

# Same People, Different Results: Categorizing Cancer Registry Cases across the Rural-Urban Continuum

**Andrea M. Schiefelbein**

Morgridge Institute for Research

**Amy K. Taylor**

University of Wisconsin–Madison

**John K. Krebsbach**

University of Wisconsin–Madison

**Jienian Zhang**

University of Wisconsin–Madison

**Chloe E. Haimson**

University of Wisconsin–Madison

**Patrick R. Varley**

University of Wisconsin–Madison

**Melissa C. Skala**

University of Wisconsin–Madison

**John M. Eason**

University of Wisconsin–Madison

**Noelle K. LoConte** (✉ [Ns3@medicine.wisc.edu](mailto:Ns3@medicine.wisc.edu))

University of Wisconsin–Madison

---

## Research Article

**Keywords:** Rurality, Rural-Urban Continuum, Methods, Cance

**Posted Date:** December 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1200114/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

Background: Many rural-urban indexes are utilized in United States cancer research. This variation introduces inconsistencies between studies with a rural-urban component. Recommendations to date on which index to utilize have prioritized index geographical unit over feasibility of index inclusion in analysis. We evaluated rural-urban indexes and recommend one index for use to increase comparability across studies. Methods: We assessed nine U.S. rural-urban indexes regarding their respective rural and urban code ranges; geographical unit, land area, and population distributions; percent agreement; suitability as continuous variables in analysis; and feasibility of integration into national, state, and local cancer research. We referenced 1,569 Wisconsin Pancreatic Cancer Registry patients to demonstrate how rural-urban index choice impacts patient categorization. Results: Six indexes categorized rural and urban areas. Indexes agreed on binary rural-urban designation for 88.8% of the U.S. population. As ternary variables, they agreed for 83.4%. For cancer registry patients, this decreased to 73.4% and 60.4% agreement, respectively. Rural-Urban Continuum Codes (RUCC) performed the best with ability to differentiate metropolitan, micropolitan, and rural counties, are available for retrospective and prospective studies, and can be coded continuously for analysis. Conclusions: Whether a patient was categorized as urban or rural changed depending on which index was used when applied to a cancer registry data set. We conclude that RUCC is an appropriate and feasible rural-urban index to include in cancer research, as it is standardly available in national cancer registries in its 9-code format and can be matched to patient's county of residence for local research and it had the least amount of fluctuation of the indices analyzed. Utilizing RUCC as a continuous variable across studies with a rural-urban component will increase reproducibility and comparability of results and eliminate the choice of rural-urban index as a potential source of discrepancy between studies. Trial registration: Not applicable

## Background

Research on cancer disparities increasingly incorporates community factors to understand variation across patient treatment and outcomes. Rurality, one such community factor, predicts later stages of cancer diagnosis,<sup>1-3</sup> lower rates of specific therapies,<sup>4-6</sup> less effective therapies,<sup>7</sup> shorter overall survival,<sup>6,8-11</sup> and higher mortality rates.<sup>12-15</sup> These trends persist across geographical regions of the United States (U.S.) and cancer types.<sup>6,7,15</sup> However, there also exists important variation in treatment and outcomes between the group of patients categorized as rural and urban.<sup>16</sup> These variations result from differences in communities,<sup>17</sup> patient demographics,<sup>18</sup> and health care organizations.<sup>19</sup>

Methodological differences in identifying rural communities and patients also produce variations and inconsistencies in measuring disparities in rural cancer patient treatment and outcomes.<sup>16,20</sup> More than 9 rural-urban indexes are used across cancer research to categorize patients. Indexes are based on differing geographical levels, including census tract, ZIP Code Tabulation Area (ZCTA),<sup>21</sup> and county. Indexes differ in terms of the unique combinations of criteria on which they are based, incorporating factors such as population size, percentage of commuting population, and adjacency to urban areas. Additionally there is confusion about terms. For example, urban and metropolitan are often used interchangeably, but these are

distinct terms. (“Urban” being the densely populated city, but “metropolitan” encompassing the city and some surrounding less populated areas.) Furthermore, in reviewing rural cancer studies, we identified gaps in how rural-urban indexes are deployed, including incorrect index identification<sup>6</sup>, omitting the index or process employed to classify patients’ communities,<sup>22</sup> and using outdated indexes.<sup>23</sup> In recommending an optimal index, researchers often focus on the desired geographical measure of a community, specifically prioritizing indexes based on census tract, the smallest areal measurement of a community. Such focus is to the detriment of index availability in data sources, including national registries and health care records, or over time.<sup>2,16</sup>

We evaluated rural-urban indexes from recent U.S. cancer research for their appropriateness in categorizing patients across the rural-urban continuum. We examined indexes across their respective rural and urban code ranges; geographical unit, land area, and population distributions; suitability to be incorporated as continuous variables in statistical analysis; and feasibility to be integrated into local-, state-, and national-level cancer research. We utilized the University of Wisconsin-Health Pancreatic Cancer Registry patient cohort as a local case study to demonstrate how index choice impacts patient rural-urban categorization.

## Methods

### Rural-urban indexes

We identified 9 rural-urban indexes from cancer research between 2000-2020 including: U.S. Census Bureau Urban Rural Classification of Urban Areas and Urban Clusters (UACE)<sup>24</sup>; U.S. Census Bureau Core Based Statistical Areas (CBSA)<sup>25</sup>; U.S. Department of Agriculture Frontier and Remote Area Codes (FAR)<sup>26</sup>; U.S. Department of Agriculture Rural-Urban Commuting Area Codes (RUCA)<sup>27</sup>; WWAMI Rural Health Research Center and Administration for Community Living: Aging, Independence, and Disability Rural-Urban Commuting Area Codes at the ZIP Code Tabulation Area level (RUCA(z))<sup>28,29</sup>; U.S. Department of Agriculture Rural-Urban Continuum Codes (RUCC)<sup>30</sup>; U.S. Department of Agriculture Urban Influence Codes (UIC)<sup>31</sup>; Centers for Disease Control and Prevention National Center for Health Statistics (NCHS) Urban-Rural Classification Scheme for Counties<sup>32</sup>; and the Purdue University Index of Relative Rurality (IRR)<sup>33</sup> (Supporting Table 1). We retrieved indexes spanning 1980-2013 from publishers’ websites. Each index has been described elsewhere.<sup>2,16,17,34-36</sup>

### Land area and population data

We acquired land area (square miles) and population information at the geographical unit of each index to compare land area and population distributions across indexes. Geographical units included census block (UACE), census tract (RUCA), ZIP Code Tabulation Area (ZCTA) (FAR, RUCA(z)), and county (RUCC, UIC, NCHS, IRR, CBSA). UACE, RUCA, and FAR indexes included 2010 population and land area variables in their source files. We obtained 2010 county-level population and land area data from the 2010 Census of Population Summary File 1 (SF1)<sup>37</sup> for the RUCC, UIC, NCHS, IRR, and CBSA indexes. The RUCA(z) index is

based on approximate boundaries of 2013 ZIP Code Tabulation Areas (ZCTA).<sup>21</sup> Since these boundaries fluctuate over time, we were unable to obtain the 2013 ZCTA population or land area data on which RUCA(z) was based. Therefore, we excluded this index from parts of our analysis.

1,569 patients from the University of Wisconsin-Health Pancreatic Cancer Registry (Registry Patients) diagnosed between 2004 to 2016 with pancreatic ductal adenocarcinoma served as a reference population to demonstrate how rural-urban index choice may impact patient categorization. We evaluated differences in rurality of the Registry patient cohort via percent agreement across county and ZCTA-based binary and ternary indexes. We compared the change in each index's median and interquartile range and mean over time for the patient cohort.

### **Comparing rural-urban indexes**

In comparing indexes, we evaluated breadth - the extent to which rural and urban communities are differentiated from one another - and depth - the extent to which distinctions are made within rural or urban-designated communities. Supporting Table 1 shows the 9 indexes by geographical unit, classification of urban and rural codes, and the amount and percentage of land area, geographical units, and population each index classifies as urban and rural (2010 versions) for the U.S., Midwestern States, and Wisconsin.

We excluded indexes that simply distinguished rural from urban communities; therefore, we excluded UACE and CBSA, as they only designate urban communities, and FAR, as it only designates rural communities (Supporting Table 1). We included the remaining 6 rural-urban indexes, RUCC, UIC, NCHS, IRR, RUCA, and RUCA(z), in the full analysis. We transformed these to binary indexes based on each index's binary categorization of metropolitan and non-metropolitan areas and to ternary indexes based on each index's ternary categorization of metropolitan, micropolitan/urban, and noncore/small town/rural. Because IRR is a continuous variable that does not subcategorize counties, we established divisions between metropolitan and non-metropolitan counties at IRR = 0.50 and further subdivided non-metropolitan counties into micropolitan/urban and rural counties at IRR = 0.60.<sup>34</sup>

We calculated Cohen's kappa, with an ordinal weight, to evaluate the level of agreement across indexes in their binary and ternary forms by geographical units, land area, and population. We also compared the percent agreement of geographical units, land area, and population across county- and census-tract-based binary and ternary indexes (Table 1 and Supporting Table 2). We compared the distribution of geographical units, land area, and population across indexes via median and inter-quartile range and mean and standard deviation. We examined these trends visually via violin plots, with indexes standardized to illustrate transitions along a rural-urban interface, for the U.S., Wisconsin, and Registry patients (Figure 1 and Supporting Figure 1).

We used STATA Version 16.1 (StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC) to complete the analysis and ArcGIS Version 10.7 (ESRI 2019. ArcGIS Desktop: Release 10.7.1, Redlands, CA: Environmental Systems Research Institute) to create maps.

## Results

# Inconsistency and agreement across binary rural-urban designations

Supporting Table 1 displays the geographical unit, binary rural-urban delineation, and rural-urban categorization of Land Area, Geographical Units, and Total Population for each of the 9 indexes. There are 2 methods to designate communities as rural or urban in RUCA and RUCA(z); both methods are shown.

In addition to a large difference in distribution across rural and urban communities, there is also a wide distribution within rural areas across indexes as the percentage of rural communities (by geographical unit) in the U.S. ranged from 17.5% of ZCTAs (RUCA (option 2)) to 63.0% of counties (IRR) (Supporting Table 1). By comparison, the difference in percentage of rural communities is even larger across the 12 Midwestern States (23.2–71.4%) and Wisconsin as a single state (13.2–63.9%). The percent of land area across rural communities ranged from 52% (FAR) to 97% (UACE). Recall that FAR and UACE indexes categorize either rural or urban areas, but not both. The variation in land area was smaller across indexes categorizing both rural and urban areas. The US population living in rural areas follows a similar pattern, with indexes categorizing rural or urban areas allocating 3.9% (FAR) to 19.3% (UA) of the US population to rural codes compared to indexes categorizing rural and urban areas allocating 11.5% (IRR) to 16.5% (RUCA (option 1)) of US population to rural codes. These trends for land area and total population were similar across the Midwest and Wisconsin.

Binary rural or urban designations agreed across RUCA, RUCC, UIC, NCHS, and IRR indexes for 88.8% of the US population (Table 1 and Supporting Table 2). RUCC and RUCA, the 2 most employed rural-urban indexes in cancer research, agreed on 94.9% of the US population. There was 73.4% agreement among Registry patients in classifying patients across binary RUCC, UIC, NCHS, IRR, and RUCA(z) indexes. This increased to 91.0% agreement among Registry patients when comparing RUCC and RUCA(z) only. We included RUCA(z) in this analysis as patient ZIP codes were known. The difference between the percent agreement at the national level compared to the local registry cohort was notable.

Cohen's Kappa ranged from 0.60 when comparing IRR with RUCC, UIC, and NCHS to 0.81 when comparing RUCA with RUCC, UIC, and NCHS, indicating moderate to very good agreement between indexes. We excluded RUCA(z) from this analysis as ZCTAs cannot be matched one to one with census tracts or counties.

## Agreement decreased across ternary metropolitan, micropolitan, and rural designations

RUCA, RUCC, UIC, NCHS, and IRR indexes agreed on ternary metropolitan, micropolitan, and rural designations for 83.4% of the US population (Table 1 and Supporting Table 2). These indexes simultaneously designate 6.0% of land area and 1.8% of US total population as rural, micropolitan, and metropolitan depending on the index used. Adding further confusion, while some indexes designate 5.3% of land area and 1.5% of total population as rural, other indexes designate these same areas and people as

metropolitan. Again, there is higher agreement across designation in comparing RUCC and RUCA indexes; 88.8% of US population agreed when considering the RUCC and RUCA indexes only. Within the Registry patients, there was 60.4% agreement across these indexes when designating ternary metropolitan, micropolitan or rural communities. This increased to 74.9% agreement when limited to RUCC and RUCA(z). RUCA(z) was included in the Registry patient analysis since patient ZIP codes were known. Cohen's Kappa ranged across indexes from 0.53 for IRR compared to UIC and NCHS to 0.77 for RUCC and RUCA compared to UIC and NCHS, indicating moderate to good agreement as ternary indexes.

## **Differences in discrete or continuous index geographical units, land area, and population distributions**

Figure 1 and Supporting Figure 1 show RUCC, UIC, NCHS, IRR, RUCA, and RUCA(z) total population, geographical units, and land area distributions across the US and Wisconsin. Maps of Wisconsin based on each index are shown in Figure 2; the discrepancy between RUCA and RUCA(z) in Wisconsin is shown in Figure 3. The median geographical unit across the US falls within the urban designation for RUCA and RUCA(z); conversely, the median falls within the rural designation for RUCC, UIC, NCHS, and IRR. The median land area distribution falls within the rural designation for the 5 county- and census-tract-based indexes, and the median population distribution falls within the urban designation for the same indexes. RUCA(z) was excluded from the land area and population distribution analysis since 2013 ZCTA-based land area and population totals were unavailable. The rural-urban distribution of Registry patients followed Wisconsin population distribution trends, where median index values tended towards the lower-end of metropolitan codes compared to the U.S. population for the 4 county-based indexes.

## **Changes in discrete or continuous index distributions over time**

RUCC, RUCA, UIC, NCHS, and RUCA(z) indexes captured changes in rural-urban community designations over time (Figure 4 and Supporting Figure 2). RUCC changes on a per-county basis for Wisconsin are mapped in Figure 5. The mean rural-urban value across counties, ZCTAs, and census tracts for each of these indexes decreased over time with new versions. The distributions highlight where indexes underwent methodological changes, such as between 1993 and 2003 for RUCC and UIC (Figure 4 and Supporting Figure 2). The mean IRR value remained constant between its 2 available versions due to IRR being a relative measure, and therefore unable to capture absolute changes in rurality, which is necessary for longitudinal studies.

## **Discussion**

### **Categorizing rural and urban communities**

Indexes must categorize both rural and urban areas to study cancer treatment and outcomes across the rural-urban continuum. UACE and CBSA indexes only categorize urban areas, and the FAR index only

categorizes rural areas, making them unsuitable for research including a spectrum of rurality. RUCC, UIC, NCHS, IRR, RUCA, and RUCA(z) categorize areas across metropolitan, micropolitan, and rural areas.

### **Comparability of research based on different indexes**

RUCC, UIC, and NCHS are county-level indexes based on OMB metropolitan and non-metropolitan definitions,<sup>30-32,38</sup> making them identical as binary variables and research based on them as binary variables comparable in terms of rurality (Supporting Table 1). UIC and NCHS further follow OMB guidelines to divide non-metropolitan counties into micropolitan and rural counties, making them identical as ternary variables, too. These 3 indexes employ different methodologies to subdivide counties within metropolitan, micropolitan, and rural categories, making them unique at individual code-levels. They also emphasize different subsets of counties; RUCC identifies 3 metropolitan levels of counties, 4 urban levels, and 2 rural levels<sup>30</sup>; UIC prioritizes rural counties by designating 7 of 12 codes as rural<sup>31</sup>; and NCHS prioritizes metropolitan counties by designating 4 of 6 codes as metropolitan.<sup>32</sup>

RUCA and RUCA(z) also stem from OMB metropolitan and non-metropolitan categories.<sup>27,29,38</sup> They are subdivided into 2, 3, or 4 categories across 10 primary codes and further divided into 21 secondary codes (2010 index). Some researchers create a binary variable based on the primary codes (option 1) and other researchers group counties with a secondary code of x.1, indicating high commuting areas, with metropolitan counties to create a different binary variable (option 2). Therefore, research based on binary RUCA or RUCA(z) variables may not be directly comparable as researchers may use different methodologies to create binary rural-urban designations. This problem is exacerbated when researchers do not disclose which method they employed to create a binary RUCA variable in manuscripts.<sup>39,40</sup>

The high population percent agreement between RUCC and RUCA at binary (94.9%) and ternary (88.8%) levels suggests that the index may introduce less variability into results than expected. However, the percent agreement between RUCC and RUCA(z) decreased to 91.0% at a binary and 74.9% at a ternary level when compared for the Registry Patients (Table 2). This may be due to this specific patient population differing from national trends or be further evidence of RUCA(z) being a poor approximation of RUCA, as demonstrated by RUCA and RUCA(z) disagreeing for 28.9% of Wisconsin land area (Figure 3). Repeating this analysis on a cohort that includes patient-specific census tract, ZIP code, and county is necessary to further explore this question. The differences in percent agreement between national and local populations highlight that national trends may not be replicated at a local health-system level.

### **Comparing indexes by geographical unit, land area, and population distributions**

Indexes varied in whether and to what extent they employed each of their individual codes to categorize geographical units, land area, and population. U.S. counties and land area were distributed across RUCC, though few counties, and therefore minimal land area and population, are categorized as RUCC 5

(Supporting Figures 1A and 1B). This creates a natural binary division within RUCC that does not follow the index's metropolitan/non-metropolitan or metropolitan/micropolitan/rural designations. Since UIC only designates 2 of 12 codes as urban, counties clustered within the urban group (Supporting Figure 1A). UIC cannot be interpreted across a continuum since micropolitan (codes 3, 5, and 8) and rural categories (codes 4, 6, 7, 9, 10, 11, 12) are not designated with sequential codes.<sup>31</sup> NCHS counties and land area clustered by its most rural code since it allocates only 1 category to rural counties (Supporting Figures 1A and 1C). IRR showed normal distributions across geographical units, land area, and population distributions, which is based on it being a relative measure of rurality (Figure 1A and Supporting Figures 1A and 1C).

Census tract and population distributions were clustered in RUCA's most urban code, which is a product of census tracts being smaller and denser in more populated urban areas (Figure 1A, Figure 2E, and Supporting Figure 1A). The opposite trend was seen in the RUCA land area distribution, with most land area clustering in its most rural code (Figure 2E and Supporting Figure 1C). RUCA(z) separated to its more urban and most rural ZCTAs. Differences between the RUCA and RUCA(z) geographical unit distribution suggest that RUCA(z) may not adequately approximate RUCA (Figure 3 and Supporting Figure 1A).

National trends were magnified when viewed for Wisconsin, especially for population distributions (Figure 1B). The population distribution was spread more evenly across RUCC and UIC metropolitan codes for Wisconsin than for the U.S. The population distribution was almost consistent across all NCHS codes, showed an urban cluster separating itself from the rest of Wisconsin for counties in and around Milwaukee County in IRR, and remained similar for RUCA in Wisconsin compared to the U.S. Within the Registry, patients were naturally divided into 2 patient populations by RUCC and UIC and into 4 patient populations by RUCA(z) (Figure 1C). These differences between national, state, and local population distributions highlight that the rural-urban composition of research participants may differ drastically based a study's geographical reach.

Maps of Wisconsin by RUCC, UIC, NCHS, IRR, RUCA, and RUCA(z) highlight similarities and differences across county-, ZCTA-, and census-tract-level indexes (Figure 2). IRR and NCHS tended to homogenize rurality status. IRR designated most counties as micropolitan and used fewer than 50% of its values to categorize Wisconsin counties (Figure 2D). Since IRR categorizes counties with a normal distribution, it draws a large distinction between the most urban and most rural counties and homogenizes counties that fall between those extremes. NCHS classified 32 of 72 counties into its 1 rural code, preventing researchers from distinguishing between groups of patients who live in different rural communities (Figure 2C). UIC showed divergence in rurality, though recall its codes do not sequentially identify metropolitan, micropolitan, and rural counties, making it incorrect to interpret the UIC map along a continuum of rurality (Figure 2B). RUCC, RUCA, and RUCA(z) showed divergences in rurality across their respective code ranges, giving weight to their utility in measuring rurality across a continuum (Figures 2A, 2E, and 2F).

## **Index suitability as a continuous variable**

As researchers move away from binary rural-urban designations and towards studying rurality across the rural-urban continuum, indexes need to be conducive to continuous or multi-level ordinal coding for inclusion in analysis. Binary rural-urban designations may mask outcome variation within rural or urban groups. Continuous or multi-level ordinal variables may expose non-linear trends in cancer outcomes across the rural-urban continuum.<sup>19,35</sup> As indexes become more commonly employed as continuous variables, it becomes more important for researchers to use one index consistently across studies since index agreement decreases as the number of rurality groups used in analysis increases (Table 2).

RUCC, NCHS, IRR, RUCA (option 1), and RUCA(z) (option 1) are ordinal indexes that may be coded as continuous variables in analysis. UIC does not divide its non-metropolitan codes into micropolitan versus rural codes sequentially,<sup>31</sup> preventing it from being used as a continuous variable. NCHS only designates 1 code for micropolitan counties and 1 code for rural/non-core counties, restricting researcher's ability to distinguish between levels of rurality within subgroups of micropolitan or rural patients.<sup>32</sup> IRR, as a relative index, designates counties along a normal distribution, effectively homogenizing rurality status such that it is difficult to distinguish between counties of different rurality levels on a regional or local scale. For example, IRR uses fewer than half of its values to categorize Wisconsin's 72 counties (Figure 2D).

RUCA (option 2) and RUCA(z) (option 2), which include the x.1 secondary code as metropolitan, introduce ambiguity as to the most appropriate way to order codes as continuous variables. If RUCA and RUCA(z) are used as continuous variables, it should be based on primary RUCA codes only. RUCC includes multiple codes for metropolitan, micropolitan, and rural designations that are ordered sequentially, making it an unambiguous index conducive to use as a continuous variable in analysis.

### **Index feasibility to be used in cancer research**

The National Cancer Database (NCDB), North American Association of Central Cancer Registries (NAACCR), and Surveillance, Epidemiology, and End Results Program (SEER) registries include RUCC and RUCA indexes. RUCC is included in its original 9-code form, and RUCA is recoded into a binary rural-urban variable. Registry inclusion makes RUCC and RUCA accessible to researchers, though recoding RUCA into a binary variable prevents researchers from studying the rural-urban continuum or variation within rural or urban subgroups. RUCA is reduced to a binary variable to protect confidentiality and prevent case identification via the combination of census tract and county level data. Therefore, RUCC is the most accessible and specific index available for registry-based cancer research.

At a health-system and local level, a patient's county or ZIP code is more readily available than their census tract.<sup>36</sup> Counties and ZIP codes are standard fields in electronic health records (EHR) and health system registries. This difference in availability means that generally researchers use county or ZIP-code based indexes (RUCC, UIC, NCHS, IRR, RUCA(z)) in EHR or local registry cancer research. However, since ZIP codes change frequently and RUCA(z) versions are only available for non-census years (1998, 2004, 2006, 2013), researchers risk excluding cases from their analysis if a patient's ZIP code does not have a match in the chosen RUCA(z) file. With this and additional limitations to RUCA(z) outlined below, it is preferable to avoid

ZIP code and ZCTA-based indexes.<sup>36,41,42</sup> Therefore, county-based indexes, namely RUCC for the reasons listed above, are preferred for health-system and local level research.

## **Indexes over time**

The choice of which index year to employ should be based on the role rurality is hypothesized to play in one's study. Rurality as an exposure may be calculated on a past version of an index, whereas rurality as an enabler or barrier to care should be calculated from a current version, relative to the year(s) of study. When rurality is investigated as an exposure, patients may be misclassified as they move. This may obstruct the rurality designation of interest.

Since IRR is relative to other counties, absolute changes in rurality over time are masked, making this index inappropriate for longitudinal studies.

## **Considerations for RUCA(z) and ZCTAs**

RUCA(z) approximates RUCA and is not calculated directly from ZCTA-level characteristics. ZCTAs approximate ZIP codes, and it is possible for a patient's ZIP code to differ from their ZCTA.<sup>21</sup> ZIP Codes are subject to change, as evidenced by the regular ZIP Code updates released by the U.S. Postal Service,<sup>43</sup> so a patient's ZIP code at diagnosis may not match their ZIP code for the year of study, irrespective of whether they have moved. These approximations and ongoing administrative changes may introduce inaccuracies into RUCA(z) and expose multiple opportunities for patient misclassification.<sup>17,41</sup> The difference between RUCA and RUCA(z) geographical unit distributions across the U.S. and Wisconsin highlight that RUCA(z) may not adequately approximate the census-tract based RUCA (Figure 3 and Supporting Figure 1A and 1B). The RUCA(z) map shows irregular ZCTA boundaries, supporting the advice that researchers should be wary of using ZCTAs as a geographical unit (Figure 2F).<sup>41,42</sup> The extent of misclassification could be further studied if land area and population data is made available for ZCTA 2013 or for the year of the next RUCA(z) release that can be compared to census-tract level data. Furthermore, as opposed to RUCC, UIC, NCHS, and RUCA, RUCA(z) is not published by a government agency, which makes its ongoing availability less assured.

## **Limitations**

We evaluated rural-urban indexes for their ability to categorize cancer patients across the rural-urban continuum, geographical unit, land area, and population distributions, and percent agreement. We did not have access to a cancer patient data set with patient-specific ZIP codes, census tracts, and counties, though, and were unable to obtain the percent agreement at a data set-level across indexes that utilize these 3 geographic units. County, ZCTA, and Census Tract land area varies by state, and we did not evaluate land

area distributions on a per-state level. This consideration is especially important for states with fewer and larger counties.

## Conclusions

Utilizing the Rural-Urban Continuum Code (RUCC) index across cancer research that includes a rural-urban component will increase reproducibility and comparability of results and eliminate index choice as a source of discrepancy across studies. Counties are a stable geographic unit of analysis and are readily available at the patient level within local, regional, and national research settings and within electronic health record and registry data sources. RUCC includes a spectrum of codes across metropolitan, micropolitan, and rural communities. If necessary, it can be grouped into a binary or ternary variable. RUCC indexes for 1993, 2003, and 2013 are available in several national registries at a discrete level, enabling researchers to study rural-urban residence across a continuum rather than as a binary factor impacting patients' treatments and outcomes. ZCTA-based indexes should be avoided as ZCTAs approximate actual ZIP code boundaries, change frequently, and represent administrative rather than geographical areas<sup>42</sup>. Government agencies should work towards a census-block level measure of rurality that is accessible to researchers without compromising patient confidentiality. The census block provides the most specific unit of geographical analysis and therefore minimizes the risk of masking disparities within larger geographical units. Finally, just as a patient is more than their age or ethnicity, a patient is more than their rural residence. Researchers should continue to include social, economic, and health-related variables alongside rurality in cancer prevention and outcomes research to understand the many factors impacting disparities in cancer treatment and outcomes and how these factors interact differently across geographical and care settings.

## Declarations

**Ethics Approval:** The pancreas cancer registry data use was exempt as human subjects research by the University of Wisconsin Health Sciences IRB, ID # 2019-0155, expiration 4/26/2024.

**Availability of data and materials:** The datasets generated and/or analyzed during the current study are not publicly available due to HIPAA restrictions with personal health information for the registry patients, but are available from the corresponding author on reasonable request. The other datasets analyzed are publicly available and are referenced as such in the manuscript.

## References

1. Gosain R, Ball S, Rana N, et al. Geographic and demographic features of neuroendocrine tumors in the United States of America: A population-based study. *Cancer*. 2020;126(4):792–799.
2. Pruitt SL, Eberth JM, Morris ES, Grinsfelder DB, Cuate EL. Rural-Urban Differences in Late-Stage Breast Cancer: Do Associations Differ by Rural-Urban Classification System? *Tex Public Health J*. 2015;67(2):19–27.

3. Zahnd W, Fogleman A, Jenkins W. Rural–Urban Disparities in Stage of Diagnosis Among Cancers With Preventive Opportunities. *American Journal of Preventive Medicine*. 2018;54.
4. Baldwin L-M, Patel S, Andrilla CHA, Rosenblatt RA, Doescher MP. Receipt of recommended radiation therapy among rural and urban cancer patients. *Cancer*. 2012;118(20):5100–5109.
5. Hao Y, Landrine H, Jemal A, et al. Race, neighbourhood characteristics and disparities in chemotherapy for colorectal cancer. *J Epidemiol Community Health*. 2011;65(3):211–217.
6. Atkins GT, Kim T, Munson J. Residence in Rural Areas of the United States and Lung Cancer Mortality. Disease Incidence, Treatment Disparities, and Stage-Specific Survival. *Annals of the American Thoracic Society*. 2017;14(3):403–411.
7. Baldwin LM, Andrilla CH, Porter MP, Rosenblatt RA, Patel S, Doescher MP. Treatment of early-stage prostate cancer among rural and urban patients. *Cancer*. 2013;119(16):3067–3075.
8. Rana N, Gosain R, Lemini R, et al. Socio-Demographic Disparities in Gastric Adenocarcinoma: A Population-Based Study. *Cancers*. 2020;12(1):157.
9. Loberiza FR, Cannon AJ, Weisenburger DD, et al. Survival Disparities in Patients With Lymphoma According to Place of Residence and Treatment Provider: A Population-Based Study. *Journal of Clinical Oncology*. 2009;27(32):5376–5382.
10. Bertens KA, Massman JD, 3rd, Helton S, et al. Initiation of adjuvant therapy following surgical resection of pancreatic ductal adenocarcinoma (PDAC): Are patients from rural, remote areas disadvantaged? *J Surg Oncol*. 2018;117(8):1655–1663.
11. Yao N, Alcalá HE, Anderson R, Balkrishnan R. Cancer Disparities in Rural Appalachia: Incidence, Early Detection, and Survivorship. *J Rural Health*. 2017;33(4):375–381.
12. Hashibe M, Kirchhoff AC, Kepka D, et al. Disparities in cancer survival and incidence by metropolitan versus rural residence in Utah. *Cancer Med*. 2018;7(4):1490–1497.
13. 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 Epidemiology Biomarkers & Prevention*. 2017;26(7):992–997.
14. Moy E, Garcia MC, Bastian B, et al. Leading Causes of Death in Nonmetropolitan and Metropolitan Areas— United States, 1999–2014. *MMWR Surveillance Summaries*. 2017;66(1):1–8.
15. Henley J, Anderson R, Thomas C, Massetti G, Peaker B, Richardson L. Invasive Cancer Incidence, 2004–2013, and Deaths, 2006–2015, in Nonmetropolitan and Metropolitan Counties — United States. *MMWR Surveillance Summaries*. 2017;66:1–13.
16. Meilleur A, Subramanian SV, Plaschak 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.
17. Yaghjyan L, Cogle C, Deng G, et al. Continuous Rural-Urban Coding for Cancer Disparity Studies: Is It Appropriate for Statistical Analysis? *International Journal of Environmental Research and Public Health*. 2019;16(6):1076.

18. Delavar A, Feng Q, Johnson KJ. Rural/urban residence and childhood and adolescent cancer survival in the United States. *Cancer*. 2019;125(2):261–268.
19. McLafferty S, Wang F. Rural reversal? *Cancer*. 2009;115(12):2755–2764.
20. Zahnd WE, Askelson N, Vanderpool RC, et al. Challenges of using nationally representative, population-based surveys to assess rural cancer disparities. *Preventive Medicine*. 2019;129:105812.
21. ZIP Code Tabulation Areas (ZCTAs). United States Census Bureau. <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html>. Published 2020. Accessed March 20, 2020.
22. Swords DS, Mulvihill SJ, Skarda DE, et al. Hospital-level Variation in Utilization of Surgery for Clinical Stage I-II Pancreatic Adenocarcinoma. *Ann Surg*. 2019;269(1):133–142.
23. Zahnd WE, Davis MM, Rotter JS, et al. Rural-urban differences in financial burden among cancer survivors: an analysis of a nationally representative survey. *Support Care Cancer*. 2019;27(12):4779–4786.
24. 2010 Census Urban and Rural Classification and Urban Area Criteria. United States Census Bureau. <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html>. Published 2019. Accessed March 13, 2020.
25. Core based statistical areas (CBSAs), metropolitan divisions, and combined statistical areas (CSAs). United States Census Bureau. <https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html>. Published 2013. Accessed March 12, 2020.
26. Frontier and Remote Area Codes. United States Department of Agriculture Economic Research Service. <https://www.ers.usda.gov/data-products/frontier-and-remote-area-codes/>. Published 2015. Accessed March 31, 2020.
27. Rural-Urban Commuting Area Codes. United States Department of Agriculture Economic Research Service. <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/>. Published 2019. Accessed December 19, 2019.
28. RUCA Data: ZIP Code RUCA Approximation. WWAMI Rural Health Research Center. <https://depts.washington.edu/uwruca/ruca-approx.php>. Published 2007. Accessed December 19, 2019.
29. Rural-Urban Commuting Area (RUCA) Codes, Zip Code File: version 3.10. Administration for Community Living: AGing, Independence, and Disability (AGID) Program. [https://agid.acl.gov/Resources/OAA\\_SPR.aspx](https://agid.acl.gov/Resources/OAA_SPR.aspx). Published 2013. Accessed March 31, 2020.
30. Rural-Urban Continuum Codes. United States Department of Agriculture Economic Research Service. <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/>. Published 2013. Accessed December 13, 2019.
31. Urban Influence Codes. United States Department of Agriculture Economic Research Service. <https://www.ers.usda.gov/data-products/urban-influence-codes/>. Published 2013. Accessed December 13, 2019.
32. NCHS Urban-Rural Classification Scheme for Counties. Centers for Disease Control and Prevention National Center for Health Statistics. [https://www.cdc.gov/nchs/data\\_access/urban\\_rural.htm](https://www.cdc.gov/nchs/data_access/urban_rural.htm). Published 2017. Accessed March 13, 2020.

33. Waldorf B, Kim A. The Index of Relative Rurality (IRR): US County Data for 2000 and 2010. In:2018.
34. Waldorf B, and Kim, Ayoung. Defining and Measuring Rurality in the US: From Typologies to Continuous Indices. 2015.  
[http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse\\_168031.pdf](http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_168031.pdf).  
Published April 2015.
35. Hall SA, Kaufman JS, Ricketts TC. Defining Urban and Rural Areas in U.S. Epidemiologic Studies. *Journal of Urban Health*. 2006;83(2):162–175.
36. Zahnd WE, Mueller-Luckey GS, Fogleman AJ, Jenkins WD. Rurality and Health in the United States: Do Our Measures and Methods Capture Our Intent? *J Health Care Poor Underserved*. 2019;30(1):70–79.
37. GCT-PH1: Population, Housing Units, Area, and Density: 2010 - United States – County by State; and for Puerto Rico. In: Bureau USC, ed: United States Census Bureau; 2014.
38. Revised Delineations of Metropolitan Statistical Areas, Micropolitan Statistical Areas, and Combined Statistical Areas, and Guidance on Uses of the Delineations of These Areas In: Budget OoMa, ed. Washington, D.C.: Office of Managment and Budget; 2013.
39. Freeman AT, Kuo M, Zhou L, et al. Influence of Treating Facility, Provider Volume, and Patient-Sharing on Survival of Patients With Multiple Myeloma. *J Natl Compr Canc Netw*. 2019;17(9):1100–1108.
40. Peppercorn J, Horick N, Houck K, et al. Impact of the elimination of cost sharing for mammographic breast cancer screening among rural US women: A natural experiment. *Cancer*. 2017;123(13):2506–2515.
41. Krieger N, Waterman P, Chen JT, Soobader M-J, Subramanian SV, Carson R. Zip Code Caveat: Bias Due to Spatiotemporal Mismatches Between Zip Codes and US Census–Defined Geographic Areas–The Public Health Disparities Geocoding Project. *American Journal of Public Health*. 2002;92(7):1100–1102.
42. Grubestic TH, Matisziw TC. On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *International Journal of Health Geographics*. 2006;5(1):58.
43. 2020 Postal Bulletin/PO Changes. United States Postal Service Postal Pro.  
<https://postalpro.usps.com/postal-bulletin-changes>. Published 2020. Updated July 7, 2020. Accessed November 18, 2020.

## Tables

**Table 1: Agreement between binary and ternary Rural-Urban Indexes at the United States Population and University of Wisconsin - Health Pancreatic Cancer Registry (1,539 patients) levels.** Count and percent agreement of population and Registry patients across Rural-Urban Continuum Codes (RUCC), Urban Influence Codes (UIC), National Center for Health Statistics (NCHS) Urban-Rural Classification Scheme for Counties, Index of Relative Rurality (IRR), Rural-Urban Commuting Area (RUCA), and ZIP Code Tabulation Area Rural-Urban

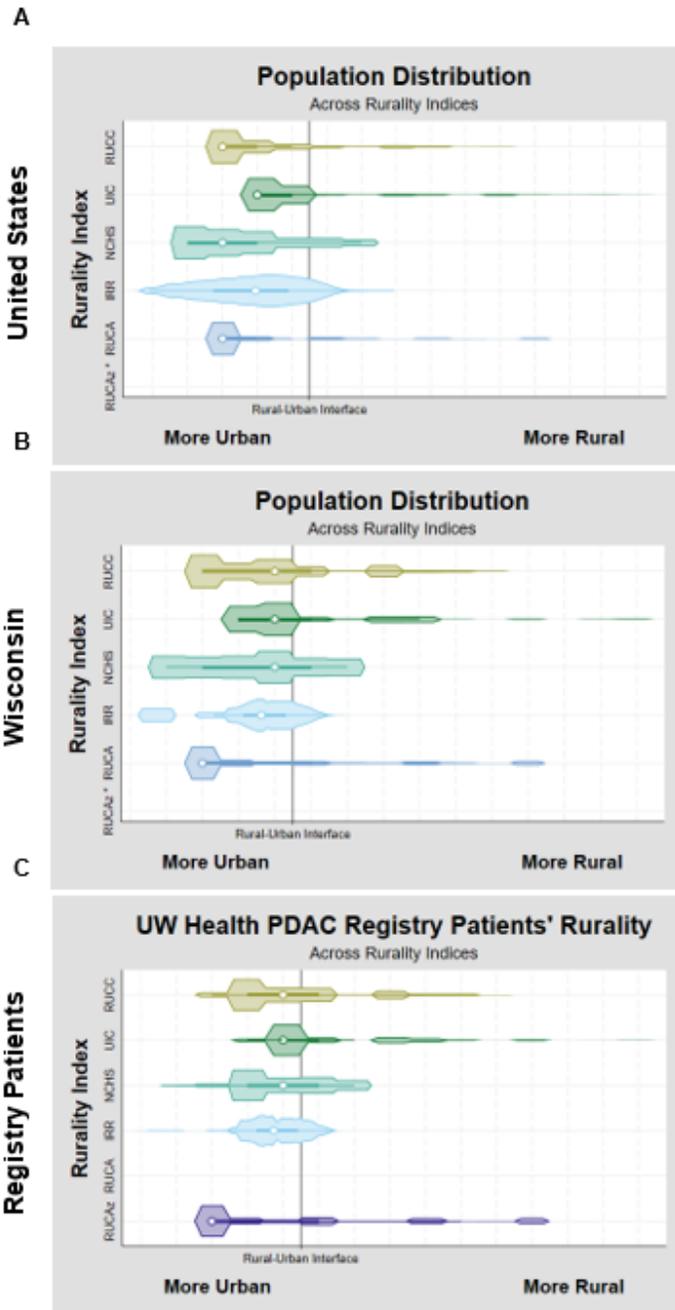
Commuting Area (RUCA(z)) when those indexes are treated as binary and ternary rural-urban variables. The Index of Relative Rurality was missing data for 0.1% of the US Population and Registry Patients.

**Table 1: Agreement between binary and ternary Rural-Urban Indexes at the United States Population and University of Wisconsin - Health Pancreatic Cancer Registry (1,539 patients) levels.**

Variable Type	Indexes Included	Agreement and Disagreement	Rural - Urban Category	United States Population			Registry Patients*		
				Count	Percent	% agree- or disagreement	Count	Percent	% agree- or disagreement
Binary	RUCC, UIC, NCHS, IRR, RUCA	Agree	Metropolitan	247,994,082	80.3%	88.8%	908	59.0%	73.4%
			Non-metropolitan	26,248,722	8.5%		222	14.4%	
		Disagree	Metropolitan & Non-metropolitan	34,347,840	11.1%	11.1%	408	26.5%	26.5%
Binary	RUCC & RUCA	Agree	Metropolitan	252,306,164	81.7%	94.9%	920	59.8%	91.0%
			Non-Metropolitan	40,789,077	13.2%		480	31.2%	
		Disagree	Metropolitan & Non-metropolitan	15,650,297	5.1%	5.1%	138	9.0%	9.0%
Ternary	RUCC, UIC, NCHS, IRR, RUCA	Agree	Metropolitan	247,994,082	80.3%	83.4%	908	59.0%	60.4%
			Micropolitan	8,722,475	2.8%		21	1.4%	
			Rural/Noncore	886,289	0.3%		1	0.1%	
		Disagree - 1 level	Metropolitan & Micropolitan	24,208,656	7.8%	13.2%	244	15.9%	28.8%
			Micropolitan & Rural/Noncore	16,639,958	5.4%		200	13.0%	
		Disagree - 2 levels	Metropolitan & Rural/Noncore	4,639,840	1.5%	1.5%	28	1.8%	1.8%
		Disagree - all levels	Metropolitan, Micropolitan, & Rural/Noncore	5,499,314	1.8%	1.8%	136	8.8%	8.8%
Ternary	RUCC & RUCA	Agree	Metropolitan	252,306,164	88.4%	88.8%	920	59.8%	74.9%
			Micropolitan	405,347	0.1%		184	12.0%	
			Rural/Noncore	827,415	0.3%		49	3.2%	
		Disagree - 1 level	Metropolitan & Micropolitan	10,777,424	3.8%	9.5%	87	5.7%	21.7%
			Micropolitan & Rural/Noncore	16,270,060	5.7%		247	16.0%	
		Disagree - 2 levels	Metropolitan & Rural/Noncore	4,872,873	1.7%	1.7%	51	3.3%	3.3%

\* RUCA(z) was used in place of RUCA for the Registry patients since patient ZIP codes were available in the registry and census tracts were not.

## Figures



**Figure 1**

**Comparison of rural-urban index distributions across the United States population (A), Wisconsin population (B), and University of Wisconsin – Health Pancreatic Cancer Registry patients (C).** Each index is standardized such that the 'Rural-Urban Interface' line lies between those values that the index designated as urban (left of line) and rural (right of line). Center points indicate the median of the index, boxes indicate the inter-quartile range, and spikes indicate the upper- and lower-adjacent values (1.5 times the inter-quartile range). Rural-Urban Continuum Codes (RUCC), Urban Influence Codes (UIC), National Center for Health Statistics (NCHS) Urban-Rural Classification Scheme for Counties, Index of Relative Rurality (IRR), Rural-Urban Commuting Area (RUCA), and ZIP Code Tabulation Area Rural-Urban Commuting Area (RUCA(z))

\* RUCA(z) is based on 2013 ZCTAs. United States and Wisconsin Population data is not available at the ZCTA-level for 2013, therefore those distributions are excluded from the figure.

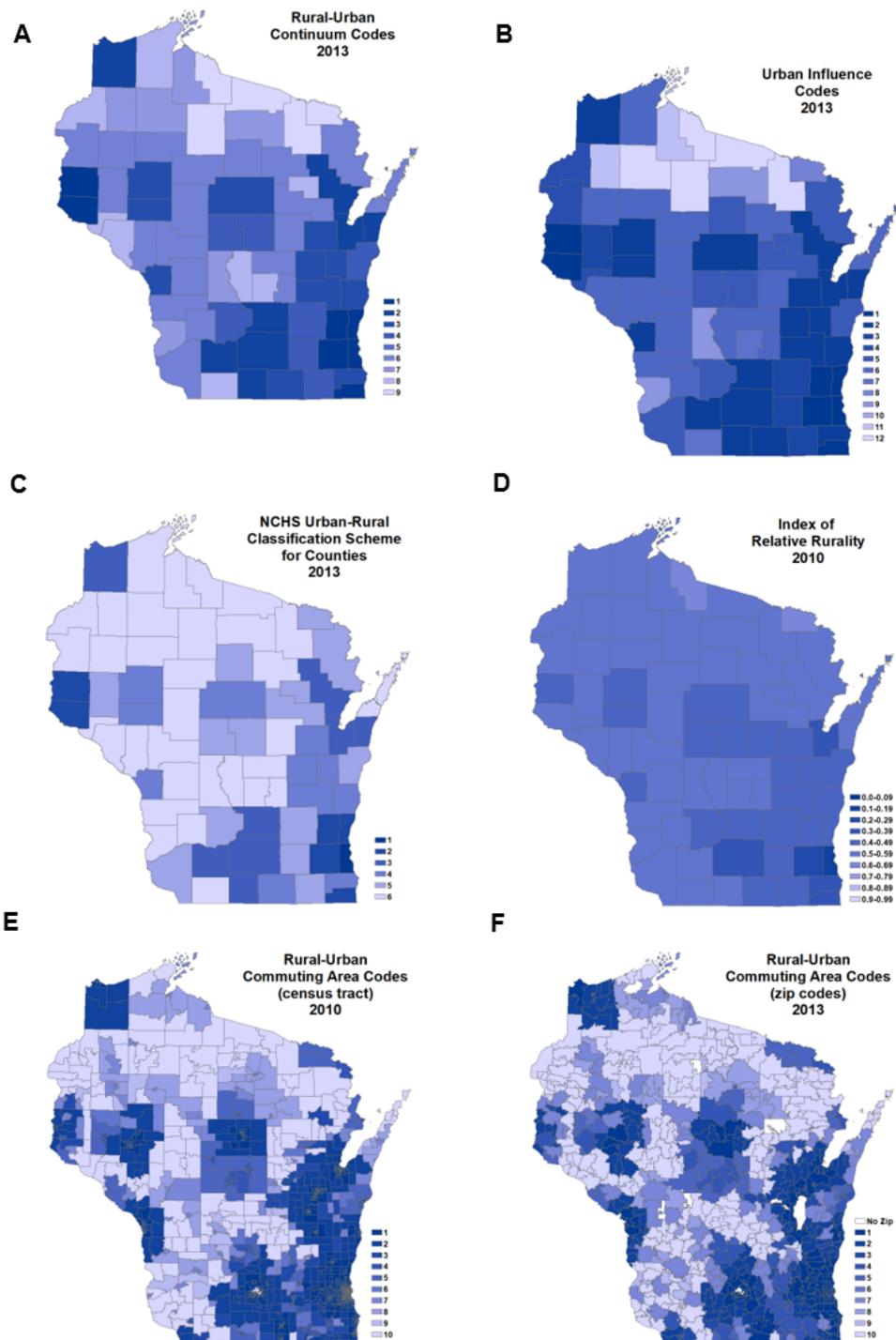
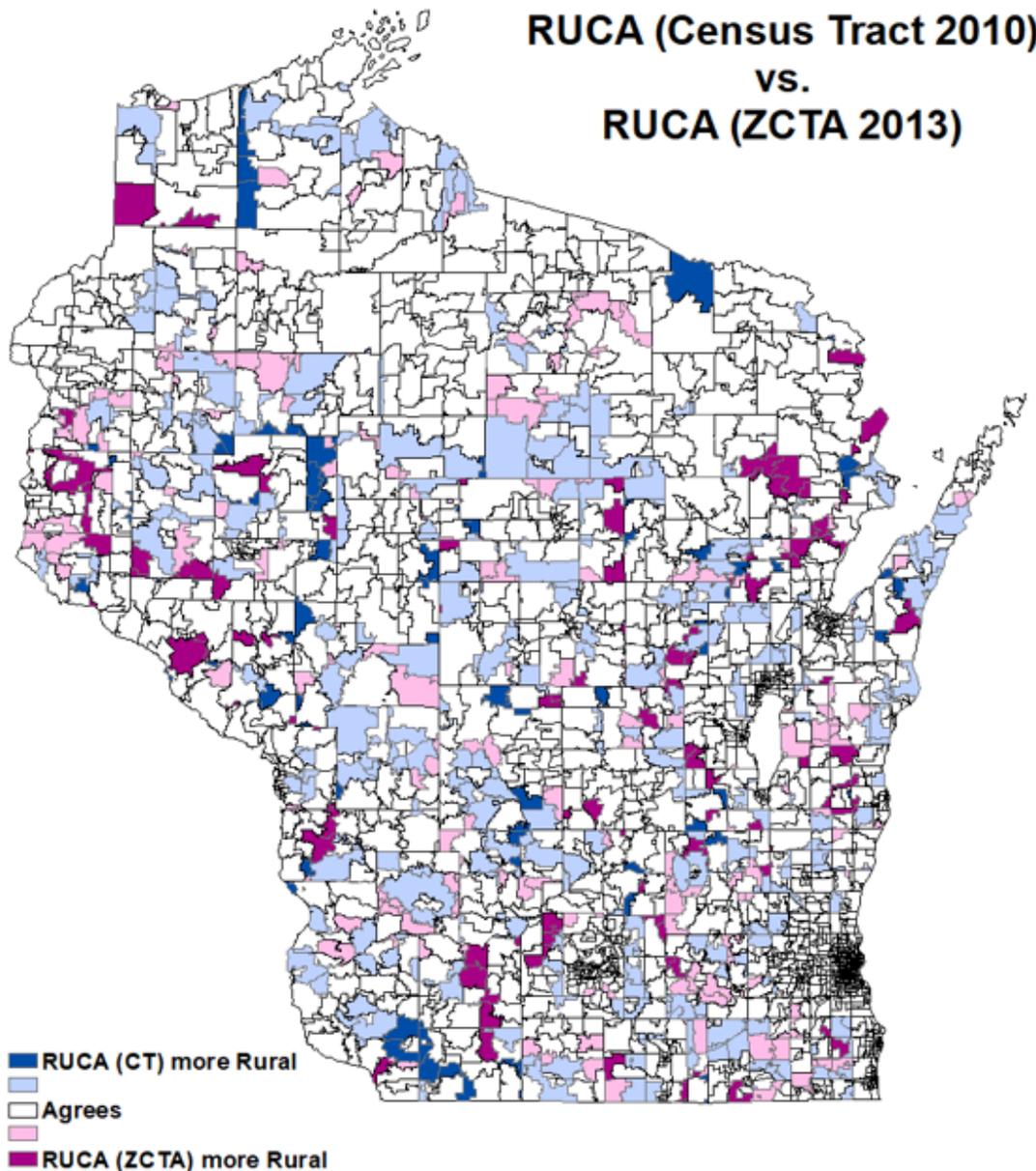


Figure 2

**Maps of Wisconsin rurality by 6 different rural-urban indexes.** Darker colors indicate more urban areas, and lighter colors indicate more rural areas. Rural-Urban Continuum Codes (RUCC) (A), Urban Influence Codes (UIC) (B), National Center for Health Statistics (NCHS) Urban-Rural Classification Scheme for Counties (C),

and Index of Relative Rurality (IRR) (D) maps are at the county level. Rural-Urban Commuting Area (RUCA) map (E) is at the census tract level. ZIP Code Tabulation Area (ZCTA) Rural-Urban Commuting Area (RUCA(z)) (F) map is at the ZCTA level.



**Figure 3**

**Comparing Rural Urban Commuting Area (RUCA) indexes (primary codes only) based on Census Tract (CT) and ZIP Code Tabulated Area (ZCTA) across Wisconsin.** Dark blue indicates Census Tract-based RUCA (RUCA (CT)) categorized areas as more rural than ZCTA-based RUCA (RUCA (ZCTA)) by a code difference of 5 to 9 (2.6% of land area) and light blue indicates RUCA (CT) categorized areas as more rural than RUCA (ZCTA) by a code difference of 1 to 4 (16.3% of land area). Dark pink indicates RUCA (ZCTA) categorized areas as more rural than RUCA (CT) by a code difference of 5 to 9 (3.9% of land area) and light pink

indicates RUCA (ZCTA) categorized areas as more rural than RUCA(CT) by a code difference of 1 to 4 (6.2% of land area). RUCA (CT) and RUCA (ZCTA) agreed for white areas (71.1% of land area).

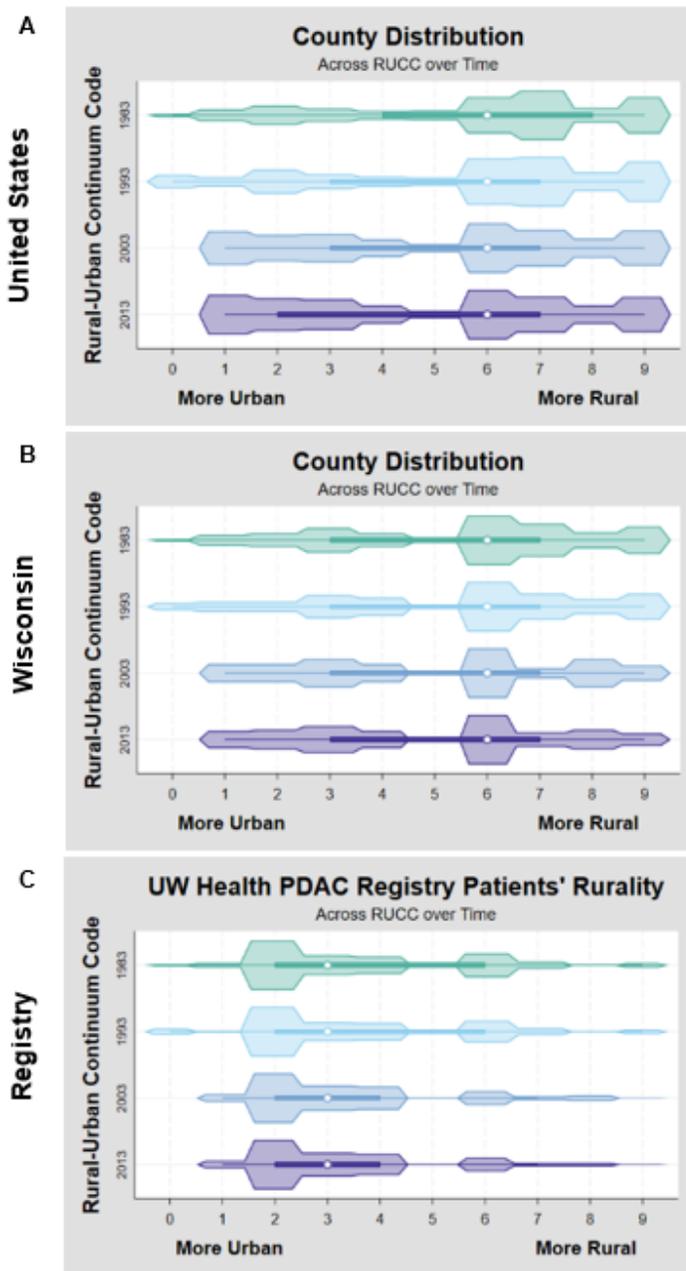


Figure 4

Rurality over time: Rural-Urban Continuum Code (RUCC) Geographical Unit distributions across the United States (A), Wisconsin (B), and University of Wisconsin – Health Pancreatic Cancer Registry Patients (C). Geographical unit is measured at the county-level for RUCC. Center points indicate the median code of the index, boxes indicate the inter-quartile range, and spikes indicate the upper- and lower-adjacent values (1.5 times the inter-quartile range). Changes in the RUCC range between 1993 and 2003 reflect methodology changes.

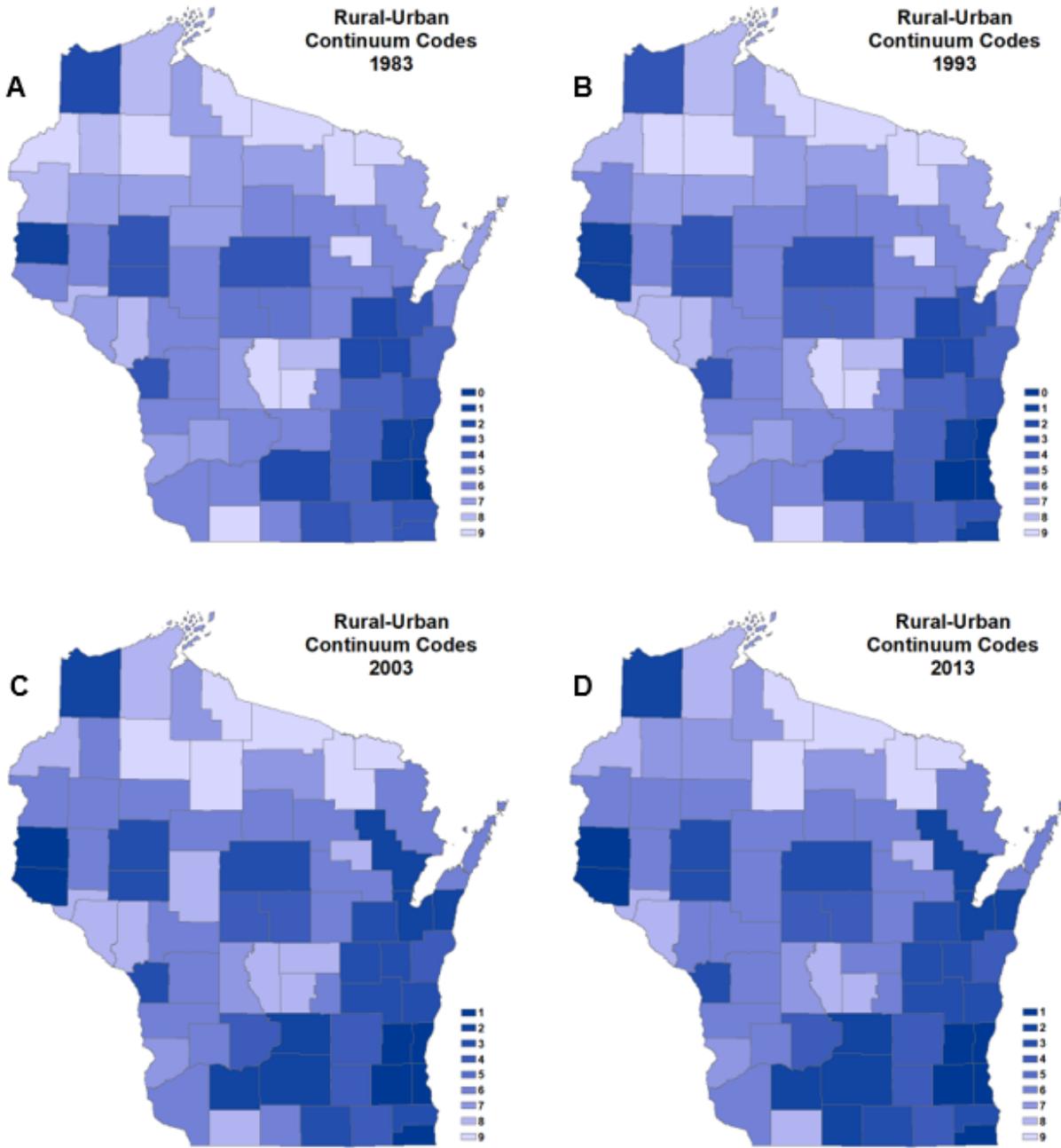


Figure 5

Rural-Urban Continuum Code (RUCC) changes across Wisconsin over time. Maps showing the rural-urban continuum in Wisconsin over time by RUCC for 1983 (A), 1993 (B), 2003 (C), and 2013 (D). Darker colors indicate more urban areas, and lighter colors indicate more rural areas. Maps are at the county level.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Manuscript.SupportingInformation.docx](#)