Categorizing community type for epidemiologic evaluation of community factors and chronic disease across the United States

Tara P. McAlexander, PhD, MPH[1], **Yasemin Algur, MPH**[1], **Brian S. Schwartz, MD, MS**[2,3,4], **Pasquale E. Rummo, PhD, MPH**[5], **David C. Lee, MD, MS**[5,6], **Karen R. Siegel**[7], **Victoria Ryan, MHS**[1], **Nora L. Lee, PhD**[1], **Gargya Malla, MD, MPH**[8], **Leslie A. McClure, PhD, MS**[1]

[1]Department of Epidemiology and Biostatistics, Drexel University Dornsife School of Public Health, Philadelphia, Pennsylvania, United States

[2]Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States

[3]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States

[4]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States

[5]Department of Population Health, NYU School of Medicine, New York, New York, United States

[6]Department of Emergency Medicine, NYU School of Medicine, New York, New York, United States

[7]Division of Diabetes Translation, Centers for Disease Control and Prevention, Atlanta, GA, United States

[8]Department of Epidemiology, The University of Alabama at Birmingham, Birmingham, Alabama, United States

Abstract

Existing classifications of community type do not differentiate urban cores from surrounding non-rural areas, an important distinction for analyses of community features and their impact on health. Inappropriately classified community types can introduce serious methodologic flaws in epidemiologic studies and invalid inferences from findings. To address this, we evaluate a modification of the United States Department of Agriculture's Rural Urban Commuting Area codes at the census tract, propose a four-level categorization of community type, and compare this with existing classifications for epidemiologic analyses. Compared to existing classifications, our method resulted in clearer geographic delineations of community types within urban areas.

Corresponding Author: Tara P. McAlexander, Dornsife School of Public Health, Drexel University, Nesbitt Hall, 3215 Market Street, Philadelphia, PA 19104, tpm58@drexel.edu.

Introduction

Geographic disparities in the prevalence of chronic diseases such as obesity and diabetes are present throughout the United States (U.S.) (Voeks et al., 2008, Singh et al., 2008). Understanding the causes of these disparities is crucial for designing effective policies and interventions to reduce the public health burden of chronic diseases and ensuring equity in health outcomes, and increasing research implicates place-based causes such as access to healthful resources including stable housing, healthy food, recreational spaces, and active modes of transportation in the prevention of chronic disease (Larson et al., 2009, Levasseur, 2015). These findings are consistent with several epidemiologic studies, motivated by strong *a priori* theory, that have demonstrated the relevance of community- or neighborhood-level factors to health disparities (Siegrist, 2000, Krieger, 2003, Macintyre et al., 2002, Boardman, 2004, Augustin et al., 2008, Kershaw et al., 2015, Merkin et al., 2007) and that these community-level factors have potential for targeted interventions (Smurthwaite and Bagheri, 2017). Communities in the U.S. range from rural to suburban, small towns and cities, and micropolitan and metropolitan cores, (Morrill et al., 1999, Adams et al., 1999, Hart et al., 2005, Economic Research Service, 2019), which may necessitate different types of interventions based on which neighborhood-level factors are most important. To effectively evaluate the role of place-based and community-level features in geographic health disparities, appropriate consideration and classification of the range of community types is needed.

Without appropriate classification of the range of communities in the U.S. in epidemiologic evaluations of place-based factors and individual-level health outcomes, several methodological challenges emerge. First, many place-based factors cluster at the community level, and different communities (e.g., urban vs. rural) differ by multiple domains of measures such as poverty, housing, education, employment, and racial composition (Messer et al., 2006). Second, the co-occurrence or clustering of place-based factors (e.g., walkability and healthy food access) might differ by community type; therefore it is important to evaluate these factors within the community contexts in which they occur. Third, the distributions of place-based measures do not necessarily overlap across all ranges of communities, introducing the methodological challenge of non-positivity when evaluating place-based measures across community types (Ahern et al., 2013). This challenge refers to the violation of the positivity assumption that observed exposure variables should vary within strata of confounding variables (Petersen et al., 2012). When individual-level exposures are highly correlated within community contexts, such as individual-level exposures to neighborhood violence, it is possible that individual-level exposures do not vary within a community type, or at least do not vary sufficiently across strata of potential confounding variables (Ahern et al., 2013). If we believe that community types could confound the association between an exposure variable and a health outcome, adjusting for community types in an inferential model would be inadequate to

account for non-overlapping distributions of exposure variables that vary by community types. Fourth, there may be differential item functioning, that is, differential measurement properties and associated errors of place-based factors, by community type (Sapawi and Said, 2012, Jones, 2019). For example, measures of high school graduation rates might have different socioeconomic significance in rural vs. urban or suburban areas. Guidance on how to address these complexities is lacking in epidemiologic research but Messer et al. recommends, as a first step, stratifying analyses by relevant classifications of community type to help mitigate these methodological challenges in population-based epidemiology studies that encompass a variety of community types (Messer et al., 2010).

Researchers have grappled with defining and classifying neighborhood and community typologies through a variety of ways (Hall et al., 2006). These strategies for classifying neighborhoods include: the use of secondary data to delineate neighborhoods (Rhew et al., 2017, Kolak et al., 2020); direct observation of neighborhoods (Masoumi et al., 2019); and reports of neighborhood perceptions from residents (Basta et al., 2010). Some researchers have used data-driven techniques such as latent profile analyses (Humphrey et al., 2019) or k-means clustering (Kolak et al., 2020) to classify neighborhoods based on several domains of interest, but limitations to these methods include a high computational burden; dependence on the domains one chooses; difficulty in interpretation; and lack of reproducibility in differing contexts. Direct observation of neighborhoods and reports of neighborhood perceptions from residents are both time and resource intensive and might therefore be best suited for small areas. Because of these limitations, obtaining accurate community classifications based on administrative boundaries (i.e., census tracts or counties) is most practical and relevant for large-scale population-based studies. However, a challenge is that existing community classifications, such as the U.S. Department of Agriculture's (USDA) Rural Urban Commuting Area (RUCA) designations (Economic Research Service, 2019) that are available at census tract boundaries in the U.S., were not generated for the purpose of evaluating features of the environment and health outcomes. For example, RUCA designations include 10 levels of community classifications: metropolitan area core; metropolitan area high commuting; metropolitan area low commuting; micropolitan area core; micropolitan high commuting; micropolitan low commuting; small town core; small town high commuting; small town low commuting; and rural areas (Economic Research Service, 2019), many of which are determined by patterns and practices of commuting into urban areas, and not the actual land area contained within the census tract boundaries. Inspection of maps of these RUCA designations at the census tract level demonstrates that large geographic areas with wide variation in population density are classified as "metropolitan core," the most urban of these 10 categories, suggesting that these classifications of community type do not differentiate urban cores from surrounding non-rural areas. This observation prompted an exploration of other approaches to characterize the original RUCA designations that have been used in prior research (Weeks et al., 2004, Euler et al., 2019, Yaghjyan et al., 2019) as well as our own, with the goal of finding a method for community type classification and stratification that would begin to address challenges of multidimensionality, differential dimension clustering, differential item functioning, and non-positivity (Messer et al., 2006, Ahern et al., 2013).

The Diabetes LEAD (Location, Environmental Attributes, and Disparities) Network (The Diabetes LEAD Network, 2020), funded by the Centers for Disease Control and Prevention, has been performing epidemiologic studies to evaluate community contributors to geographic disparities in individual risk of type 2 diabetes across the U.S. To classify communities for the LEAD Network, we first examine existing approaches to classifying the neighborhoods on a national level and compare these to a new approach that we developed for this purpose, which has additional criteria concerning the locations of Census-designated urbanized areas (UA) and urban clusters (UC). We then compare these approaches by geographic distributions and by several U.S. census-derived variables of relevance to type 2 diabetes. Our goal in this analysis was to find a community classification that minimized within-category variation and maximized between-category differences by these variables nationally. We hypothesize that this classification method will better differentiate distributions of key variables related to both community type and access to healthful resources (e.g., percent of developed land, street connectivity, educational attainment, and percent of households without car ownership) in urban cores from surrounding non-rural areas. In so doing, this method will lead to minimizing the aforementioned methodological challenges (i.e., multi-dimensionality, differential item functioning, non-overlapping distributions) when used as a variable to stratify analyses of features of the environment with individual-level health outcomes.

Materials and Methods

Existing classifications

We first evaluated three categorizations of RUCA designations used in published epidemiologic studies (Weeks et al., 2004) (Euler et al., 2019) (Yaghjyan et al., 2019). In the first study, the 10 RUCA categories were collapsed into three categories of urban, suburban, and rural (Weeks et al., 2004). The second study collapsed the 10 RUCA categories into four categories of metropolitan, micropolitan, small town, and rural (Euler et al., 2019). The third study also collapsed the 10 RUCA categories into four categories referred to as urban, large rural town, small rural town, and isolated rural (Yaghjyan et al., 2019). We compared these three classifications by mapping the spatial distributions nationally and in select cities in different U.S. regions (Atlanta, Georgia; Chicago, Illinois; Philadelphia, Pennsylvania; Seattle, Washington) using ArcGIS® Pro (ESRI, 2019).

Modifications to RUCA

We obtained the 2010 primary RUCA designations for all census tracts in the contiguous U.S. (Economic Research Service, 2019) and tract-level measures of land area from the 2010 U.S. Census (US Census Bureau, 2010c) for all census tracts in the contiguous U.S. and collapsed these into four categories. To do so, we first calculated the percentage of land area contained within Census-designated UAs or UCs (US Census Bureau, 2010b). The U.S. Census defines UAs as areas of 50,000 or more people, and UCs as areas with at least 2,500 and less than 50,000 people; each is distributed across the U.S. (Figure 1). RUCA designates census tracts as metropolitan core if more than 30% of the tract's population is living in an UA. Because we wanted to distinguish urban from more sprawling tracts, we modified this criterion so that tracts were categorized as urban if 100% of their land area

was contained within the UA. Because UCs are much smaller than UAs, we required that 50% or more of a tract's land area was in an UC in order to be classified as micropolitan or small town core. Tracts that did not meet our inclusion criteria in an UA or UC were assigned their original RUCA code. Combining RUCA codes with these modifications, we categorized tracts into three groups—urban, suburban/small town, and rural. The urban category consisted of tracts that met our modified criteria for urban as described above. The suburban/small town category consisted of tracts that either fulfilled our UC criteria, or were originally classified as metropolitan core but did not meet our UA criteria to be considered urban, while the remaining tracts were assigned to the rural category.

After examining distribution of census tract assignments by our algorithm, we noticed that there was large variation in the sizes of urban tracts; we therefore further divided the urban category into two based on the distribution of the land area (i.e., geographic size) of these urban census tracts. Tracts with land area below the 40th percentile of the land area distribution (i.e., geographically smaller tracts) were classified as "higher density urban" and tracts with land area in the higher 60th percent of the land area distribution (i.e., geographically larger tracts) were classified as "lower density urban." We chose the 40th percentile of the land area distribution to divide urban tracts after comparing several cut-points (i.e., 25th vs. 75th, 30th vs. 70th percentile) and examining how these each classified tracts within cities in several regions in the U.S. Figure 2 illustrates how we included land area in our classification process. Additionally, we compared the geographic distributions of community type classifications by the three previously used modified RUCA methods to ours, nationally and in the selected U.S. cities of differing sizes and from different regions of the country.

Comparison of relevant domains by community classification

Last, we examined distributions of relevant variables across these community type classification methods for the ability of each to maximize between-group differences. We obtained data from the 2010 1-year American Community Survey (US Census Bureau, 2010a) for several measures: poverty, households with income of less than $30,000/year, households receiving public assistance, households without a car, unemployment, and educational attainment. Similarly, we obtained data from the Environmental Systems Research International (ESRI) 2009 Vintage Street data to compute measures of intersection density, average block length, and street connectivity using ArcGIS Pro® (ESRI, 2019). We examined the distributions (mean, standard deviation, and median) of these variables across all community type classifications, and further examined these variables among tracts that were classified as rural in the LEAD classification method but not in the other methods.

Results

The distribution of census tracts in the contiguous U.S. (n = 72,539) in various categories differed greatly by the four approaches that we evaluated: Weeks et al., Euler et al., Yaghjyan et al., and our method, the LEAD classification (Table 1). Notably, the frequencies of tracts in the LEAD community type classification method show a greater proportion of tracts classified as rural (24.4% vs. 9.2%, 4.7% and 4.5% for the Weeks, Euler, and

Yaghjyan methods, respectively, Table 1). In contrast, the three previous methods had a greater proportion of tracts (> 70%) in their most urban categories than did the LEAD community type classification (23.6%). This is an expected result that is consistent with the locations of UAs and UCs throughout the U.S. (Figure 1). The LEAD community type classification method also resulted in fewer census tracts being "unclassified" compared to the other three methods (174 vs. 259, Table 1), which is due to our additional criteria for classifying tracts based on the proportion of land area contained within UAs or UCs (Figure 2, step 1).

We observed differing spatial distributions of community type categories nationally, with a greater proportion of tracts being classified as rural using the LEAD community type method compared to the others (Figure 3) and in selected cities (Figure 4), where we see greater differentiation of urban cores from surrounding areas. In Figures 3 and 4, we omitted a visual display of results from one of these methods (Euler et al.) because the spatial distributions of these community types were overlapped by that of the Yagjyhan method. However, we display the Euler et al. results in the supplement (Figures S1–S3). Differences in the spatial distributions of community types across methods were most visible when viewing selected cities in differing regions across the U.S. (Figure 4), suggesting that the LEAD community type classifications appear to better differentiate community types within urban and metropolitan areas than the other methods shown.

We examined discrepancies across classifications, highlighting in green the tracts that were classified as rural in our method but not in the other methods (Figure 5). Tracts that had this discrepancy are largely clustered around urban areas (Figure 5) and are consistent with the locations of UAs and UCs (Figure 1). To understand more about the community variables within these tracts, we compared estimates of several variables for tracts that were classified as rural using our classification method only to those that were classified as rural using all four methods. On average, tracts that were classified as rural in the LEAD community type classification only (i.e., highlighted green tracts in Figure 5, n = 14,469 tracts) had a lower percentage of households earning less than $30,000 annually (31.7% vs. 36.6%), persons living below the poverty level (13.9% vs. 15.5%), and households without car ownership (5.1% vs. 5.6%), and greater household density (61.9 households per mile (standard deviation [SD] 84.9) vs. 16.7 households per mile (SD 33.6)) compared to tracts that were classified as the most rural category in all four methods (n = 3,254 tracts). This suggests that the tracts that were only classified as rural in the LEAD community type method had characteristics that were more consistent with suburban areas (e.g., greater household density) compared to tracts that were classified as rural in all the methods evaluated. We include the means, standard deviations, and medians of all land use and ACS variables considered in this analysis by all four community type classification methods evaluated (Table S1).

To examine how distributions of these variables differed by community type classifications across all census tracts evaluated (n = 72,539), we created histograms for several variables relevant to land use and population characteristics stratified by community type for all of the four methods evaluated (Figure 6). Similar to Figures 3 and 4, we omit the display of the Euler et al. results because they are visually similar to those shown from the Yaghjyan

et al. categories, but these are included in Figure S3. Compared to the two previously published methods displayed in Figure 6, the LEAD categories illustrate greater between-group differences of percent developed land, street connectivity, and percent of persons with less than a high school degree. We did not see this with the percent of households without a car, where distributions across community types for all methods evaluated did not appear to be distinctly different. However, this may not be visually clear because of the relatively high proportion of tracts that had 0 % of households without a car.

Conclusions

We evaluated a modification to the USDA 2010 RUCA census tract designations and compared this modification to previously published classifications of RUCA categories used for epidemiologic research. Our method, which relies on classifying communities based on the land area contained within a UA or UC, resulted in clearer geographic delineations of community types within urban areas. We believe that this is an improvement on previous methods and can be used as a stratification variable for epidemiologic analyses of community characteristics and obesity and type 2 diabetes. Because stratifying analyses by relevant classifications of community type can help to mitigate the methodological challenges of differential item functioning, different co-occurrence of community dimensions, and non-positivity that can arise when evaluating community features and health outcomes across heterogenous communities, we argue that classifications resulting in greater between-group differences of distributions of these community dimensions would better mitigate these challenges (Messer et al., 2010). We therefore recommend the use of our new approach to classifying communities for evaluation of community and environmental domains relevant to obesity and type 2 diabetes, associations which might have been obfuscated due to previous methodological limitations.

Compared to the modifications to RUCA classifications that were used in three previous epidemiologic studies (Weeks et al., 2004, Euler et al., 2019, Yaghjyan et al., 2019), our community type variable appeared to better distinguish distributions of land characteristics (e.g., street connectivity, percent developed land) of census tracts across community type categories. The main reason for the distinction in classification methods is that RUCA's classifications were based on commuting flows into/out of metropolitan areas, which had the effect of classifying suburban areas with high proportions of commuters as a metropolitan area. For example, areas are classified as "Rural" according to RUCA when the primary commuting flows for that tract do not cross into urbanized areas or urban clusters, whereas tracts are classified as "Metropolitan Core" if primary commuting flows are contained within an urbanized area (Economic Research Service, 2019). This is especially evident in Figure 4, where much of the greater Atlanta region was classified as "metropolitan" or "urban" when using the Yaghjyan classification; the Weeks classification designated large portions of the area surrounding Atlanta as suburban, whereas our proposed classification focused its differentiation within Atlanta, allowing for higher density urban, lower density urban, and suburban/small town classifications all to exist within areas that were considered the same category (i.e., either "metropolitan" or "urban") in the other methods. This pattern is similar across other large cities in the U.S., and this distinction was also evident in the distributions of percent developed land across community type categories and classification

methods. By modifying tract-level criteria pertaining to urbanized areas and urban clusters from the original RUCA designations, our classification method is more reflective of the land characteristics within a tract than the original RUCA designations (Figure 6). This has important implications for use in epidemiology studies, as it is likely that individuals living within an urban area have different interactions with their environment, particularly through physical activity and active commuting (Fan et al., 2017) than do those who primarily commute to an urban area but live in a surrounding non-rural area.

A limitation of this work is inherent to the use of administrative boundaries for approximation of communities and neighborhoods: individuals often interact with their environment in ways that are not bound by census tract lines on a map. However, using administrative boundaries at a larger geographic unit than the census tract (e.g., county) would introduce additional measurement concerns and potential misclassification of communities. Although there are measures that classify communities at the county level, we chose to use an administrative unit that reflects a smaller geography and is more frequently used in neighborhood research because geographic disparities in health outcomes often exist at a small geographic scale (Basta et al., 2010, Lee et al., 2018). We acknowledge that the geographic scale of rural census tracts is larger than those in urban areas, and that rural areas in our categorization likely encompass a wide variety of communities consisting of individuals that may or may not consider their community to be rural (Bennett et al., 2019). We also acknowledge that census tracts are not necessarily behaviorally or culturally relevant to how communities define themselves. That said, it is logistically challenging to obtain individual-level perceptions to define communities for a broad geographic scope, and there is heterogeneity in the ways in which individuals conceptualize their neighborhood or local community (Basta et al., 2010), so like many researchers, we opted to use neighborhood typologies defined by census boundaries (Kolak et al., 2020, Rhew et al., 2017). Lastly, we acknowledge that there are many reasons for defining community typologies, and the purpose of the LEAD classification is most appropriately used when stratifying epidemiologic analyses that encompass a variety of place types in the US. These classifications should not be used, for example, to determine resource allocation to communities or for other non-analytic purposes.

This new community type measure has several implications. First, stratifying analyses of community-level features and individual-level health outcomes by this variable can mitigate the aforementioned methodologic challenges that would emerge without a community type stratification variable, or with a community type variable that classified rural and suburban areas as more urban than they actually are due to reliance on commuting patterns. From both theoretical and methodological perspectives, classifying communities in a way that is agnostic to individual-level behavioral patterns can help epidemiologic researchers appropriately identify community- or individual- level interventions to improve health outcomes. Second, with respect to the research activities of the Diabetes LEAD Network, because we were able to distinguish community types within major U.S. cities better than previous methods, we can better characterize the relationships of interest between community-level features and individual-level health outcomes such as type 2 diabetes in stratified analyses. Given that our community type classification appeared to distinguish distributions of key variables better than other available measures, we believe that our

community type variable, when used to stratify analyses, will mitigate the challenges of non-positivity, differential item functioning, multi-dimensionality and collinearity that are inherent to neighborhood-level research that encompasses a variety of community types. Third, other researchers who have an interest in community-level determinants of health in the U.S. can use this variable, which is based on publicly available data. An advantage of this methodology is that it is constructed at the census tract level, which is consistent with the boundaries of many available community-level data sources, facilitating their linkage with this community type variable for epidemiologic studies. Finally, this community type classification method is relevant to the entire contiguous U.S., which can facilitate comparison across studies that include both small populations (i.e., subsets of the U.S.) or the entire contiguous U.S. for epidemiologic inference.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References:

ADAMS JS, VANDRASEK BJ & PHILLIPS EG 1999. METROPOLITAN AREA DEFINITION IN THE UNITED STATES. Urban Geography, 20, 695–726.

AHERN J, CERDÁ M, LIPPMAN SA, TARDIFF KJ, VLAHOV D & GALEA S 2013. Navigating non-positivity in neighbourhood studies: an analysis of collective efficacy and violence. Journal of epidemiology and community health, 67, 159–165. [PubMed: 22918895]

AUGUSTIN T, GLASS TA, JAMES BD & SCHWARTZ BS 2008. Neighborhood Psychosocial Hazards and Cardiovascular Disease: The Baltimore Memory Study. American Journal of Public Health, 98, 1664–1670. [PubMed: 18633086]

BASTA LA, RICHMOND TS & WIEBE DJ 2010. Neighborhoods, daily activities, and measuring health risks experienced in urban environments. Social science & medicine, 71, 1943–1950. [PubMed: 20980088]

BENNETT KJ, BORDERS TF, HOLMES GM, KOZHIMANNIL KB & ZILLER E 2019. What Is Rural? Challenges And Implications Of Definitions That Inadequately Encompass Rural People And Places. Health Affairs, 38, 1985–1992. [PubMed: 31794304]

BOARDMAN JD 2004. Stress and physical health: the role of neighborhoods as mediating and moderating mechanisms. Social Science & Medicine, 58, 2473–2483. [PubMed: 15081198]

ECONOMIC RESEARCH SERVICE. 2019. 2010 Rural-Urban Commuting Area (RUCA) Codes [Online]. United States Department of Agriculture. Available: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/ [Accessed October 3 2019].

ESRI I 2019. ArcGIS Pro 2.4.2.

EULER R, JIMENEZ EY, SANDERS S, KUHLEMEIER A, VAN HORN ML, COHEN D, GONZALES-PACHECO D & KONG AS 2019. Rural-Urban Differences in Baseline Dietary Intake and Physical Activity Levels of Adolescents. Preventing chronic disease, 16, E01–E01. [PubMed: 30605423]

FAN JX, WEN M & WAN N 2017. Built environment and active commuting: Rural-urban differences in the U.S. SSM - Population Health, 3, 435–441. [PubMed: 29124104]

HALL SA, KAUFMAN JS & RICKETTS TC 2006. Defining Urban and Rural Areas in U.S. Epidemiologic Studies. Journal of Urban Health, 83, 162–175. [PubMed: 16736366]

HART LG, LARSON EH & LISHNER DM 2005. Rural Definitions for Health Policy and Research. 95, 1149–1155.

HUMPHREY J, LINDSTROM M, BARTON K, SHRESTHA P, CARLTON E, ADGATE J, MILLER S & ROOT E 2019. Social and Environmental Neighborhood Typologies and Lung Function in a Low-Income, Urban Population. International Journal of Environmental Research and Public Health, 16, 1133.

JONES RN 2019. Differential Item Functioning and its Relevance to Epidemiology. Current Epidemiology Reports, 6, 174–183. [PubMed: 31840016]

KERSHAW KN, OSYPUK TL, DO DP, DE CHAVEZ PJ & DIEZ ROUX AV 2015. Neighborhood-Level Racial/Ethnic Residential Segregation and Incident Cardiovascular Disease. Circulation, 131, 141–148. [PubMed: 25447044]

KOLAK M, BHATT J, PARK YH, PADRÓN NA & MOLEFE A 2020. Quantification of Neighborhood-Level Social Determinants of Health in the Continental United States. JAMA Network Open, 3, e1919928. [PubMed: 31995211]

KRIEGER N 2003. Place, Space, and Health: GIS and Epidemiology. Epidemiology, 14.

LARSON NI, STORY MT & NELSON MC 2009. Neighborhood Environments: Disparities in Access to Healthy Foods in the U.S. American Journal of Preventive Medicine, 36, 74–81.e10. [PubMed: 18977112]

LEE DC, GALLAGHER MP, GOPALAN A, OSORIO M, VINSON AJ, WALL SP, RAVENELL JE, SEVICK MA & ELBEL B 2018. Identifying Geographic Disparities in Diabetes Prevalence Among Adults and Children Using Emergency Claims Data. Journal of the Endocrine Society, 2, 460–470. [PubMed: 29719877]

LEVASSEUR M 2015. Importance of proximity to resources, social support, transportation and neighborhood security for mobility and social participation in older adults: results from a scoping study. BMC public health., 15, 503. [PubMed: 26002342]

MACINTYRE S, ELLAWAY A & CUMMINS S 2002. Place effects on health: how can we conceptualise, operationalise and measure them? Social Science & Medicine, 55, 125–139. [PubMed: 12137182]

MASOUMI HE, TERZI F & SERAG YM 2019. Neighborhood-scale urban form typologies of large metropolitan areas: Observations on Istanbul, Cairo, and Tehran. Cities, 85, 170–186.

MERKIN SS, DIEZ ROUX AV, CORESH J, FRIED LF, JACKSON SA & POWE NR 2007. Individual and neighborhood socioeconomic status and progressive chronic kidney disease in an elderly population: The Cardiovascular Health Study. Social Science & Medicine, 65, 809–821. [PubMed: 17499411]

MESSER LC, LARAIA BA, KAUFMAN JS, EYSTER J, HOLZMAN C, CULHANE J, ELO I, BURKE JG & O'CAMPO P 2006. The Development of a Standardized Neighborhood Deprivation Index. Journal of Urban Health, 83, 1041–1062. [PubMed: 17031568]

MESSER LC, OAKES JM & MASON S 2010. Effects of Socioeconomic and Racial Residential Segregation on Preterm Birth: A Cautionary Tale of Structural Confounding. American Journal of Epidemiology, 171, 664–673. [PubMed: 20139129]

MORRILL R, CROMARTIE J & HART G 1999. METROPOLITAN, URBAN, AND RURAL COMMUTING AREAS: TOWARD A BETTER DEPICTION OF THE UNITED STATES SETTLEMENT SYSTEM. Urban Geography, 20, 727–748.

PETERSEN ML, PORTER KE FAU - GRUBER S, GRUBER S FAU - WANG Y, WANG Y FAU - VAN DER LAAN MJ & VAN DER LAAN MJ 2012. Diagnosing and responding to violations in the positivity assumption. Stat Methods Med Res, 21, 31–54. [PubMed: 21030422]

RHEW IC, KOSTERMAN R & LEE JO 2017. Neighborhood Typologies Associated with Alcohol Use among Adults in Their 30s: a Finite Mixture Modeling Approach. Journal of Urban Health, 94, 542–548. [PubMed: 28484960]

SAPAWI R & SAID I 2012. Constructing Indices Representing Physical Attributes for Walking in Urban Neighborhood Area. Procedia - Social and Behavioral Sciences, 50, 179–191.

SIEGRIST J 2000. Place, social exchange and health: proposed sociological framework. Social Science & Medicine, 51, 1283–1293. [PubMed: 11037217]

SINGH GK, KOGAN MD & VAN DYCK PC 2008. A Multilevel Analysis of State and Regional Disparities in Childhood and Adolescent Obesity in the United States. Journal of Community Health, 33, 90–102. [PubMed: 18049885]

SMURTHWAITE K & BAGHERI N 2017. Using Geographical Convergence of Obesity, Cardiovascular Disease, and Type 2 Diabetes at the Neighborhood Level to Inform Policy and Practice. Preventing chronic disease, 14, E91–E91. [PubMed: 29023234]

THE DIABETES LEAD NETWORK. 2020. About the Diabetes LEAD Network [Online]. Available: http://diabetesleadnetwork.org/ [Accessed March 5 2020].

US CENSUS BUREAU. 2010a. 2010 American Community Survey 1-year estimates [Online]. Available: https://www.census.gov/programs-surveys/acs/technical-documentation/table-and-geography-changes/2010.html [Accessed December 5 2019].

US CENSUS BUREAU. 2010b. 2010 Census Urban and Rural Classification and Urban Area Criteria [Online]. Available: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/2010-urban-rural.html [Accessed December 5 2019].

US CENSUS BUREAU 2010c. Census Summary File 1 (Contiguous US). https://factfinder.census.gov/.

VOEKS JH, MCCLURE LA, GO RC, PRINEAS RJ, CUSHMAN M, KISSELA BM & ROSEMAN JM 2008. Regional Differences in Diabetes as a Possible Contributor to the Geographic Disparity in Stroke Mortality. Stroke, 39, 1675–1680. [PubMed: 18388336]

WEEKS WB, KAZIS LE, SHEN Y, CONG Z, REN XS, MILLER D, LEE A & PERLIN JB 2004. Differences in Health-Related Quality of Life in Rural and Urban Veterans. American Journal of Public Health, 94, 1762–1767. [PubMed: 15451747]

YAGHJYAN L, COGLE CA-O, DENG G, YANG J, JACKSON P, HARDT N, HALL J & MAO LA-O 2019. Continuous Rural-Urban Coding for Cancer Disparity Studies: Is It Appropriate for Statistical Analysis? International Journal of Environmental Research and Public Health, 16.
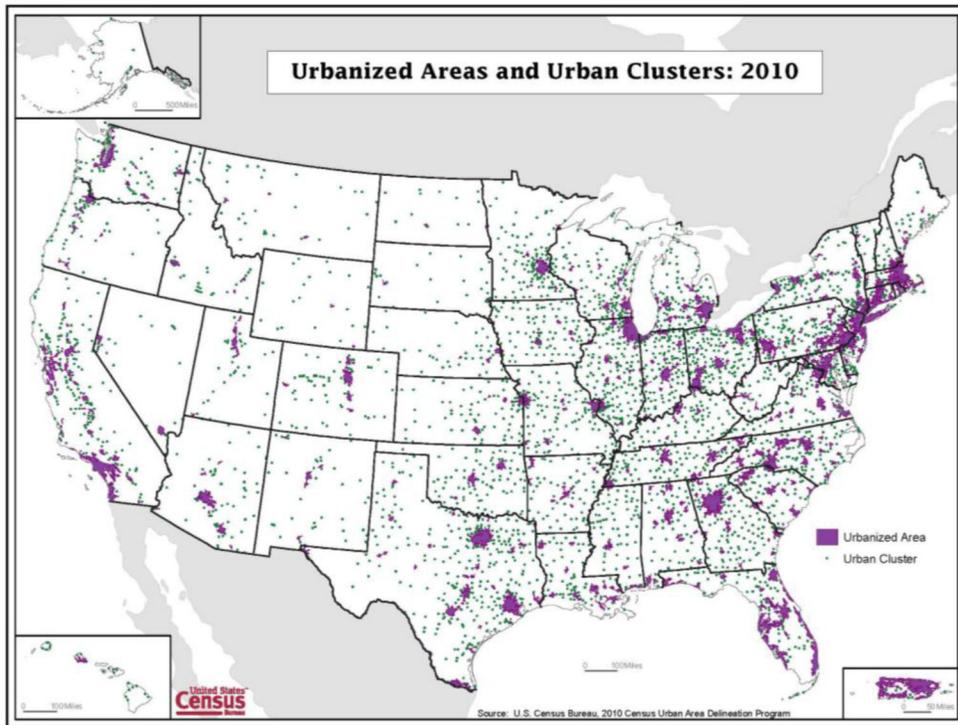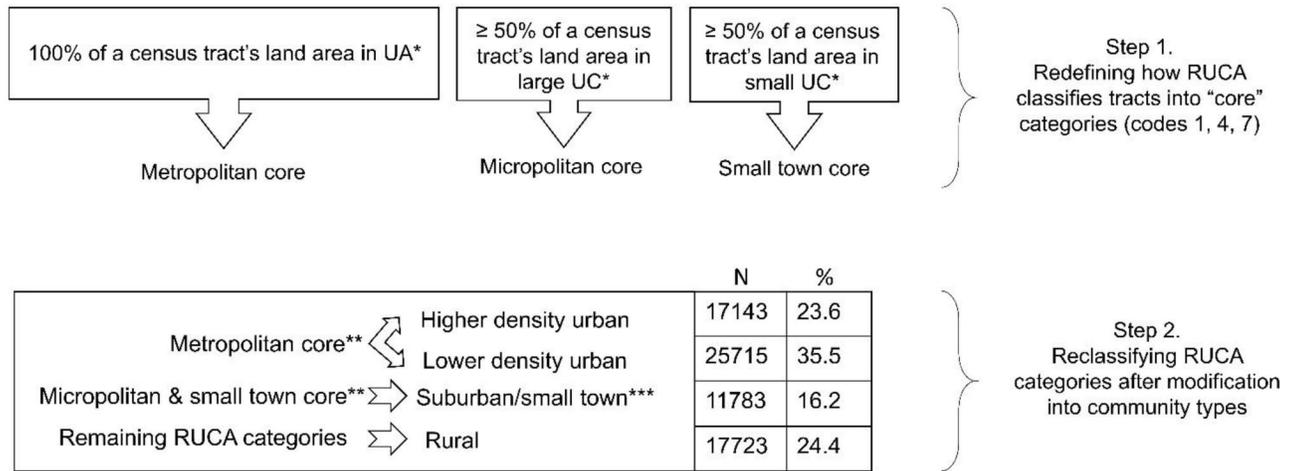
Figure 1.
Map of Urbanized Areas and Urban Clusters

*UA = Urbanized Area, UC = Urban Cluster
**RUCA categories represent redefined categories after modification by UA and UC land area criteria.
***Of the 51,630 (71.2%) tracts with an initial metropolitan core RUCA category (code 1), 9,149 (12.6%) did not meet our modified criterion and were grouped under the suburban/small town category.

Figure 2.
Flow chart of census tract classifications and frequencies of tracts by LEAD community type variable
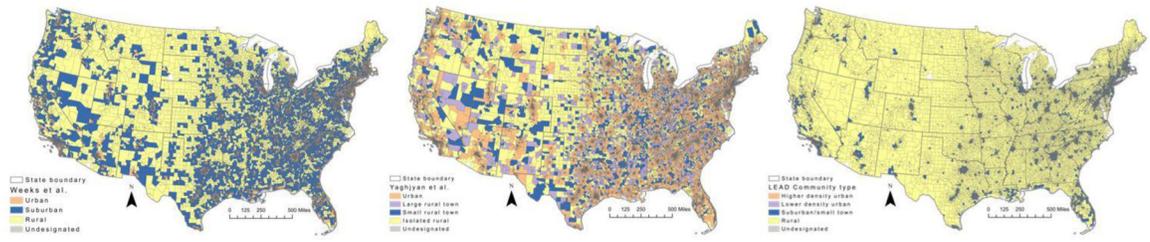
Figure 3.
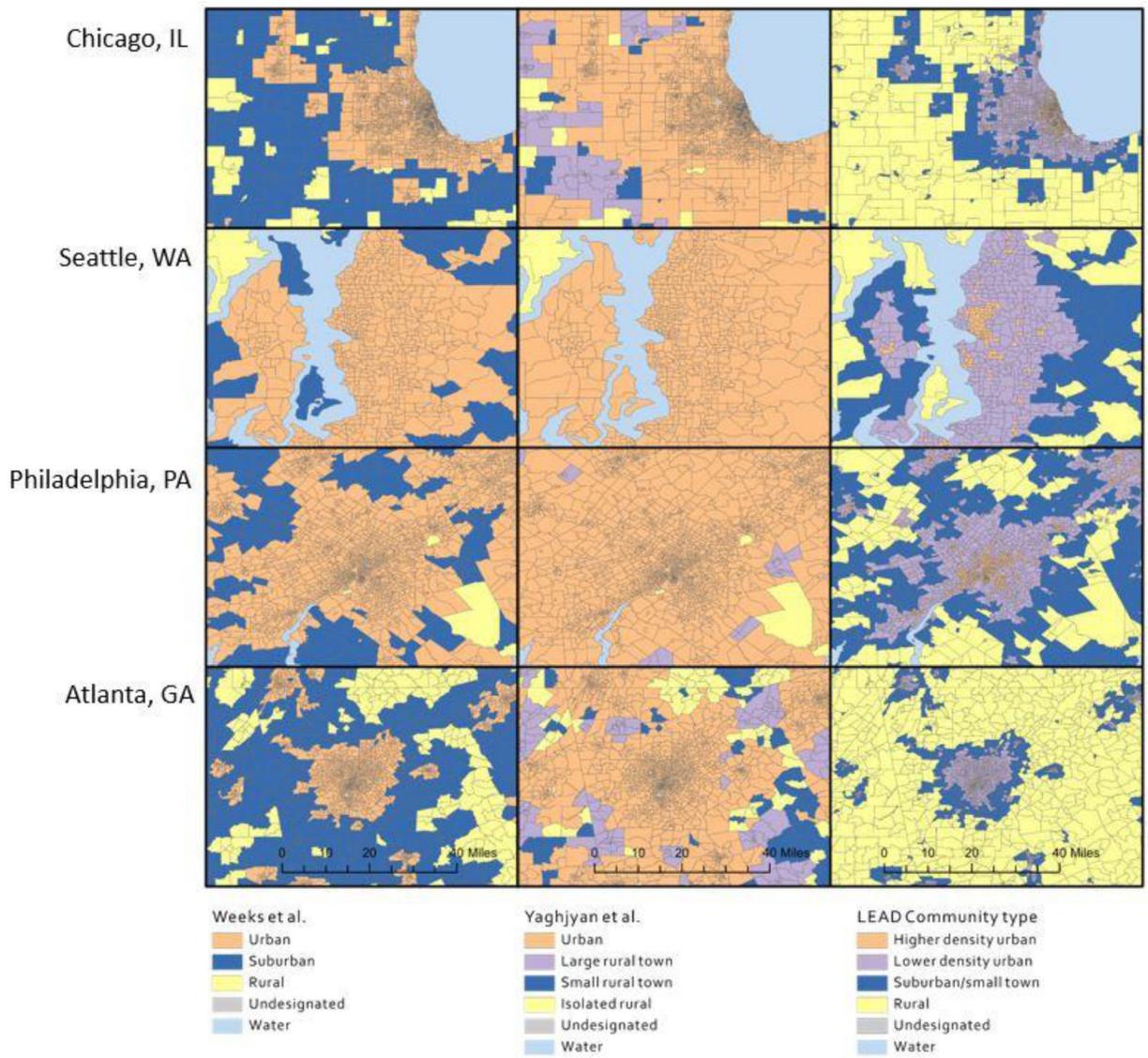National maps of two modified RUCA designations and LEAD community type designations

Figure 4.
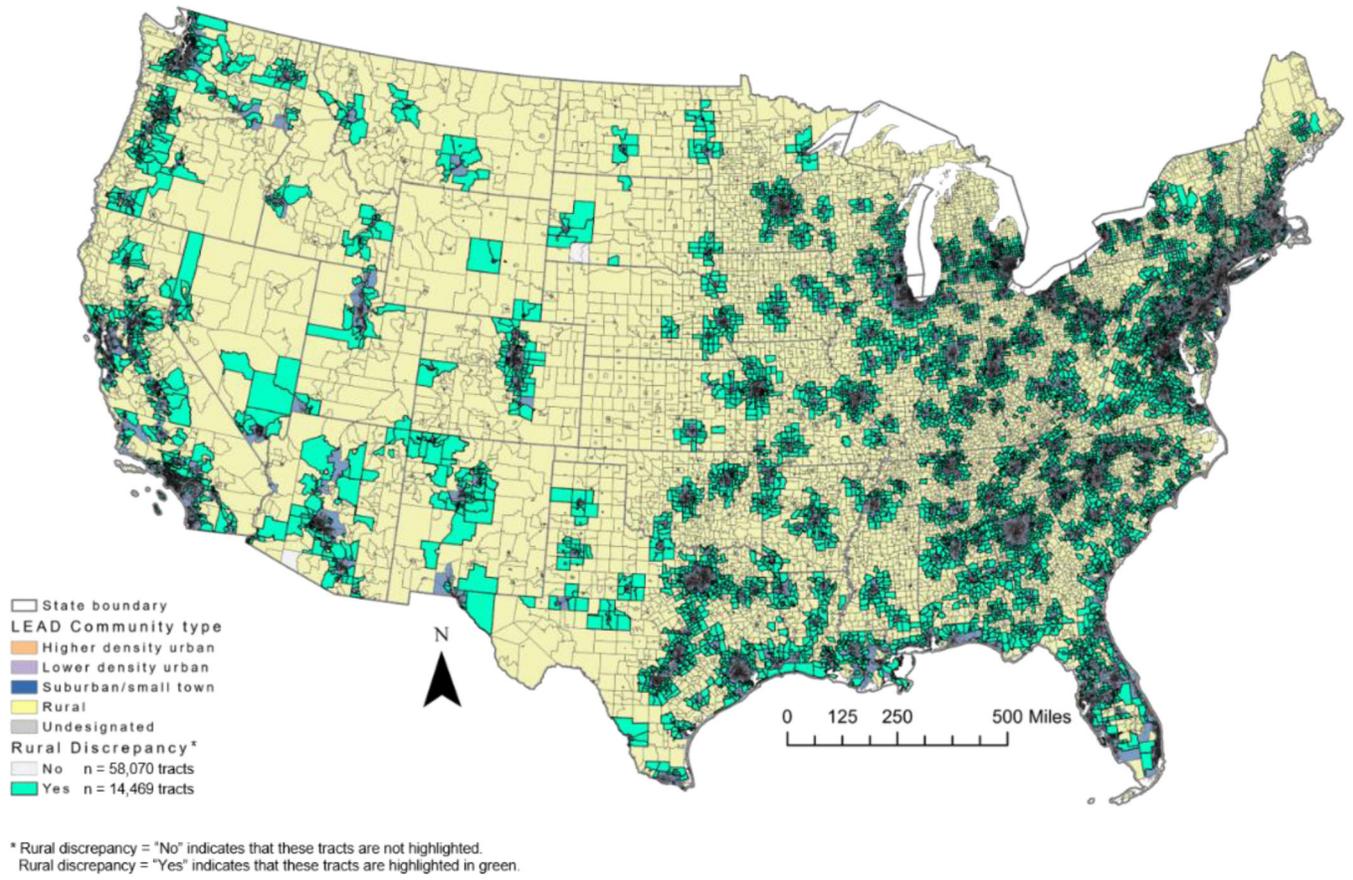Maps of two modified RUCA designations and LEAD community type designations in selected cities

Figure 5.
Map of tract discrepancies for tracts classified as "rural" in LEAD classification but a
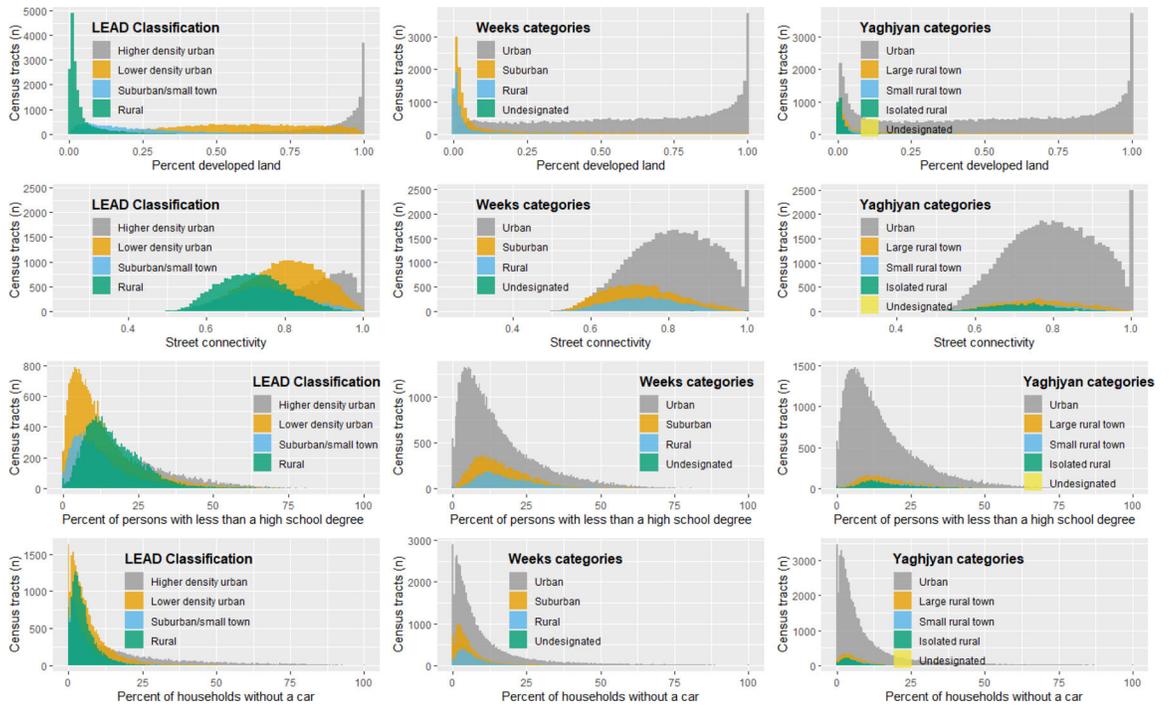non-rural category in different classifications

Figure 6.
Histograms of distributions of select community variables, stratified by community type classification method

Table 1.

Frequencies of census tracts in each community type classification method's categories

Method	Community type categories Census tracts, n (%)					Total
Weeks et al.	**Urban** 51,630 (71.2)		**Suburban** 13,973 (19.3)	**Rural** 6,677 (9.2)	**Undesignated** * 259 (0.4)	72,539
Euler et al.	**Metropolitan** 59,065 (9.0)		**Small town** 3,303 (4.6)	**Rural** 3,374 (4.7)	**Undesignated** 259 (0.4)	72,539
Yaghjyan et al.	**Urban** 59,682 (82.3)	**Large rural town** 6,201 (8.6)	**Small rural town** 3,128 (4.3)	**Isolated rural** 3,269 (4.5)	**Undesignated** 259 (0.4)	72,539
LEAD classification	**Higher density urban** 17,143 (23.6)	**Lower density urban** 25,715 (35.5)	**Suburban/small town** 11,783 (16.2)	**Rural** 17,723 (24.4)	**Undesignated** 174 (0.2)	72,539

*
Undesignated census tracts are those that lack an original RUCA designation. The LEAD classification has fewer undesignated tracts because of the differing land area criteria for urbanized areas and urban clusters.