Geospatial Tools for Social Medicine: Understanding Rural-Urban Divide

Steven A. Cohen, Mary L. Greaney, Elizabeth Erdman, and Elena N. Naumova

Introduction

Place impacts population health. Increasing evidence suggests that one's place of residence plays a substantial role in determining one's health status in the USA and many other nations across the globe. As a result, health disparities based on geography can and do occur. Among the multitude of studies that have demonstrated geographic health disparities, examples include, but are not limited to, cancer (Krieger et al. 2002), physical activity and obesity (Gordon-Larsen et al. 2006), and healthcare quality and access (Baicker et al. 2005; Stiel et al. 2017; Walker and Crotty 2015). The examination of the causes of place-based health disparities has focused primarily on social determinants of health, such as wealth, education, environmental factors, crime, and others, on a defined geographic level, such as the region, state, or county (Woolf and Braveman 2011). Recently, there has been increasing interest in assessing smaller geographic areas to examine how small-area, place-based neighborhood characteristics influence health. Policies, demographics, natural resources, and economic conditions on the local level may affect availability and quality of resources, development, and economic opportunities (Braveman et al. 2011). A growing body of research suggests that understanding if and how small-area social determinants, including education, wealth, crime, environmental factors, and housing, influence population health is critical to reducing health disparities that may often occur within these areas (Beck et al. 2017; Benach et al. 2001; Diez-Roux 1998; Grow et al. 2010; Kruger et al. 2007; Kulkarni et al. 2011; Lippert et al. 2017; Marmot and Bell 2011).

To identify, understand, and address any potential mechanisms through which place-based factors influence population health and lead to geographic health disparities, it is important to understand how the notion of "place" is conceptualized in population health research. In a seminal paper by Macintyre and colleagues, the authors suggest that there are three categories of geographic variation in health—compositional, contextual, and collective—and that these categories are not mutually exclusive (Macintyre et al. 2002). Compositional factors are attributes of the individuals living in a particular area, such as socioeconomic status, race/ethnicity, and other factors. Contextual factors refer to those in the local environment with emphasis on sociocultural and historical features of the community, such as changing demographics, business, and crime. Collective factors are the collective norms, traditions, values, and needs of the community (Macintyre 1997).

Understanding place-based drivers of health is critical to address health disparities. Increasing research suggests that much of the variability in population health is not due to medical-related factors but to geographic differences in non-medical factors, including social, economic, and demographic factors related to geography. For example, a recent analysis found that social, economic, and physical factors account for nearly 54% of population health (Park et al. 2015), factors that are explicitly linked or related to place-based factors (Fig. 1) (Minnesota Department of Health). Similarly,

S. A. Cohen (✉) · M. L. Greaney
Department of Health Studies, University of Rhode Island, Kingston, RI, USA
e-mail: steven_cohen@uri.edu; mgreaney@uri.edu

E. Erdman
Department of Civil and Environmental Engineering, School of Engineering, School of Engineering, Tufts University, Medford, MA, USA
e-mail: lizerdman505@gmail.com

E. N. Naumova
Division of Nutrition Epidemiology and Data Science, Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA
e-mail: elena.naumova@tufts.edu

F. S. Faruque (ed.), *Geospatial Technology for Human Well-Being and Health*,
https://doi.org/10.1007/978-3-030-71377-5_10

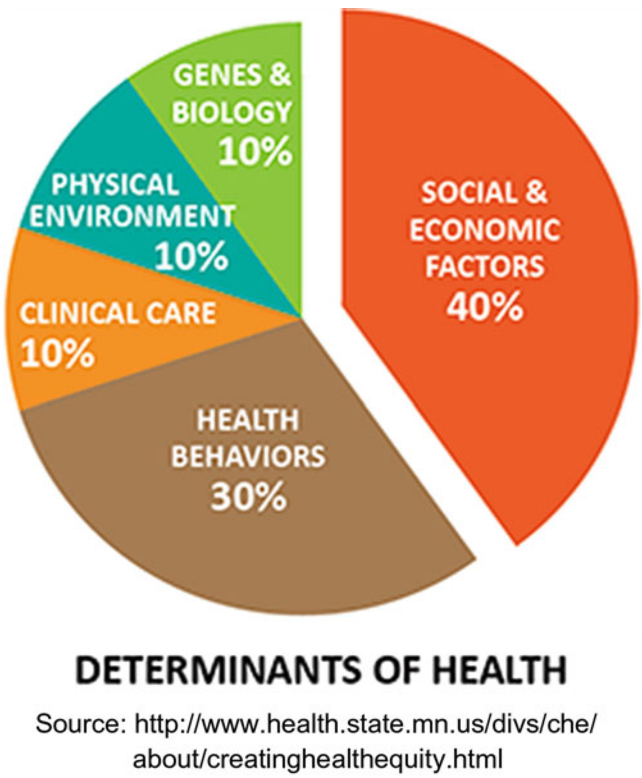

Fig. 1 Determinants of health. (Source: http://www.health.state.mn.us/divs/che/about/creatinghealthequity.html)

analyses from the County Health Rankings and Roadmaps model suggest that approximately 50% of population health is attributable to social and economic factors and the physical environment (Remington et al. 2015; Tarlov 1999; Adler and Newman 2002; McGinnis et al. 2002). Therefore, addressing population-level, place-based drivers of health disparities may have a substantially greater impact on population health than medical and medical-related interventions alone.

Given that place-based factors are associated with population health, a logical question to ask is how do place-based factors influence health. Many such potential mechanisms have been discussed in the literature. The oldest and most well-known example is the Broken Windows theory developed initially by James Wilson that suggests that the appearance of a community's physical environment influences individual behaviors, which ultimately impacts the individual's health and the collective health of the population (Wilson 1987). This theory suggests a dynamic relationship between the environment, health behaviors, and health status. Therefore, based on this theory, as neighborhoods deteriorate in physical appearance, so-called social buffers that could reduce high-risk behaviors may gradually disappear. As the overall health behaviors of a community start to worsen, population health declines, and this is why population health outcomes are often worse in areas of substantial neighborhood degradation (Cohen et al. 2000).

In addition to the Broken Windows theory, other mechanisms by which place may influence population health have been hypothesized. While it is unlikely that place-based factors, such as socioeconomic status (SES), directly impact health, these types of factors shape conditions that ultimately impact population health (Adler and Rehkopf 2008). It is posited that a place-based factor may impact population health through indirect pathways across the lifespan. For example, low SES conditions may contribute to poor nutrition, exposure to harmful environmental contaminants, discrimination, and other aspects of life that may impact health behaviors, access to healthcare, and health status (Adler and Ostrove 1999; Williams and Collins 1995). Such potentially harmful place-based conditions may lead to increased allostatic load throughout life (Seeman et al. 2010) so that stress accumulates over time upon continual exposure to harmful living conditions (McEwen 1998; Juster et al. 2010; Lupien et al. 2015).

The biological, psychological, and sociological pathways of place-based factors which influence population health are just beginning to be understood. Cummins et al. argue that much of the existing research on how "place" influences population health through contextual factors, focusing predominantly on deprivation, has not focused on the contextual and compositional factors (Cummins et al. 2007). Adler and Rehkopf suggest that appropriately combining multiple types of data on SES, demographic, psychosocial, and biological factors on multiple levels will facilitate the creation of causal models that identify direct and indirect pathways that lead to critical but addressable health disparities (Williams and Collins 1995; Adler and Rehkopf 2008).

Exploration of place-based contextual factors has currently centered on a fixed population in a clearly delineated geographic area, such as a county, state, or neighborhood, at fixed points in time. The current view of how place-based characteristics influence population health centers on readily quantifiable, often static measures, such as SES, availability of resources, existence of and proximity to resources, segregation, etc. (Diez-Roux 1998). Some argue that this conventional approach to understanding specific mechanisms through which "place" affects health is integral to population health not just for strengthening causal inferences about place-based risk factors but also for identifying potential avenues for intervention on the population level (Cummins et al. 2007). They suggest moving from the "contextual and compositional" approach to a "relational" approach. In a relational approach, place-based characteristics are viewed somewhat differently. For instance, this approach relies more on socio-relational distances and networks than on physical distance and boundaries and uses area definitions that are more dynamic and fluid. Additionally, a relational approach tends to focus on the cultural aspects of place rather than the resource or deprivation-based aspects of place. There are, however, numerous methodological challenges in utilizing a relational approach to understand place-based influences on population health which make undertaking such studies more difficult. In this chapter, we focus on the more conventional approach toward understanding the influence of place on health while identifying potential areas where researchers can integrate the more relational approach to understanding these complex pathways.

Why Focus on Defining Place-Based Characteristics in Geospatial Models?

Defining place-based characteristics is integral for population health and assessment of health disparities in geospatial models. Geospatial models have been used extensively for a wide variety of research investigating geographic variation in population health and healthcare, health inequalities, and healthcare service needs. Wennberg and colleagues were among the first to conduct a comprehensive geospatial analysis of variation in Medicare services across the USA (Wennberg et al. 2002). Numerous other studies have followed and have examined critical geographic variability in healthcare service using GIS and geospatial models (Gilmer and Kronick 2011; Matlock et al. 2013; Nicholas et al. 2011; Newhouse and Garber 2013a; Hanchate et al. 2017; Chui et al. 2011). Geospatial models have been widely used to assess health inequalities across populations (Krieger et al. 2002; Weich et al. 2003; McDonald et al. 2012; Suzuki et al. 2012; Newhouse and Garber 2013b; Dwyer-Lindgren et al. 2017) and the need for healthcare services (Black et al. 2004; Padilla et al. 2016). Geospatial modeling provides a critical tool for the analysis and planning of health services and infrastructure to reduce inequalities and promote population health, regardless of geography (McLafferty 2003). Geospatial modeling provides insights into the spatial organization of health services and population health, which allows researchers to incorporate multiple dimensions of place-based factors and potentially multiple levels of observation (e.g., individual and different levels of geographic influence) into the analysis. Examples include calculating travel time to health facilities (Branas et al. 2000; Pearce et al. 2006), assessing optimal locations for healthcare centers (Jia et al. 2007), health behaviors and social capital (Mohnen et al. 2012), and obesity prevalence (Prince et al. 2012).

To optimize the effectiveness and utility of any geospatial model, researchers should delve into what and how place-based factors truly drive and mold potential relationships between health and place. Many geospatial models have focused on assessing specific place-based factors that exist in the geographic regions (e.g., state) and how they contribute to spatial disease patterns and health behaviors. Geospatial models of health outcomes were originally developed to quantify spatial patterns across different geographies by incorporating place-based characteristics (Waller and Gotway 2004; Banerjee et al. 2014). However, incorporating place-based factors in geospatial models requires a multitude of considerations that are often overlooked in such modeling procedures. Therefore, the remainder of this chapter will focus on discussing several important challenges to properly identifying and utilizing place-based spatial characteristics in geospatial modeling for applied population health research, with an emphasis on measuring rural-urban gradients. This chapter also will provide insight into strengths and limitations of different approaches, as well as opportunities for future research into these potentially powerful analytical tools to maximize their utility in research and policy.

Challenges in Determining Place-Based Spatial Characteristics

Selection of Characteristics

In geospatial models of health outcomes, the researchers' assumptions related to a spatial distribution of a sociodemographic indicator of interest, specifically its homogeneity, or consistency within the study area, are critical to proper interpretation and use in developing policies and interventions to address health issues. Building geospatial models starts with the selection of place-based characteristics. Characteristics commonly used in such geospatial models may include socioeconomic status (e.g., wealth and income), demographic composition (e.g., racial/ethnic composition, age, and gender), environmental factors (e.g., climate, air/water/soil quality), and education. These characteristics and how they are measured are typically dynamic, location-specific, multidimensional, and potentially highly correlated and could be costly. Thus, rarely does a universally accepted, singular measure of place-based characteristics suffice for geospatial modeling. This raises several issues, some of which are addressed in the subsequent sections.

Spatial Heterogeneity

An important challenge in determining which place-based spatial characteristics to use, particularly when analyzing spatially aggregated data, is spatial heterogeneity of a selected characteristic. Spatial heterogeneity generally refers to uneven and often heavily skewed distributions of characteristics within an area or between adjacent areas. This section focuses on two specific aspects of spatial heterogeneity,

heterogeneity both within and between observational units in geospatial models of aggregated data.

Heterogeneity Within Observational Units

Place-based characteristics can be defined on multiple geographic levels. Oftentimes, there can be notable heterogeneity in terms of these place-based characteristics, especially when aggregating on a large geographic level such as a county, state, or country both in terms of central tendency and variability. Consider the case of two adjacent counties on the western side of the San Francisco, California Bay Area: San Mateo County and the City and County of San Francisco. Both counties are highly urbanized and are part of the metropolitan statistical area of San Francisco. Both are among the wealthiest counties in the USA in terms of per capita income. San Mateo County has the thirteenth-highest per capita income, $50,262, while the City and County of San Francisco has the sixth-highest per capita income of $55,567 (based on the most recent data from the US Census Bureau 2016 American Community Survey 5-year estimates).

Functionally, in studies examining county-level effects on population health, both counties would be considered as individual units of observation of equal importance. However, a deeper examination of the counties themselves reveals substantial differences between the two. First, there are notable differences in the size and composition. San Francisco has a total land area of 46.9 square mile, while San Mateo County has a total land area of 448.4 square miles. The population size of the counties is comparable, with San Francisco (805,235) having a slightly higher population than San Mateo County (718,451). Due to their differences in land area, the population density of San Francisco is considerably higher (8042 per square mile) than that of San Mateo County (604 per square mile). A simple comparison of housing units reveals further distinctions between the two counties. In San Francisco, there are 376,942 total housing units, whereas in San Mateo County, there are 271,031 total housing units. The resultant data represent notable differences between the two county units in terms of the number of people per household, with the average number of people per household being 24.1% higher in San Mateo County than in San Francisco (2.65 vs 2.14 people per household, respectively). Although this difference appears small, San Francisco has one of the lowest average household sizes for all US counties, whereas San Mateo's average household size is above the US average of 2.54 people per household.

The heterogeneity of these two counties also can be compared using smaller geographic units. Although both county-equivalent units of San Mateo County and San Francisco are similar in terms of per capita income (as shown above), a closer examination of the census tracts within each county reveals stark differences between the counties. It should be noted that San Francisco has 195 census tracts while San Mateo County has 156. There is more variability in per capita income among the census tracts of San Francisco than among the census tracts of San Mateo County (Fig. 2). Although San Francisco's per capita income is just over 10% higher than that of San Mateo County, San Francisco has a wider

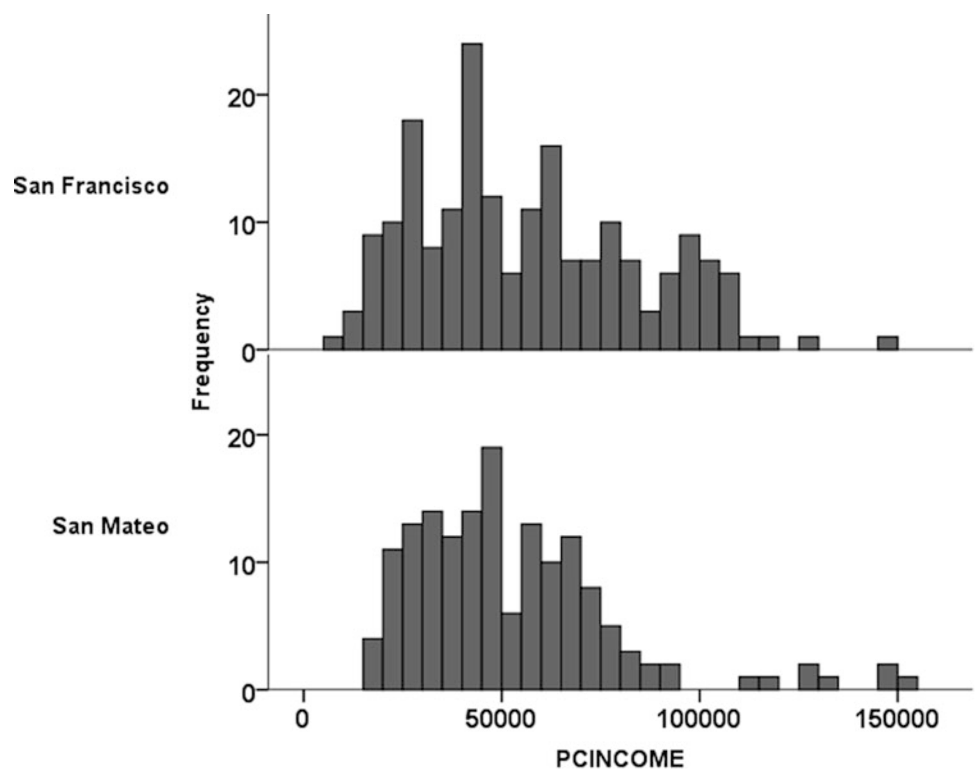

Fig. 2 Distribution of per capita income by census tract in San Francisco City County and San Mateo County (2010)

range of census tract-level per capita income values than San Mateo County does. Despite having the sixth-highest county-equivalent per capita income in the USA, San Francisco contains some of the poorest census tracts in the nation, with the lowest census tract per capita income of $7355. The lowest tract-level per capita income in San Mateo County is $16,732. Considering all census tracts in both counties, the six lowest ones in terms of per capita income are all found in San Francisco. Thirteen of San Francisco's 195 census tracts (6.7%) have a per capita income below $20,000, compared to just 4 of San Mateo's 156 census tracts (2.6%). On the opposite end of the income spectrum, six of the top seven census tracts in terms of per capita income are found in San Mateo County.

This San Francisco and San Mateo County example illustrates that the sum of the parts does not necessarily represent the whole due to spatial heterogeneity within the units of observation. In terms of average per capita income, the two adjacent county-equivalent units appear similar. However, a comparison of additional sociodemographic factors reveals that despite their geographic proximity and similarity in per capita income, these two counties have notably different distributions of the study characteristics. Examining the census tracts within these two county-equivalent units with respect to per capita income reveals further spatial heterogeneity that is masked when examining per capita income on the county level in isolation. It is important to consider such spatial heterogeneity when conducting spatial analyses. In such analyses, where possible, identifying and perhaps quantifying such spatial heterogeneity and underling spatial distributions would enrich the analysis of key socioeconomic, demographic, and other place-based characteristics.

Level of Aggregation

A closely related topic linked to spatial heterogeneity specific to geospatial analyses of aggregated data is the choice of the geographic level of aggregation such as the state or province, county, census tract, or any other unit. There are strengths and drawbacks to each level of aggregation. In this section we discuss a few important issues I to consider when selecting an appropriate level of aggregation in a geospatial analysis. Importantly, this discussion is not exhaustive assessment of the strengths and drawbacks of different levels of aggregation but rather a "jumping-off point" to consider how the selected level of aggregation may influence the results and interpretation of findings and their implications for research and policy.

Data Quality and Confidentiality

There are a variety of valid geographic scales, and the choice of geographic level can lead to different but equally valid results that emphasize different data features (Elliott and Wartenberg 2004). The challenge of selecting a proper scale and aggregation method is referred to as the modifiable area unit problem (Openshaw 1984). The goal of selecting variables to use in geospatial modeling using aggregated data is to choose the smallest geographic units possible to simultaneously maximize sample size and minimize spatial heterogeneity. Yet, the choice is often dictated by available data. Due to limited available data, there is a trade-off between homogeneity within selected geographic units and precision of the estimated associations or disease frequencies. Therefore, a key issue in in geospatial analyses is selecting the most reasonable geographic unit of observation while recognizing its limitation with respect to accuracy and bias this aggregation may introduce.

If they are found to be valid and reliable, a large number of units of observation such as county, state, or region typically increases the likelihood that the geographic information contained within the study area has broad coverage. However, it may reduce the ability to detect potentially critical small-scale trends and associations. Since there is no singular industry standard in terms of data source or protocol for evaluating data quality at different levels of spatial aggregation, the data user must assess the benefits and drawbacks of each potential level of aggregation and, at minimum, identify and discuss the drawbacks in any publically disseminated research project.

When analysis includes records on a fine scale, issues of confidentiality and privacy may also arise, especially when a research question addresses vulnerable populations or people with unique demographic characteristics. Such issues are most pronounced in spatial analysis using small geographic units, such as street address, the census tract, or block group (Clapp and Wang 2006). Methods that attempt to address data confidentiality and privacy include geographic masking, the process of altering the coordinates of geographic data to limit the risk of re-identification in the released data to make it difficult to accurately reverse geocode the released data (Zandbergen 2014). Masking techniques are especially useful in non-aggregated data and also apply to aggregated data (Armstrong et al. 1999). It is worth noting that aggregation itself may mask problematic issues of confidentiality that occur with point-source data (Kounadi and Leitner 2014). Nonetheless, data confidentiality and privacy issues remain a highly debated issue in geospatial modeling (Fefferman et al. 2005; O'Keefe and Rubin 2015), and there is an ongoing need to develop and test statistical methods to address this issue.

Policy Relevance

Another issue to consider when using geospatial models for health research is the utility of the geographic level of aggregation in terms of informing policy. Many health policies are set on the state level by state governments, which make analyses comparing state-level differences appealing and useful for this purpose. The results of state-level research can immediately inform individual states as to which states are better and which are worse in terms of whatever health outcome or risk factor is examined. In the USA, however, this approach is often limited simply by the limited number of units of observation (50, or slightly more if the District of Columbia and US territories are included), which greatly reduces statistical power, especially in geospatial models where some of the error is explained by spatial correlations.

Counties offer more granularity and greatly increase the number of units of observation (3142 counties and county-equivalents in the USA). There are several important caveats to consider when using county as the level of spatial aggregation. First, the function of county governments varies by state. Some states do not have an active county government system, and all governance is done on the state or municipality level (e.g., New England states). Second, counties and county-equivalent units vary in terms of size and structure within and between states. For example, all independent cities in Virginia, regardless of population size or area, are treated as county-equivalents. Consider the case of Norton, Virginia, an independent city in the rural western part of the state with a population of under 4000 as of the 2010 US Census and a geographic area of 7.5 square miles. County-equivalent units such as Norton and other small, independent cities are considered on the same level of observation as actual counties in Virginia that may be orders of magnitude larger, either based on geographic size, population size, or both, such as Fairfax County, with a population of 1.14 million and a land area of 396 square miles. In many other states, all municipalities, regardless of size, are considered to be part of a county. In Massachusetts, for example, the major city of Boston is part of the larger Suffolk County. All municipalities in Massachusetts, even minor cities with population sizes over 100,000, are part of a larger county. Similarities are found in states such as California, where only large cities, such as San Francisco, are considered county-equivalent units while all l smaller cities and towns are part of the California county system. Numerous other similar examples can be found across the USA. Collectively, these are just a few examples and illustrations of the heterogeneity in terms of function, size, and composition, within and between US counties, especially evident when comparing counties across different states.

Smaller units of observation, such as the census tract, municipality, and block group, offer additional gains in terms of the number of units of observation and offer an increasing amount of granularity and the ability to detect key neighborhood and other area-level differences. Geographic levels of observation created by the Census Bureau, such as the census tract and block group, are designed to be relatively homogeneous with respect to population size and function across the small areas they represent. Nonetheless, data aggregated to fine levels of geography such as these may be subject to issues of data reliability, privacy, and confidentiality, as noted previously. Furthermore, policies and interventions designed to address population health issues assessed at a fine geographic level may be difficult to implement due to a variety of factors, including, but not limited to, spillover effects from one area to another and population migration and movement among these small geographic units.

Example: The Swiss Paradox

A key example of how the level of aggregation can affect the findings of geospatial models and therefore impact downstream policies and programs is known as the "Swiss paradox" (Clough-Gorr et al. 2015). It has been widely established in the public health and social medicine and public health literature that higher income inequality is generally associated with worse population health outcomes (Kawachi and Kennedy 1999; Krieger et al. 2002; Lynch et al. 2000). Examples of this association are numerous and include obesity (Zhang and Wang 2004; Wilkinson and Pickett 2006), self-reported health (Kondo et al. 2009), and overall mortality (Kennedy et al. 1996; Vincens and Stafström 2015). Although there are a variety of theories and empirical evidence to support these associations, the precise reasons for them are not entirely clear. A seminal article by Kawachi and Kennedy (1997) suggested that income inequality promotes poorer health outcomes by reducing social cohesion. Further studies have suggested other potential complementary mechanisms through which income inequality affects health outcomes. One hypothesis is that income inequality is a correlate of other structural, demographic inequalities, such as racial segregation, whereby spatial concentrations of race and poverty influence individual and population health outcomes (Subramanian and Kawachi 2004).

The term "Swiss paradox" was coined by Clough-Gorr and colleagues in a 2015 article, one of the first studies to formally investigate how level of spatial aggregation may influence the associations between income inequality and health. When measured on the state level, income inequality is associated with poorer health outcomes (Kahn et al. 2000; Kennedy et al. 1998; Subramanian and Kawachi 2004; Subramanian and Kawachi 2003). However, with lower levels of aggregation, such as the census tract and county, the findings are mixed (Fiscella and Franks 1997; Soobader and LeClere 1998; Eckenrode et al. 2014). Clough-Gorr and colleagues observed that higher income inequality in Swiss municipalities was consistently associated with lower mortality risk, except for certain health outcomes, even after accounting for sex, marital status, nation of origin, rural-urban status, and other potential confounding factors. Their results challenge current beliefs about the effect of income inequality on health

Table 1 Parameter estimates from the association between Gini index and each of the five listed health outcomes on the county and state levels

Model type and Gini level	Obesity	Diabetes	Current smoker	Poor/fair SRH	Sedentary lifestyle
Unadjusted					
County	−0.33 (−0.54, −0.13)	−0.08 (−0.12, 0.27)	0.05 (−0.10, 0.20)	0.82 (0.59, 1.04)	0.19 (−0.04, 0.42)
State	−0.01 (−0.25, 0.23)	0.55 (0.32, 0.78)	0.24 (0.07, 0.41)	1.11 (0.84, 1.38)	0.66 (0.39, 0.94)
Income-adjusted					
County	−0.39 (−0.59, −0.19)	0.03 (−0.16, 0.23)	0.01 (−0.14, 0.15)	0.66 (0.45, 0.88)	0.05 (−0.17, 0.28)
State	−0.09 (−0.33, 0.15)	0.50 (0.27, 0.73)	0.18 (0.01, 0.36)	0.89 (0.64, 1.15)	0.48 (0.21, 0.75)
Fully adjusted					
County	−0.42 (−0.63, −0.20)	−0.10 (−0.31, 0.10)	0.01 (−0.14, 0.17)	0.63 (0.41, 0.86)	0.23 (−0.01, 0.47)
State	−0.25 (−0.50, 0.01)	0.30 (0.06, 0.55)	0.12 (−0.06, 0.30)	0.71 (0.44, 0.98)	0.58 (0.30, 0.85)

on a fine geographic scale. The reasons for such findings, however, remain unclear and merit further research.

A direct comparison between income inequalities based in the USA that examined the effect of aggregating at the state level versus the county level further corroborates the Swiss paradox. To illustrate the challenges, data from the 2012 Behavioral Risk Factor Surveillance System (BRFSS) were utilized. The BRFSS is a nationally representative phone survey of nearly 500,000 US residents in all 50 states, plus districts and overseas territories. The 2012 BRFSS sample was selected because it was the last year in which county of residence was publicly available in the data set. The association between income inequality and the county-level prevalence of five representative health behaviors and outcomes—obesity, diabetes, current smoking, sedentary lifestyle, and fair/poor self-reported health—was assessed using generalized linear models. The analysis was conducted using income inequality on two levels of spatial aggregation, the state and county, adjusting for income and other sociodemographic factors. Findings identified three distinct patterns of associations (Table 1). First, for fair/poor self-reported health, higher income inequality on both the state and county levels was associated with an increase in the prevalence of this health outcome, which is what might be expected. Second, higher income inequality was associated with a higher prevalence of both diabetes and having a sedentary lifestyle when income inequality was measured on the state level, but not when measured on the county level. Similar results were obtained for current smoking status, except the association between state-level income inequality and prevalence of current smoking became nonsignificant in the fully adjusted models. Third, and perhaps most interestingly, for obesity, higher income inequality on the county level was actually associated with a decreased prevalence of obesity, while there were no significant associations observed when income inequality was measured on the state level. In this analysis, several challenges are apparent. This analysis considered each geographic unit as spatially independent and did not test for potential spatial dependency among geographic units using Moran's I or other statistic. This is likely a more important problem for counties than for states (Manley et al. 2006) in terms of ability to distinguish local patterns of spatial autocorrelation. Additionally, there is a considerable difference in sample size and the number of units of observation between states (51, including DC) and counties (3143), resulted from spatial aggregation. Related to this caveat, the distribution of Gini index is notably different when measured on the state and county levels (Table 2).

Table 2 Descriptive statistics for Gini index on the state and county levels (2012, source: US Census Bureau)

Statistic	States	Counties
N	51	3143
Mean (SD)	0.4552	0.4350
Median	0.4559	0.4325
Min	0.4132	0.3161
Max	0.5315	0.5994
Skewness	0.7190	0.3573
Kurtosis	1.9461	0.3150

Study findings underscore the notion that level of aggregation matters. Why the association between income inequality and health varied based on the level of aggregation is not entirely clear. Uncovering some of the potential mechanisms through which these social characteristics affect health on these and other geographic levels is integral to creating effective policies and programs designed to reduce health inequalities and improve population health, regardless of geography.

Case Study: Rural-Urban Status

Examining place-based factors that drive population health and promote health disparities requires careful attention to the place-based factors and characteristics studied. There are many instances in which there is no scientific consensus as to the best measure of a certain social, demographic, environmental, or economic factors as each measure may have its own benefits and drawbacks. One key example of this is the measure of rural-urban status (Cohen et al. 2018b). Whether describing rural health issues, assessing rural-urban health disparities, and examining the process of urbanization or any other issues pertaining to the rural-urban divide, it is essential to understand, utilize, and interpret appropriate measures of rural-urban status to properly characterize the place-based characteristics the researcher seeks to address. Furthermore, it is valuable to note that place-based characteristics, especially those concerning measures such as rural-urban status, depend heavily on environmental factors that have a meaningful impact on health both in and surrounding the areas of study. Such factors include, but are not limited to, the use of agricultural land, roads, landfills, presence of bodies of water, forests, national preserves, parks, and even concentrations of man-made structures such as buildings (Erdman et al. 2015; Jagai et al. 2010). The following exemplar case studies illustrate some of the many options and considerations of measuring rural-urban status in geospatial and other related models.

Defining Rural-Urban Status

One basic issue to consider when using rural-urban status in geospatial models is which definition of rural-urban status to use. As is the case of many sociodemographic measures, there is no scientific consensus as to the “best” measure of rural-urban status, and each one has unique strengths and weaknesses that should be taken into account (Hart et al. 2005). Furthermore, each measure requires a unique interpretation and may reflect different aspects of the geographies under study. It may be useful to note that some measures are only defined and available on a certain geographic level of aggregation.

Commonly used measures or rural-urban status in social medicine and public health studies include, but are not limited to, population density, percent urban population, Urban Influence Codes (UIC), Rural-Urban Continuum Codes (RUCC), and Rural-Urban Commuting Areas (RUCA). Population density and percent urban population are available from the US Bureau of the Census and have the flexibility to be used at the state, county, census tract, and block group. The UIC, RUCC, and RUCA are produced and maintained by the US Department of Agriculture. These three measures—UIC, RUCC, and RUCA—are only available on certain geographic level of aggregation. The UIC and RUCC are available only at the county level, while the RUCA is available on the census tract level, which could be scaled up to other geographic levels with appropriate weighting schemes.

Consider the case of the percent urban variable that is used in many studies of rural-urban health and health disparities. This variable, defined as the percentage of the area population that is deemed by the Census Bureau to live in an urbanized place, has far-reaching research caveats. Take, for example, an in-depth examination of the 29 counties or county-equivalent places with a 100% urban population. Among those 29 counties are large metropolitan counties, such as Denver County, Colorado, one of the largest cities and counties in the country, with a population of 285,797, as well as far smaller counties and county-equivalents, such as Covington, Virginia, with a population of just 3067. Covington, Virginia, is situated in a highly rural, mountainous area with no major population centers within several hundred miles. Yet, both Denver and Covington would be considered to be equally “urban” according to the percent urban variable. For comparison, San Diego County, California, which comprises the majority of the second-largest city in California with just over 1.1 million county residents, would be considered less urban (96.5%) than Covington, Virginia (100%), using percent urban as the measure of rural-urban status.

As a result of this limitation, composite indices of rural-urban status take into account multiple aspects of the rural-urban gradient and are gaining traction in public health and biomedical research (Naumova et al. 2009). An example of a composite measure is the Index of Relative Rurality (IRR) (Waldorf 2007), which is a continuous measure (0 to 1) of rural-urban status that takes into account population density, population size, proximity to metropolitan areas, and percent urban population. This measure was originally used at the county level but can easily be calculated for other geographic levels, such as the census tract or block group. The IRR and other related measures have clear strengths, such as they are continuous, take into account multiple aspects of the rural-urban gradient, and are flexible on different geographic scales. The central drawback of using this type of measure is in its interpretation. As in the example of the IRR, since, by definition, the measure is a relative measure of rurality, differences between geographic units have no immediate, obvious, and easy-to-comprehend interpretation. For example, the difference in IRR between San Diego County, California (0.24), and Covington, Virginia (0.31), is 0.07 IRR units. The scale of the IRR ranges from 0.04 for New York City Manhattan Borough to 0.89 for Northwest Arctic Borough in Alaska.

The choice of how to measure rural-urban status affects the potential associations observed between rural-urban status and health outcomes. While several of the individual rural-urban status measures are strongly correlated to each other, others are not. Further complicating this issue is that the magnitude of some of the correlations varies substantially by geographic region. For instance, the rank correlation among the RUCC, UIC, population density, percent urban, and IRR was as high as 0.917 ($p < 0.001$) for the RUCC-UIC correlation, to as low as 0.521 ($p < 0.001$) for the UIC-percent urban correlation for US counties (Cohen et al.

2015). When stratified into nine Census divisions, the range of the UIC-percent urban correlation varied substantially from 0.802 in the Pacific states to as low as 0.384 in the West South Central states. The same study found that, as a result, the magnitude and direction of the association between rural-urban status and the health outcome of obesity varied considerably by the choice of rural-urban status measure as well as the geographic region of analysis.

Consideration of Variable Type in Assessing Rural-Urban Differences

Another important element of assessing and using rural-urban status in geospatial models is what type of rural-urban measurement to use (i.e., dichotomous, ordinal, discrete, or continuous). There is no scientific consensus as to which type of variable to use (Hart et al. 2005). One type of variable commonly used is a rural-urban dichotomy (Haque and Telfair 2000; Dahly and Adair 2007) which has several advantages. Perhaps the most obvious advantage to dichotomizing rural-urban status is the ease of interpretation and dissemination in research and practice. When a dichotomous measure, such as metropolitan vs nonmetropolitan, or when a continuous measure of rural-urban status, such as population density or percent urban, is dichotomized, it is straightforward to interpret in the context of disparities and can facilitate easy comparison. The concepts of "rural" and "urban" can be directly compared for interpretation, statistical analysis, and subsequent dissemination in research and to the general public.

As would be the case for converting any continuous measure to a dichotomous measure, there is a critical issue of deciding which cut point to use when delineating "rural" from "urban." Consider the example of population density and obesity among a sample of older adults aged 65 and above abstracted from the 2012 BRFSS. In this analysis, nine different cut points are used to delineate "rural" from "urban" counties in the USA at each decile of population density (Fig. 3). If the tenth decile is used, which indicates the lowest 10% of population density (extremely rural) versus all other counties, the prevalence of obesity in those counties considered to be rural is significantly lower (24.2%) than in those counties considered to be urban (27.7%). However, if the 90th percentile of population density is used, which would separate counties into highly urban (top 10%) versus all others, the prevalence of obesity in the rural counties is significantly higher (27.6%) than that of the urban counties (25.2%). Similar results are observed when using the 80th percentile of population density as the cutoff value: the prevalence of obesity is significantly higher in the rural counties (27.6%) than in urban counties (26.0%). Using the median county population density or any of the surrounding deciles as cutoffs (20th through 70th), there would be no significant differences between rural and urban counties in the prevalence of obesity. Therefore, in this example, it is evident that when dichotomizing a continuous variable to obtain a measure of rural-urban status, the choice of cutoff value makes a substantial difference in the conclusions reached about the health outcome of study. In this case, the selection of two different cutoff values—at 10% and 90%—to delineate "rural" from "urban" results in completely op-

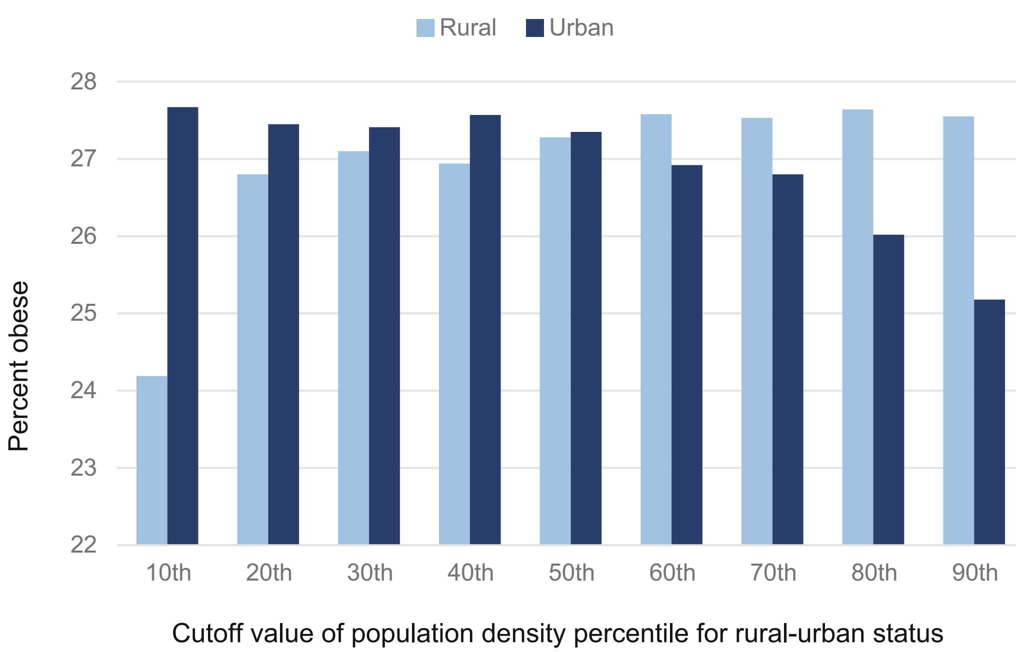

Fig. 3 Percent obese using ten different cut points for rural-urban status based on population density decile

posite findings. These results are simply an application or extension of the problem of dichotomization used in other, non-geospatial models.

Ordinal variables, such as the RUCC, RUCA, and UIC, discussed previously, are advantageous for a variety of reasons, but they also have several drawbacks that should be taken into consideration. Ordinal variables may be preferred over dichotomous variables because of their ability to distinguish finer gradations of rural-urban status. For example, the RUCC is a classification scheme that delineates metropolitan counties from nonmetropolitan counties. Metropolitan counties are classified by the population size of their metro area and nonmetropolitan counties by degree of urbanization and proximity to a metro area. The RUCC ranks counties on a scale from 1 to 9 based on these characteristics. An advantage of using an ordinal variable, such as the RUCC, is the flexibility to treat it either as a continuous or discrete predictor variable or as a series of dummy or indicator variables if there is enough statistical power to do so. The advantage of the former is to assess to see if there is a quasi-linear association between rural-urban status and the outcome, while the advantage of the latter is to assess potential nonlinear associations between rural-urban status.

Nonetheless, there are some inherent drawback to using ordinal variables, some unique to variables such as the RUCC, UIC, and RUCA. The first pertains to using RUCC, for example, as a continuous or discrete predictor in models. This assumes that there is a linear association between the RUCC and the outcome of interest (whether continuous, ordinal, or dichotomous). If there is a nonlinear component to the relationship, i.e., curvilinear, j-shaped, etc., the model may not adequately quantify this association. A more serious issue may be is the construction of the measure itself. While the RUCC is presented as an ordinal variable (1 to 9), the gradations between each unit do not reflect an ordinal process. Consider a RUCC value of 3, which represents "counties in metro areas of fewer than 250,000 population," whereas a RUCC value of 4 represents "counties with an urban population of 20,000 or more, adjacent to a metro area." A RUCC value of 3 is considered more urban than a value of 4. Yet, there are numerous examples of counties with population levels well below the threshold of 250,000 that lie in a metropolitan area (considered a 3), while much more populous counties lie just outside and immediately adjacent to one or two metropolitan areas with large urban populations (considered a 4). Although the county considered a 3 on the RUCC scale might appear less urban than the county considered a 4, the former would be considered to be more urban than the latter. Similar issues exist with the RUCA and UIC measures as well.

It is important to note that no classification system, dichotomous, ordinal, discrete, continuous, or other, is free of issues and caveats. Ordinal measures such as the RUCC, UIC, and RUCA provide a robust array of options for assessing rural-urban status above and beyond many traditional unidimensional measures, such as population density or percent urban population. Given that there is no universal, standard measure of rural-urban status, there are a variety of available measures and variable types to use to suit the needs of researchers interested in assessing place-based characteristics, such as rural-urban status. There is value in understanding the strengths and weaknesses of each one, but if they are used properly, their use will not render an analysis invalid.

Assessment of Nonlinearity in the Rural-Urban Gradient

Example: Rural-Urban Status and Health Outcomes

As an example, we consider linear measures of rural-urban status and assess potential nonlinearity of an association between rural-urban status and health outcomes. In the case study highlighted here, we assessed the associations between rural-urban status and multiple health outcomes from a national survey of older adults using seven commonly used measures of rural-urban status: RUCC, UIC, RUCA, Euclidean distance to nearest metropolitan area, population size, population density, and percent of the population that is urban, with each measure being stratified into quintiles. The association between quintile of rural-urban status measures and the examined health outcomes (obesity and missing annual medical checkup) was assessed through logistic regression modeling, accounting for complex sampling and controlling for confounding variables. We also examined linear trends by treating quintile of each rural-urban status measure as a discrete variable. Details are outlined in the article (Cohen et al. 2018b).

Study results emphasize some of the points made previously. First, compared to the most urban quintile of each measure (reference group), generally speaking, the odds of each outcome—obesity and missing an annual checkup—were significantly higher in the more rural areas (Fig. 4), with some key exceptions. For population density, the odds of obesity were significantly lower in the most rural quintile and significantly higher in the third and fourth quintiles compared to the most urban quintile. Analyses revealed a significant monotonic association and population density quintile (increasing urbanity was associated with an increased likelihood of obesity). However, a linear or monotonic association was not evident for any of the other six measures (RUCC, UIC, RUCA, Euclidian distance, population size, and percent urban), likely to the curvilinear relationship between rural-urban status and obesity for many of the measures.

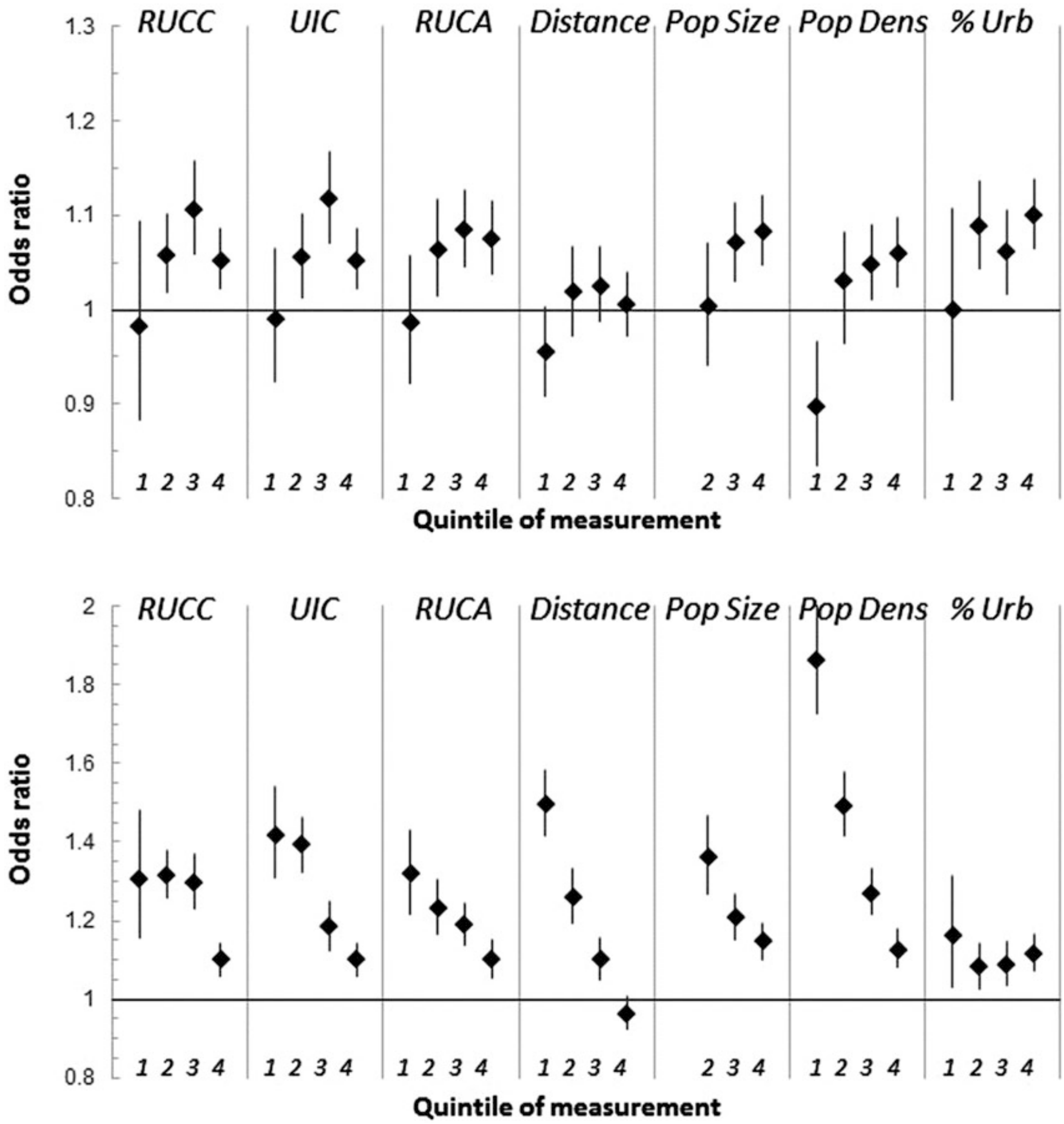

Fig. 4 Odds ratio of obesity (top panel) and missed annual checkup (bottom panel) for seven measures of rural-urban status in quintiles: RUCC, UIC, RUCA, distance to nearest metropolitan area, population size, population density, and percent urban population. Reference group is the highest (most urban) quintile. (*Adapted from Cohen et al. 2018a)

Therefore, using a non-dichotomous measure of rural-urban status revealed a nuanced, U- or J-shaped association between rural-urban status and obesity that might have been masked had a dichotomous measure of rural-urban status been used. Also, the associations depended upon the specific measure of rural-urban status. Had population density quintile been a discrete variable, we would have been able to assess a potential monotonic relationship between it and the log odds of obesity. A monotonic relationship would have been observed: increasing population density is associated with greater odds of obesity. However, the results show that this is not entirely true, based on the data. There may be a positive monotonic relationship between population density and obesity but only among the four most rural quintiles of population density. In other words, the monotonic association does not hold for the most urban quintile. The odds ratios of the association between both the third and fourth quintiles of population density and obesity were above 1. Therefore, the risk of obesity is highest in the intermediate (third and fourth quintiles) of population density, and not in the highest (most urban) quintile, and the association between obesity and rural-urban status was curvilinear and non-monotonic.

Analogous findings also were observed for missing an annual checkup. In the case of this measure, six of the seven measures of rural-urban status were inversely and monotonically related to the likelihood of missing an annual checkup: as rurality increased, respondents were significantly more likely to have missed an annual checkup for all measures, except percent urban population. For percent urban, although respondents in each of the four most rural quintiles were significantly more likely to have missed their annual checkup, the associations varied in magnitude, which likely precluded a monotonic association.

Example: Rural-Urban Status, Vegetation, and Asthma in Older Adults

Here we further illustrate the effect of nonlinearity in the rural-urban gradient in exploring the relationship between hospitalizations among older adults due to asthma in the New England states (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont) and New York State (Erdman et al. 2015). These associations can be clearly affected by population density in many ways. The seven states included in this study range from densely populated New York City and southern Connecticut to rural Maine and New Hampshire. All of the seven states have relatively similar climates and other environmental factors. In 2006, the population of these states was 11% of the entire US population (33,576,172), 13% (4,439,893) of whom were over 65 years of age (census.gov). To assess the associations between the rates of disease and environmental characteristics, it was necessary to spatially and temporally align multiple data sets. We used satellite imagery to assess degree of greenness and created a measure for the percent of green cover for 16-day time periods during 2005–2006. Census data (2010) were used to abstract population size and calculated hospitalization rates based on patients' zip code of residence.

The first step of the analysis was creating the data set to match the hospitalization data to the satellite imagery and detecting alignment and identify misclassification issues. For example, hospitalization records were arranged by residential zip code, whereas for imagery data, we used a shape file to align with zip codes boundaries. The shape file generally had fewer pixels of data on smaller areas that tend to also have larger population densities. The hospitalization records listed 3109 zip codes; after merging those with the census data resulted in a data set of 2864 zip codes, a net loss of 245 zip codes with low population sizes. To reduce spuriously high rates, zip codes with older adult populations fewer than 100 residents were merged with adjoining zip codes providing these were in the same state. Neighboring zip codes with the most similar population size were merged until the joined county population exceeded 100 residents aged 65 + .

Linking medical claims and satellite imagery along with spatial alignment should consider temporal alignment as well. While medical records were complete for the study period, some satellite images were missing or unusable during that period. When we linked imagery data by zip code to the hospitalization records, we lost additional 424 zip codes resulting in 2876 complete matches. We then merged zip codes that had missing values with neighbors that are likely to be similar in environmental exposures. Both spatial and temporal alignments within and between data sets are time-consuming, but the effect of missing data may compound across multiple data sets and may influence the final analysis and its findings. Thus far, the data linkage procedures are rarely described in the epidemiological literature, and a system of checks and balances to identify data discrepancies does not yet exist.

As we explored the associations between degree of greenness and asthma rates, we noted that the relationships were influenced by population density and that association was not monotonic. We applied simple cutoffs and marked a zip code as urban if a zip code had >830.7 persons per square mile, rural if a zip code had <107 persons per square mile, and semi-urban otherwise. In the studied seven states, the average elderly population was 13.1% and an average log population density of 2.1 (131 people per square mile). The number of zip code falling in a rural category was high for Vermont, New Hampshire, and Maine while Connecticut, New York, Massachusetts, and Rhode Island had almost equal mix of rural and urban zip codes. Overall the relationship between hospitalization rates and population density was U-shaped with a marked increase at both extremes: for heavily populated and the least populated zip codes (Figs. 5 and 6). This nonlinearity requires exploring the relationship separately for urban, rural, and semi-urban zip codes. After adjusting for income and percent elderly population, higher evergreen vegetation in urban areas demonstrated a small yet protective effect.

Summary of Examples

The provided or included examples are not intended to imply that all measures of rural-urban status are invalid and inconsistent. Rather, they highlight the need to consider the specific rural-urban status measure being used and what aspect or aspects of the rural-urban continuum the selected measure is intended to emphasize. Moreover, as with any predictor variable used in modeling health outcomes, whether it is geospatial or traditional, non-spatial models, it is important to consider the trade-offs of using one variable type over another. For example, as discussed, treating rural-urban status as a continuous or discrete variable reduces a model degrees of freedom and may optimize statistical power. However, this use assumes a monotonic association between rural-urban status and the health outcome(s) under study. Using indicator variables, as illustrated, can allow for the assessment of non-monotonic associations but require additional model degrees of freedom. There is no one valid way to use rural-urban

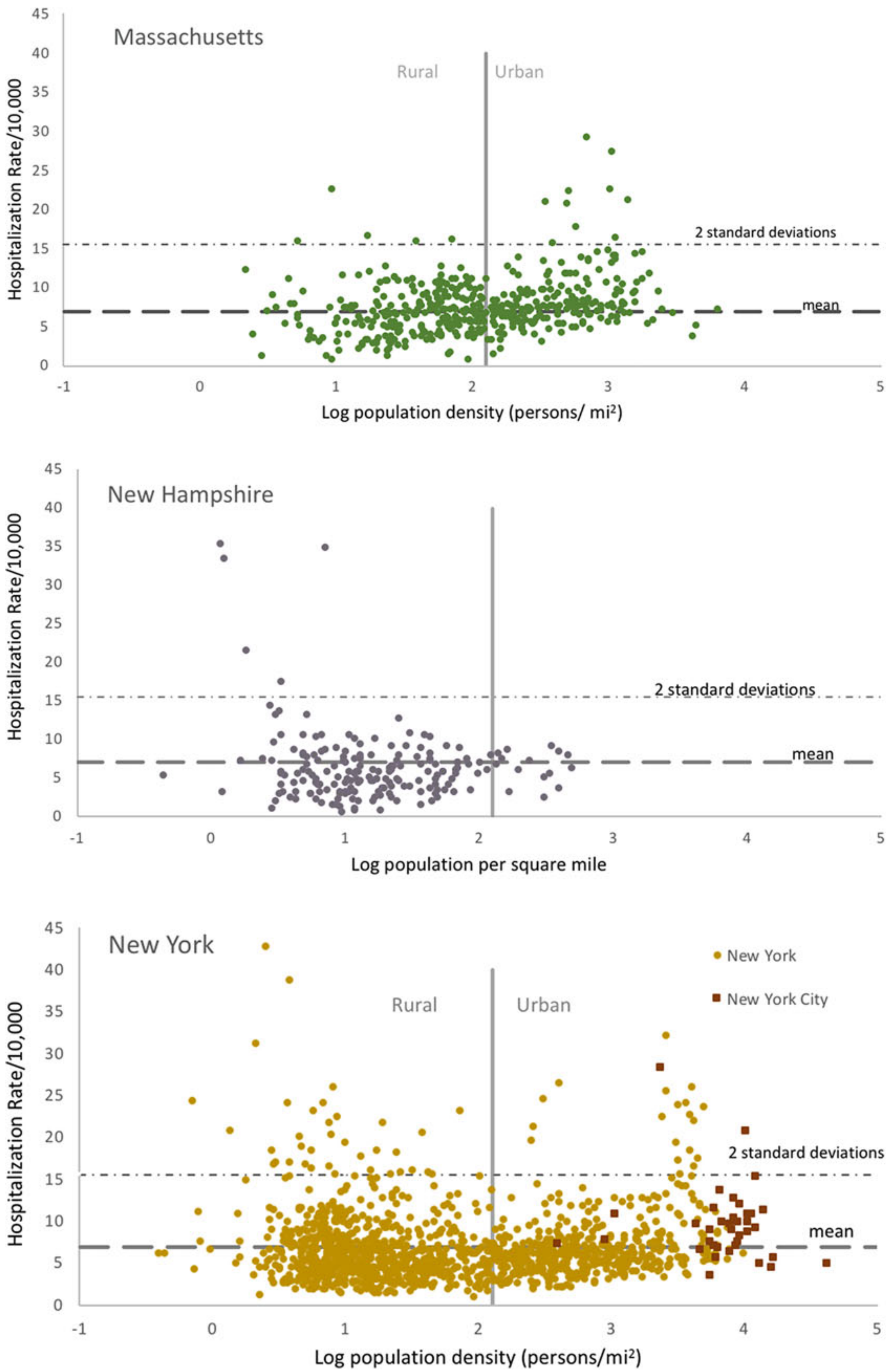

Fig. 5 Relationship between population density of asthma hospitalization rates among older adults in selected states, 2005–2006

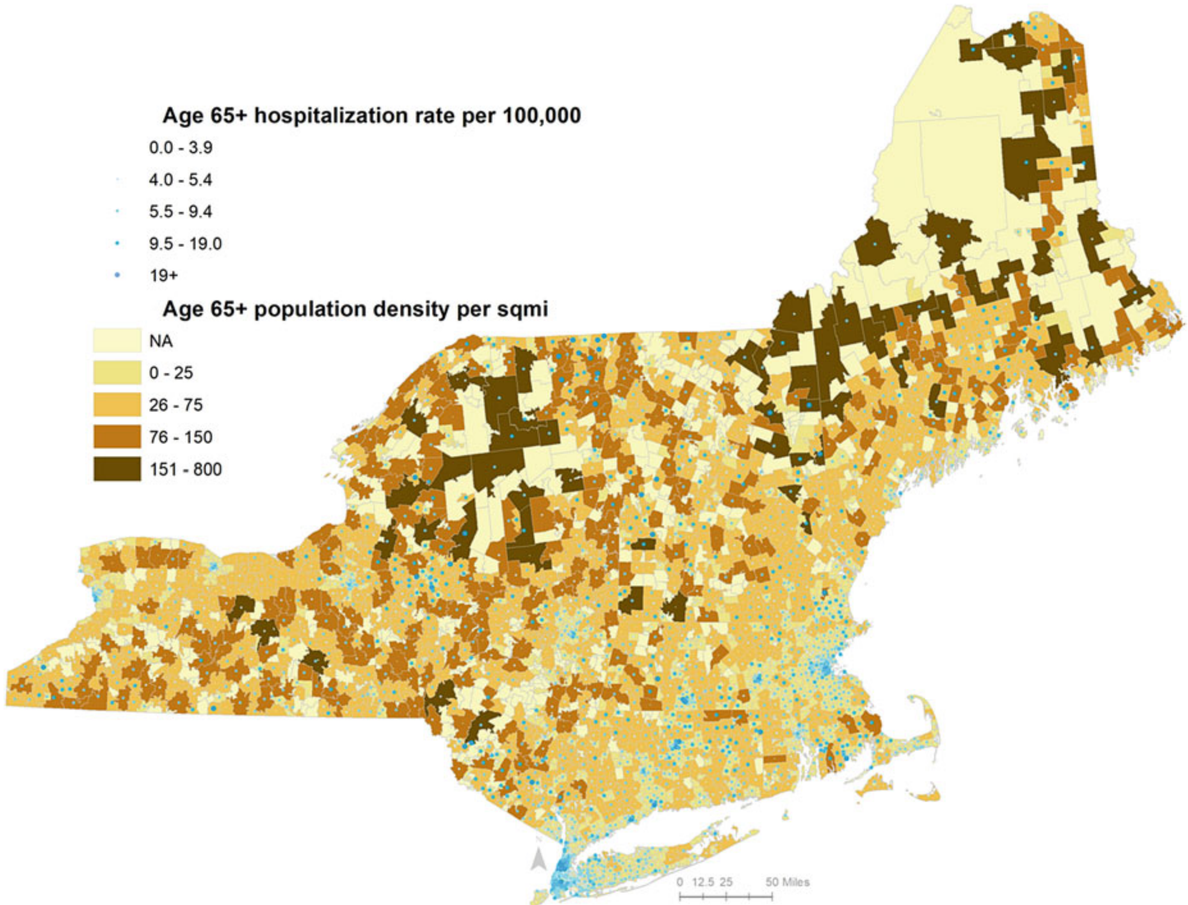

Fig. 6 Relationship between population density of asthma hospitalization rates among older adults in New York and six New England states, 2005–2006

status, but its use requires attention to these and other issues to properly characterize the relationship between rural-urban status and the health outcome under study.

Recommendations and Conclusions

This chapter addresses a handful of the many issues associated with selecting and utilizing variables to address key place-based social determinants of health, with examples to the rural-urban gradient. This chapter may raise more questions than it answers regarding decisions around selection of measures to use and how to use them. There remains no scientific consensus as to best practices, and it is left to the researcher to decide which measures to use based on study questions, on what geographic scale or scales to use them, and interpretation of findings.

When using place-based characteristics in geospatial models and, by extension, in non-spatial models, it is important to consider the following questions: First, on what geographic scale will the characteristic be measured and analyzed? Different scales provide certain benefits and drawbacks in terms of statistical stability, policy relevance, sample size, availability of data, and other considerations. Second, what factors or determinants are most relevant for answering the research question? This question raises the issue of policy relevance, ability to take action upon significant findings, accuracy of the measure, and numerous other issues. Many measures are multidimensional, and selecting one over the other may have meaningful implications for the directionality, magnitude, and overall nature of any observed association. Third, what type of variable will be used in the analysis? This question is relevant to all types of models, not just geospatial models. Different types of variables offer trade-offs in terms of modeling the type of association, interpretability of findings, and statistical power. In the example of rural-urban status, there is a need to use the concept of a power law to incorporate rural-urban metrics that take into account population distribution and population density measures that could be sustainable and valid across different geographic aggregation schemes.

This chapter discussed the rural-urban gradient as an example of a social determinant of health explored in a growing body of population health studies that is intrinsically linked to geography. Although the discussion, particularly the case studies, focuses on issues pertaining to measuring rural-urban health status specifically, the broader concepts of geographic scale, policy relevance, statistical power, the modifiable area unit problem, and many of the other issues described above extend to other social determinants of health, such as SES, household composition, education, income inequality, and demographic structure (e.g., age, race/ethnicity, gender, etc.). Furthermore, for rural-urban status, income inequality and other social determinants of health are intrinsically tied to the concept of place. The processes we are trying to capture are dynamic, yet we are limited by the preponderance of static tools and measures available to researchers. Therefore, all of these measures have a critical temporal component that may be challenging to address in standard geospatial modeling. What we observed today with respect to these social determinants is not necessarily can be observed in the past yet quite likely reflects the consequences of the past, including historical nature-made and man-made events and other perhaps ongoing measurable and unmeasurable processes.

The place-based factors discussed in this chapter and other social determinants of health often represent the ultimate or distal causes of disease and health disparities. On the other hand, they also provide opportunities on which base policies, programs, and interventions can be designed to promote healthy behaviors, improve population health, and ultimately reduce health disparities. Awareness of the issues surrounding measurement of these determinants is integral to conduct meaningful and impactful research through which population health can be improved.

References

Adler, N.E., and K. Newman. 2002. Socioeconomic disparities in health: Pathways and policies. *Health Affairs* 21 (2): 60–76.
Adler, N.E., and J.M. Ostrove. 1999. Socioeconomic status and health: What we know and what we don't. *Annals of the New York Academy of Sciences* 896 (1): 3–15.
Adler, N.E., and D.H. Rehkopf. 2008. US disparities in health: Descriptions, causes, and mechanisms. *Annual Review of Public Health* 29: 235–252.
Armstrong, M.P., G. Rushton, and D.L. Zimmerman. 1999. Geographically masking health data to preserve confidentiality. *Statistics in Medicine* 18: 497–525.
Baicker, K., A. Chandra, and J. Skinner. 2005. Geographic variation in health care and the problem of measuring racial disparities. *Perspectives in Biology and Medicine* 48 (1): 42–S53.
Banerjee, S., B.P. Carlin, and A.E. Gelfand. 2014. *Hierarchical modeling and analysis for spatial data*. CRC press.
Beck, A.F., M.T. Sandel, P.H. Ryan, and R.S. Kahn. 2017. Mapping neighborhood health geomarkers to clinical care decisions to promote equity in child health. *Health Affairs* 36 (6): 999–1005.
Benach, J., Y. Yasui, C. Borrell, M. Sáez, and M.I. Pasarin. 2001. Material deprivation and leading causes of death by gender: Evidence from a nationwide small area study. *Journal of Epidemiology & Community Health* 55 (4): 239–245.
Black, M., S. Ebener, P.N. Aguilar, M. Vidaurre, and Z. El Morjani. 2004. Using GIS to measure physical accessibility to health care. *World Health Organization*: 3–4.
Branas, C.C., E.J. MacKenzie, and C.S. ReVelle. 2000. A trauma resource allocation model for ambulances and hospitals. *Health Services Research* 35 (2): 489.
Braveman, P.A., S.A. Egerter, S.H. Woolf, and J.S. Marks. 2011. When do we know enough to recommend action on the social determinants of health? *American Journal of Preventive Medicine* 40 (1): S58–S66.
Chui, K.K., S.A. Cohen, and E.N. Naumova. 2011. Snowbirds and infection–new phenomena in pneumonia and influenza hospitalizations from winter migration of older adults: A spatiotemporal analysis. *BMC Public Health* 11 (1): 444.
Clapp, J.M., and Y. Wang. 2006. Defining neighborhood boundaries: Are census tracts obsolete? *Journal of Urban Economics* 59 (2): 259–284.
Clough-Gorr, K.M., M. Egger, and A. Spoerri. 2015. A Swiss paradox? Higher income inequality of municipalities is associated with lower mortality in Switzerland. *European Journal of Epidemiology* 30 (8): 627–636.
Cohen, D., S. Spear, R. Scribner, P. Kissinger, K. Mason, and J. Wildgen. 2000. "Broken windows" and the risk of gonorrhea. *American Journal of Public Health* 90 (2): 230.
Cohen, S.A., L. Kelley, and A.E. Bell. 2015. Spatiotemporal discordance in five common measures of rurality for US counties and applications for health disparities research in older adults. *Frontiers in Public Health* 3: 267.
Cohen, S.A., S.K. Cook, T.A. Sando, and N.J. Sabik. 2018a. What aspects of rural life contribute to rural-urban health disparities in older adults? Evidence from a national survey. *Journal of Rural Health* 34: 293–303.
Cohen, S.A., M.L. Greaney, and N.J. Sabik. 2018b. Assessment of dietary patterns, physical activity and obesity from a national survey: Rural-urban health disparities in older adults. *PLoS One* 13 (12): e0208268.
Cummins, S., S. Curtis, A.V. Diez-Roux, and S. Macintyre. 2007. Understanding and representing 'place' in health research: A relational approach. *Social Science & Medicine* 65 (9): 1825–1838.
Dahly, D.L., and L.S. Adair. 2007. Quantifying the urban environment: A scale measure of urbanicity outperforms the urban–rural dichotomy. *Social Science & Medicine* 64 (7): 1407–1419.
Diez-Roux, A.V. 1998. Bringing context back into epidemiology: Variables and fallacies in multilevel analysis. *American Journal of Public Health* 88 (2): 216–222.
Dwyer-Lindgren, L., A. Bertozzi-Villa, R.W. Stubbs, C. Morozoff, J.P. Mackenbach, F.J. van Lenthe, and C.J. Murray. 2017. Inequalities in life expectancy among US counties, 1980 to 2014: Temporal trends and key drivers. *JAMA Internal Medicine* 177 (7): 1003–1011.
Eckenrode, J., E.G. Smith, M.E. McCarthy, and M. Dineen. 2014. Income inequality and child maltreatment in the United States. *Pediatrics*. peds-2013.
Elliott, P., and D. Wartenberg. 2004. Spatial epidemiology: Current approaches and future challenges. *Environmental Health Perspectives* 112 (9): 998.
Erdman, E., A. Liss, D. Gute, C. Rioux, M. Koch, and E.N. Naumova. 2015. Does the presence of vegetation affect asthma hospitalizations among the elderly? A comparison between rural, suburban, and urban areas. *International Journal of Environment and Sustainability* 4 (1).
Fefferman, N., E. O'Neil, and E.N. Naumova. 2005. Confidentiality and confidence: Is data aggregation a means to achieve both? *Journal of Public Health Policy* 26 (3): 430–449.

Fiscella, K., and P. Franks. 1997. Poverty or income inequality as predictor of mortality: Longitudinal cohort study. *BMJ* 314 (7096): 1724.

Gilmer, T.P., and R.G. Kronick. 2011. Differences in the volume of services and in prices drive big variations in Medicaid spending among US states and regions. *Health Affairs* 30 (7): 1316–1324.

Gordon-Larsen, P., M.C. Nelson, P. Page, and B.M. Popkin. 2006. Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics* 117 (2): 417–424.

Grow, H.M.G., A.J. Cook, D.E. Arterburn, B.E. Saelens, A. Drewnowski, and P. Lozano. 2010. Child obesity associated with social disadvantage of children's neighborhoods. *Social Science & Medicine* 71 (3): 584–591.

Hanchate, A.D., M.K. Paasche-Orlow, K.S. Dyer, W.E. Baker, C. Feng, and J. Feldman. 2017. Geographic variation in use of ambulance transport to the emergency department. *Annals of Emergency Medicine* 70 (4): 533–543.

Haque, A., and J. Telfair. 2000. Socioeconomic distress and health status: The urban-rural dichotomy of services utilization for people with sickle cell disorder in North Carolina. *The Journal of Rural Health* 16 (1): 43–55.

Hart, L.G., E.H. Larson, and D.M. Lishner. 2005. Rural definitions for health policy and research. *American Journal of Public Health* 95 (7): 1149–1155.

Jagai, J.S., J.K. Griffiths, P.H. Kirshen, P. Webb, and E.N. Naumova. 2010. Patterns of protozoan infections: Spatiotemporal associations with cattle density. *EcoHealth* 7 (1): 33–46.

Jia, H., F. Ordóñez, and M. Dessouky. 2007. A modeling framework for facility location of medical services for large-scale emergencies. *IIE Transactions* 39 (1): 41–55.

Juster, R.P., B.S. McEwen, and S.J. Lupien. 2010. Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews* 35 (1): 2–16.

Kahn, R.S., P.H. Wise, B.P. Kennedy, and I. Kawachi. 2000. State income inequality, household income, and maternal mental and physical health: Cross sectional national survey. *BMJ* 321 (7272): 1311.

Kawachi, I., and B.P. Kennedy. 1997. Socioeconomic determinants of health: Health and social cohesion: Why care about income inequality? *BMJ* 314 (7086): 1037.

———. 1999. Income inequality and health: Pathways and mechanisms. *Health Services Research* 34 (1 Pt 2): 215.

Kennedy, B.P., I. Kawachi, and D. Prothrow-Stith. 1996. Income distribution and mortality: Cross sectional ecological study of the Robin Hood index in the United States. *BMJ* 312 (7037): 1004–1007.

Kennedy, B.P., I. Kawachi, R. Glass, and D. Prothrow-Stith. 1998. Income distribution, socioeconomic status, and self rated health in the United States: Multilevel analysis. *BMJ* 317 (7163): 917–921.

Kondo, N., G. Sembajwe, I. Kawachi, R.M. van Dam, S.V. Subramanian, and Z. Yamagata. 2009. Income inequality, mortality, and self rated health: Meta-analysis of multilevel studies. *BMJ* 339: b4471.

Kounadi, O., and M. Leitner. 2014. Why does geoprivacy matter? The scientific publication of confidential data presented on maps. *Journal of Empirical Research on Human Research Ethics* 9 (4): 34–45.

Krieger, N., J.T. Chen, P.D. Waterman, M.J. Soobader, S.V. Subramanian, and R. Carson. 2002. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: Does the choice of area-based measure and geographic level matter? The Public Health Disparities Geocoding Project. *American Journal of Epidemiology* 156 (5): 471–482.

Kruger, D.J., T.M. Reischl, and G.C. Gee. 2007. Neighborhood social conditions mediate the association between physical deterioration and mental health. *American Journal of Community Psychology* 40 (3–4): 261–271.

Kulkarni, S.C., A. Levin-Rector, M. Ezzati, and C.J. Murray. 2011. Falling behind: Life expectancy in US counties from 2000 to 2007 in an international context. *Population Health Metrics* 9 (1): 16.

Lippert, A.M., C.R. Evans, F. Razak, and S.V. Subramanian. 2017. Associations of continuity and change in early neighborhood poverty with adult cardiometabolic biomarkers in the United States: Results from the National Longitudinal Study of Adolescent to Adult Health, 1995–2008. *American Journal of Epidemiology* 185 (9): 765–776.

Lupien, S.J., I. Ouellet-Morin, A. Hupbach, M.T. Tu, C. Buss, D. Walker, et al. 2015. Beyond the stress concept: Allostatic load—A developmental biological and cognitive perspective. *Developmental Psychopathology: Volume Two: Developmental Neuroscience*: 578–628.

Lynch, J.W., G.D. Smith, G.A. Kaplan, and J.S. House. 2000. Income inequality and mortality: Importance to health of individual income, psychosocial environment, or material conditions. *BMJ* 320 (7243): 1200–1204.

Macintyre, S. 1997. What are spatial effects and how can we measure them? In *Exploiting national survey data: The role of locality and spatial effects*, ed. A. Dale, 1–17. Manchester: Faculty of Economic and Social Studies, University of Manchester.

Macintyre, S., A. Ellaway, and S. Cummins. 2002. Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science & Medicine* 55 (1): 125–139.

Manley, D., R. Flowerdew, and D. Steel. 2006. Scales, levels and processes: Studying spatial patterns of British census variables. *Computers, Environment and Urban Systems* 30 (2): 143–160.

Marmot, M., and R. Bell. 2011. Social determinants and dental health. *Advances in Dental Research* 23 (2): 201–206.

Matlock, D.D., P.W. Groeneveld, S. Sidney, S. Shetterly, G. Goodrich, K. Glenn, et al. 2013. Geographic variation in cardiovascular procedure use among Medicare fee-for-service vs Medicare advantage beneficiaries. *JAMA* 310 (2): 155–161.

McDonald, D.C., K. Carlson, and D. Izrael. 2012. Geographic variation in opioid prescribing in the US. *The Journal of Pain* 13 (10): 988–996.

McEwen, B.S. 1998. Protective and damaging effects of stress mediators. *New England Journal of Medicine* 338 (3): 171–179.

McGinnis, J.M., P. Williams-Russo, and J.R. Knickman. 2002. The case for more active policy attention to health promotion. *Health Affairs* 21 (2): 78–93.

McLafferty, S.L. 2003. GIS and health care. *Annual Review of Public Health* 24 (1): 25–42.

Mohnen, S.M., B. Völker, H. Flap, and P.P. Groenewegen. 2012. Health-related behavior as a mechanism behind the relationship between neighborhood social capital and individual health-a multilevel analysis. *BMC Public Health* 12 (1): 116.

Naumova, E.N., S.M. Parisi, D. Castronovo, M. Pandita, J. Wenger, and P. Minihan. 2009. Pneumonia and influenza hospitalizations in elderly people with dementia. *Journal of the American Geriatrics Society* 57 (12): 2192–2199.

Newhouse, J.P., and A.M. Garber. 2013a. Geographic variation in Medicare services. *New England Journal of Medicine* 368 (16): 1465–1468.

———. 2013b. Geographic variation in health care spending in the United States: Insights from an Institute of Medicine report. *JAMA* 310 (12): 1227–1228.

Nicholas, L.H., K.M. Langa, T.J. Iwashyna, and D.R. Weir. 2011. Regional variation in the association between advance directives and end-of-life Medicare expenditures. *JAMA* 306 (13): 1447–1453.

O'Keefe, C.M., and D.B. Rubin. 2015. Individual privacy versus public good: Protecting confidentiality in health research. *Statistics in Medicine* 34 (23): 3081–3103.

Openshaw, S. 1984. The modifiable areal unit problem. *Concepts and Techniques in Modern Geography*.

Padilla, C.M., W. Kihal-Talantikit, V.M. Vieira, and S. Deguen. 2016. City-specific spatiotemporal infant and neonatal mortality clusters: Links with socioeconomic and air pollution spatial patterns in France. *International Journal of Environmental Research and Public Health* 13 (6): 624.

Park, H., A.M. Roubal, A. Jovaag, K.P. Gennuso, and B.B. Catlin. 2015. Relative contributions of a set of health factors to selected health outcomes. *American Journal of Preventive Medicine* 49 (6): 961–969.

Pearce, J., K. Witten, and P. Bartie. 2006. Neighbourhoods and health: A GIS approach to measuring community resource accessibility. *Journal of Epidemiology & Community Health* 60 (5): 389–395.

Prince, S.A., E.A. Kristjansson, K. Russell, J.M. Billette, M.C. Sawada, A. Ali, et al. 2012. Relationships between neighborhoods, physical activity, and obesity: A multilevel analysis of a large Canadian city. *Obesity* 20 (10): 2093–2100.

Remington, P.L., B.B. Catlin, and K.P. Gennuso. 2015. The county health rankings: Rationale and methods. *Population Health Metrics* 13 (1): 11.

Seeman, T., E. Epel, T. Gruenewald, A. Karlamangla, and B.S. McEwen. 2010. Socio-economic differentials in peripheral biology: Cumulative allostatic load. *Annals of the New York Academy of Sciences* 1186 (1): 223–239.

Stiel, L., S. Soret, and S. Montgomery. 2017. Geographic patterns of change over time in mammography: Differences between Black and White US Medicare enrollees. *Cancer Epidemiology* 46: 57–65.

Subramanian, S.V., and I. Kawachi. 2003. The association between state income inequality and worse health is not confounded by race. *International Journal of Epidemiology* 32 (6): 1022–1028.

———. 2004. Income inequality and health: What have we learned so far? *Epidemiologic Reviews* 26 (1): 78–91.

Suzuki, E., S. Kashima, I. Kawachi, and S.V. Subramanian. 2012. Social and geographic inequalities in premature adult mortality in Japan: A multilevel observational study from 1970 to 2005. *BMJ Open* 2 (2): e000425.

Tarlov, A.R. 1999. Public policy frameworks for improving population health. *Annals of the New York Academy of Sciences* 896 (1): 281–293.

United States Bureau of the Census. Decennial Census 2010. Website: https://data.census.gov/cedsci/advanced. Accessed August 22, 2021.

Vincens, N., and M. Stafström. 2015. Income inequality, economic growth and stroke mortality in Brazil: Longitudinal and regional analysis 2002-2009. *PLoS One* 10 (9): e0137332.

Waldorf, B. 2007. What is rural and what is urban in Indiana. *Purdue Center for Regional Development Report* 4.

Walker, K.E., and S.M. Crotty. 2015. Classifying high-prevalence neighborhoods for cardiovascular disease in Texas. *Applied Geography* 57: 22–31.

Waller, L.A., and C.A. Gotway. 2004. *Applied spatial statistics for public health data*. Vol. 368. Wiley, Hoboken, New Jersey, USA.

Weich, S., G. Holt, L. Twigg, K. Jones, and G. Lewis. 2003. Geographic variation in the prevalence of common mental disorders in Britain: A multilevel investigation. *American Journal of Epidemiology* 157 (8): 730–737.

Wennberg, J.E., E.S. Fisher, and J.S. Skinner. 2002. Geography and the debate over Medicare reform. *Health Affairs* 21 (2): 10–10.

Wilkinson, R.G., and K.E. Pickett. 2006. Income inequality and population health: A review and explanation of the evidence. *Social Science & Medicine* 62 (7): 1768–1784.

Williams, D.R., and C. Collins. 1995. US socioeconomic and racial differences in health: Patterns and explanations. *Annual Review of Sociology* 21 (1): 349–386.

Wilson, W.J. 1987. *The truly disadvantaged: The inner city, the underclass, and public policy*. University of Chicago Press.

Woolf, S.H., and P. Braveman. 2011. Where health disparities begin: The role of social and economic determinants—And why current policies may make matters worse. *Health Affairs* 30 (10): 1852–1859.

Zandbergen, P.A. 2014. Ensuring confidentiality of geocoded health data: assessing geographic masking strategies for individual-level data. *Advances in Medicine*: 2014.

Zhang, Q., and Y. Wang. 2004. Socioeconomic inequality of obesity in the United States: Do gender, age, and ethnicity matter? *Social Science & Medicine* 58 (6): 1171–1180.