Applied Geography 57 (2015) 22–31

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Classifying high-prevalence neighborhoods for cardiovascular disease in Texas

Kyle E. Walker*, Sean M. Crotty

Texas Christian University, Department of History and Geography, 2850 S University Dr, Fort Worth, TX, 76129, USA

ARTICLE INFO

Article history:
Available online

Keywords:
Cardiovascular disease
Texas
Cluster analysis
Demographics
Public health

ABSTRACT

Cardiovascular disease (CVD) is the number one cause of death in the state of Texas. In order to develop effective healthcare policies to combat CVD, it is essential to understand what types of communities are most affected. In this paper, we develop a classification scheme to segment high-prevalence communities based on selected social and demographic characteristics. We find that while many high-prevalence areas reflect traditional relationships between socioeconomic deprivation and poor health outcomes, a subset of more affluent communities in Texas are also affected. This suggests the importance of tailored approaches to prevention that account for this diversity among high-prevalence neighborhoods.

Introduction

Cardiovascular disease (CVD), which refers to a suite of diseases that impact the heart and blood vessels, is the top cause of death in Texas, claiming over 38,000 lives in 2010. Current forecasts for CVD in Texas predict that direct healthcare costs associated with CVD will triple, reaching nearly 1 trillion dollars per year by 2030 (Heidenreich et al., 2011). Both public and private sector groups are working to improve CVD health in Texas and beyond. The American Heart Association's 2020 impact goal is "to improve the cardiovascular health of all Americans by 20 percent, while reducing deaths from cardiovascular diseases and stroke by 20 percent" (Go et al., 2014). Considering the significant social and economic costs of CVD, it is imperative that policy interventions for CVD take into account its variable impacts in different types of populations and neighborhoods.

While rates of CVD have declined in recent years, it continues to disproportionately impact a variety of demographic groups, including African-Americans and individuals living in poverty (Ang et al., 2013). The links between disparities in CVD prevalence and demographic characteristics are well-established in the public health and epidemiology literature. In a review of this literature, Kurian and Cardarelli (2007) find a consensus that hypertension and diabetes tend to be elevated in the African-American community, and leisure-time inactivity is higher amongst Mexican-Americans. Such findings, they argue, help explain disparities in mortality from CVD across racial and ethnic groups. Scholars have similarly established relationships between socio-economic marginality and CVD risk. For example, Jones et al. (2009) observe an elevated risk for CVD amongst impoverished and homeless populations in the US and Canada, and Ford (2013) finds a significant relationship between food insecurity and CVD in the US.

Such demographic disparities in CVD prevalence in turn manifest themselves geographically, with some neighborhoods bearing a disproportionate burden of heart disease. In turn, scholars have employed a variety of approaches to investigate a relationship between location and cardiovascular health outcomes. In one example, Hu and Rao (2009) used remote sensing techniques and county-level mortality data to demonstrate positive spatial correlation between aerosol air pollution and Chronic Ischemic Heart Disease (CHID) mortality rates, warranting targeted policy interventions to decrease aerosol pollution. Similarly, Michimi, Ellis-Griffith, Nagy, and Peterson (2013) examined the relationship between the employment base of US Metropolitan/Micropolitan Statistical Areas (MMSA) and Coronary Heart Disease (CHD). Their study showed that CHD prevalence is lower in MMSAs where tourism or quaternary sector economic activities are more prevalent; conversely, CHD prevalence is higher in places with economies dominated by secondary and primary sector industries such as manufacturing and mining. CVD hospitalization and mortality also varies spatially based on accessibility to hospital services and

---

* Corresponding author. Tel.: +1 817 257 5241.
E-mail addresses: kyle.walker@tcu.edu (K.E. Walker), sean.crotty@tcu.edu (S.M. Crotty).

http://dx.doi.org/10.1016/j.apgeog.2014.11.011
0143-6228/© 2014 Elsevier Ltd. All rights reserved.

facilities (Hare & Barcus, 2007). Each of these studies demonstrates the role of a single factor in driving spatial variations in CVD. Considered together, however, these studies suggest that *multiple or overlapping factors* generate spatial variations in CVD prevalence. For this reason, interest in the relationship between neighborhood or local context on health has increased considerably from the 2000's through today (Diez-Roux, 2001).

Interest in neighborhood health effects is not limited to research examining CVD. Other studies have identified relationships between social or demographic characteristics of neighborhood residents and general health outcomes. Bosma, Dike van de Mheen, Borsboom, and Mackenbach (2001) for example, undertook a longitudinal health survey of more than 8000 people living in the Netherlands to examine the relationship between neighborhood characteristics and mortality. They found that independent of individual socioeconomic characteristics, mortality rates were higher for people living in neighborhoods with high rates of unemployment, disability, or poverty. Among living survey-respondents, social disintegration and unhealthy psychological profiles and behaviors were more common among those living in economically marginal neighborhoods. These findings suggest that certain neighborhood characteristics cause higher stress levels for all inhabitants that lead to poor health outcomes.

The salience of such neighborhood-level disparities in health outcomes have led some researchers to adopt classification methods to explore these geographic variations. Lalloué et al. (2013) provide an example of this approach. Their study analyzed tract-level demographic data from the French census using a variety of statistical approaches, including principal components analysis and hierarchical clustering, to generate a numerical index by which neighborhoods can be grouped according to health risk. The variables included in the study were all shown to be important predictors of health outcomes in previous studies: income, ethnicity, immigrant status, education, etc. However, Lalloué's classification system uses no direct measures of health, such as hospitalization records or mortality data. As such, the index and resulting groups have a high potential for producing a spatialized ecological fallacy (Grubesic, Miller, & Murray, 2014). In contrast, Shortt, Richardson, Mitchell, and Pearce (2011, 2012) employ principal components and cluster analysis to develop a neighborhood classification scheme for the UK and New Zealand based on a range of environmental factors hypothesized to have a relationship with health outcomes. They then incorporate age-adjusted mortality data into negative binomial regression models to identify variation in mortality from various diseases (including CVD) amongst the different cluster types. Bellis et al. (2012) adopt a similar methodology to analyze health disparities in England. They use *k*-means clustering to sort English localities into 5 distinct clusters, and find that health outcomes vary significantly amongst the clusters, with the worst outcomes found in Northern England. In turn, they argue that health policy must take into account the distinct variations between these localities to be effective.

In this paper, we contribute to the growing body of work applying geographic techniques and analysis to public health research (Grubesic et al., 2014; Lin, Schootman, & Zhan, 2015; Pierce, Martin, Scherr, & Greiner, 2012; Weeks et al., 2012). This literature has and continues to push forward our collective understanding of the spatiality of disease risk and burden, while creating new research methods and adapting long-respected research methods to meet the topical and data challenges of health-science research (Diez-Roux, 2000; Krieger et al., 2002; Root, 2012). Our research further supports research within geography and public health disciplines that examines the relationships between neighborhood characteristics and social outcomes (Bell, Wilson, Bissonnette, & Shah, 2012; Grady & Darden, 2012; Hernandez, Roy Chowdhury, Fleming, & Griffith, 2011; LeDoux & Vojnovic, 2014; Lee & Cubbin, 2002). Such geographical methods, we believe, have the potential to offer a more detailed and nuanced portrait of the demography of cardiovascular disease. While much of the health sciences literature suggests general relationships between certain socio-demographic characteristics (e.g. poverty) and CVD prevalence, spatial variations in CVD revealed by a geographical approach can show additional complexity in these relationships and potentially challenge general, aspatial conclusions about the drivers of CVD prevalence. Such an approach is particularly important in a state like Texas, which exhibits considerable demographic and economic diversity by geography.

The specific intent of our research is to identify high-prevalence neighborhoods for CVD in the state, and then analyze how such neighborhoods vary both demographically and geographically. Using direct health measures to select the neighborhoods in our study allows for a more nuanced analysis of high-CVD prevalence neighborhoods than would be possible by classifying all the neighborhoods in our study area based on their social, economic, and behavioral characteristics and then attempting to link health outcomes to the neighborhood types after the classification was complete. We use hospital discharges for CVD in Texas in 2006 to identify neighborhoods that exhibit high prevalence of CVD. We then use factor analysis and *k*-means clustering to group high-CVD neighborhoods according to shared social, economic, and behavioral characteristics. We observe significant variation amongst neighborhoods with elevated CVD prevalence. While some neighborhoods are representative of commonly-identified relationships between socioeconomic deprivation and CVD, we find that a subset of high-prevalence neighborhoods is not deprived in the traditional sense, with relatively high incomes and low unemployment. In turn, our findings raise questions about how risk factors for CVD might manifest themselves differently in different types of neighborhoods.

Data and methods

Our data for this paper come from the Texas Hospital Inpatient Discharge Public Use Data File (PUDF) for 2006, available from the Texas Department of State Health Services. The PUDF contains information on all inpatient hospital stays in non-exempt hospitals licensed by the state of Texas[1] (Texas DSHS, 2006). We use data from 2006 as it was the most recent year for which data were freely available at the time of the study; the 2006 data file includes a total of 2,917,188 hospitalization records. To prepare the data for analysis, we wrote scripts to import the data into Stata and identify hospitalization cases in which patients received a heart disease-related diagnosis. To sort out these cases, we retained records with an ICD-9-CM diagnosis code corresponding to cardiovascular disease, comprising codes 390–459 and 745–747 (Go et al., 2014). In sum, 1,323,297 hospitalization cases had at least one diagnosis code corresponding to cardiovascular disease, comprising approximately 45 percent of the dataset.

Our next step was to calculate age-adjusted hospitalization rates by zip code (the finest level of geographic detail available in the dataset) to measure geographic disparities in hospitalization across the state. To preserve patient confidentiality, the PUDF suppresses

---

[1] Hospitals exempt from reporting their data “include those located in a county with a population less than 35,000, or those located in a county with a population more than 35,000 and with fewer than 100 licensed hospital beds and not located in an area that is delineated as an urbanized area by the United States Bureau of the Census. Exempt hospitals also include hospitals that do not seek insurance payment or government reimbursement” (Texas DSHS, 2006).

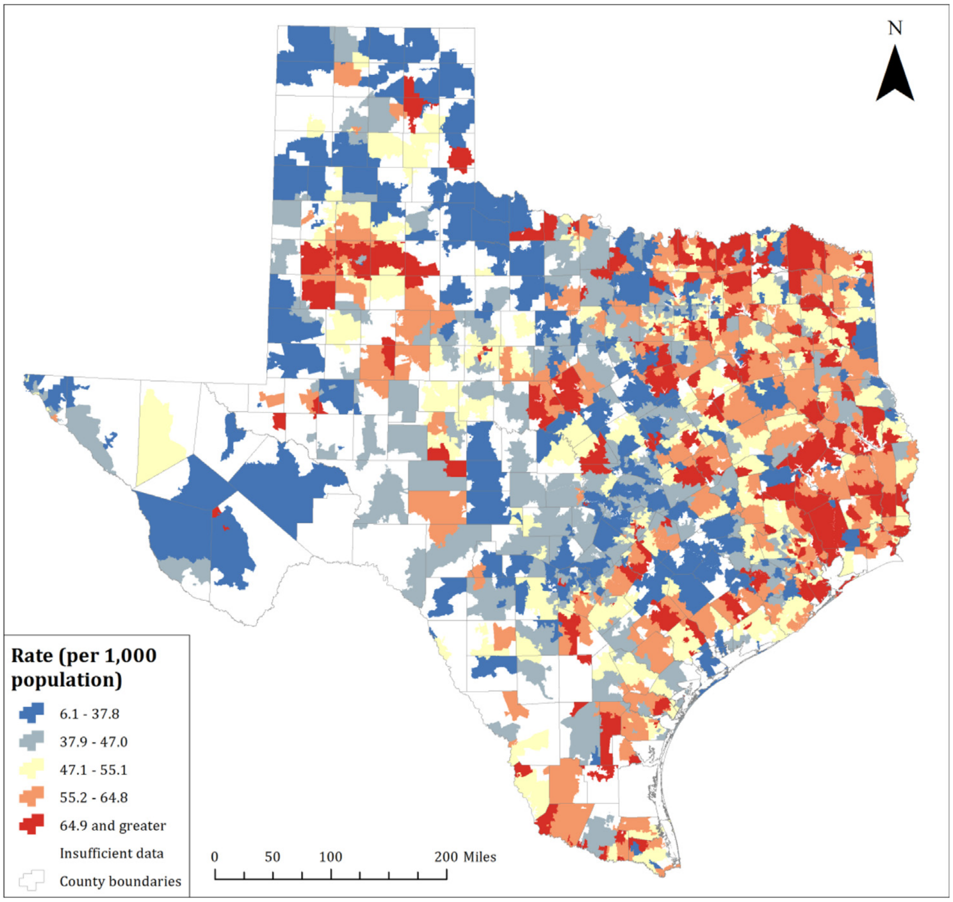

Fig. 1. Age-adjusted heart disease hospitalization rates in Texas by zip code, 2006.

the patient's zip code in certain instances.[2] In turn, we dropped cases that did not have zip-code level information encoded, or that had zip codes outside the state of Texas. This left us with 1,189,589 cases. We then used information on patient age from the PUDF to aggregate the data to the zip code level by age group (0–17, 18–44, 45–64, 65–74, and 75+), in order to facilitate the calculation of age-specific rates. As overall population counts for zip codes in 2006 were unavailable from the Census Bureau, we estimated age group counts for 2006 based on Census data from 2000 and 2010. Using GIS software, we identified Census blocks from 2000 and 2010 with their centroids in the boundaries of ZCTAs (the Census proxy for zip code areas) and used linear interpolation to estimate age group counts for 2006 based on these endpoints. We then matched these interpolated age group counts to our aggregated hospitalization data, and calculated age-adjusted hospitalization rates for each zip code using Texas as the standard population.[3] Of the 1906 ZCTAs in Texas in the 2006 file, we calculated age-adjusted rates in this way for 1,511, or 79.3 percent.

Fig. 1 displays the geographic distribution of age-adjusted hospitalization rates by ZCTA in Texas, with blue shades (in the web version) representing lower rates and red shades (in the web version) representing higher rates. This map suggests some general geographic trends in CVD hospitalization rates across the state. High rates are found in east and northeast Texas, as well as in the Rio Grande Valley and in west Texas near Lubbock. Lower rates are found in parts of central and northwest Texas. Figs. 2 and 3 show within-metropolitan variations in CVD hospitalizations with a focus on ZCTAs in and around Dallas and Houston.

Both the Dallas-Fort Worth and Houston maps exhibit striking inequities in CVD hospitalization rates in both metropolitan areas. In Dallas-Fort Worth, higher rates tend to be located in the neighborhoods to the south and east of both downtowns, with hospitalization rates exceeding 65 per 1000 residents in several of these ZCTAs, and in some instances exceeding 80 per 1000 residents in parts of South Dallas and the suburb of Balch Springs. The lowest rates are found in the neighborhoods to the north of downtown Dallas, with rates below 30 per 1000 in parts of north Dallas and the city of Highland Park, and to the northeast of the city of Fort Worth. Disparities are similarly present in the Houston area. The highest rates in Houston, sometimes exceeding 80 per 1000 residents, tend to be found in neighborhoods to the south and northeast of downtown. Conversely, the lowest rates are generally located in neighborhoods stretching from just west of downtown Houston to the western suburbs of the metropolitan area.

At first glance, these disparities appear to relate to socioeconomic inequalities within these metropolitan areas; for example,

---

[2] The complete five-digit zip code is suppressed in the PUDF in the following circumstances: if there are fewer than 30 patients in the zip code; if a hospital has fewer than fifty discharges in a quarter; if the ICD-9-CM code pertains to alcohol/drug abuse or HIV; and if a hospital has fewer than five discharges of a gender (Texas DSHS, 2006).

[3] Following convention in the literature (e.g. Ahmad et al., 2001), an age-adjusted rate AAR for a zip code a is calculated as follows: $\mathrm{AAR}_a = \sum r_{ia}(n_{is}/\sum n_{is})$, where r_{ia} is the age-specific rate for age group i in zip code a, and n_{is} is the overall population of age group i in the standard population s.

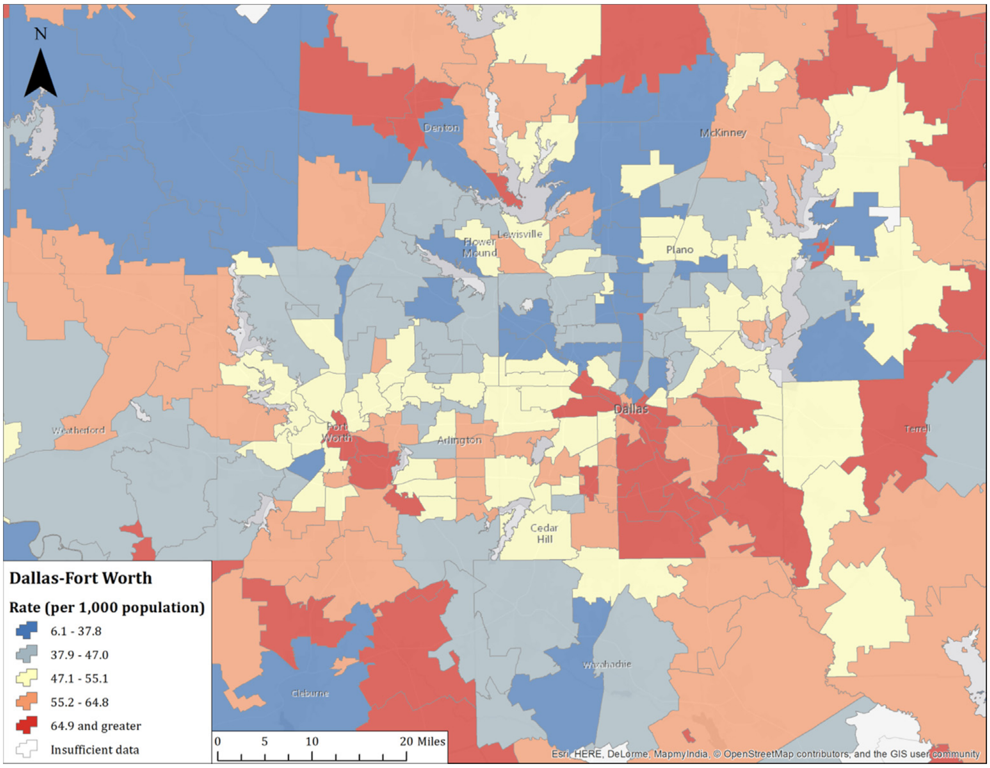

Fig. 2. CVD hospitalization rates in the Dallas-Fort Worth, Texas area.
Basemap Sources: Esri, DeLorne, MapmyIndia, OpenStreetMap contributors, and the GIS user community. Basemaps in this article were accessed using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com.

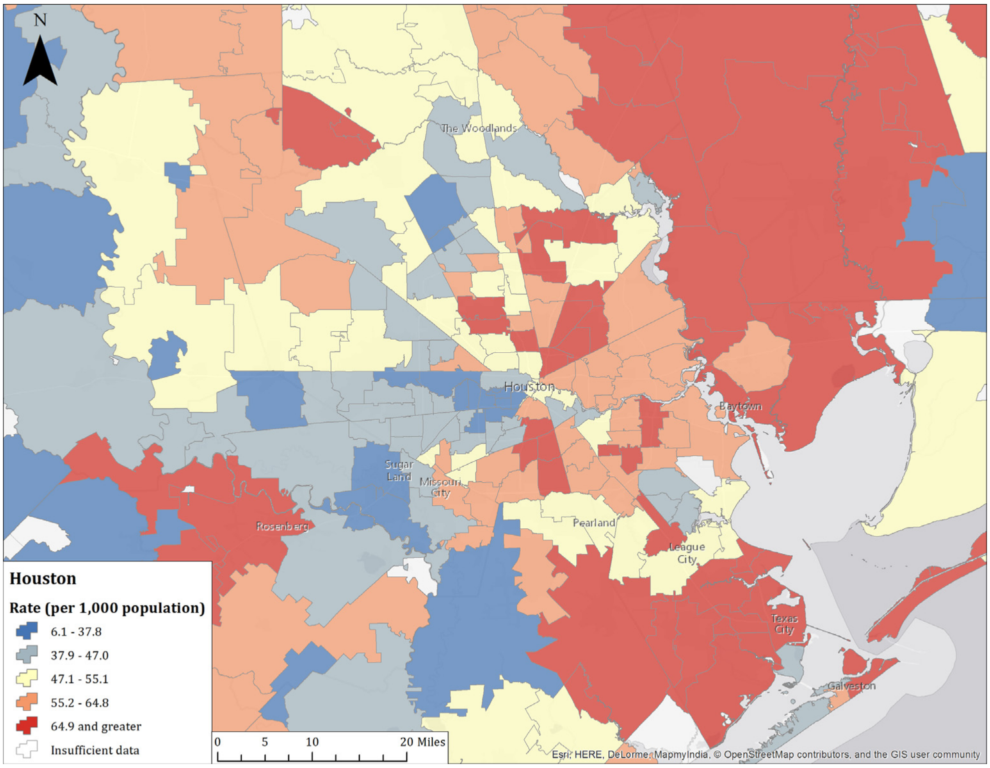

Fig. 3. CVD hospitalization rates in the Houston, Texas area.
Basemap Sources: Esri, DeLorne, MapmyIndia, OpenStreetMap contributors, and the GIS user community.

Table 1
Rotated factor loadings (orthogonal varimax rotation). Loadings <0.3 are omitted.

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Percent non-Hispanic white	0.86			−0.43
Percent non-Hispanic black				0.98
Percent Hispanic	−0.96			
Percent of population with a bachelor's degree		0.83		
Median household income	0.37	0.47	0.54	
Percent of the population in poverty	−0.53	−0.30	−0.57	
Percent driving alone to work			0.52	
Average commute time	0.37			
Percent unemployment		−0.32	−0.42	
Percent owner-occupied housing	0.31	−0.39	0.56	−0.39
Percent in white-collar workforce		0.72		
Percent with health insurance	0.57	0.41	0.33	
Population density	−0.43			0.35
Proportion variance explained	35%	25%	19%	19%

the ZCTAs to the north of downtown Dallas and to the west of downtown Houston are also among the wealthiest in their respective cities. While we do observe geographic trends in ZCTAs with high rates of CVD hospitalization, these high rates are not limited to specific areas within cities. In Houston, high hospitalization rates stretch from urban neighborhoods to the south and west of downtown to more rural areas on the eastern fringes of the metropolitan area. In turn, our analysis seeks to make sense of this diversity by identifying what types of neighborhoods in Texas are experiencing elevated rates of CVD hospitalization.

To conduct this analysis, we kept the quarter of ZCTAs with the highest hospitalization rates and merged this dataset with several social and demographic variables from the 2007–2011 American Community Survey, the closest period for which such variables were available at the ZCTA level.[4] We selected several variables to measure the general social and economic characteristics of these communities, which are listed in Table 1[5]. After dropping a small number of ZCTAs for which ACS data were unavailable, we retained 368 ZCTAs for further analysis.

To classify the ZCTAs based on these variables, we turned to the *k*-means clustering algorithm, which is commonly used in classification analyses in both health research and geography more generally. *K*-means is an unsupervised, iterative algorithm that seeks to sort a dataset into *k* sub-groups, or clusters, such that the within-group variation is minimized (James, Witten, Hastie, & Tibshirani, 2013). Before performing the *k*-means cluster analysis, we reduced the dimensions of our dataset into orthogonal factors, as is standard practice in classification research in social geography (Vicino, Hanlon, & Short, 2011; Wyly & DeFilippis, 2010) and public health (Odoi et al., 2005; Pedigo, Seaver, & Odoi, 2011). These factors are then used as inputs to our *k*-means cluster analysis and inform our classification of the high-rate ZCTAs.

Based on our factor analysis (principal factors method, varimax rotation), we retained four factors that collectively account for 99.7 percent of the variance in our dataset. Rotated factor loadings for the four factors are found in Table 1. **Factor 1** predominantly measures the extent to which a ZCTA is homogenously non-Hispanic white or homogenously Hispanic; positive scores suggest a larger white population, whereas negative scores reflect a larger Hispanic population. **Factor 2** reflects the educational and professional backgrounds of the ZCTA population, with higher scores indicating a more educated, white-collar workforce. **Factor 3** can be interpreted as a measure of economic well-being; higher scores suggest higher incomes and higher percentages of owner-occupied housing, whereas lower scores point to higher poverty and unemployment. Finally, **Factor 4** is largely a measure of the size of the non-Hispanic black population in the ZCTA; however, negative scores also reflect lower population density, more owner-occupied housing, and a larger non-Hispanic white population.

We then performed *k*-means clustering on the ZCTAs using scores for these four factors, requiring us to select the number of clusters *k* in advance. We first employed the common "elbow method" and plotted the within-group sums of squares for different values of *k*, looking for a "bend" in the plot that indicated diminishing returns in within-group variability (Claude, 2008). We observed a bend around 5 or 6 clusters. We then mapped the cluster solutions for $k = 4$, 5, and 6, in order to examine which cluster solution made the most practical sense in terms of geographically partitioning ZCTAs (James et al., 2013). This analysis led us to settle on a six-cluster solution.

Results

In the space that follows, we discuss the general characteristics of the six clusters in turn, and then consider the geographic distribution of these clusters in Texas. In Table 2, we provide descriptive statistics for each of the six clusters regarding the variables that served as inputs to our factor and cluster analyses. Additionally, we provide maps of the geographic location of cluster members in the Dallas-Fort Worth and Houston metropolitan areas in Figs. 4 and 5. Areas not covered by clusters in the maps are in the bottom three-quarters of CVD hospitalization rates for Texas, and in turn not part of the cluster analysis.

Cluster 1 is characterized by high population densities; the average population density of ZCTAs in this cluster is over 2800 persons per square mile, far eclipsing that of other clusters. This cluster is also very racially diverse – on average, its ZCTAs are 47 percent white, 16 percent black, and 31 percent Hispanic – and has much higher educational attainment than other clusters, as the mean percentage of its ZCTA's population over 25 with at least a bachelor's degree is 33 percent. Geographically, ZCTAs in this cluster tend to be located in and around the downtowns of major cities, as reflected in the low percentage of owner-occupied housing in these ZCTAs, and in other high-density areas. In Dallas and Fort Worth, the ZCTAs covering both downtowns fall into this cluster, as does part of Galveston to the south of Houston. In particular, these ZCTAs appear to be characterized by high levels of income inequality, as evidenced by the high levels of educational attainment and high levels of poverty in these areas.

Cluster 2 is similarly very ethnically diverse, as on average its ZCTAs are 42 percent non-Hispanic white and 43 percent Hispanic. Educational attainment in these ZCTAs is relatively low, as on average around 12.5 percent of their adult populations have a college degree, and less than one quarter of the population is employed in white-collar jobs. Geographically, these clusters often consist of older suburban communities just outside of major cities that have transitioned from predominantly non-Hispanic white to heavily Hispanic, with Hispanics making up over 50 percent of the population in many of these ZCTAs. In Dallas, for example, several ZCTAs to the east of the city fall into Cluster 2, as does one in Grand

---

[4] 2007–2011 was the first period for which ACS data were available at the ZCTA level; further, given the small population size of most ZCTAs, this information is only available as part of the 5-year ACS. We acknowledge that this is an imperfect proxy for the demographic conditions around our 2006 hospital discharge data; however, we felt that this more recent data was preferable to data from the 2000 Census (the alternative) given the significant demographic shifts and population growth that have taken place in Texas during the 2000s.

[5] As the health insurance variable was first available in the 2008–2012 ACS, we use information from this dataset.

Table 2
Descriptive statistics for the six clusters. All statistics are means unless otherwise indicated. "Other" refers to ZCTAs with CVD rates in the bottom 75 percent.

Variable	1	2	3	4	5	6	Other
Median age-adjusted hospitalization rate	73.34	71.09	67.85	71.16	69.03	76.51	45.81
Percent non-Hispanic white	47.07	41.92	76.95	79.39	9.65	13.61	57.06
Percent non-Hispanic black	15.73	12.88	5.42	8.5	3.7	60.69	8.08
Percent Hispanic	31	43.04	15	9.89	85.82	23.98	30.72
Percent of population with a bachelor's degree	33.22	12.59	21.19	12.02	7.97	10.87	23.29
Average median household income	$43,530	$41,836	$59,171	$41,783	$29,338	$32,315	$52,614
Percent of the population in poverty	25.18	19.82	9.71	17.57	34.54	28.24	15.74
Percent driving alone to work	68.8	78.23	82.5	79.98	74.01	75.14	78.98
Average commute time (min)	20.3	23.1	25.7	28.39	22.35	25.82	23.66
Percent unemployment	9.19	7.33	4.94	8.01	11.07	13.24	6.80
Percent owner-occupied housing	28.8	69.53	81.41	77.05	68.39	56.48	70.22
Percent in white-collar workforce	37.85	23.84	34.16	24.22	18.58	20	32.45
Percent with health insurance	75.53	73.87	83.51	77.33	65.37	71.15	78.48
Population density (pers/sqmi)	2839.28	687.73	231.6	104.47	1494.81	1913.93	1264.07
n	23	64	57	133	53	38	1125

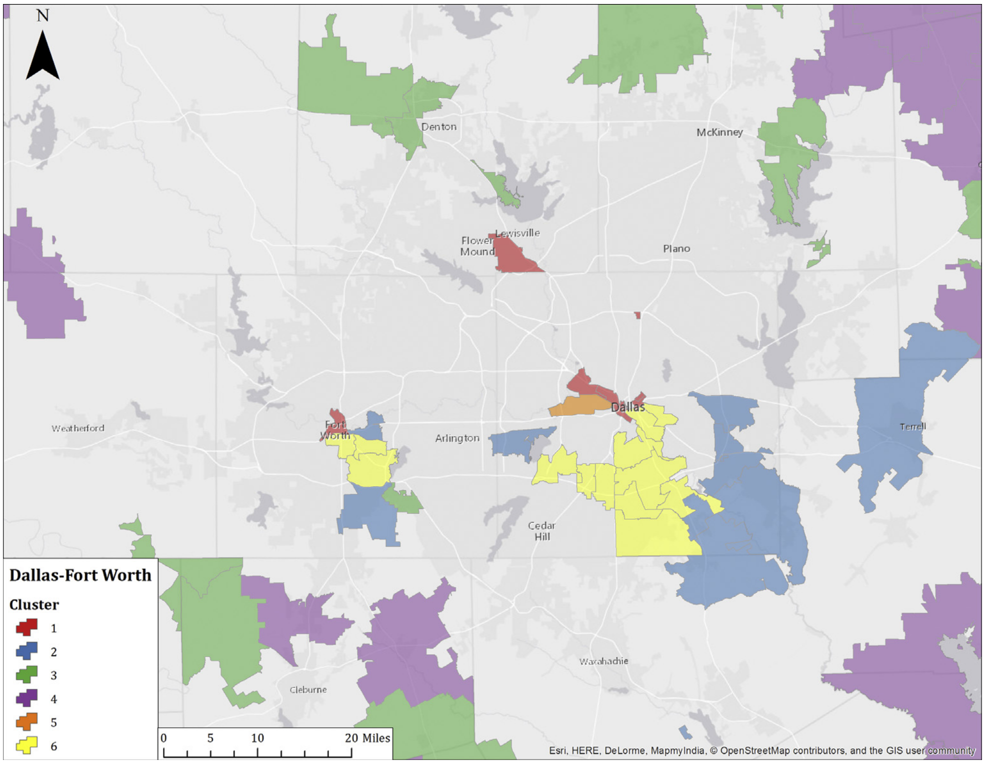

Fig. 4. Cluster locations in the Dallas-Fort Worth, Texas area.
Basemap Sources: Esri, DeLorne, MapmyIndia, OpenStreetMap contributors, and the GIS user community.

Prairie to the west of the city. In Houston, eastern suburbs such as Jacinto City and Baytown are classified in Cluster 2.

Cluster 3 stands out amongst the other clusters in a number of key areas. Household incomes in these ZCTAs are significantly higher than those in the other clusters; on average, the median household income of these ZCTAs is nearly $60,000, which is especially notable as it exceeds the average of ZCTAs that are not in the top 25 percent of hospitalization rates. Additionally, over 83 percent of the ZCTA populations, on average, have health insurance, and these ZCTAs also have a high average level of bachelor's degree attainment (21 percent) and low unemployment (under 5 percent) compared to the other clusters. Members of Cluster 3 are often located on the outskirts of major metropolitan areas, representing outer suburban or "exurban" communities. This is reflected in the relatively low population density (232 persons per square mile) and high proportion of the population driving alone to work (82.5 percent) on average in these communities. In Dallas-Fort Worth, ZCTAs in Cluster 3 include communities outside of Denton, to the north of Fort Worth, and far north of Dallas in Celina and Princeton. In Houston, suburbs such as La Porte and Dickinson fall into this cluster.

Cluster 4 is the largest cluster by numbers of ZCTAs, as 133 ZCTAs are classified in this cluster (36 percent of the ZCTAs in our analysis). This cluster consists largely of predominantly non-Hispanic white ZCTAs in rural Texas. On average, nearly 80 percent of the populations of ZCTAs in Cluster 4 are non-Hispanic white, with average black and Hispanic populations below 10 percent. These ZCTAs have the longest average commute times in the sample, at over 28 min, and the lowest population densities, averaging 104 persons per square mile. As indicated in the Dallas-

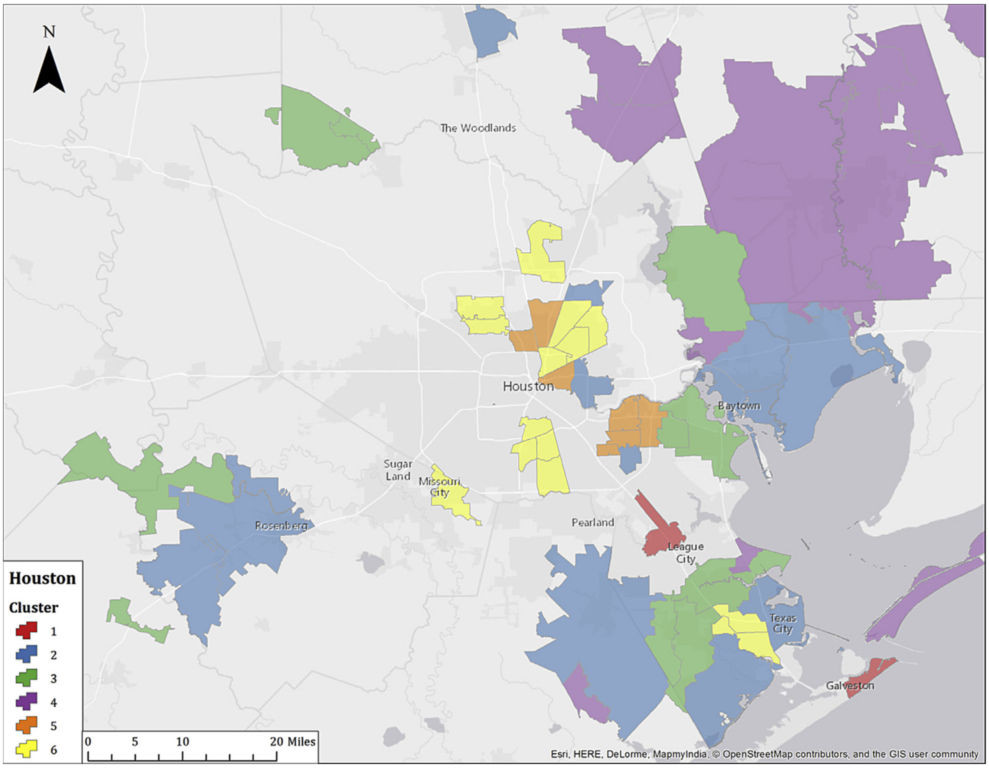

Fig. 5. Cluster locations in the Houston, Texas area.
Basemap Sources: Esri, DeLorne, MapmyIndia, OpenStreetMap contributors, and the GIS user community.

Fort Worth and Houston cluster maps, some of these ZCTAs are located on the far outskirts of the respective metropolitan areas; most are in non-metropolitan parts of northeast and east Texas.

Cluster 5 is the most Hispanic cluster, with an average ZCTA Hispanic population of nearly 86 percent. These ZCTAs also have the lowest average incomes ($29,338, on average) and the highest rates of poverty (34.5 percent, on average) of all of the cluster groups. Insurance rates are also the lowest of any of the clusters; the mean percent of ZCTA population with health insurance in Cluster 5 is 65 percent. ZCTAs in Cluster 5 tend to be found in two principal areas in Texas. First, many of these ZCTAs are located in the Rio Grande Valley of South Texas, along the border of the United States with Mexico. Second, as indicated in the metropolitan cluster maps, ZCTAs Cluster 5 are also located in heavily Hispanic sections of major cities, such as north Houston and Pasadena in the Houston area, and West Dallas.

Cluster 6 has the highest median age-adjusted hospitalization rate of any of the clusters, with 76.51 hospitalizations per 1000 residents in 2006. ZCTAs in this cluster are the most heavily African-American, with an average black population over 60 percent. Poverty and unemployment are high in these ZCTAs; on average, over 28 percent of the population in these ZCTAs are living in poverty, and the average unemployment rate of these ZCTAs is over 13 percent, the highest of any of the clusters. This cluster also has the second-lowest average ZCTA household income, at $32,315. ZCTAs in Cluster 6 are predominantly located in and around major cities, as reflected in Figs. 4 and 5. Several ZCTAs in South Dallas, Southeast Fort Worth, and the southern suburbs of Dallas are classified in Cluster 6; these are the traditional cores of the African-American communities in these cities. In Houston, ZCTAs in the southern and northeastern parts of the city are part of Cluster 6, as is Missouri City, a predominantly African-American suburb to the southwest of Houston with a comparatively low poverty rate (10 percent).

This examination of the demographic characteristics of the six clusters reveals considerable socioeconomic heterogeneity amongst the ZCTAs in Texas experiencing elevated levels of CVD hospitalization. While many ZCTAs have characteristics consistent with those identified in the literature that are associated with CVD (e.g. low education, high poverty), this is not uniformly the case, and indeed there are some clusters, like Cluster 3, that appear to disrupt this general relationship entirely given its high average incomes and low levels of poverty. This heterogeneity is illustrated by Fig. 6, a scatterplot showing one dimension of the relationships informing the cluster analysis, with Factor 1 (the factor largely reflecting a gradient between predominantly white and predominantly Hispanic) on the *x*-axis and Factor 3 (the factor measuring income, poverty, and unemployment) on the *y*-axis. The chart reveals the importance of Factor 1 (which explained the most total variance of all the factors) in sorting the clusters, given the divide between Cluster 4 (the rural white cluster) and Cluster 5 (the Hispanic cluster) along the *x*-axis.

The scatterplot also shows how most ZCTAs in Cluster 3, the exurban cluster, record factor scores above 1 for both dimensions, with many of these ZCTAs scoring amongst the highest in the sample for Factor 3, which measures economic well-being. While these scores simply mean that these ZCTAs in Cluster 3 are amongst the highest in the sample of high-rate zips, many of these communities are nonetheless stable economically in broader context. In the Dallas-Fort Worth area, the area around the city of Lake Dallas (just north of Lewisville on the map) has a median household income of $74,489 and a poverty rate of 4.9 percent; additionally, one of the zip codes to the west of Denton in the northern reaches of the metropolitan area has a median household income of $70,256, and only 3.5 percent of its residents are living in poverty. In turn, the high CVD hospitalization rates in the communities in Cluster 3 do not appear to be explained by economic instability.

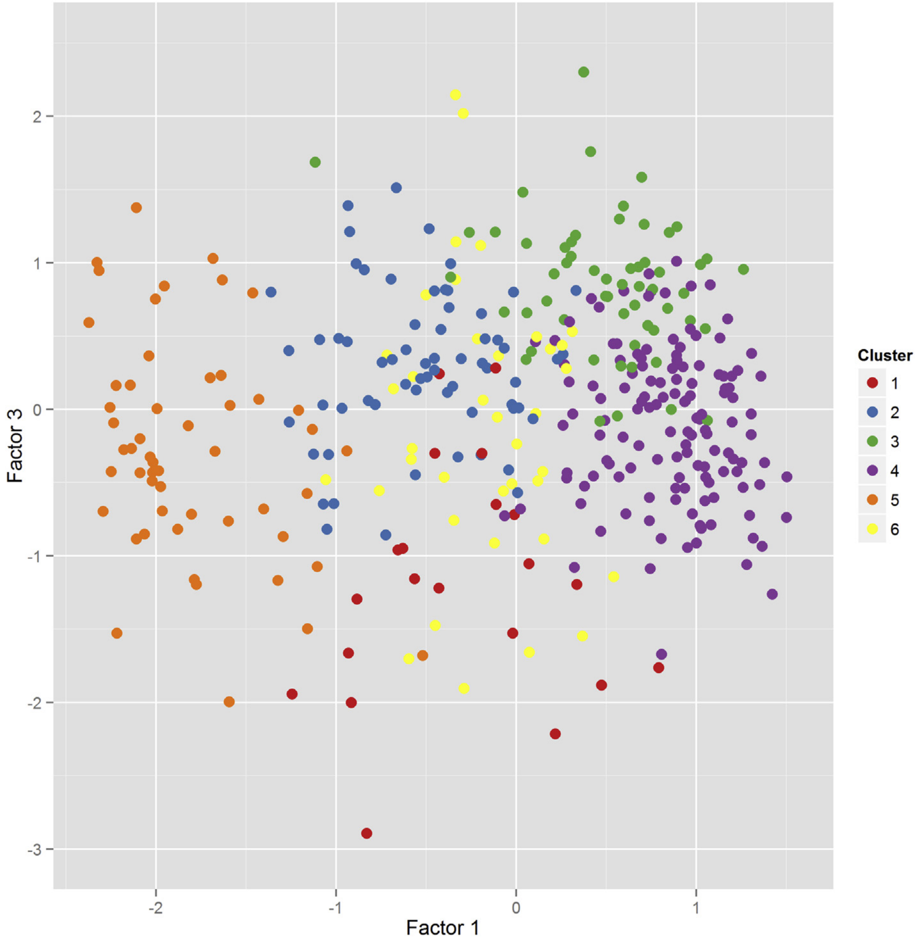

Fig. 6. Scatterplot of Factor 1 vs. Factor 3, by cluster membership. One outlier in Cluster 1 at $x = 1.56$, $y = -6.19$ is omitted from the plot.

Discussion

The classification scheme we developed allows us to examine the demographic characteristics of ZCTAs in Texas experiencing high rates of CVD hospitalization. Some of the clusters reflect expected relationships between certain demographic variables and CVD, given the general consensus in the literature. For example, our analysis reveals a subset of predominantly African-American and low-income, high-unemployment communities with elevated levels of CVD hospitalization (Cluster 6); these ZCTAs tend to be concentrated in major cities, such as in South Dallas, as revealed in Fig. 4. Similarly, Cluster 5 represents predominantly-Hispanic ZCTAs with low levels of education and health insurance coverage; while some of these communities are found in major metropolitan areas (as evidenced in Figs. 4 and 5), they are largely located in South Texas near the US-Mexico border. Our identification of these clusters was unsurprising, given the consistent findings in the literature that show blacks, Hispanics, and low-income individuals at higher risk for CVD and other ailments (Foraker et al., 2011; Jones et al., 2009; Kurian & Cardarelli, 2007; Lalloué et al., 2013).

However, our analyses reveal additional heterogeneity in the geographic profile of CVD in Texas, as revealed by both the maps in Figs. 4 and 5 and the scatter plot in Fig. 6. High rates of CVD hospitalization are not limited to economically marginal areas; they are also found in communities like Cluster 3, which represents exurban areas with relatively low levels of poverty and unemployment. The size of this cluster ($n = 57$) suggests to us that these communities are not simply statistical outliers, but rather represent a meaningful subset of communities experiencing high rates of CVD hospitalization. Further, given their socio-demographic profiles, they are also communities that would not be identified in an analysis that attempts to determine general statistical relationships between demographics and CVD.

In this study, we rely on hospital discharge data for 2006 to determine high-prevalence areas, the most recent data that were freely available at the time. We acknowledge that the dataset represents only one point in time as well as a temporal context from the previous decade, which demands further examination in future research. This research will incorporate hospitalization data from 1999 to the present day to assess how CVD prevalence has changed over time in Texas, and examine shifts or stability in these clusters over time. We also acknowledge that hospitalization data can be an imprecise proxy for disease prevalence given geographic variations in whether patients use hospitals or office-based physicians for primary care (Gamble et al., 2011). Further, as small hospitals are exempt from reporting to the Texas Department of State Health Services, the data necessarily under-report cases in rural areas, which impacts our analysis. In turn, we plan to replicate the study with other metrics for measuring CVD, such as mortality information, to allow for validation of the robustness of our findings. Additional research could also incorporate methods that directly look for ways in which community-specific risk factors influence cardiovascular disease in the different clusters. One such way to address this is through on-the-ground, qualitative studies that compare health behaviors in high-prevalence communities that belong to different clusters. Such studies, as well as longitudinal analyses and replication studies with additional data, would allow for deeper analysis of CVD prevalence in neighborhoods like those in Cluster 3, where traditional characteristics associated with CVD are absent.

An improved understanding of the relationship between access to high-quality CVD preventative care and high-risk neighborhoods is critical for public policy in Texas. CVD creates a substantial, and growing, financial burden within the state. The per-capita fiscal impacts of CVD are most severe in the high-prevalence zip codes identified in this study, as cost of treatment is substantially higher when patients' primary care facility is the hospital (Braunschweig, Cowie, & Auricchio, 2011). As such, these zip codes also offer the highest potential return on investment for stakeholders interested in reducing CVD incidence, mortality, and cost. The most effective policy interventions, however, will be tailored to meet neighborhood-specific needs and targeted to neighborhood-specific audiences. The demographic heterogeneity we identify with our cluster analysis suggests that CVD risk factors might vary in different types of communities. Such heterogeneity has substantive implications for any state-wide policy interventions that target CVD, as well as for health practitioners working to treat the disease.

For example, the literature on the relationships between psychosocial stress and CVD is voluminous (e.g. Dimsdale, 2008; Rozanski, Blumenthal, & Kaplan, 1999), with studies pointing to the role of low socioeconomic status in this relationship. In turn, anti-poverty measures and initiatives to promote economic development in low-income communities in Clusters 5 and 6 might be considered as part of a prevention initiative. However, while stress as a risk factor for CVD may be influenced by economic marginality in some communities, other communities like those in Cluster 3 might have a higher prevalence of other types of stressors, e.g. volatility in the housing market, or long commutes to work, that pose similar health risks (e.g. Steptoe & Kivimaki, 2012).

Further, while geography might serve as a proxy for other factors that influence CVD (Cook & Lauer, 2011), these geographic disparities might also represent inequalities in residents' abilities to access organized health care services. For example, high rates of CVD hospitalization in the heavily-Hispanic Cluster 5 communities may reflect low levels of insurance coverage and high socioeconomic disadvantage. However, they may also represent ways in which socioeconomic inequalities are themselves spatial. This is perhaps reflected by a paucity of health care programs designed to assist new immigrants and persons of low English proficiency, and a lower quality of health care provision in general for Hispanic residents (Baicker, Chandra, & Skinner, 2005).

Additionally, Gamble et al. (2011) find that mortality due to heart failure is more prevalent amongst rural men than urban men; additionally, re-hospitalization amongst rural patients who have experienced heart failure is more likely than urban patients. Given the preponderance of rural ZCTAs in our analysis, particularly in Cluster 4, this is especially salient for prevention initiatives in Texas. The low population density of Cluster 4 is notable, as people living in these areas are significantly more dispersed than in the other clusters. This raises questions about geographic accessibility to health care and emergency services that may not be as significant in more densely-populated areas with similarly high CVD rates; as such, effective policy interventions in Cluster 4 might require a focus on accessibility. In turn, we believe that efforts to prevent cardiovascular disease cannot adopt a “one-size-fits-all” approach; instead, they will need to be sensitive to the different characteristics and factors that may influence CVD prevalence in these different communities.

Our future research will seek to inform public policies that directly address the variegated nature of CVD prevalence in Texas. To do so effectively, this work will require an evaluation of how CVD risk factors manifest themselves in diverse communities across the state, with both quantitative and qualitative analytical approaches that can reveal these risk factors. Any policy intervention to prevent cardiovascular disease requires an understanding of the types of communities in which to intervene, which can be accomplished through a classification analysis like the approach used in this paper. In turn, the clusters identified can provide a good starting point for framing a research program to inform such targeted healthcare policies.

Acknowledgments

This research was supported by a Junior Faculty Summer Research Program fellowship at Texas Christian University, and an Institute for Urban Living and Innovation Small Grant from AddRan College, Texas Christian University.

References

Ahmad, O. B., Boschi-Pinto, C., Lopez, A. D., Murray, C. J. L., Lozano, R., & Inoue, M. (2001). *Age standardization of rates: A new WHO standard*. GPE Discussion Paper Series No. 31.

Ang, R., Benedict, M., Bhakta, N., Brea, C., Spencer, B., Stagg, J., et al. (2013). *Texas plan to reduce cardiovascular disease and stroke 2013–2017*. Report for the Texas Department of State Health Services.

Baicker, K., Chandra, A., & Skinner, J. S. (2005). Geographic variation in health care and the problem of measuring racial disparities. *Perspectives in Biology and Medicine, 48*(Suppl. 1), S42–S53.

Bell, S., Wilson, K., Bissonnette, L., & Shah, T. (2012). Access to primary health care: does neighborhood of residence matter? *Annals of the Association of American Geographers, 103*(1), 85–105.

Bellis, M. A., Jarman, I., Downing, J., Perkins, C., Beynon, C., Hughes, K., et al. (2012). Using clustering techniques to identify localities with multiple health and social needs. *Health & Place, 18*(2), 138–143.

Bosma, H., Dike van de Mheen, H., Borsboom, G. J. J. M., & Mackenbach, J. P. (2001). Neighborhood socioeconomic status and all-cause mortality. *American Journal of Epidemiology, 153*(4), 363–371.

Braunschweig, F., Cowie, M. R., & Auricchio, A. (2011). What are the costs of heart failure? *Europace, 13*, ii13–ii17.

Claude, J. (2008). *Morphometrics with R*. New York: Springer.

Cook, N. L., & Lauer, M. S. (2011). The socio-geography of heart failure: why it matters. *Circulation: Heart Failure, 4*, 244–245.

Diez-Roux, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health, 21*(1), 171–192. http://dx.doi.org/10.1146/annurev.publhealth.21.1.171.

Diez-Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health, 91*(11), 1783–1789.

Dimsdale, J. E. (2008). Psychological stress and cardiovascular disease. *Journal of the American College of Cardiology, 51*(13), 1237–1246.

Foraker, R. E., Rose, K. M., Suchindran, C. M., Chang, P. P., McNeill, A. M., & Rosamond, W. D. (2011). Socioeconomic status, Medicaid coverage, clinical comorbidity, and rehospitalization or death after an incident heart failure hospitalization. *Circulation: Heart Failure, 4*, 308–316.

Ford, E. S. (2013). Food security and cardiovascular disease risk among adults in the United States: findings from the National Health and Nutrition Examination Survey, 2003–2008. *Preventing Chronic Disease, 10*, 1–10.

Gamble, J. M., Eurich, D. T., Ezekowitz, J. A., Kaul, P., Quan, H., & McAlister, F. A. (2011). Patterns of care and outcomes differ for urban versus rural patients with newly diagnosed heart failure, even in a universal health care system. *Circulation: Heart Failure, 4*, 317–323.

Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Blaha, M. J., et al., , on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. (2014). Heart disease and stroke statistics – 2014 update: a report from the American Heart Association. *Circulation, 129*, e28–e292.

Grady, S., & Darden, J. (2012). Spatial methods to study local racial residential segregation and infant health in Detroit, Michigan. *Annals of the Association of American Geographers, 102*(5), 922–931.

Grubesic, T. H., Miller, J. A., & Murray, A. T. (2014). Geospatial and geodemographic insights for diabetes in the United States. *Applied Geography, 55*, 117–126.

Hare, T. S., & Barcus, H. R. (2007). Geographical accessibility and Kentucky's heart-related hospital services. *Applied Geography, 27*(3–4), 181–205.

Heidenreich, P. A., Trogdon, J. G., Khavjou, O. A., Butler, J., Dracup, K., Ezekowitz, M. D., et al. (2011). Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association. *Circulation, 123*(8), 933–944.

Hernandez, M. N., Roy Chowdhury, R., Fleming, L. E., & Griffith, D. A. (2011). Colorectal cancer and socioeconomic status in Miami-Dade County: neighborhood-level associations before and after the Welfare Reform Act. *Applied Geography, 31*(3), 1019–1025.

Hu, Z., & Rao, K. R. (2009). Particulate air pollution and chronic ischemic heart disease in the eastern United States: a county level ecological study using satellite aerosol data. *Environmental Health, 8*, 26–36.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. New York: Springer.

Jones, C. A., Perera, A., Chow, M., Ho, I., Nguyen, J., & Davachi, S. (2009). Cardiovascular disease risk among the poor and homeless – what we know so far. *Current Cardiology Reviews, 5*, 69–77.

Krieger, N., Chen, J. T., Waterman, P. D., Soobader, M.-J., Subramanian, S. V., & Carson, R. (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. http://dx.doi.org/10.1093/aje/kwf068.

Kurian, A. K., & Cardarelli, K. M. (2007). Racial and ethnic differences in cardiovascular disease risk factors: a systematic review. *Ethnicity & Disease, 17*, 143–152.

Lalloué, B., Monnez, J. M., Padilla, C., Kihal, W., Le Meur, N., Zmirou-Navier, D., et al. (2013). A statistical procedure to create a neighborhood socioeconomic index for health inequalities analysis. *International Journal for Equity in Health, 12*(21). http://www.equityhealthj.com/content/12/1/21.

LeDoux, T. F., & Vojnovic, I. (2014). Examining the role between the residential neighborhood food environment and diet among low-income households in Detroit, Michigan. *Applied Geography, 55*, 9–18.

Lee, R. E., & Cubbin, C. (2002). Neighborhood context and youth cardiovascular health behaviors. *American Journal of Public Health, 93*(3), 428–437.

Lin, Y., Schootman, M., & Zhan, F. B. (2015). Racial/ethnic, area socioeconomic, and geographic disparities of cervical cancer survival in Texas. *Applied Geography, 56*, 21–28.

Michimi, A., Ellis-Griffith, G., Nagy, C., & Peterson, T. (2013). Coronary heart disease prevalence and occupational structure in U.S. metropolitan areas: a multilevel analysis. *Health & Place, 21*(0), 192–204.

Odoi, A., Wray, R., Emo, M., Birch, S., Hutchinson, B., Eyles, J., et al. (2005). Inequalities in neighborhood-based socioeconomic characteristics: potential evidence-base for neighborhood health planning. *International Journal of Health Geographics, 4*, 20.

Pedigo, A., Seaver, W., & Odoi, A. (2011). Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and discriminant analysis approaches. *PLoS One, 6*(7), e22693.

Pierce, J., Martin, D. G., Scherr, A. W., & Greiner, A. (2012). Urban politics and mental health: an agenda for geographic research. *Annals of the Association of American Geographers, 102*(5), 1084–1092.

Root, E. D. (2012). Moving neighborhoods and health research forward: using geographic methods to examine the role of spatial scale in neighborhood effects on health. *Annals of the Association of American Geographers, 102*(5), 986–995.

Rozanski, A., Blumenthal, J. A., & Kaplan, J. (1999). Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. *Circulation, 99*, 2198–2217.

Shortt, N. K., Richardson, E. A., Mitchell, R., & Pearce, J. (2011). Re-engaging with the physical environment: a health-related environmental classification of the UK. *Area, 43*(1), 76–87.

Shortt, N. K., Richardson, E. A., Pearce, J., & Mitchell, R. (2012). Mortality inequalities by environment type in New Zealand. *Health & Place, 18*(5), 1132–1136.

Steptoe, A., & Kivimaki, M. (2012). Stress and cardiovascular disease. *Nature Reviews Cardiology, 9*(6), 360–370.

Texas Department of State Health Services, Center for Health Statistics. (2006). *User manual: Texas Hospital Inpatient Discharge Public Use Data File (PUDF)*. Available from http://www.dshs.state.tx.us/THCIC/Hospitals/UserManual2006.pdf.

Vicino, T. J., Hanlon, B., & Short, J. R. (2011). A typology of urban immigrant neighborhoods. *Urban Geography, 32*(3), 383–405.

Weeks, J. R., Getis, A., Stow, D. A., Hill, A. G., Rain, D., Engstrom, R., et al. (2012). Connecting the dots between health, poverty and place in Accra, Ghana. *Annals of the Association of American Geographers, 102*(5), 932–941.

Wyly, E., & DeFilippis, J. (2010). Mapping public housing: the case of New York City. *City & Community, 9*(1), 61–86.