



# HHS Public Access

Author manuscript

*Cancer Epidemiol Biomarkers Prev.* Author manuscript; available in PMC 2020 March 01.

Published in final edited form as:

*Cancer Epidemiol Biomarkers Prev.* 2019 September ; 28(9): 1409–1416. doi:  
10.1158/1055-9965.EPI-19-0021.

## Providing higher resolution indicators of rurality in the Surveillance, Epidemiology, and End Results (SEER) database: Implications for patient privacy and research

Jennifer L. Moss<sup>1,2</sup>, David G. Stinchcomb<sup>3</sup>, Mandi Yu<sup>4</sup>

<sup>1</sup>Cancer Prevention Fellowship Program, National Cancer Institute, Bethesda, MD 20892, USA;

<sup>2</sup>Department of Family and Community Medicine, Penn State College of Medicine, Hershey, PA 17033, USA;

<sup>3</sup>Westat, Inc., Rockville, MD 20850, USA;

<sup>4</sup>Surveillance Research Program, National Cancer Institute, Bethesda, MD 20892 USA

### Abstract

**Background.**—The burden of cancer is higher in rural areas than urban areas. The National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) database currently provides county-level information on rurality for cancer patients in its catchment area, but more nuanced measures of rurality would improve etiologic and surveillance studies.

**Methods.**—We analyzed disclosure risk and conducted a sample utility analysis of census tract-level measures of rurality, using (1) U.S. Department of Agriculture’s Rural Urban Commuting Area (RUCA) codes and (2) U.S. Census data on percent of the population living in non-urban areas. We evaluated the risk of disclosure by calculating the percentage of census tracts and cancer cases that would be uniquely identified by a combination of these two rurality measures with a census tract-level socioeconomic status (SES) variable. The utility analyses examined SES disparities across levels of rurality for lung and breast cancer incidence and relative survival.

**Results.**—Risk of disclosure was quite low: <0.03% of census tracts and <0.03% of cancer cases were uniquely identified. Utility analyses demonstrated an SES gradient in lung and breast cancer incidence and survival, with relatively similar patterns across rurality variables.

**Conclusions.**—The RUCA and Census rurality measures have been added to a specialized SEER 18 database. Interested researchers can request access to this database to perform analyses of urban/rural differences in cancer incidence and survival.

**Impact.**—Such studies can provide important research support for future interventions to improve cancer prevention and control.

---

**Corresponding author:** Jennifer L. Moss, PhD, Department of Family and Community Medicine, Penn State College of Medicine, 134 Sipe Ave., #205, P.O. Box 850, MC HS72, Hershey, PA 17033, Phone: 1-717-531-0003 × 323429, Fax: 1-717-531-3696, jmoss1@pennstatehealth.psu.edu.

Conflict of interest disclosure:

The authors declare no potential conflicts of interest.

## Keywords

cancer; registry; surveillance; epidemiology; disparities; Rural; urban; socioeconomic status (SES); confidentiality; geography

People living in rural areas of the United States (US) are facing a public health crisis<sup>1</sup>, with lower life expectancy<sup>2</sup> and higher rates of chronic diseases<sup>3-5</sup> than their counterparts in more urban areas. For example, cancer mortality rates are at least 8% higher in the most rural compared to the most urban areas.<sup>6-8</sup> In addition, rural areas are characterized by elevated cancer risk based on demographic (e.g., older median age, lower socioeconomic status<sup>5</sup>), behavioral (e.g., higher tobacco use<sup>9</sup>, lower uptake of cancer screening<sup>10-12</sup>), and diagnostic (e.g., later stage of diagnosis for colorectal<sup>13,14</sup>, lung<sup>14</sup>, and cervical<sup>15</sup> cancers) factors, suggesting that cancer disparities will persist into the future. About 14% of the US population lives in rural (non-metro) counties (using the definition from the Office of Management and Budget)<sup>16</sup>, so researching cancer burden in these areas could have great impacts on population health. Increasingly, rural health has become a priority for national research agencies, including the National Cancer Institute (NCI), but additional infrastructure is needed to support epidemiologic and surveillance work around cancer in rural areas.

A key challenge facing cancer surveillance and rural health research is selecting among various definitions and measures of rurality<sup>17-19</sup> (see also: <https://www.ruralhealthinfo.org/topics/what-is-rural>). Rural environments can be characterized as areas having agricultural activity, low population density, or limited access to public and private sector services, with each definition having different implications for health research.<sup>20</sup> In addition, the geographic scale of measures of rurality is an important consideration. County-level measures of rurality are used in many US studies<sup>1,18,21,22</sup>, and these measures have been available in cancer databases, including the NCI's Surveillance, Epidemiology, and End Results (SEER) database, for some time.<sup>23</sup> However, many US counties include both urban and rural areas, and studies done at the county level can, therefore, suffer from classification bias.<sup>21</sup> Several measures of rurality are available at sub-county levels, such as census tracts.<sup>18,21,24,25</sup> Making tract-level measures of rurality more readily available for health research will provide analytic benefits including reducing misclassification and increasing potential to identify significant urban-rural health disparities.

Despite the analytic benefits, including detailed census tract-based rurality information in the SEER cancer incidence database can confer a substantial risk of identifying individual patients. Already, a census tract-based socioeconomic status (SES) quintile<sup>26,27</sup> was added to the SEER incidence database for studying SES disparities in cancer incidence and survival.<sup>23</sup> Adding rurality information at the census tract-level would provide valuable opportunities for studying intertwined effects of SES and rurality on cancer outcomes, but would concomitantly increase the risk of deductive disclosure. In this paper, we describe the addition of two indicators of census tract-level rurality to NCI's SEER incidence database<sup>23</sup>, demonstrate their utility for research on geographic disparities in cancer burden, and outline the procedures in place to maintain patient confidentiality. These variables improve upon

existing resources in SEER by offering more detailed insight into the relationship between rurality and cancer risks.

## Materials and Methods

We evaluated adding two tract-level measures of rurality to the SEER cancer cases using the census tract of residence at the time of diagnosis: one based on the US Department of Agriculture (USDA)'s Rural Urban Commuting Area (RUCA) codes (<https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx>), and one based on the Census Bureau's percent of the population living in non-urban areas (<https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html>) (Table 1). The RUCA measure reflects proximity to a large urban center and, thus, may be most relevant for studies that focus on access to care, e.g., cancer treatment and survival studies. There are 10 primary RUCA codes that range from 1 ("metropolitan area core: primary flow within an urbanized area (UA)") to 10 (rural areas: primary flow to a tract outside a UA or urban cluster (UC)"); within each primary code, there are additional secondary codes (total secondary codes=23) that describe the commuting patterns. The Census measure reflects the rural nature of the immediate environment and may be most relevant for studies that focus on behaviors and risk, e.g., cancer prevention and screening studies; the original data from Census are population counts that result in continuous percentages, which are categorized based on predefined cutoffs.

### Tests of uniqueness to assess impact on confidentiality

To understand the implications of adding these two tract-level measures to the SEER cancer incidence database on patient confidentiality, we sought to assess the potential change in risk of identification. Our analyses assumed that the novel rurality measures would be made available in conjunction with the existing measure of SES and that no additional geographic area identification would be available when releasing rurality data. We evaluated combinations of (1) two- or four-category versions of the RUCA-based measure (<https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx>) and (2) two-, three-, or four-category versions of the Census-based measure (<https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html>), in addition to (3) the quintile SES measure<sup>26</sup> (Table 1). Under each of these conditions, we calculated the number of uniquely-identified census tracts in SEER areas, excluding Alaska, and then estimated the percent of cancer cases in SEER that were located in these tracts. We excluded Alaska because of special confidentiality constraints for the Alaska Native Tumor Registry.

In addition, we examined the degree of difference in classification of census tract using the two rurality measures. Specifically, we calculated the percent of cases in census tracts that had discordant classifications, i.e., were either (1) classified as "urban" by the RUCA measure and as "mostly rural"/"all rural" by the Census measure, and (2) classified as "rural" by the RUCA measure and as "mostly urban"/"all urban" by the Census measure. We calculated this indicator of mismatch across cancer sites (i.e., lung, breast), by diagnosis year (i.e., 2005–2014), and by registry (i.e., among the SEER 18 registries).

## Demonstration of utility of rurality variables for research

Next, we undertook a brief analysis to demonstrate the utility of these variables for studies of geographic disparities in cancer. The rurality variables were linked to the incidence database, which also allows for survival analysis by examining outcomes among the incident cancer cases. To calculate age-standardized rates, the incidence database uses annual population estimates derived from the U.S. Census Bureau's Population Estimates Program in collaboration with the National Center for Health Statistics (for more details, see <https://seer.cancer.gov/popdata/methods.html>). We examined urban/rural differences in the socioeconomic gradient in incidence and survival of lung and breast cancers (separately). As others have demonstrated, lung cancer burden is greater in areas that are rural<sup>6,7</sup> and low SES<sup>28</sup> than in other areas, potentially due to individual-level differences in tobacco use.<sup>9,29</sup> In contrast, breast cancer incidence is higher in areas that are urban<sup>6,7</sup> and high SES<sup>30</sup> than in other areas, potentially due to lifestyle factors and access to mammography screening.<sup>10,31</sup> However, few studies have examined the interaction of area-level rurality and SES in their association with these cancer outcomes.

**Data sources and variables.**—Cancer incidence and relative survival data came from the SEER 18 database, which covers about 28% of the US population (<https://seer.cancer.gov/>). Age-standardized cancer incidence rates in SEER 18 were calculated for 2005–2014, expressed as cases per 100,000 people (lung cancer) or per 100,000 women (female breast cancer). Five-year relative survival was calculated as the percent of patients diagnosed between 2005 and 2009 whose vital status was “alive” by sixty months after their diagnosis. We used SEER\*Stat software<sup>23</sup> to generate estimates and standard errors of incidence and survival across census tract characteristics (separately for lung and breast cancer).

We examined differences in these outcomes across tract-level SES quintile<sup>26</sup>, as well as across rurality categories. We categorized SES into five quintiles with Q1 indicating lowest SES and Q5 indicating highest SES. The RUCA-based measure was categorized into two levels: “urban” (metropolitan area with commuting flow to urbanized areas); or “rural” (micropolitan, small town, or rural area with commuting flow to urbanized clusters or to non-urbanized areas) (<https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx>). The Census-based measure was categorized into four levels: “all urban” (100% of population living in urban areas); “mostly urban” ( 50% but <100% in urban areas); “mostly rural” (>0% but <50% in urban areas); or “all rural” (0% in urban areas) (<https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html>).

**Statistical analysis.**—To examine disparities in cancer burden across these census tract groups, we imported incidence and survival data into HD\*Calc (<https://seer.cancer.gov/hdcalc/>), a program supported by the NCI that generates summary measures of disparities for population-based health data. We extracted estimates and standard errors of four commonly-used disparities measures<sup>32,33</sup> comparing cancer burden in different SES groups across categories of rurality: (1) range difference (RD), the arithmetic difference between cancer rates in the lowest versus highest SES groups; (2) absolute concentration index (ACI), which summarizes the concentration of cancer burden among all SES groups on the

absolute scale; (3) range ratio (RR), the cancer rate for the lowest SES group divided by the cancer rate for the highest SES group; and (4) relative concentration index (RCI), which is a relative version of the ACI, using the average cancer burden in the population as the reference. Additional details on the calculation of these variables are available (<https://seer.cancer.gov/hdcalc/>).<sup>34</sup>

## Results

### Tests of disclosure risk to assess impact on confidentiality

The risks of identification resulting from adding the census tract-level rurality measures are quite small. Across combinations of SES and rurality measures, fewer than 5 out of 18,909 census tracts in relevant SEER catchment areas were uniquely identified by combinations of these variables (Table 2). The risk of identification was quite small: across combinations, <0.03% of census tracts and <0.03% of cases were uniquely identified. Using the proposed categorization scheme (combining five categories of SES, two RUCA-based rurality categories, and four Census-based rurality categories), only one census tract was uniquely identified (0.005% of tracts and 0.004% of cases).

**Mismatch of rurality measures.**—Overall, 91.0% of cases lived in census tracts that were classified as urban using the RUCA measure, and 86.7% lived in census tracts classified as urban using the Census measure (Figure 1). Across SEER registries, the disagreement of rurality classifications ranged from 1.8% in Detroit to 34.8% in Iowa (Figure 1). Across cases, 12.4% lived in census tract that had different classifications for the RUCA-based rurality variable versus the Census-based rurality variable. As a result, 13.8% of lung cancer cases and 10.5% of breast cancer cases were in census tracts that had different rurality classifications using the two different measures. Over the study years, mismatch of rurality for lung and breast cancer cases was relatively similar, ranging from 11.6–11.8% from 2005–2014.

### Demonstration of utility of rurality variables for research

**Lung cancer incidence.**—Lung cancer incidence was higher in more rural tracts than other areas for both the RUCA- and Census-based rurality variables, and incidence was higher in low- compared to high-SES census tracts (Figure 2a). For the RUCA variable, annual mean lung cancer incidence ranged from 46.4 (in high-SES, urban tracts) to 81.6 per 100,000 people (in low-SES, rural tracts). Similarly, for the Census variable, annual mean lung cancer incidence ranged from 44.6 (in high-SES, all urban tracts) to 82.9 (in low-SES, all rural tracts).

The absolute disparities measures indicated greater SES disparities in rural versus urban areas when using the RUCA variable (Table 3). The RD in rural areas (28.14, standard error[SE]=1.71) was 42% greater than in urban areas (19.81, SE=0.30), indicating a wider spread in lung cancer incidence rates in rural areas ( $p<.05$ ). Similarly, the ACI in rural areas (−4.47, SE=0.17) was 15% greater than in urban areas (−3.88, SE=0.05), indicating that while the disparity in lung cancer incidence favored those of higher SES across rurality, this tendency was greater in rural areas than in urban areas ( $p<.05$ ). In contrast, the absolute

disparities measures demonstrated inconsistent results across rurality levels when using the Census variable. That is, both the RD and ACI indicated greater SES disparities in mostly urban compared to all urban areas, but no statistical differences for mostly rural or all rural (compared to all urban areas) emerged.

The relative disparities measures found mixed results for the SES disparities in lung cancer incidence by either rurality variable (Table 3). When using the RUCA variable, the RR indicated a 7% greater SES disparity in rural than urban areas ( $p<.05$ ), but the RCI indicated a 7% greater disparity in urban than rural areas ( $p<.05$ ). When using the Census variable, the RR and RCI indicated greater SES disparities in mostly urban compared to all urban areas (both  $p<.05$ ), but the mostly/all rural areas had smaller SES disparities compared to all urban areas (all  $p<.05$ ).

**Lung cancer survival.**—In contrast, lung cancer survival was higher in more urban census tracts than in other areas for both the RUCA- and Census-based rurality variables and in high-SES census tracts (Figure 2b). For the RUCA variable, mean lung cancer survival ranged from 13.2% (in low-SES, rural tracts) to 22.7% (in high-SES, urban tracts). Similarly, for the Census variable, mean lung cancer survival ranged from 12.7% (in low-SES, all rural tracts) to 23.1% (in high-SES, all urban tracts). The absolute and relative disparities measures indicated that the SES disparities in lung cancer survival were generally greater in urban than in more rural census tracts (Table 3). For example, the RD for lung cancer survival using the RUCA variable was 57% smaller in rural areas ( $RD=3.81$ ,  $SE=0.95$ ) than in urban areas ( $RD=8.86$ ,  $SE=0.35$ ) ( $p<.05$ ).

**Breast cancer incidence.**—Breast cancer incidence was higher in more urban tracts than in other areas for both the RUCA- and Census-based rurality variables, and incidence was generally higher in high- compared to low-SES census tracts (Figure 3a). For the RUCA variable, annual mean breast cancer incidence ranged from 106.0 (in low-SES, urban tracts) to 142.5 per 100,000 women (in high-SES, urban tracts). Similarly, for the Census variable, annual mean breast cancer incidence ranged from 103.6 (in low-SES, all rural tracts) to 143.5 (in high-SES, all urban tracts). The absolute and relative disparities measures indicated that the SES disparities in breast cancer incidence were generally larger in urban than in more rural census tracts (Table 3). For example, the RD for breast cancer incidence using the RUCA variable was 49% smaller in rural areas ( $RD=18.68$ ,  $SE=3.67$ ) than in urban areas ( $RD=36.44$ ,  $SE=0.57$ ) ( $p<.05$ ).

**Breast cancer survival.**—Breast cancer survival was also higher in more urban census tracts than in other areas for both the RUCA- and Census-based rurality variables and in high-SES census tracts (Figure 3b). For the RUCA variable, mean breast cancer survival ranged from 83.1% (in low-SES, urban tracts) to 94.8% (in high-SES, urban tracts). Similarly, for the Census variable, mean breast cancer survival ranged from 80.7% (in low-SES, all rural tracts) to 95.0% (in high-SES, all urban tracts). The absolute and relative disparities measures indicated that the SES disparities in breast cancer survival were generally larger in urban than in more rural census tracts (Table 3). For example, the RD for breast cancer survival using the RUCA variable was 42% smaller in rural areas ( $RD=6.78$ ,  $SE=1.93$ ) than in urban areas ( $RD=11.68$ ,  $SE=0.34$ ) ( $p<.05$ ).



## Discussion

Adding indicators of rurality/urbanicity at the census tract-level to the NCI SEER database offers several conceptual and methodological advantages to researchers interested in geographic patterns in cancer surveillance. Importantly, in our analysis of uniqueness, we demonstrated that making measures of census tract-level rurality available does not substantially increase the risk of identifying patients, thereby preserving confidentiality. In addition, there was relatively high concordance in the cases defined as rural or urban according to each variable (i.e., 91.0% of cases were classified as urban using the RUCA-based measure, and 86.7% were classified as urban using the Census-based measure; 12.4% of cases had discordant classifications across the two measures).

Based on these results and conversations with SEER leadership, the two-category RUCA-based measures and the four-category Census-based measures have been added to SEER database. There is minimal difference in disclosure risk among the measures. These two measures have been used extensively in the health geography research literature.<sup>12,21,26,35</sup> The two-category RUCA measure (Categorization C<sup>33</sup>) is most commonly used in health research papers that use RUCA-based measures. The four-category Census-based measure can be collapsed into the two- or three-category versions in several ways and, thus, provides a good deal of flexibility to the researcher.<sup>36</sup> These measures are also compatible with the rurality measure available with the NAACCR Cancer in North America database.<sup>36</sup>

Therefore, a specialized SEER 18 cancer incidence database, containing census tract-level SES quintile and the two rurality measures, is available for research use upon request. All sub-state geographic identifiers, such as county and registry/state, are excluded from this specialized database to limit the risk of disclosure. To further prevent disclosure, users will be required to sign a confidentiality agreement prior to be given the access to this specialized database. More details about how to request this database can be found on the SEER website: <https://seer.cancer.gov/data-software/specialized.html>. Data from the Alaska Native Tumor Registry are not included because of additional confidentiality constraints. This specialized database contains cancer cases diagnosed from 2000 to 2015, thus providing valuable opportunities to evaluate prevalence as well as temporal changes in disparities of cancer incidence and survival. Although census tract definitions change over time, cases are assigned rurality values for the area where they live at diagnosis. Both RUCA and Census-based rurality designations are updated every decade, so cases diagnosed in 2000–2005 are assigned a rurality designation based on data from 2000, and cases diagnosed in 2006–2015 are assigned a rurality designation based on data from 2010.

In our demonstration analysis using this specialized database, we illustrated the utility of these new measures. When considering census tract-level rurality (along with SES disparities), the range of lung and breast cancer incidence rates was much wider than observed using only county-level rurality measures.<sup>6,7</sup> Importantly, the interaction between SES and rurality varied across cancer types and outcomes (e.g., incidence was higher in rural and low-SES areas for lung cancer, but incidence was higher in urban and high-SES areas for breast cancer). Contrary to our hypothesis, the patterns were generally consistent across rurality variables (except for lung cancer incidence rates); however, absolute and

relative disparities measures tended to be of greater magnitude for the RUCA-based compared to the Census-based rurality comparisons. Compared to the RUCA-based rurality, which has only two categories, the four-category Census-based rurality provides opportunities for detecting possible non-monotone relationship between the size of SES disparity and rurality. Notably, SES and rurality are correlated with each other (and many other demographic characteristics), so additional modeling procedures (e.g., multilevel modeling) may be useful in understanding the unique and interacting associations of these variables with cancer outcomes. Future researchers should choose the rurality variable most pertinent to their research question (as described above, the RUCA measure reflects proximity to urban centers, and the Census measure reflects the nature of the immediate environment).

Limitations should be noted for the use of these census tract-level rurality measures in studies of cancer surveillance. By definition, these rurality measures are ecological, and should not be used for individual-level inferences. In addition, the rurality measures refer to the census tract the patient lived in when diagnosed with cancer, and does not necessarily reflect exposure to rurality during the time preceding diagnosis. Finally, in order to limit the risk of disclosure and maintain patient confidentiality, some variables were removed from the specialized SEER 18 database. Namely, this specialized database does not include indicators of the patient's county or registry/state, so spatial analyses will not be possible using these data.

In conclusion, census tract-level measures of rurality add specificity to studies attempting to understand the relationship between rurality and cancer incidence and survival. The additional risk of uniquely identifying patients is limited, and will be further controlled programmatically by the Surveillance Research Program at NCI. Future studies should continue to examine how rurality may influence geographic disparities in cancer outcomes to identify potential targets for improving public health.

## Acknowledgements

This manuscript was prepared or accomplished by the authors in their personal capacity as part of official duty at the National Institutes of Health. The opinions expressed in this article are the authors' own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Financial support:

This work was completed as part of authors' official duty at the National Institutes of Health. No external financial support was provided.

## References

1. Eberhardt MS, Pamuk ER. The importance of place of residence: examining health in rural and nonrural areas. *American Journal of Public Health*. 2004;94(10):1682–1686. [PubMed: 15451731]
2. Singh GK, Siahpush M. Widening rural–urban disparities in life expectancy, US, 1969–2009. *American Journal of Preventive Medicine*. 2014;46(2):e19–e29. [PubMed: 24439358]
3. Kulshreshtha A, Goyal A, Dabhadkar K, Veledar E, Vaccarino V. Urban-rural differences in coronary heart disease mortality in the United States: 1999–2009. *Public Health Rep*. 2014;129(1): 19–29.

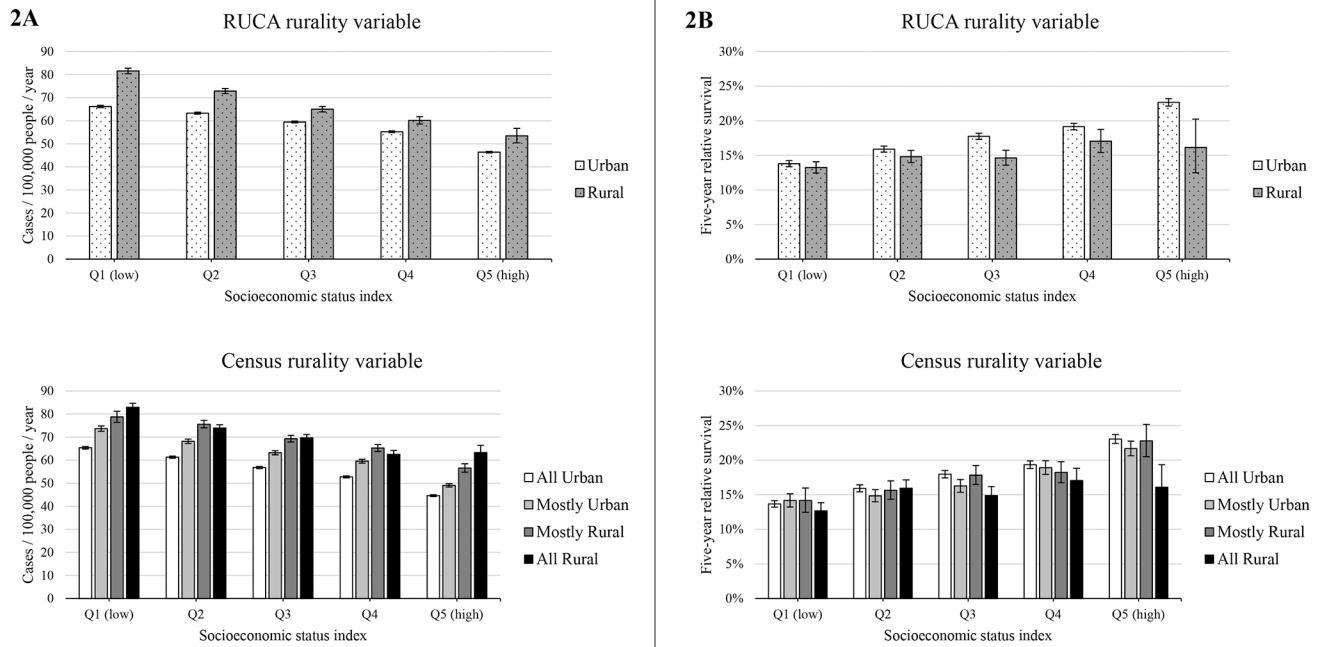


4. Bolin JN, Bellamy GR, Ferdinand AO, et al. Rural healthy people 2020: new decade, same challenges. *The Journal of Rural Health*. 2015;31(3):326–333. [PubMed: 25953431]
5. Moy E, Garcia MC, Bastian B, et al. Leading Causes of Death in Nonmetropolitan and Metropolitan Areas-United States, 1999–2014. *Morbidity and mortality weekly report Surveillance summaries*. 2017;66(1):1–8.
6. Blake KD, Moss JL, Gaysynsky A, Srinivasan S, Croyle RT. Making the case for investment in rural cancer control: An analysis of rural cancer incidence, mortality, and funding trends. *Cancer Epidemiol Biomarkers Prev*. 2017.
7. Henley SJ, Anderson RN, Thomas CC, Massetti GM, Peaker B, Richardson LC. Invasive Cancer Incidence, 2004–2013, and Deaths, 2006–2015, in Nonmetropolitan and Metropolitan Counties - United States. *Morbidity and mortality weekly report Surveillance summaries (Washington, DC : 2002)*. 2017;66(14):1–13.
8. Singh GK, Williams SD, Siahpush M, Mulhollen A. Socioeconomic, Rural-Urban, and Racial Inequalities in US Cancer Mortality: Part I-All Cancers and Lung Cancer and Part II-Colorectal, Prostate, Breast, and Cervical Cancers. *Journal of cancer epidemiology*. 2011;2011:107497. [PubMed: 22496688]
9. Roberts ME, Doogan NJ, Kurti AN, et al. Rural tobacco use across the United States: how rural and urban areas differ, broken down by census regions and divisions. *Health & place*. 2016;39:153–159. [PubMed: 27107746]
10. Moss JL, Liu B, Feuer EJ. Urban/Rural Differences in Breast and Cervical Cancer Incidence: The Mediating Roles of Socioeconomic Status and Provider Density. *Womens Health Issues*. 2017;27(6):683–691. [PubMed: 29108988]
11. Davis MM, Renfro S, Pham R, et al. Geographic and population-level disparities in colorectal cancer testing: A multilevel analysis of Medicaid and commercial claims data. *Preventive medicine*. 2017;101:44–52. [PubMed: 28506715]
12. Henry KA, McDonald K, Sherman R, Kinney AY, Stroup AM. Association between individual and geographic factors and nonadherence to mammography screening guidelines. *Journal of Women's Health*. 2014;23(8):664–674.
13. Parikh-Patel A, Bates JH, Campleman S. Colorectal cancer stage at diagnosis by socioeconomic and urban/rural status in California, 1988–2000. *Cancer*. 2006;107(S5):1189–1195. [PubMed: 16835910]
14. Paquette I, Finlayson SRG. Rural versus urban colorectal and lung cancer patients: differences in stage at presentation. *Journal of the American College of Surgeons*. 2007;205(5):636–641. [PubMed: 17964438]
15. Singh GK. Rural-Urban Trends and Patterns in Cervical Cancer Mortality, Incidence, Stage, and Survival in the United States, 1950–2008. *Journal of community health*. 2011.
16. Kusmin L Rural America at a glance. United States Department of Agriculture, Economic Research Service;2016.
17. National Cancer Institute. Understanding definitions of rural/rurality: Implications for rural cancer control. 2017; Rockville, MD.
18. Hall SA, Kaufman JS, Ricketts TC. Defining urban and rural areas in US epidemiologic studies. *Journal of urban health : bulletin of the New York Academy of Medicine*. 2006;83(2):162–175. [PubMed: 16736366]
19. Ingram DD, Franco SJ. 2013 NCHS Urban-Rural Classification Scheme for Counties. *Vital and health statistics Series 2, Data evaluation and methods research*. 2014(166):1–73.
20. Hart LG, Larson EH, Lishner DM. Rural definitions for health policy and research. *American Journal of Public Health*. 2005;95(7):1149–1155. [PubMed: 15983270]
21. Meilleur A, Subramanian SV, Plascak JJ, Fisher JL, Paskett ED, Lamont EB. Rural residence and cancer outcomes in the United States: issues and challenges. *Cancer epidemiology, biomarkers & prevention*. 2013;22(10):1657–1667.
22. Smith ML, Dickerson JB, Wendel ML, et al. The utility of rural and underserved designations in geospatial assessments of distance traveled to healthcare services: implications for public health research and practice. *Journal of environmental and public health*. 2013;2013:960157. [PubMed: 23843803]

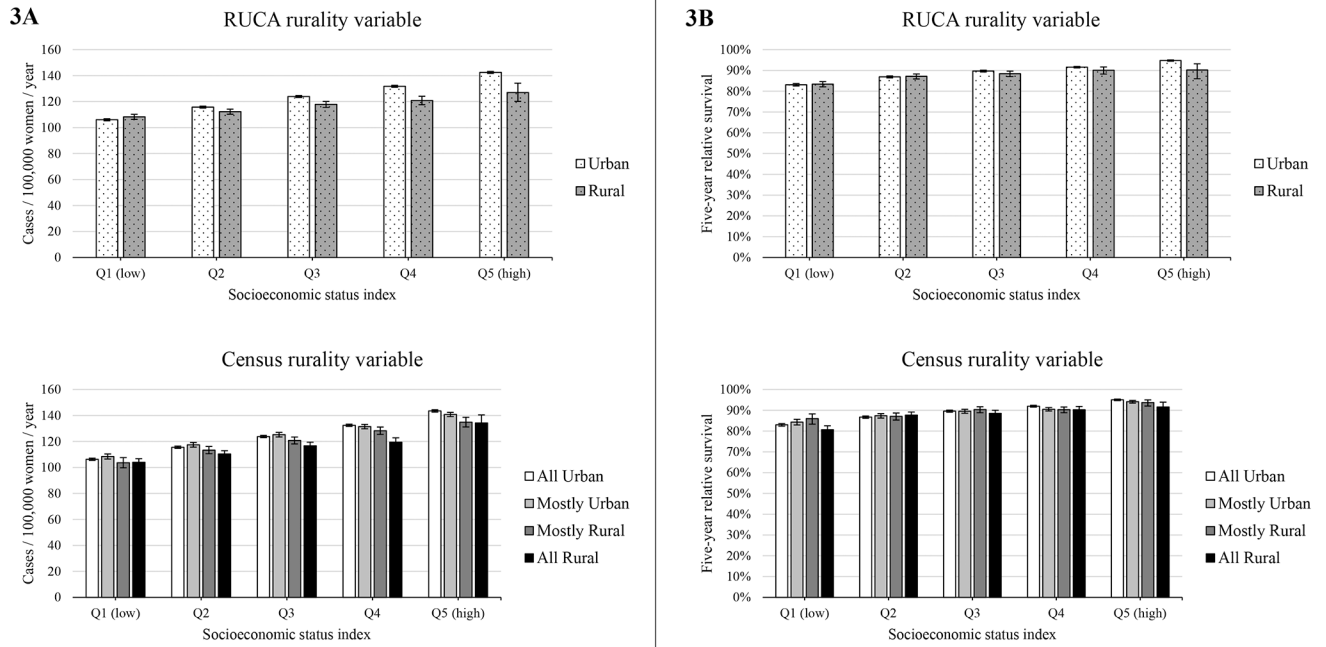
23. Surveillance Research Program NCI. SEER\*Stat software (seer.cancer.gov/seerstat) version 8.3.2.
24. Coughlin SS, Leadbetter S, Richards T, Sabatino SA. Contextual analysis of breast and cervical cancer screening and factors associated with health care access among United States women, 2002. *Social science & medicine* (1982). 2008;66(2):260–275. [PubMed: 18022299]
25. Cromartie J, Nulph D, Hart G. Mapping frontier and remote areas in the US. *Amber Waves*. 2012;10(4):1D.
26. Yu M, Tatalovich Z, Gibson JT, Cronin KA. Using a composite index of socioeconomic status to investigate health disparities while protecting the confidentiality of cancer registry data. *Cancer causes & control : CCC*. 2014;25(1):81–92.
27. Boscoe F Towards the use of a census tract poverty indicator variable in cancer surveillance. *J Registry Manag*. 2010;37(4):148–151. [PubMed: 21688744]
28. Hastert TA, Beresford SA, Sheppard L, White E. Disparities in cancer incidence and mortality by area-level socioeconomic status: a multilevel analysis. *Journal of epidemiology and community health*. 2015;69(2):168–176. [PubMed: 25288143]
29. Centers for Disease Control and Prevention. Vital signs: current cigarette smoking among adults aged > or= 18 years---United States, 2009. *MMWR Morb Mortal Wkly Rep*. 2010;59(35):1135. [PubMed: 20829747]
30. Akinyemiju TF, Pisu M, Waterbor JW, Altekruse SF. Socioeconomic status and incidence of breast cancer by hormone receptor subtype. *SpringerPlus*. 2015;4:508-015-1282-1282. eCollection 2015.
31. Robert SA, Strombom I, Trentham-Dietz A, et al. Socioeconomic risk factors for breast cancer: distinguishing individual- and community-level effects. *Epidemiology (Cambridge, Mass)*. 2004;15(4):442–450.
32. Breen N, Lewis DR, Gibson JT, Yu M, Harper S. Assessing disparities in colorectal cancer mortality by socioeconomic status using new tools: health disparities calculator and socioeconomic quintiles. *Cancer Causes Control*. 2017;28(2):117–125. [PubMed: 28083800]
33. Harper S, Lynch J, Meersman SC, Breen N, Davis WW, Reichman MC. Trends in area-socioeconomic and race-ethnic disparities in breast cancer incidence, stage at diagnosis, screening, mortality, and survival among women ages 50 years and over (1987–2005). *Cancer epidemiology, biomarkers & prevention*. 2009;18(1):121–131.
34. Harper S, Lynch J. *Methods for measuring cancer disparities: Using data relevant to Healthy People 2010 cancer-related objectives*. Bethesda, MD: National Cancer Institute;2005.
35. Gomez SL, Glaser SL, McClure LA, et al. The California Neighborhoods Data System: a new resource for examining the impact of neighborhood characteristics on cancer incidence and outcomes in populations. *Cancer Causes Control*. 2011;22(4):631–647. [PubMed: 21318584]
36. Henry KA, Sherman R, Stinchcomb DG. Assessing fitness for use of two indicators of the rural-urban environment in the NAACCR data files. *NAACCR Conference*; 2015; Charlotte, NC.

Registry	Percent of cases from urban areas		Disagreement Percent
	RUCA measure	Census measure	
California (all)	96.2	94.6	5.4
Connecticut	99.3	90.2	9.2
Detroit	99.9	98.1	1.8
Georgia (all)	89.0	73.2	23.7
Hawaii	79.0	94.1	16.9
Iowa	59.3	63.6	34.8
Kentucky	62.3	57.7	30.2
Louisiana	89.9	73.5	23.5
New Jersey	99.9	95.8	4.0
New Mexico	68.2	77.9	28.6
Seattle	93.2	86.7	15.3
Utah	90.3	90.7	10.1
Totals	91.0	86.7	12.4

**Figure 1.** Mismatch of the two rurality measures by Surveillance, Epidemiology, and End Results registry area.



**Figure 2.** Lung cancer incidence (panel A) and survival (panel B) across quintiles of socioeconomic status (SES) using two measures of rurality; Surveillance, Epidemiology, and End Results, 2005–2014. Survival indicates five-year relative survival among cases diagnosed from 2005–2009. Top figures in each panel indicate outcomes using the United States Department of Agriculture’s (USDA) Rural Urban Commuting Areas (RUCA), and bottom figures in each panel indicate outcomes using the Census Bureau’s classification of percent of residents living in non-urban areas. Leftmost columns indicate outcomes for people living in the most impoverished census tracts (Q1 for SES), while rightmost columns indicate outcomes for people living in the most affluent census tracts (Q5 for SES).



**Figure 3.** Breast cancer incidence (panel A) and survival (panel B) across quintiles of socioeconomic status (SES) using two measures of rurality; Surveillance, Epidemiology, and End Results, 2005–2014. Survival indicates five-year relative survival among cases diagnosed from 2005–2009. Top figures in each panel indicate outcomes using the United States Department of Agriculture’s (USDA) Rural Urban Commuting Areas (RUCA), and bottom figures in each panel indicate outcomes using the Census Bureau’s classification of percent of residents living in non-urban. Leftmost columns indicate outcomes for women living in the most impoverished census tracts (Q1 for SES), while rightmost columns indicate outcomes for women living in the most affluent census tracts (Q5 for SES).

**Table 1.**

Categorical alternatives for tested measures of census tract-level rurality.

Categories of RUCA-based measures	Categories of Census-based measures
<p>Four categories – Categorization A:</p> <ul style="list-style-type: none"> <li>• Urban area commuting focused (codes 1.0, 1.1, 2.0, 2.1, 3.0, 4.1, 5.1, 7.1, 8.1, and 10.1)</li> <li>• Large rural city or town focused (codes 4.0, 4.2, 5.0, 5.2, 6.0, and 6.1)</li> <li>• Small rural town focused (codes 7.0, 7.2, 7.3, 7.4, 8.0, 8.2, 8.3, 8.4, 9.0, 9.1, and 9.2)</li> <li>• Isolated small rural town focused (codes 10.0, 10.2, 10.3, 10.4, 10.5, and 10.6)</li> </ul>	<p>Four categories:</p> <ul style="list-style-type: none"> <li>• 100% urban</li> <li>• 50% but &lt;100% urban</li> <li>• &gt;0% but &lt;50% urban</li> <li>• 100% rural</li> </ul> <p>Three categories:</p> <ul style="list-style-type: none"> <li>• 100% urban</li> <li>• Mixed urban and rural</li> <li>• 100% rural</li> </ul> <p>Two categories:</p> <ul style="list-style-type: none"> <li>• 50% urban</li> <li>• &lt;50% urban</li> </ul>
<p>Two categories – Categorization C:</p> <ul style="list-style-type: none"> <li>• Urban area commuting focused (codes 1.0, 1.1, 2.0, 2.1, 3.0, 4.1, 5.1, 7.1, 8.1, and 10.1)</li> <li>• Not urban area commuting focused (all other codes)</li> </ul>	

*Note.* The specialized SEER 18 cancer incidence database includes census tract-level rurality using (1) four categories of the Census-based variable and (2) two categories of the RUCA-based variable. RUCA=Rural Urban Commuting Areas.



**Table 2.**

Uniqueness test results for combinations of categorical measures of census tract-level socioeconomic status (SES) and rurality.

SES	RUCA-based rurality	Census-based rurality	Unique tracts ( <i>n</i> )	Unique tracts (%)	Cancer cases in unique tracts (SEER 2006–2014) (%)
5 categories	n/a	n/a	0	0.000	0.000
5 categories	4 categories	4 categories	4	0.021	0.027
5 categories	2 categories	4 categories	1	0.005	0.004
5 categories	4 categories	3 categories	2	0.011	0.015
5 categories	2 categories	3 categories	1	0.005	0.004
5 categories	4 categories	2 categories	2	0.011	0.012
5 categories	2 categories	2 categories	0	0.000	0.000

*Note.* Analysis limited to census tracts located in SEER 18 catchment areas excluding Alaska. Shaded categorization scheme was selected for incorporation into a specialized SEER 18 cancer incidence database, available upon request. SES=socioeconomic status; RUCA=Rural Urban Commuting Areas; SEER=Surveillance, Epidemiology, and End Results.

Measures of disparities in lung cancer incidence and survival, and breast cancer incidence and survival, across socioeconomic status (SES) quintiles for urban and rural areas.

**Table 3.**

	SES disparities in lung cancer incidence																									
	RUCAs (2 levels)				Census (4 levels)				RUCAs (2 levels)				Census (4 levels)													
	Urban (ref <sup>§</sup> )	SE	est.	% diff.	Rural	SE	est.	% diff.	All Urban (ref <sup>§</sup> )	SE	est.	% diff.	Mostly Urban	SE	est.	% diff.	Mostly Rural	SE	est.	% diff.	All Rural	SE	est.	% diff.		
<b>ABSOLUTE MEASURES</b>																										
Range Difference (RD)	19.81	0.30	28.14	1.71	42	20.74	0.33	24.62	0.69	19	22.16	1.53	7	20.45	1.23	-1										
Absolute Concentration Index (ACI)	-3.88	0.05	-4.47	0.17	15	-4.04	0.06	-4.72	0.12	17	-3.81	0.22	-6	-3.91	0.22	-3										
<b>RELATIVE MEASURES</b>																										
Range Ratio (RR)	1.43	0.01	1.53	0.05	7	1.47	0.01	1.50	0.02	3	1.39	0.03	-5	1.33	0.02	-9										
Relative Concentration Index (RCI)	-0.07	0.00	-0.06	0.00	-7	-0.07	0.00	-0.08	0.00	7	-0.06	0.00	-24	-0.05	0.00	-25										
<b>SES disparities in lung cancer survival</b>																										
	RUCAs (2 levels)																									
	RUCAs (2 levels)				Census (4 levels)				RUCAs (2 levels)				Census (4 levels)													
	Urban (ref <sup>§</sup> )	SE	est.	% diff.	Rural	SE	est.	% diff.	All Urban (ref <sup>§</sup> )	SE	est.	% diff.	Mostly Urban	SE	est.	% diff.	Mostly Rural	SE	est.	% diff.	All Rural	SE	est.	% diff.		
<b>ABSOLUTE MEASURES</b>																										
Range Difference (RD)	8.86	0.35	3.81	0.95	-57	9.38	0.41	7.52	0.73	-20	8.62	1.49	-8	4.39	1.06	-53										
Absolute Concentration Index (ACI)	1.62	0.06	0.55	0.14	-66	1.73	0.07	1.51	0.13	-12	1.24	0.20	-28	0.67	0.18	-61										
<b>RELATIVE MEASURES</b>																										
Range Ratio (RR)	1.64	0.03	1.29	0.08	-22	1.69	0.04	1.53	0.07	-9	1.61	0.13	-5	1.35	0.09	-20										
Relative Concentration Index (RCI)	0.09	0.00	0.04	0.01	-58	0.10	0.00	0.09	0.01	-9	0.07	0.01	-27	0.04	0.01	-54										
<b>SES disparities in breast cancer incidence</b>																										
	RUCAs (2 levels)																									
	RUCAs (2 levels)				Census (4 levels)				RUCAs (2 levels)				Census (4 levels)													
	Urban (ref <sup>§</sup> )	SE	est.	% diff.	Rural	SE	est.	% diff.	All Urban (ref <sup>§</sup> )	SE	est.	% diff.	Mostly Urban	SE	est.	% diff.	Mostly Rural	SE	est.	% diff.	All Rural	SE	est.	% diff.		
<b>ABSOLUTE MEASURES</b>																										
Range Difference (RD)	36.44	0.57	18.68	3.67	-49	37.22	0.64	32.39	1.28	-13	31.24	2.74	-16	30.28	3.40	-19										

SES disparities in lung cancer incidence															
RUCA (2 levels)				Census (4 levels)				Census (4 levels)							
Urban (ref <sup>a</sup> )		Rural		All Urban (ref <sup>a</sup> )		Mostly Urban		Mostly Rural		All Rural					
est.	SE	est.	SE	est.	SE	est.	SE	est.	SE	est.	SE				
Absolute Concentration Index (ACI)															
7.14	0.10	2.68	0.31	-62	7.39	0.12	6.18	0.22	-16	5.27	0.39	-29	4.05	0.39	-45
<b>RELATIVE MEASURES</b>															
Range Ratio (RR)															
1.34	0.01	1.17	0.03	-13	1.35	0.01	1.30	0.01	-4	1.30	0.03	-4	1.29	0.04	-4
Relative Concentration Index (RCI)															
0.06	0.00	0.02	0.00	-58	0.06	0.00	0.05	0.00	-18	0.04	0.00	-27	0.04	0.00	-40
SES disparities in breast cancer survival															
RUCA (2 levels)				Census (4 levels)				Census (4 levels)							
Urban (ref <sup>a</sup> )		Rural		All Urban (ref <sup>a</sup> )		Mostly Urban		Mostly Rural		All Rural					
est.	SE	est.	SE	est.	SE	est.	SE	est.	SE	est.	SE				
Absolute Concentration Index (ACI)															
11.68	0.34	6.78	1.93	-42	12.03	0.37	9.79	0.76	-19	7.73	1.46	-36	10.89	1.71	-10
Relative Concentration Index (RCI)															
2.12	0.05	1.30	0.19	-39	2.29	0.06	1.75	0.12	-24	1.28	0.20	-44	1.77	0.23	-23
<b>RELATIVE MEASURES</b>															
Range Ratio (RR)															
1.14	0.00	1.08	0.02	-5	1.15	0.01	1.12	0.01	-3	1.09	0.02	-5	1.14	0.02	-1
Relative Concentration Index (RCI)															
0.02	0.00	0.02	0.00	-38	0.03	0.00	0.02	0.00	-24	0.01	0.00	-44	0.02	0.00	-20

Note. Bolded estimates indicate statistically-significant ( $p < .05$ ) differences in SES-associated disparities for rurality categories compared to the most urban group (the reference categories). RUCA=Rural Urban Commuting Areas; ref=reference; est.=estimate; SE=standard error; diff.=difference.

\* Reference groups were only used in the calculation of RD and RR.