

## Problem of Low 2020 Census Participation Will Vary with Sociodemographic Factors and Distance from Metro Areas

Zheng Tian, Stephan J. Goetz, and Charlie French

*JEL Classifications: C20, Y10*

*Keywords: Broadband Internet access, Decennial census, Low response score, Non-metro counties, Sociodemographic factors*

### Introduction

With billions of federal grant dollars potentially at stake, every community has a vested interest in ensuring that its residents are accurately counted in the U.S. Decennial Census of Population and Housing. In the 2010 Census, 20.7% of eligible households failed to return their census forms, implying a response rate of only 79.3%. That amounts to about 22 million households not reached in the last census, the number of which not only affects the quality of the census but also may lead businesses and government officials to make inaccurate decisions when targeting specific populations.

The goal for the 2020 survey is to raise this response rate significantly through outreach and by using on-line survey forms, which rely heavily on broadband Internet access. Researchers have identified key sociodemographic factors associated with low participation in the census. However, differences in how these factors affect responses across metropolitan and non-metropolitan counties have not been adequately addressed. Lack of sufficient broadband Internet in rural areas could make the nonresponse problem even worse. Knowing these factors and the urban–rural differences provides a basis for selecting communities that would benefit from additional outreach to help improve census participation. We find that the effects of race, housing, and other characteristics—such as marital status and even Internet access—on census participation show subtle and sometimes surprising differences depending on whether the non-metro county is adjacent to a metro county.

### What Is the Low Response Score (LRS)?

The Census Low Response Score (LRS) identifies places where populations were difficult to enumerate in the 2010 Census as “block groups and tracts whose characteristics predict low census mail return rate and are highly correlated (negatively) with census and survey participation” (U.S. Census Bureau, 2019, p. 4). The LRS uses a statistical model to predict how far the actual return rate falls below 100% using 25 socioeconomic and demographic variables.<sup>1</sup> The first version of the LRS was computed using mail responses to the 2010 Census and data from the 2008–2012 American Community Survey (ACS). The earlier LRS was then updated using more current explanatory variables to predict where low responses would be a problem in the 2020 Census. These data can be downloaded from the Census Bureau’s Planning Database (<https://www.census.gov/topics/research/guidance/planning-databases.html>). The U.S. Census Bureau also provides the Response Outreach Area Mapper (ROAM) (<https://www.census.gov/roam>), an interactive web mapping tool that allows users to zoom in to the tract level.

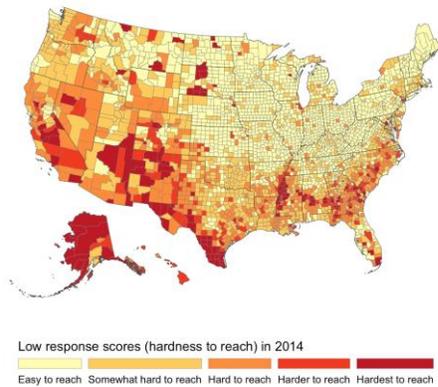
### Where Is the LRS Especially Low?

To compare the LRS across different types of counties, we aggregate the original tract-level LRS to the county level. The county-level LRS is the average of the LRS of all tracts in a county, weighted by the number of households in each tract. We also rank the counties by LRS, using these categories to show increasing difficulty of participation: “easy to reach” (the top 50% of easiest-to-reach counties), “somewhat hard to reach” (the next 25%), “hard to reach” (the next 15%), “harder to reach” (next 5%), and “hardest to reach” (top 5% of counties with worst participation rates). Figure 1 shows for the

<sup>1</sup> See Erdman and Bates (2017) for details of the construction of the LRS.

2014 LRS that most hard-to-reach counties lie in the South, especially in Texas, followed by Mississippi and Georgia. Many of the counties that are the hardest to reach are those where the majority of population are Hispanic, black Americans, or Native Americans, according to a report published by the Pew Research Center (Shaeffer, 2019). As for regional differences in the average LRS, the West has the worst score (21.0), ahead of the South (20.3); the Northeast (18.3) and Midwest (17.4) have the lowest or best LRS (17.4).

**Figure 1. Low Response Scores, 2014**



Notes: The colors from light to dark represent the percentile categories of low response scores. Source: 2014 Planning Database of the U.S. Census Bureau and authors' compilation.

**Table 1. Count and Proportion of Counties for Each LRS Category, 2014**

	Metro Counties	Rural, Metro-Adjacent Counties	Rural, Non-Metro-Adjacent Counties
Easy to reach	545 (46.7%)	515 (50.1%)	512 (54.0%)
Somewhat hard to reach	342 (29.3%)	245 (23.9%)	198 (20.9%)
Hard to reach	179 (15.3%)	163 (15.9%)	129 (13.6%)
Harder to reach	55 (4.71%)	55 (5.36%)	47 (4.95%)
Hardest to reach	46 (3.94%)	49 (4.77%)	63 (6.64%)
Total	1167 (100%)	1027 (100%)	949 (100%)

Notes: We categorize the LRS by the 0th–50th, 50th–75th, 75th–90th, 90th–95th, and 95th–100th percentiles intervals, and label the categories as “easy to reach,” “somewhat hard to reach,” “hard to reach,” “harder to reach,” and “hardest to reach,” respectively. Parentheses indicate the proportion of counties in each category along the county types.

## Do Rural Counties Have A Worse LRS?

To answer this question we use the county typology prepared by the USDA’s Economic Research Service, known as the 2013 Rural–Urban Continuum Code (RUCC) (<https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/>). For ease of analysis, we separate the nine levels of RUCC into metro (RUCC 1–3) and rural counties (RUCC 4–9). Rural counties are further classified as rural counties adjacent to a metro county (RUCC 4, 6, and 8), and rural counties not adjacent to a metro county (RUCC 5, 7, and 9). We label these as rural, metro-adjacent and rural, non-metro-adjacent, respectively.

A simple comparison of the average 2014 LRS across the three types of counties does not reveal a statistically significant difference: The LRS ranges from 19.15% to 19.32%. However, a more refined analysis reveals that the shares of “hardest to reach” counties are highest in rural, non-metro-adjacent counties (6.64%), followed by rural, metro-adjacent counties (4.77%) and metro counties (3.94%). Below we investigate this in more detail.

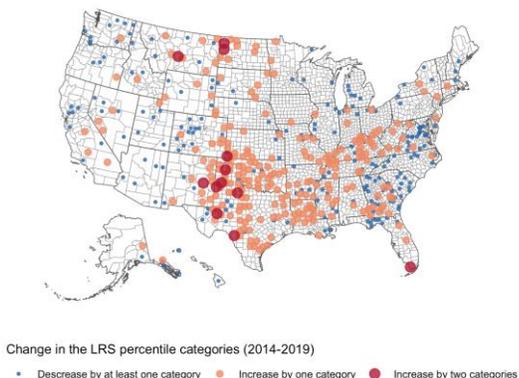
## Where Is the LRS Expected to Worsen in 2020 Compared to 2010?

The Census Bureau published the 2019 LRS with the release of the 2013–2017 ACS, which allows us to predict changes in LRS as we approach the 2020 Census. In particular, we can tell where the LRS is likely to have improved and, more importantly, where it has likely worsened as local demographic factors have changed. We compare each county’s 2014 and 2019 LRS rankings; specifically, we determine whether it moved up or down in the ranking and, if it moved up (worse LRS), where it jumped by one or two categories. Figure 2 reveals that counties in the South are at increased risk of receiving lower survey responses in the 2020 Census, of which Texas, Oklahoma, and Kentucky are the top three states. Conversely, several counties (blue dots) across the nation are expected to improve their LRS in 2020.

The contrast in the changes between 2010 and 2020 in the LRS in metro and rural counties is notable (Table 2). Although small (less than 1%), the share of counties that are expected to worsen their LRS is greater for rural than for metro counties, and the situation is considerably worse for rural, non-metro-adjacent counties (0.95%) than it is for rural, metro-adjacent counties (0.29%). No metro county experienced an increase in the LRS by two

ranking categories (such as from hard to reach to hardest to reach).

**Figure 2. Change in LRS, 2014-2019**

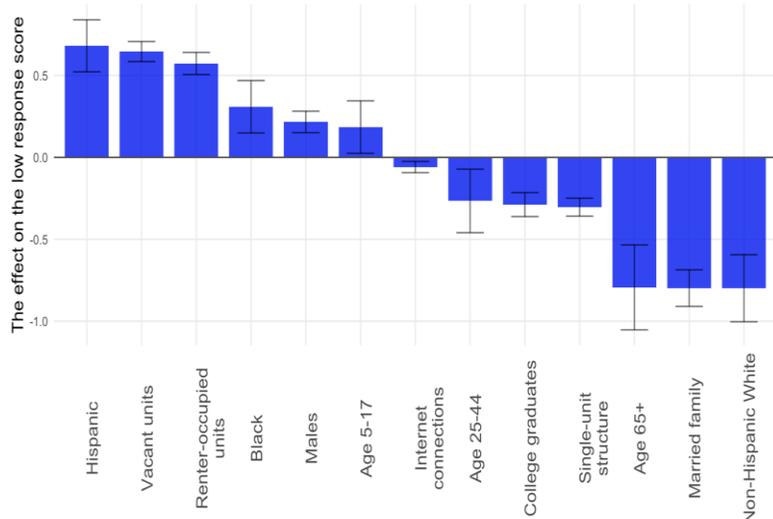


Notes: Orange and red dots indicate an increase of the low response scores by one and two categories, respectively; blue dots indicate a decrease by at least one category. Counties that show no change in categories are omitted. Source: 2014 and 2017 Planning Database of the Census Bureau and authors' compilation.

## What Socioeconomic Variables Influence the LRS?

Overall, the strongest predictors of low response scores are race- and housing-related. In particular, higher shares of Hispanic and black populations, as well as vacant and renter-occupied housing units, are strongly

**Figure 3. Independent Impacts of the Top 13 Factors on the Low Response Scores**



Notes: The y axis represents the magnitude of how much the score would increase when a variable on the x axis increases by one standard deviation from the mean. The error bar represents the 95% confidence interval.

**Table 2. Count and Proportion of Counties for Each Change in the LRS, 2014–2019**

	Metro Counties	Rural, Metro-Adjacent Counties	Rural, Non-Metro-Adjacent Counties
Decrease in the LRS percentile category	94 (8.06%)	69 (6.73%)	82 (8.65%)
No change in the LRS percentile category	1,017 (87.2%)	881 (85.9%)	735 (77.5%)
Increase by one LRS percentile category	55 (4.72%)	73 (7.12%)	122 (12.9%)
Increase by two LRS percentile categories	0 (0%)	3 (0.292%)	9 (0.949%)
Total	1,166 (100%)	1,026 (100%)	948 (100%)

Notes: Percentile categories are defined in Table 1.

associated with a lower response score on average. In contrast, places with higher shares of elderly (65 years and older), married family households, and non-Hispanic whites have lower low response scores (i.e., populations in these counties are more likely to be counted in the census).

Figure 3 presents the top six variables (out of 25) that have an independent effect in terms of increasing LRS, the top six variables that do the opposite, and the variable of Internet connections. The height of the bars represents the estimated effect of each variable on the LRS, and the error bars are the 95% confidence interval.<sup>2</sup>

## Do Effects of the Sociodemographic Variables also Differ by County Type?

To answer this question, we look at whether each variable has a varying effect on the actual 2010 Census mail nonresponse rate across county types, independent of the other variables considered. We find a few subtle differences between metro, rural, metro-adjacent, and rural, non-metro-adjacent counties. To put the following discussion into context, the 2010 Census targeted 110 million valid household addresses nationwide, of which 94 million households are in metro counties, 11 million are in rural, metro-adjacent counties, and 5 million are in rural, non-metro-adjacent counties. So, a 1-percentage-point increase in the mail

<sup>2</sup> See the appendix for a description of the regression model used to rank the variables.

nonresponse rate is equivalent to about 1 million households nationwide that are surveyed, of which 940,000 are in metro counties, 110,000 are in rural, metro-adjacent counties, and 50,000 are in rural, non-metro-adjacent counties. Although there are fewer households in rural counties than in metro counties, the cost of reaching out to rural families would be higher.

### **Race**

In metro counties, the mail nonresponse rate rises 0.13 percentage point for each 1-percentage-point increase in Hispanic population. Rural, metro-adjacent and rural, non-metro-adjacent counties experience additional increases of 0.02 (for a total of 0.15) and 0.02 (for a total of 0.17) percentage points. The difference between metro counties and rural, non-metro-adjacent counties is statistically significant. Thus, if the concern is to ensure more complete population counts, resources should be targeted first to rural, non-metro-adjacent counties. One additional percentage point of black populations would increase the nonresponse rate by 0.14 percentage point, which is basically the same for both types of rural counties. In contrast, a higher share of non-Hispanic white population reduces the nonresponse rate by 0.14, 0.18, and 0.20 percentage points, respectively, in metro, rural, metro-adjacent, and rural, non-metro-adjacent counties.

### **Housing**

Metro counties with an increase of 1 percentage point in vacant units have an increase in mail nonresponse rate by 0.11 percentage points. However, there are no statistically significant differences in this effect between metro and rural counties, regardless of adjacency status. A higher share of renter-occupied units would cause the nonresponse rate to increase by 0.18, 0.25, and 0.31 percentage points in metro, rural, metro-adjacent, and rural, non-metro-adjacent counties, respectively; these differences are significant. As opposed to vacant and renter housing, the presence of single-unit housing lowers the nonresponse rate; its effect is bigger in rural counties than in metro counties. As pointed out in the 2020 Census Operational Plan (U.S. Census Bureau, 2017, p. 8), some tactics would be employed to identify vacant households, but, more importantly, more resources should be used to increase visits to renter households, especially in rural areas.

### **Internet Access**

Internet access is measured as the share of households with broadband Internet, which is associated with an improvement in the census responses in all county types. Internet access is important because the census will move away from mail surveys in 2020. In particular, the 2020 Census will be “encouraging the population to respond to the 2020 Census using the Internet, reducing the need for more expensive paper data capture” (U.S. Census Bureau, 2017, p. 15). Internet access reduces

nonresponse rates in all counties. In metro counties, an additional 1-percentage point increase in the share of households with broadband would reduce the nonresponse rate by 0.07 percentage point. In rural counties, it provides an even greater benefit, reducing the nonresponse rate by an additional 0.11 percentage points in rural, metro-adjacent counties and 0.12 percentage points in rural, non-metro-adjacent counties. This underscores the critical importance of broadband access to ensuring an accurate and representative count of the population in 2020.

### **Other Sociodemographics**

A few other variables also stand out. Higher education plays a positive role in improving survey responses. A higher share of college graduates would lower the nonresponse rate by 0.06 percentage point; the effect is significantly larger in rural, metro-adjacent counties (0.19 percentage points) and rural, non-metro-adjacent counties (0.15 percentage points). In contrast, higher shares of populations who are not high school graduates are associated with higher nonresponse rates. Marriage status is also important. The greater presence of households with single persons or a female head but no husband relative to married couples lowers survey responses. Most aforementioned factors associated with lower response rates seems to be related to poverty. Indeed, we find that a higher share of households below the poverty line would increase the nonresponse rate by 0.42 percentage points in metro counties and 0.51 and 0.43 percentage points in rural, metro-adjacent and rural, non-metro-adjacent counties, respectively.

### **Conclusion**

Our findings suggest that demographic factors, as well as geographic and household characteristics, play a significant role in census participation. These factors include housing vacancy rates, race, Internet access, education level, and marriage status. Certain of these factors have statistically different impacts across county types—metro, rural, metro-adjacent, and rural, non-metro-adjacent—and thus could help to inform Low Response Score projections.

As is the case with many socioeconomic processes and concerns, the devil is often in the details. Given the stated goal of counting more of the population in 2020, scarce public resources will have to be deployed strategically to communities where under-participation problems are especially pronounced. Strong predictors of participation include race, housing arrangements, and other sociodemographic variables such as poverty rates. Further complicating the impacts of these factors is the fact that their importance varies across the metro-rural county spectrum, as a function of distance from or adjacency to metro areas.

Perhaps more importantly, this research raises new questions for further inquiry. What policies or

mechanisms could help boost census participation? And, if targeted policies and incentives are put in place, or

shown to be currently effective, what system could help to assess whether they are working?

## For More Information

Erdman, C., and N. Bates. 2017. "The Low Response Score (LRS): A Metric to Locate, Predict, and Manage Hard-to-Survey Populations." *Public Opinion Quarterly* 81(1):144–156.

Schaefer, K. 2019. "In a rising number of U.S. counties, Hispanic and black Americans are the majority." Pew Research Center. Available online: <https://www.pewresearch.org/fact-tank/2019/11/20/in-a-rising-number-of-u-s-counties-hispanic-and-black-americans-are-the-majority/#counties>

U.S. Census Bureau. 2017. *2020 Census Operational Plan: A New Design for the 21<sup>st</sup> Century*. Washington, D.C.: U.S. Census Bureau. Available online: <https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan3.pdf>.

U.S. Census Bureau. 2019a. *Planning Database with 2010 Census and 2013–2017 American Community Survey Data at the Tract Level*. Washington, D.C.: U.S. Census Bureau. Available online: [https://www.census.gov/content/dam/Census/topics/research/2019\\_Tract\\_PDBDocumentationV3.pdf](https://www.census.gov/content/dam/Census/topics/research/2019_Tract_PDBDocumentationV3.pdf).

U.S. Department of Agriculture. 2013. *Rural-Urban Continuum Codes*. Washington, DC: U.S. Department of Agriculture, Economic Research Service. Available online: <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/>.

## Appendix

### Descriptive Statistics of the 2014 Low Response Scores

Table A1 shows the descriptive statistics of the 2014 LRS of all counties and across three types of counties. The average Low Response Scores are not statistically significant across three types of counties.

### Regression Analysis to Generate Figure 3

The regression model uses the county-level 2019 LRS as the dependent variable. The independent variables include the dummy variables representing county types and regions and the 25 variables used to construct the LRS at the census-tract level. The goal of the regression is not to repeat the practice of constructing the LRS but to examine differences in LRS across county types when all other determining variables are controlled for. As a by-product, we reevaluate the importance of the 25 variables to the LRS at the county level. To do so, we standardize these variables by their z-scores, which is the original value minus its average, divided by its standard deviation. The coefficient on each of them can be interpreted as how much the LRS would change when X increases by 1 standard deviation from the mean. Table A2 reports the results of the baseline regression model. The top six variables with positive coefficients and the top six with negative coefficients, plus internet connections, are presented in Figure 3.

**Table A1. Descriptive Statistics of the Low Response Scores**

Counties	Mean	Std	Min	25th	50th	75th	Max
All	19.27	2.97	8.17	17.11	18.64	21.11	34.19
Metro	19.32	2.72	8.17	17.42	18.87	21.00	32.45
Rural, metro-adjacent	19.32	2.88	13.28	17.20	18.62	21.21	34.19
Rural, non-metro-adjacent	19.15	3.34	12.59	16.70	18.32	21.21	33.06

**Table A2. Baseline Regression Results****Dependent Variable: 2019 LRS**

Constant	18.994*** (0.042)
Non-metro: adjacent	0.271*** (0.041)
Non-metro: not adjacent	0.444*** (0.052)
Northeast	-0.159** (0.071)
South	0.659*** (0.048)
West	0.755*** (0.081)
Renter-occupied units	0.573*** (0.035)
Vacant units	0.646*** (0.031)
Age 18–24	0.087 (0.098)
Female head, no husband	-0.123*** (0.047)
Non-Hispanic white	-0.799*** (0.104)
Age 65+	-0.794*** (0.132)
Related child	0.115*** (0.036)
Males	0.216*** (0.033)
Married family	-0.798*** (0.057)
Age 25–44	-0.266*** (0.099)
College graduates	-0.288*** (0.037)
Age 45–64	-0.077 (0.089)
Persons per household	0.165*** (0.050)
Moved-in household	-0.086** (0.038)
Hispanic	0.681*** (0.081)
Single-unit structure	-0.304*** (0.028)
Population density	-0.002 (0.047)
Below poverty	0.003 (0.042)
Different housing unit 1 year ago	-0.163*** (0.028)
Age 5–17	0.185** (0.082)
Black	0.309*** (0.082)
Single-person households	-0.212*** (0.041)
Not high school graduate	-0.232*** (0.038)
Median household income	-0.135** (0.059)
Median house value	0.036 (0.046)
Internet connections	-0.059*** (0.018)
<i>R</i> <sup>2</sup>	0.939
Adj. <i>R</i> <sup>2</sup>	0.938
Num. obs.	3086

Notes: Single, double, and triple asterisks (\*, \*\*, \*\*\*) indicate significance at the 10%, 5%, and 1% levels, respectively. The independent variables are standardized by z-scores.

**Author Information:** Zheng Tian ([zzt36@psu.edu](mailto:zzt36@psu.edu)) is Postdoctoral Scholar, Northeast Regional Center for Rural Development, Department of Agricultural Economics, Sociology, and Education, Pennsylvania State University, University Park, PA. Stephan J. Goetz ([sig16@psu.edu](mailto:sig16@psu.edu)) is Professor, Northeast Regional Center for Rural Development, Department of Agricultural Economics, Sociology, and Education, Pennsylvania State University, University Park, PA. Charlie French ([Charlie.French@unh.edu](mailto:Charlie.French@unh.edu)) is Extension Associate Professor in Community and Economic Development, University of New Hampshire, Durham, NH.

**Acknowledgments:** We thank anonymous *Choices* reviewers and Amy Bekkerman for valuable edits, comments and suggestions. In addition, this work was supported in part through funding from NIFA/USDA as well as the Pennsylvania State University Agricultural Experiment Station. Any opinions expressed are those of the authors.

---

©1999–2020 CHOICES. All rights reserved. Articles may be reproduced or electronically distributed as long as attribution to Choices and the Agricultural & Applied Economics Association is maintained. Choices subscriptions are free and can be obtained through <http://www.choicesmagazine.org>.