied	58%	79%	79%	66%	73%	76%
Median housing value	$856,600	$87,600	$110,100	$132,800	$121,900	$77,200
Median monthly housing costs	$1,646	$567	$520	$689	$641	$458
Median gross rent	$1,463	$589	$603	$719	$668	$547
Poverty						
Poverty rate	10%	17%	18%	7%	15%	25%

Source: American Community Survey 2013–17 five-year estimates.

Opportunity Insights

Opportunity Insights estimates of measures that are comparable to ACS estimates of neighborhood characteristics paint a similar picture of the four places (table A.2). Estimates of median household income, poverty rate, and rent are fairly similar between the two datasets, which is expected because the Opportunity Insights measures are built off ACS estimates. Opportunity Insights data provide some measures that are not directly available from the ACS because Opportunity Insights pulls from various data sources.

TABLE A.2

Selected Measures from Opportunity Insights Neighborhood Characteristics Data

	Rural census tract in California	Rural census tract in Iowa	Hardy County, West Virginia	Choctaw County, Mississippi
Mean household income in 2000	$122,254	$62,127	$56,223	$50,539
Share of people 25 years or older with a bachelor's degree, master's degree, professional school degree, or doctorate degree in 2010	44%	16%	10%	11%
Median household income in 2016	$72,308	$47,994	$41,991	$32,320
Poverty rate in 2010	10%	11%	15%	21%
Average rent for a two-bedroom unit in 2015	$1,242	$530	$617	$583
Share of single-headed households with children in 2010	16%	11%	42%	22%
Share of working adults with commute time of 15 minutes or less in 2010	61%	38%	43%	41%
Census form return rate in 2010	76%	80%	79%	79%
Number of primary jobs within five miles in 2015	8,230	463	…	…
Average annual job growth rate between 2004 and 2013	5%	2%	-3%	0.2%
Job density (in square miles) in 2013	54	1	8	8

Source: Opportunity Insights.

Some Opportunity Insights county characteristics data are similar to those of the ACS, such as the unemployment rate and median house value (table A.3).

TABLE A.3

Selected Measures from Opportunity Insights County Characteristics Data

	Hardy County, West Virginia	Choctaw County, Mississippi
Uninsured rate in 2010	19%	22%
Medicare spending per enrollee	$8,374	$9,199
30-day hospital mortality rate index	-0.1	1.2
Income segregation	0.0018	0.0077
Racial segregation in 2000	0.0154	0.0540
Unemployment rate in 2000	4%	9%
Median house value	$99,575	$73,715
School expenditure per student	$5.55	$4.61
Student-to-teacher ratio	16:1	18:1
Total crime rate	0.0039	0.0035
Local government expenditures	1418.88	1601.59
Local tax rate	0.0097	0.0099

Source: Opportunity Insights.

Notes: Opportunity Insights' county characteristics data do not for most selected measures provide the year of the measure or how the measure was constructed. For more information, see the dataset's data dictionary at https://opportunityinsights.org/wp-content/uploads/2018/04/health_ineq_online_table_12_readme.pdf.

Some data, such as those from the Census of Governments and the NCCS, offer measures beyond those in the ACS, providing additional context on government and nonprofit capacity, educational outcomes, racial and income segregation, and more.

Census of Governments

Various groups use Census of Governments data. The Census Bureau identifies two primary users at the federal level: the Bureau of Economic Analysis and the Federal Reserve System Board of Governors. It also notes that state and local governments use the data as inputs for budgeting, policy development, and financial assessment. Researchers use the data to assess changes in the government sector at a national level. The Census of Governments provides measures that can be used to assess the relative strength of governments as employers, institutions, and potential catalysts in economic development and prosperity for states, counties, and incorporated municipalities and townships. The dataset also includes special districts and school districts.

Table A.4 shows Census of Governments data for Bottineau, Hardy County, and Choctaw County. The ACS has no directly comparable measures of government capacity. The Census of Governments provides this additional input for understanding a locality's or region's capacity for economic development. In the table, we provide a calculated ratio of full-time equivalents to population to provide a comparable indicator for the relative scale of each government in its jurisdiction.

TABLE A.4

Selected Measures from Census of Governments Data

	Full-time employees	Full-time equivalencies	Monthly payroll total (full- and part-time)
Bottineau, North Dakota			
Municipal government	24	26	$107,737
Park district (special district)	0	0	$0
Rural Fire Protection District 1 (special district)	0	0	$0
Municipal airport authority (special district)	0	0	$0
Bottineau public school district	77	102	$382,209
Total	*101*	*128*	*$489,946*
Population-to-employees ratio	*86.7*		
Hardy County, West Virginia			
County government	56	64	$182,896
Moorefield/Hardy County Wastewater Authority (special district)	0	0	$0
Hardy County Public Library board (special district)	2	4	$9,143
Hardy County Public Service District (special district)	6	6	$24,541
Hardy County Rural Development Authority (special district)	2	2	$9,240
Hardy County school district	379	379	$1,201,886
Total	*445*	*455*	*$1,427,706*
Population-to-employees ratio	*214.3*		
Choctaw County, Mississippi			
County government	57	63	$183,317
Choctaw County Soil Conservation District (special district)	0	0	$0
Ackerman Choctaw County Airport board (special district)	0	0	$0
Choctaw County School District	267	277	$766,929
Total	*324*	*340*	*$950,246*
Population-to-employees ratio	*131.4*		

Source: Census of Governments.

The National Center for Charitable Statistics

NCCS data provide insight into the size, capacity, and scope of the nonprofit sector in an area. Similar to the Census of Governments data, NCCS data can be used to assess the relative strength of the nonprofit sector as an employer, institution, and potential catalyst for economic development and prosperity. Breakdowns of nonprofits by type allow for a more detailed view of the civic and social service infrastructures available in an area.

Table A.5 shows NCCS data for Hardy County, Choctaw County, the Florida zip code, and Bottineau. As with the Census of Governments, the ACS has no directly comparable measures. Relative

to the ACS, the NCCS provides additional information to understand a locality's or region's capacity for economic development and ability to leverage cultural, human, and natural capital.

TABLE A.5

Selected Measures from the National Center for Charitable Statistics

Number of nonprofit organizations by nonprofit type and nonprofit sector net assets, net income, and revenue

	Hardy County, West Virginia	Choctaw County, Mississippi	Zip code in rural Florida county	Bottineau, North Dakota
Arts, culture, and humanities	1	0	0	0
Education	4	1	0	1
Environmental quality, protection, and beautification	1	0	0	0
Health	2	0	0	3
Housing, shelter	2	0	0	0
Public safety	2	0	0	0
Recreation, sports, leisure, athletics	1	0	0	0
Youth development	0	1	1	0
Human services—multipurpose and other	2	0	0	1
Community improvement, capacity building	2	0	0	1
Philanthropy, volunteerism, and grantmaking foundations	0	0	0	1
Public, society benefit—multipurpose and other	0	0	0	1
Religion-related, spiritual development	3	1	1	0
Total nonprofits	**20**	**3**	**2**	**8**
Total net assets	$11,794,898	$38,409,763	$344,108	$9,565,678
Total net income	$3,251,245	$1,793,126	$24,398	$1,028,088
Total revenue	$21,151,263	$9,760,654	$99,783	$10,709,272

Source: National Center for Charitable Statistics.

US Financial Pulse Survey

The US Financial Health Pulse provides insights into the financial situation of US households. The survey—administered to a nationally representative, probability-based internet panel—asks households to assess their financial health, level of financial stress, income and spending patterns, saving and borrowing, and financial plans. The US Financial Health Pulse offers various ways to measure household financial vulnerability not available in other datasets, including the ACS. We selected several survey questions related to household financial health and summarized them by whether an area is "rural," "urban," or "mixed," the only geographic identifiers available in the dataset.[20]

The calculations show that respondents from all three geographic levels generally reported similar financial health trends (table A.6). Respondents from rural areas were slightly more likely to report having a "fair or poor" credit score and slightly less likely to report being able to pay all or nearly all bills on time. The US Financial Health Pulse also collects information on several characteristics that allow, for example, exploration of how these trends vary within rural areas. Those characteristics include the following:

- receipt of public benefits, including Medicaid, Medicare, Social Security, Supplemental Security Income, Social Security Disability Insurance, Temporary Assistance for Needy Families, the Supplemental Nutrition Assistance Program, the Children's Health Insurance Program, housing assistance, the earned income tax credit, and the Special Supplemental Nutrition Program for Women, Infants, and Children
- types and dollar amounts of accounts or financial assets that anyone in the household has, including checking accounts, savings accounts, employer-provided retirement accounts, individual retirement accounts, employer-provided traditional pensions, other personal savings or investments, and savings accounts
- types and dollar amounts of debt that anyone in the household has, including auto loans, student loans, small business loans, mortgages, medical debts, and outstanding credit card balances
- insurance owned, including car insurance, homeowners' insurance, renters' insurance, life insurance, short- or long-term disability insurance, and health insurance

TABLE A.6

Selected Measures from US Financial Health Pulse

Share of respondents from a survey fielded to all members of the nationally representative "Understanding America Study" between April 26, 2018, and July 4, 2018

	Rural	Mixed	Urban	Total
Satisfaction with current financial situation				
Extremely or very satisfied	29	32	31	31
Somewhat satisfied	47	46	44	46
Not very or not at all satisfied	24	22	25	24
Self-reported credit score rating				
Excellent or very good	49	53	53	52
Good	18	17	18	17
Fair or poor	28	26	25	26
Don't know	6	4	5	5
How often respondent feels finances control his or her life				
Always or often	28	29	26	28
Sometimes	34	35	36	35
Rarely or never	38	37	38	37
Stress caused by finances				
High	18	18	16	17
Moderate	26	25	28	26
Some	36	38	38	38
None	20	19	18	19
How much household income varied from month to month				
Roughly the same	64	65	64	65
Occasionally varied	27	27	27	27
Varied quite often	9	8	8	8
Household bills paid on time				
All or nearly all	81	84	84	83
Most	10	8	9	9
Some or very few	8	7	7	7
How manageable was level of debt				
Manageable amount of debt	54	56	51	54
A bit more debt than is manageable	19	20	21	20
Far more debt than is manageable	9	10	10	10
No debt	17	15	18	16
How frequently family struggled financially				
Regularly	30	28	27	28
Sometimes	39	38	36	38
Rarely	19	18	20	19
Never	9	11	11	10
Don't know	4	5	6	5

Source: US Financial Health Pulse.

Notes: Some percentages do not sum to 100 because of rounding. The analysis presented in this publication relies on data from survey(s) administered by the Understanding America Study, which is maintained by the Center for Economic and Social Research at the University of Southern California. These data are part of the US Financial Health Pulse, run by the Financial Health Network. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of USC or the Financial Health Network.

Zillow Property and Assessment Data

The Zillow Transaction and Assessment Database (ZTRAX) is a real estate database that contains information on property transactions, such as deed transfers or mortgages, as well as geographic information and characteristics of individual properties. Zillow reports that ZTRAX contains more than 400 million records across more than 2,750 counties, as well as property characteristics for about 150 million parcels across more than 3,100 counties. Property assessment files and transaction files are stored separately in ZTRAX, organized by state. The files themselves are organized by type of information. For example, transaction files include separate files for buyer name, property characteristics, and seller name. The ZTRAX publicly available data dictionary[21] details the structure of the assessment and transaction files.

Across the "main" data files for both the assessment and transaction data, we selected several measures and summarized them for the two rural counties from our six deep-dive geographies, Hardy County, West Virginia and Choctaw County, Mississippi. For the assessment data, we selected measures that detail property or land classification, average lot size in acres, and the property tax levied on the property. The property or land classification is missing for many properties across both counties; when the measure is not missing, we see that Hardy County has a number of properties classified as mobile home and as split parcels and that Choctaw County has a number of properties classified as master parcel. For the transaction data, we selected measures that detail property use, loan types, the sales price or transfer amount, and the amount of the loan. As with the property or land classification data, the property use classifications are also missing for many properties in the two counties. In each case, these high levels of missing data (more than 70 percent missing) mean the data are not useful for understanding characteristics of these rural communities. This is not a failure of Zillow or its data, but instead a result of the transaction rates for low-population areas.

Notes

[1] Amanda Gold and Yipeng Su, "Rural Communities Aren't Immune from a Census Undercount. Here's How They Can Prepare for 2020," *Urban Wire* (blog), Urban Institute, October 31, 2019, https://www.urban.org/urban-wire/rural-communities-arent-immune-census-undercount-heres-how-they-can-prepare-2020.

[2] Paul Krugman, "Getting Real About Rural America," *New York Times*, March 18, 2019, https://www.nytimes.com/2019/03/18/opinion/rural-america-economic-decline.html.

[3] David Swenson, "Most of America's Rural Areas Are Doomed to Decline," *CityLab*, May 7, 2019, https://www.citylab.com/perspective/2019/05/most-of-americas-rural-areas-are-doomed-to-decline/588883/.

[4] Brian Thiede, Lillie Greiman, Stephan Weiler, Steven C. Beda, and Tessa Conroy, "Six Charts That Illustrate the Divide between Rural and Urban America," The Conversation, March 16, 2017, https://theconversation.com/six-charts-that-illustrate-the-divide-between-rural-and-urban-america-72934.

[5] America Counts Staff, "One in Five Americans Live in Rural Areas," US Census Bureau, August 9, 2017, https://www.census.gov/library/stories/2017/08/rural-america.html.

[6] See Chapter 13, "Metropolitan Areas," in US Census Bureau, "Geographic Areas Reference Manual," last revised May 16, 2018, https://www.census.gov/programs-surveys/geography/guidance/geographic-areas-reference-manual.html.

[7] US Department of Agriculture, "Atlas of Rural and Small-Town America," last updated October 23, 2019, https://www.ers.usda.gov/data-products/atlas-of-rural-and-small-town-america/.

[8] See US Department of Agriculture, "Rural-Urban Commuting Area Codes," last updated October 24, 2019, https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx; and US Department of Agriculture, "Frontier and Remote Area Codes," last updated August 20, 2019, https://www.ers.usda.gov/data-products/frontier-and-remote-area-codes/.

[9] National Conference of State Legislatures, "Federal and State Recognized Tribes," updated October 2019, http://www.ncsl.org/research/state-tribal-institute/list-of-federal-and-state-recognized-tribes.aspx.

[10] US Census Bureau, "Census Bureau Releases Estimates of Undercount and Overcount in the 2010 Census," news release CB12-95, May 22, 2012, https://www.census.gov/newsroom/releases/archives/2010_census/cb12-95.html.

[11] See "ZTRAX Frequently Asked Questions," Zillow, accessed December 30, 2019, https://www.zillow.com/research/ztrax/ztrax-faqs/.

[12] See Opportunity Project, "Past Projects in Action," Census Bureau, accessed December 24, 2019, https://opportunity.census.gov/projects/.

[13] "Office of the CTO," US Department of Health and Human Services, accessed December 24, 2019, https://www.hhs.gov/cto/index.html.

[14] Meta S. Brown, "States Offer Information Resources: 50+ Open Data Portals," Forbes, April 30, 2018, https://www.forbes.com/sites/metabrown/2018/04/30/us-states-offer-information-resources-50-open-data-portals/#637f8dcd5225.

[15] Graham MacDonald and Kathy Pettit, "Announcing the Urban Institute Data Catalog," *Data@Urban* (blog), Urban Institute, September 10, 2019, https://medium.com/@urban_institute/announcing-the-urban-institute-data-catalog-ce2c787e38e9.

[16] Amanda Gold and Yipeng Su. "Rural Communities Aren't Immune from a Census Undercount. Here's How They Can Prepare for 2020."

[17] Persistent poverty counties are defined by USDA's Economic Research Service as counties where 20 percent or more of county residents were poor (living at or below the official federal poverty line), measured across at least three consecutive decades. See "Rural Poverty & Well-Being," USDA, accessed February 17, 2020, https://www.ers.usda.gov/topics/rural-economy-population/rural-poverty-well-being/.

[18] The analysis presented in this publication relies on data from survey(s) administered by the Understanding America Study, which is maintained by the Center for Economic and Social Research at the University of Southern California. These data are part of the US Financial Health Pulse, run by the Financial Health Network. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the University of Southern California or the Financial Health Network.

[19] Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information on accessing the data can be found at www.zillow.com/ztrax. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group.

[20] The analysis presented in this publication relies on data from survey(s) administered by the Understanding America Study, which is maintained by the Center for Economic and Social Research at the University of Southern California. These data are part of the US Financial Health Pulse, run by the Financial Health Network. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of USC or the Financial Health Network.

[21] See "ZTRAX Frequently Asked Questions," Zillow, accessed December 30, 2019. https://www.zillow.com/research/ztrax/ztrax-faqs/.

References

Beaulieu, Lionel J. 2014. "Promoting Community Vitality and Sustainability: The Community Capitals Framework." West Lafayette, IN: Purdue University Center for Regional Development.

Brumfield, Cara. 2018. "Why the 2020 Census Matters for Rural America: Defining, Understanding, and Investing in Rural Communities." Washington, DC: Georgetown Law Center on Poverty and Inequality.

Derian, Alexandra. 2016. "People and Place Matter: Using Integrated Data Systems to Understand Chronic Absenteeism." Washington, DC: Urban Institute.

Goetz, Stephan J., Mark Partridge, Steven C. Deller, and David A. Fleming. 2010. "Evaluating U.S. Rural Entrepreneurship Policy." *Journal of Regional Analysis and Policy* 40 (1): 20–33.

Housing Assistance Council. 2010. "Native American Lands" in *Taking Stock: Rural People, Poverty, and Housing in the 21st Century*. Washington, DC: Housing Assistance Council.

Johnson, Kenneth M. 2018. "Domestic Migration and Fewer Births Reshaping America." Durham, NH: University of New Hampshire Carsey School of Public Policy.

Kretzmann, John P., and John L. McKnight. 1993. *Building Communities from the Inside Out: A Path toward Finding and Mobilizing a Community's Assets*. Chicago: ACTA Publications.

Lichter, Daniel T. 2012. "Immigration and the New Racial Diversity in America." *Rural Sociology* 77 (1): 3–35.

National Congress of American Indians. 2018. "The State of Tribal Data Capacity in Indian Country: Key Findings from the Survey of Tribal Data Practices." Washington, DC: National Congress of American Indians Policy Research Center.

O'Hare, William P. 2017. "2020 Census Faces Challenges in Rural America." Durham, NH: University of New Hampshire Carsey School of Public Policy.

Schmidt, Ryan W. 2011. "American Indian Identity and Blood Quantum in the 21st Century: A Critical Review." *Journal of Anthropology* (2011), Article ID 549521.

Theodos, Brett, and Eric Hangen. 2017. "Expanding Community Development Financial Institutions." Washington, DC: Urban Institute.

US Census Bureau. 2017. *Uses of Census Bureau Data in Federal Funds Distribution: A New Design for the 21st Century*. Washington, DC: US Census Bureau.

USDA (US Department of Agriculture). 2018. "Rural America at a Glance: 2018 Edition." Economic Information Bulletin 200. Washington, DC: USDA, Economic Research Service.

About the Authors

Corianne Payton Scally is a principal research associate in the Metropolitan Housing and Communities Policy Center at the Urban Institute, where she explores housing as a platform for opportunity and community and economic development in urban and rural places. From 2015 to 2016, she led data and research initiatives at the US Department of Agriculture's Rural Housing Service. Scally received her PhD in urban planning and policy development from Rutgers University.

Eric Burnstein is a research associate in the Metropolitan Housing and Communities Policy Center, where his research focuses on regional inequality, governance, and the built environment. Burnstein received his MUP and MPP from the University of Michigan.

Matthew Gerken is a research analyst in the Metropolitan Housing and Communities Policy Center. His areas of interest include affordable housing, homelessness, and community development. Gerken holds a BS in mathematics and a BA in nonprofit administration from Stetson University and an MPP from Duke University.

Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.

URBAN
INSTITUTE · ELEVATE · THE · DEBATE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org