Published in final edited form as:

J Expo Sci Environ Epidemiol. 2017 May ; 27(3): 281–289. doi:10.1038/jes.2016.53.

A novel approach for measuring residential socioeconomic factors associated with cardiovascular and metabolic health

Jaime E. Mirowsky[1,2], **Robert B. Devlin**[3], **David Diaz-Sanchez**[3], **Wayne Cascio**[3], **Shannon C. Grabich**[3], **Carol Haynes**[4], **Colette Blach**[4], **Elizabeth R. Hauser**[4,5,6], **Svati Shah**[4,7], **William Kraus**[4,7], **Kenneth Olden**[8], and **Lucas Neas**[3]

[1]Curriculum in Toxicology, University of North Carolina, Chapel Hill, North Carolina, USA

[2]Center for Environmental Medicine, Asthma, and Lung Biology, University of North Carolina, Chapel Hill, North Carolina, USA

[3]National Health and Environmental Effects Laboratory, US Environmental Protection Agency, Chapel Hill, North Carolina, USA

[4]Duke Molecular Physiology Institute, Duke University Medical Center, Durham, North Carolina, USA

[5]Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, North Carolina, USA

[6]Cooperative Studies Program Epidemiology Center, Durham Veterans Affairs Medical Center, Durham, North Carolina, USA

[7]Division of Cardiology, Department of Medicine, School of Medicine, Duke University, Durham, North Carolina, USA

[8]National Center for Environmental Assessment, US Environmental Protection Agency, Chapel Hill, North Carolina, USA

Abstract

Individual-level characteristics, including socioeconomic status, have been associated with poor metabolic and cardiovascular health; however, residential area-level characteristics may also independently contribute to health status. In the current study, we used hierarchical clustering to aggregate 444 US Census block groups in Durham, Orange, and Wake Counties, NC, USA into six

Correspondence: Dr. Lucas Neas, US EPA Human Studies Facility, 104 Mason Farm Road, Chapel Hill, NC 27599, USA. Tel.: +1 919 966 9961. Fax: +1 919 966 6271. neas.lucas@epa.gov.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Supplementary Information accompanies the paper on the Journal of Exposure Science and Environmental Epidemiology website (http://www.nature.com/jes)

DISCLAIMER

Research described in this article was conducted under contract to the Health Effects Institute (HEI), and organization jointly funded by the United States Environmental Protection Agency (EPA) (Assistance Award No. R-82811201) and certain motor vehicle and engine manufacturers. The contents of this article do not necessarily reflect the views of HEI, or its sponsors, nor do they necessarily reflect the view and policies of the EPA or motor vehicle and engine manufacturers. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

homogeneous clusters of similar characteristics based on 12 demographic factors. We assigned 2254 cardiac catheterization patients to these clusters based on residence at first catheterization. After controlling for individual age, sex, smoking status, and race, there were elevated odds of patients being obese (odds ratio (OR) = 1.92, 95% confidence intervals (CI) = 1.39, 2.67), and having diabetes (OR = 2.19, 95% CI = 1.57, 3.04), congestive heart failure (OR = 1.99, 95% CI = 1.39, 2.83), and hypertension (OR = 2.05, 95% CI = 1.38, 3.11) in a cluster that was urban, impoverished, and unemployed, compared with a cluster that was urban with a low percentage of people that were impoverished or unemployed. Our findings demonstrate the feasibility of applying hierarchical clustering to an assessment of area-level characteristics and that living in impoverished, urban residential clusters may have an adverse impact on health.

Keywords

cardiovascular disease; CATHGEN; hierarchical clustering; metabolic disease; neighborhood; socioeconomic status

INTRODUCTION

In the United States, the prevalence of cardiovascular disease (CVD) is expected to rise 10% between 2010 and 2030, which can be attributed to increases in aging, obesity, diabetes, and physical inactivity.[1] Researchers have identified several risk factors that are associated with having poor cardiovascular health, including an individual’s socioeconomic status (SES). In a recent review, educational attainment is the most utilized marker of SES;[2] people with more education had a reduced prevalence of cardiovascular risk factors, morbidity, and mortality. Further, income level and job classification have also been associated with cardiovascular health.[3] In addition, there is also an unequal racial burden with CVD, where researchers found that people who were Black had elevated blood pressure, more frequent diagnoses of diabetes, and were more likely to experience stroke-related death compared with those who were White.[4] Thus, there is strong evidence linking individual-level socioeconomic and racial status to increased incidence of CVD and disease risk factors.

In addition to individual-level SES, there is a growing body of literature suggesting that neighborhood-level SES contributes to adverse health outcomes. In these studies, neighborhood SES was associated with cardiovascular-related mortality,[5,6] stroke,[7,8] coronary heart disease,[9,10] and myocardial infarction,[11] and in lower income neighborhoods associations between neighborhood SES and metabolic-related health effects were also observed.[12–14] Studies have identified two major mediators of these effects: psychosocial stress[2,15,16] and limited access to resources.[17–20] These factors can put communities at an increased risk for poor health by influencing mental wellbeing, diet, and physical activity. Other possible mechanisms, such as higher pollutant levels[21] and the built environment,[22] have also been shown to influence adverse outcomes.

Previous studies of neighborhood-level SES and health have used various methods to assign neighborhoods into levels of deprivation, yet there is little consistency with how deprivation is defined. Examples of discrepancies include the use of different US Census geographies (tracts verse block groups), how many (and which variables) are best for analyses, and which statistical methods to use.[23] In general, the choice of variables has not been well defined by researchers, and there is no consensus as to the best variables to be used for these studies. With respect to statistical methods, the most commonly employed methods are the construction of a neighborhood deprivation index (NDI) or *z*-score.[6,8,14,23–25] Some of the published indices which measure neighborhood deprivation use principal component analysis (PCA) as a data reduction technique to find non-correlated linear combinations with maximum variance. From this, components are used to weigh the variable contributions.[23]

However, rather than define neighborhoods by levels of deprivation, one could cluster neighborhood environments together based on having similar demographic information. This method does not give an index of deprivation or separate neighborhoods into those of low or high deprivation. Rather, the output from clustering gives a summary of all the selected attributes that make up the clustered geographical areas and allows researchers to compare the levels of those attributes between the clusters. Clustering is also advantageous in minimizing the number of geographical areas to be researched and increasing the power of the study to assess for health differences across similar areas, particularly when the sample size may be limited. Thus, in the current manuscript we are offering an alternate way of defining neighborhoods than that done previously.

If populations are at an increased risk of cardiovascular and metabolic diseases due to their socioeconomic status, the increased prevalence of CVD and obesity over the next several decades may not be equally distributed throughout the US. As we are interested in better understanding the relationship between neighborhood-level factors and cardio-metabolic disease, we used hierarchical clustering to aggregate US Census block groups into homogenous clusters. Using our clusters, we can then determine whether there is an association between neighborhood residence and cardio-metabolic disease. With the creation of these residential area clusters, future work will include looking at how air pollution concentrations and green spaces are distributed across our area. Additional research looking at mechanisms associated with disease (i.e., epigenetics, genetics, biomarkers, and stress) will be assessed in the same population in subsequent studies.

MATERIALS AND METHODS

Study Population

The source population for this study was the CATHeterization GENetics (CATHGEN) cohort. CATHGEN combines clinical data with biological samples from 9334 individuals who underwent cardiac catheterizations at Duke University Medical Center between 2001 and 2010.[26] Patient demographics, medical history, and health data were obtained from the Duke Databank for Cardiovascular Disease. Subject level information was obtained from clinical medical records and held in the Duke Information System for Cardiovascular Diseases. Residential addresses were obtained from medical records and geocoded to the street level by the Children’s Environmental Health Initiative (http://cehi.snre.umich.edu/) for 8017 of the 9334 study participants (ArcGIS 10.1, Esri, Redlands, CA, USA).[27] Of these patients, 7118 resided in North Carolina and ~ 5600 patients were catheterized between 2002 and 2009, the only years for which we have comprehensive air pollution data for the state of North Carolina. Although not limiting for this current analysis, a major goal of a future study using these clusters is to understand whether air pollution contributes to different health outcomes in different clusters. Of the 5600 patients, 2254 patients lived in Durham, Orange, and Wake Counties.[27] As this three county area represented the highest density of the CATHGEN cohort, we selected this location for the current study. As there are very few CATHGEN patients from Charlotte or the Piedmont Triad (the other two large urban centers in North Carolina)—likely because they went to hospitals in those centers—the majority of the other CATHGEN patients reside in low density rural areas. Each patient was assigned to a block group according to the 2000 US Census. All patients provided written informed consent prior to enrollment; CATHGEN was approved by the Duke University Institutional Review Board.

Individual-level health measurements for the participants included body mass index (BMI) and fasting glucose concentrations. In addition, information about whether the patients had diabetes, hypertension, congestive heart failure (CHF), and CVD at the time of their first catheterization was obtained. An index of coronary artery disease severity, the coronary artery disease (CAD) index, was also used. This index ranges from 1 to 100, with values >23 representing at least one hemodynamically significant lesion in one epicardial coronary artery in the patient; a score >23 is considered of clinical significance.[28]

Study Location

We used US Census block groups as a proxy for a residential area. Block groups are the smallest level of categorization for which social characteristics are reported in the US Census. Given their relatively small land area and population size (i.e., a block group represents ~ 400 households), we believe that it is suitable for describing residential characteristics.[29] By looking at clustered areas, we increased the power of the study to look at differences between residential areas made up of people with similar demographics.

There were 448 block groups that encompassed the selected area for this study; Orange, Durham, and Wake Counties comprised of 56, 129, and 263 block groups, respectively. One block group had no residents living within it; this block group consisted of a portion of the North Carolina State University campus and was removed from analysis. Another block group contained both a correctional institution and a prison within its borders. As prisoners are required to participate in the US Census and use the location of the prison as their current address, this block group, as well as three other block groups that contained a correctional institution and/or a prison, were removed from all analyses.[30] The removal of the four block groups left a total of 444 block groups available for clustering.

Defining Residential SES Clusters using US Census Data

The residential SES (R-SES) variables used to construct the clusters were obtained from the 2000 US Census.[31] As Census block group demographic information is not available for 2010, data from the year 2000 was used. Seven categories (education, wealth, income, race, employment, housing, and land-use), defined by 12 variables from the Census (Table 1) and cited in previous work on neighborhood-level SES and health, were identified as being influential.[6,7,11,12,14,21,25,32,33] These variables were chosen without previous knowledge of their distribution across our study area, and no variables were excluded in a sensitivity analysis to try to determine the most influential attributes or bias the results.

Single parent housing is referring to the percentage of male or female only (no spouse present) family households in owner and renter occupied housing units divided by the total number of owner and renter occupied housing units. We also defined the percentage of the population in non-managerial positions as the percentage of both sexes of employed civilian population 16 years and over not in management, professional, and related occupations. Detailed definitions for several of the other 12 variables can be found in the Supplementary Information.

Hierarchical Clustering

Using Ward's hierarchical clustering method,[34] (R Version 3.2.1 (ref. 35) and the *hclust* function), the Census variables were transposed and the block groups were assembled into residential clusters based on the 12 Census factors. Ward's clustering technique uses a bottom-up approach to look for similarities in a group of observations with respect to several variables.[36] Ward's method was chosen for this analysis because the pooled with-in group sum of squares is minimized, and the cluster distances using Ward's method are defined as the squared Euclidean distance between points. The output for hierarchical clustering is a dendrogram. To spatially identify the block groups making up the clusters, ArcGIS (ESRI, Version 10.3.1, Redlands, CA, USA) was used. Determining the optimal number of clusters is a fundamental and challenging problem, and to help us determine the optimal cluster number, we used the Friedman method[37] (R Version 3.2.1 and the *NbClust* function). We also wanted to ensure that each cluster had enough CATHGEN participants for appropriate statistical analyses.

Statistical Analyses

Using Prism 4.0 (GraphPad, San Diego, CA, USA), descriptive statistics of the 12 factors contributing to the clusters were derived. Descriptive statistics were also run on the characteristics of the total patient population as well as the characteristics of the patient population residing within each cluster. To ensure none of the 12 originally identified factors were highly correlated with each other, Pearson correlation coefficients were calculated between all the R-SES factors. A correlation matrix and clustering dendrogram of the variables can be seen in the Supplementary Information.

Logistic regression models estimated the odds ratio (OR) and 95% confidence intervals (CI) among the residential clusters for BMI >25 (representing patients that are obese and overweight), BMI >30 (representing patients that are obese), diabetes, CHF, hypertension, CVD, and the CAD index >23. Linear regression models estimated risk differences (RD) and 95% CI among clusters for fasting glucose concentrations. All linear and logistic regression models controlling for age, sex, smoking, and race, were run using R Version 3.2.1.[35] The R code used to generate the tables and figures are available upon request.

RESULTS

Formation of Residential Clusters using 12 US Census Variables Residential clusters were formed using US Census demographic information. Our cluster analysis identified six unique residential clusters for this study as shown in the dendrogram (Figure 1); the number of clusters formed is consistent with other studies using similar methodology to examine neighborhood-associated SES.[38,39]

In the dendrogram, the outer ring is made up of 444 individual leaves, with each leaf representing a Census block group. Each leaf was also given a color, and the colors correspond to the cluster designation. The lines inside the dendrogram joining the leaves correspond to the differences between clusters or block groups, and the shorter the lines the more similar the groups. The red circle in the center of the dendrogram highlights where the cut-off for six clusters lies; each of the clusters is labeled alongside it. For example, the block groups that are in yellow text belong to cluster 1, and the gray block groups are part of cluster 4. By increasing or decreasing the diameter of the circle, more (or less) clusters are formed. With seven residential clusters, one cluster was comprised of only five block groups. This new cluster differed from its previously formed cluster by containing more people with higher education degrees, less people living in owner-occupied housing, more people with income below the poverty level, and a greater population density. All these attributes suggest that people residing in this cluster were primarily undergraduate and graduate students, and the location of the block groups in this cluster were adjacent to major Universities, which also support this claim.

For each of the 12 R-SES factors obtained from the US Census, the demographic information was averaged for the block groups making up the clusters. The mean and SEM of the factors can be seen in Figure 2. Low standard errors for the factors were observed, which suggest a limited spread in the data distribution. The locations of the block groups comprised of each cluster are shown in Figure 3; the downtown areas of Chapel Hill, Durham, and Raleigh have been highlighted. The colors of the block groups correspond to those of the dendrogram for cluster designation (Figure 1).

Characteristics of the Residential Clusters

The residential clusters differed on the 12 selected Census variables (Figure 2). With 12 variables contributing to the formation of each cluster, the clusters are best described by the sum of their attributes. Residential cluster 1 (71 block groups) was urban and had high percentages people who were Black and worked in non-managerial positions. Cluster 2 (42 block groups) was urban, impoverished, and the population living in this cluster were on public assistance, unemployed, and working in non-managerial occupations. This cluster also had higher percentages of single parent homes and people who were Black. Residential cluster 3 (155 block groups) was less urban and had a low percentage of the population living below the poverty level and unemployed; this cluster also had a high percentage of people in non-managerial positions and obtaining at least a Bachelor’s degree. Residential cluster 4 (73 block groups) had the highest percentage of residents not identified as Black or White, and a high percentage of people with a Bachelor’s degree, but low in unemployment status. Residential cluster 5 (45 block groups) was rural with a low percentage of the population living below the poverty line, unemployed, and Black, with the highest average percentage of owner-occupied housing. Residential cluster 6 (58 block groups) was similar to that of residential cluster 5, but urban. The characteristics of these clusters are consistent with our knowledge of the geographical area for this study. A table summarizing the data in Figure 2 can be seen in the Supplementary Information.

Residential Clusters and Health Outcomes from the CATHGEN Cohort

The CATHGEN participants were subdivided into their respective clusters using their geocoded addresses. The average age of the patients in each cluster did not greatly deviate from the average mean of the cohort; however, differences in sex, race, and smoking status were observed based on the cluster designation (Table 2).

Compared with the other clusters, patients living in cluster 2 had the highest BMI, fasting glucose, and prevalence of diabetes, hypertension, CHF, and CVD. Patients living in cluster 3 had the lowest BMI and prevalence of diabetes and CHF. Patients living in cluster 6 had the lowest prevalence of hypertension and CVD.

Cluster 3 had the largest sample size and represented the healthiest cluster (Table 2). Therefore, cluster 3 was used as the reference cluster for the regression models. When controlling for individual-level characteristics (age, sex, race, and smoking status), the CATHGEN patients living in clusters 1, 2, and 6 had an increased odds of having metabolic and cardiovascular-related diseases (Table 3). Those living in cluster 2 had higher odds of being overweight or obese (BMI >25 kg/m^2, OR = 1.65, 95% CI = 1.10, 2.54), (BMI >30 kg/m^2, OR = 1.92, 95% CI = 1.39, 2.67), and having diabetes (OR = 2.19, 95% CI = 1.57, 3.04). In addition, more patients from cluster 2 had CHF (OR = 1.99, 95% CI = 1.39, 2.83) and hypertension (OR = 2.05, 95% CI = 1.38, 3.11). Elevated odds for diabetes were observed in cluster 6 patients (OR = 2.04, 95% CI = 1.29, 3.17) and cluster 1 patients (OR = 1.31, 95% CI = 1.00, 1.72) compared with patients from cluster 3. Elevated odds were also found for BMI >30, CHF, and hypertension for the CATHGEN patients in cluster 1. None of the clusters had elevated coronary artery disease (CAD index >23) or cardiovascular disease compared with cluster 3; this probably reflects the origin and disease bias of the cohort—those referred for cardiac catheterization for suspected disease.

Fasting glucose was also measured in the CATHGEN patients immediately before catheterization (Table 3). Using a linear regression model, elevated risk differences in fasting glucose levels were observed in the patients residing in cluster 1 (RD = 8.25, 95% CI = 2.34, 14.15), cluster 2 (RD = 14.23, 95% CI = 6.58, 21.88) and cluster 6 (RD = 16.54, 95% CI = 5.86, 27.22) when compared with residential cluster 3.

DISCUSSION

In the current study we adopted a hierarchical clustering technique to explore residential-level SES in Central North Carolina and its relation to disease and disease-related biomarkers. For our analysis, we aggregated 444 US Census block groups in Orange, Durham, and Wake Counties into six clusters based on 12 residential SES factors. After controlling for individual-level demographic factors, significant differences in disease status were found based on the residents' cluster designation.

We observed an uneven distribution of health outcomes across our six residential clusters, with elevated odds of metabolic and cardiovascular disease in clusters 1, 2, and 6 compared with cluster 3. Many of the neighborhoods belonging to cluster 2 were located centrally in the cities of Durham and Raleigh (Figure 3). In Durham County, surrounding the cluster 2-associated block groups were the block groups belonging to cluster 1. The cluster 1 block groups were then surrounded by block groups associated with cluster 3. Thus, in a bulls-eye pattern, it appears that cluster 1 may act as a transition between the least healthy (cluster 2) and most healthy (cluster 3) clusters. This idea was corroborated with the Census demographic variables, as the average values for the variables in cluster 1 fell in between those calculated for clusters 2 and 3. When superimposed with the CATHGEN cohort health data, cluster 2 residents had the greatest odds of cardiovascular and metabolic disease, with cluster 1 residents also having elevated odds but lower effect estimates.

Previous studies have found poorer health in people living in locations such as those having attributes similar to those found in cluster 2.[5–14,17] However, this does not support our findings for cluster 6. Cluster 6 had a low percentage of people living below the poverty line, unemployed, and Black. We also compared the odds ratios between clusters 5 and 6; these clusters had similar characteristics for 11 of the Census variables but were dissimilar in their urban/rural classification (cluster 5 was rural whereas cluster 6 was urban). Upon further investigation, the block groups making up cluster 6 were located near major roadways in Wake County. Thus, cluster 6 appears to have some unique contributors to dysglycemia that may be related to proximity to roadways;[27] however, it is also possible that social conditions, access to fast food, or the built environment that may be driving poor metabolic health of the CATHGEN subjects in cluster 6.

Many previously conducted studies of neighborhood-level SES and health assessed area-level deprivation by using PCA followed by the construction of either a NDI or *z*-score.[6,8,14,23–25] For these studies, the researchers sought to identify the most and least deprived neighborhoods. However, in the current work our focus was not in comparing deprivation levels between neighborhoods, but instead to combine neighborhoods of similar attributes to increase the power of our study. Our output, compared to the NDI or *z*-score, is a list of characteristics of each cluster that collectively can be used to define a neighborhood. Using NDI or *z*-scores would not be fitting for our research goal, as it is possible that two heterogeneous neighborhoods may have the same NDI- or *z*-score,[40] and whereas this may not be a problem in defining deprivation, those methods would not be ideal for our work.

A handful of studies have also used combinations of PCA and hierarchical clustering to help define neighborhood SES.[25,38,39] Using PCA, researchers mathematically determine the most important factors representing the greatest variability in the geographic area. However, the cut-off for determining the importance of the factor components can be arbitrary. Further, PCA is then combined with clustering, combining two techniques that involve researcher-based interpretation. Rather than use PCA, we chose our SES factors based on a literature review. For example, education level has been used extensively as a factor impacting neighborhood-level SES and health outcomes;[6,7,12,19,24,25] therefore, we included a variable representing education in our clustering analysis. For this work, given the quantity of universities in the three county area and the amount of high tech industry in the Research Triangle Park area, we selected having at minimum a Bachelor's degree as our education-associated variable. Similarly, occupation has also been widely cited by researchers to impact health,[6,7,11,24,25] and studies have shown that people in higher status jobs were less hypertensive than their peers.[41,42] Thus, similar to past work, we used non-managerial positions as a measure of occupation. Housing may also influence neighborhood SES. Percent of owner-occupied housing was chosen as a variable in our work, as opposed to median home value, because it separates renters from owners. We believed that in an area with many universities and apartment complexes that this metric would be more useful in defining housing. Other metrics such as unemployment status[6,12,14,19,25] and race[6,12,14] have also been widely cited by researchers to impact health when looking at neighborhood-level characteristics, and we used these variables as well. We believe that by using an *a priori* approach based on our knowledge of the geographical area in question, a literature review,[21,32,33] and equal factor balancing, we have encompassed the appropriate variables that contribute to neighborhood SES in our study.

The present study has several limitations. First, we did not measure psychosocial stress or the built environment.[2,43–46] However, the importance of missing psychosocial stress may be minimal, as the relationship between psychosocial stress and disease have been inconsistent,[5,18,24,47–50] and healthy behavior is linked with both the built environment and SES.[51] In addition, we did not survey our cohort to determine why the patients lived in their neighborhoods and we were also missing information on the past residences of some the participants.[14,15] Further, we were unable to obtain SES information for each resident, and individual-level SES factors have previously been associated with health outcomes.[3] Although we were unable to obtain this information, several studies looking at neighborhood SES and health that controlled for individual SES factors still found significant outcomes.[10,19,52] Regardless, we acknowledge that lacking this information is a major limitation to our work. We also acknowledge that there might be some selection or referral bias, as health insurance influences referrals for medical testing,[2] and that the population described in this study is not generalizable to the general population. We do not know if having a residence in specific neighborhoods would make someone more or less likely to use the Duke University hospital compared to other hospitals in the area. However, we did not eliminate any participant from the analysis based on the neighborhood they resided in beyond the three county area. In this study we assumed that Census variables were similar at the block group level between 2001 and 2010—the enrollment period of CATHGEN. Due to no block group demographic data being available for 2010, we used data from 2000. Last, it is possible that using different R-SES input factors might change our cluster designations; past studies using similar clustering techniques have found this method to be fairly sensitive.[36] Unfortunately, there is no consensus in the literature as to the best variables to describe neighborhood SES, and this is further complicated by the possibility that variables that may be valid indicators in one location are less relevant in another.[53] However, we believe that our input variables adequately covered the main categories associated with area-level SES in our area, and these variables have highly been supported in the literature.[6,7,11,12,14,21,25,32,33]

There are several strengths of this study that should be noted. We used US Census block group demographic data. Block groups are smaller geographic areas than census tracts—there are ~ 2.5 times more block groups in our three county area compared to tracts. In a side-by-side comparison of tracts and block groups, we found greater resolution for several of our variables using the block groups (data not shown). Thus, we believe using block groups was superior to using census tracts. In addition, several studies examining neighborhood SES and health status used only one SES variable for their work;[25,54] we used 12 variables to define our residential clusters. Having more than one variable can provide a superior assessment of the population.

In conclusion, we used hierarchical clustering in a novel manner to identify six unique and homogeneous residential clusters in Central North Carolina based on 12 *a priori*-selected demographic factors. This clustering method delivered a more direct way to examine residential area SES. Further, using 2254 patients in the CATHGEN cohort, we estimated the prevalence of cardiovascular and metabolic health outcomes of these patients by cluster. In a cluster that was highly urban, Black, impoverished, and unemployed, there were greater odds of the residents of the cluster that were in CATHGEN being overweight, obese, and having diabetes, congestive heart failure, and hypertension compared with a cluster that was highly urban but had a low percentage of people that were Black, below the poverty levels, on public assistance, and unemployed. As there are factors beyond individual-level SES contributing to health, assessment of neighborhood environment becomes both necessary and complicated. With this method, we are offering a novel way to assess area-level SES associate using hierarchical clustering. In subsequent work we will look at genetic/epigenetic differences in the populations based on their assigned clusters as well as the distribution of air pollutants between the clusters. We also aim to look at a subset of the population in more detail to assess differences in the participants' microenvironments and whether multi-level approaches involving individual-level SES influences our outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to acknowledge Drs Marie Lynn Miranda and Cavin K Ward-Caviness for their involvement in the geocoding of the CATHGEN patients. We thank all of the participants in the CATHGEN study, and we acknowledge the essential contributions of the faculty and staff of the Duke Cardiac Catheterization Lab, the Duke Databank for Cardiovascular Disease, and the Duke Center for Human Genetics for their contributions to this manuscript. This work was supported by United States Environmental Protection Agency internal funds; the EPA Cooperative Agreement with the Center for Environmental Medicine, Asthma, and Lung Biology at the University of North Carolina (CR83346301); the National Institutes of Environmental Health Sciences (T32ES007126); a UNC Golberg Fellowship; the National Institutes of Health (HL73042, HL36587, HL095987); an award from the Neurosciences Education and Research Foundation (Encinitas, CA); and the Health Effects Institute (4946-RFPA10-3/14-7).

References

1. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail. 2013; 6:606–619. [PubMed: 23616602]
2. Havranek EP, Mujahid MS, Barr DA, Blair IV, Cohen MS, Cruz-Flores S, et al. Social determinants of risk and outcomes for cardiovascular disease: a scientific statement from the American Heart Association. Circulation. 2015; 132:873–898. [PubMed: 26240271]
3. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature. Circulation. 1993; 88:1973–1998. [PubMed: 8403348]
4. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015; 131:e29–322. [PubMed: 25520374]
5. Ramsay SE, Morris RW, Whincup PH, Subramanian SV, Papacosta AO, Lennon LT, et al. The influence of neighbourhood-level socioeconomic deprivation on cardiovascular disease mortality in older age: longitudinal multilevel analyses from a cohort of older British men. J Epidemiol Community Health. 2015; 69:1224–1231. [PubMed: 26285580]
6. Major JM, Doubeni CA, Freedman ND, Park Y, Lian M, Hollenbeck AR, et al. Neighborhood socioeconomic deprivation and mortality: NIH-AARP diet and health study. PLoS One. 2010; 5:e15538. [PubMed: 21124858]
7. Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, et al. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011; 42:3363–3368. [PubMed: 21940966]
8. Aslanyan S, Weir CJ, Lees KR, Reid JL, McInnes GT. Effect of area-based deprivation on the severity, subtype, and outcome of ischemic stroke. Stroke. 2003; 34:2623–2628. [PubMed: 14576369]
9. Diez Roux AV, Merkin SS, Arnett D, Chambless L, Massing M, Nieto FJ, et al. Neighborhood of residence and incidence of coronary heart disease. N Engl J Med. 2001; 345:99–106. [PubMed: 11450679]
10. Kershaw KN, Diez Roux AV, Bertoni A, Carnethon MR, Everson-Rose SA, Liu K. Associations of chronic individual-level and neighbourhood-level stressors with incident coronary heart disease: the Multi-Ethnic Study of Atherosclerosis. J Epidemiol Community Health. 2015; 69:136–141. [PubMed: 25271247]
11. Deguen S, Lalloue B, Bard D, Havard S, Arveiler D, Zmirou-Navier D. A small-area ecologic study of myocardial infarction, neighborhood deprivation, and sex: a Bayesian modeling approach. Epidemiology. 2010; 21:459–466. [PubMed: 20489648]
12. Geraghty EM, Balsbaugh T, Nuovo J, Tandon S. Using Geographic Information Systems (GIS) to assess outcome disparities in patients with type 2 diabetes and hyperlipidemia. J Am Board Fam Med. 2010; 23:88–96. [PubMed: 20051547]
13. Li X, Memarian E, Sundquist J, Zoller B, Sundquist K. Neighbourhood deprivation, individual-level familial and socio-demographic factors and diagnosed childhood obesity: a nationwide multilevel study from Sweden. Obes Facts. 2014; 7:253–263. [PubMed: 25096052]
14. Powell-Wiley TM, Cooper-McCann R, Ayers C, Berrigan D, Lian M, McClurkin M, et al. Change in neighborhood socioeconomic status and weight gain: Dallas Heart Study. Am J Prev Med. 2015; 49:72–79. [PubMed: 25960394]
15. Yen IH, Michael YL, Perdue L. Neighborhood environment in studies of health of older adults: a systematic review. Am J Prev Med. 2009; 37:455–463. [PubMed: 19840702]
16. Richardson S, Shaffer JA, Falzon L, Krupka D, Davidson KW, Edmondson D. Meta-analysis of perceived stress and its association with incident coronary heart disease. Am J Cardiol. 2012; 110:1711–1716. [PubMed: 22975465]
17. Brown P, Guy M, Broad J. Individual socioeconomic status, community socioeconomic status and stroke in New Zealand: a case control study. Soc Sci Med. 2005; 61:1174–1188. [PubMed: 15970229]

18. Yan T, Escarce JJ, Liang LJ, Longstreth WT Jr, Merkin SS, Ovbiagele B, et al. Exploring psychosocial pathways between neighbourhood characteristics and stroke in older adults: the cardiovascular health study. Age Ageing. 2013; 42:391–397. [PubMed: 23264005]
19. Sundquist K, Malmstrom M, Johansson SE. Neighbourhood deprivation and incidence of coronary heart disease: a multilevel study of 2. 6 million women and men in Sweden. J Epidemiol Community Health. 2004; 58:71–77. [PubMed: 14684730]
20. Diez Roux AV, Mair C. Neighborhoods and health. Ann N Y Acad Sci. 2010; 1186:125–145. [PubMed: 20201871]
21. Hajat A, Diez-Roux AV, Adar SD, Auchincloss AH, Lovasi GS, O'Neill MS, et al. Air pollution and individual and neighborhood socioeconomic status: evidence from the Multi-Ethnic Study of Atherosclerosis (MESA). Environ Health Perspect. 2013; 121:1325–1333. [PubMed: 24076625]
22. Coogan PF, White LF, Evans SR, Adler TJ, Hathaway KM, Palmer JR, et al. Longitudinal assessment of urban form and weight gain in African-American women. Am J Prev Med. 2011; 40:411–418. [PubMed: 21406274]
23. Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, et al. The development of a standardized neighborhood deprivation index. J Urban Health. 2006; 83:1041–1062. [PubMed: 17031568]
24. Kim D, Diez Roux AV, Kiefe CI, Kawachi I, Liu K. Do neighborhood socioeconomic deprivation and low social cohesion predict coronary calcification?: the CARDIA study. Am J Epidemiol. 2010; 172:288–298. [PubMed: 20610467]
25. Lalloue B, Monnez JM, Padilla C, Kihal W, Le Meur N, Zmirou-Navier D, et al. A statistical procedure to create a neighborhood socioeconomic index for health inequalities analysis. Int J Equity Health. 2013; 12:21. [PubMed: 23537275]
26. Kraus WE, Granger CB, Sketch MH Jr, Donahue MP, Ginsburg GS, Hauser ER, et al. A guide for a cardiovascular genomics biorepository: the CATHGEN experience. J Cardiovasc Transl Res. 2015; 8:449–557. [PubMed: 26271459]
27. Ward-Caviness CK, Kraus WE, Blach C, Haynes CS, Dowdy E, Miranda ML, et al. Association of roadway proximity with fasting plasma glucose and metabolic risk factors for cardiovascular disease in a cross-sectional study of cardiac catheterization patients. Environ Health Perspect. 2015; 123:1007–1014. [PubMed: 25807578]
28. Bart BA, Shaw LK, McCants CB Jr, Fortin DF, Lee KL, Califf RM, et al. Clinical determinants of mortality in patients with angiographically diagnosed ischemic or nonischemic cardiomyopathy. J Am Coll Cardiol. 1997; 30:1002–1008. [PubMed: 9316531]
29. Auchincloss AH, Van Nostrand JF, Ronsaville D. Access to health care for older persons in the United States: personal, structural, and neighborhood characteristics. J Aging Health. 2001; 13:329–354. [PubMed: 11813730]
30. U.S. Census Bureau. US Census Bureau Plans and rules for taking the Census. 2000. Available from https://www.census.gov/population/www/censusdata/resid_rules.html#Students
31. U.S. Census Bureau. Summary Files 1 and 3 [Internet]. 2000. [cited Sep 26, 2015]. Available from http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
32. Purser JL, Kuchibhatla MN, Miranda ML, Blazer DG, Cohen HJ, Fillenbaum GG. Geographical segregation and IL-6: a marker of chronic inflammation in older adults. Biomark Med. 2008; 2:335–348. [PubMed: 19655043]
33. Black JL, Macinko J. The changing distribution and determinants of obesity in the neighborhoods of New York City, 2003–2007. Am J Epidemiol. 2010; 171:765–775. [PubMed: 20172920]
34. Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc. 1963; 58:236–244.
35. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
36. Austin E, Coull BA, Zanobetti A, Koutrakis P. A framework to spatially cluster air pollution monitoring sites in US based on the PM2. 5 composition. Environ Int. 2013; 59:244–254. [PubMed: 23850585]
37. Friedman HP, Rubin J. On some invariant criteria for grouping data. J Am Stat Assoc. 1967; 62:1159–1178.

38. Pedigo A, Seaver W, Odoi A. Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and discriminant analyses approaches. PLoS One. 2011; 6:e22693. [PubMed: 21829481]

39. Roussot A, Cottenet J, Gadreau M, Giroud M, Bejot Y, Quantin C. The use of national administrative data to describe the spatial distribution of inhospital mortality following stroke in France, 2008–2011. Int J Health Geogr. 2016; 15:2. [PubMed: 26754188]

40. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. J Epidemiol Community Health. 2001; 55:111–122. [PubMed: 11154250]

41. Leigh JP, Du J. Hypertension and occupation among seniors. J Occup Environ Med. 2009; 51:661–671. [PubMed: 19415032]

42. Davila EP, Kuklina EV, Valderrama AL, Yoon PW, Rolle I, Nsubuga P. Prevalence, management, and control of hypertension among US workers: does occupation matter? J Occup Environ Med. 2012; 54:1150–1156. [PubMed: 22885710]

43. Rollings KA, Wells NM, Evans GW. Measuring physical neighborhood quality related to health. Behav Sci (Basel). 2015; 5:190–202. [PubMed: 25938692]

44. Schaefer-McDaniel N, Caughy MO, O'Campo P, Gearey W. Examining methodological details of neighbourhood observations and the relationship to health: a literature review. Soc Sci Med. 2010; 70:277–292. [PubMed: 19883966]

45. Kroeger GL, Messer L, Edwards SE, Miranda ML. A novel tool for assessing and summarizing the built environment. Int J Health Geogr. 2012; 11:46. [PubMed: 23075269]

46. Mujahid MS, Diez Roux AV, Morenoff JD, Raghunathan T. Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. Am J Epidemiol. 2007; 165:858–867. [PubMed: 17329713]

47. Ogilvie RP, Everson-Rose SA, Longstreth WT Jr, Rodriguez CJ, Diez-Roux AV, Lutsey PL. Psychosocial factors and risk of incident heart failure: the multi-ethnic study of atherosclerosis. Circ Heart Fail. 2016; 9:e002243. [PubMed: 26699386]

48. Christine PJ, Auchincloss AH, Bertoni AG, Carnethon MR, Sanchez BN, Moore K, et al. Longitudinal associations between neighborhood physical and social environments and incident type 2 diabetes mellitus: the Multi-Ethnic Study of Atherosclerosis (MESA). JAMA Intern Med. 2015; 175:1311–1320. [PubMed: 26121402]

49. Miranda ML, Edwards SE, Anthopolos R, Dolinsky DH, Kemper AR. The built environment and childhood obesity in Durham, North Carolina. Clin Pediatr (Phila). 2012; 51:750–758. [PubMed: 22563061]

50. Sundquist K, Theobald H, Yang M, Li X, Johansson SE, Sundquist J. Neighborhood violent crime and unemployment increase the risk of coronary heart disease: a multilevel study in an urban setting. Soc Sci Med. 2006; 62:2061–2071. [PubMed: 16203075]

51. Carroll-Scott A, Gilstad-Hayden K, Rosenthal L, Peters SM, McCaslin C, Joyce R, et al. Disentangling neighborhood contextual associations with child body mass index, diet, and physical activity: the role of built, socioeconomic, and social environments. Soc Sci Med. 2013; 95:106–114. [PubMed: 23642646]

52. Chaix B, Rosvall M, Merlo J. Recent increase of neighborhood socioeconomic effects on ischemic heart disease mortality: a multilevel survival analysis of two large Swedish cohorts. Am J Epidemiol. 2007; 165:22–26. [PubMed: 16973762]

53. Clark AM, DesMeules M, Luo W, Duncan AS, Wielgosz A. Socioeconomic status and cardiovascular disease: risks and implications for care. Nat Rev Cardiol. 2009; 6:712–722. [PubMed: 19770848]

54. Villanueva C, Aggarwal B. The association between neighborhood socioeconomic status and clinical outcomes among patients 1 year after hospitalization for cardiovascular disease. J Community Health. 2013; 38:690–697. [PubMed: 23468321]

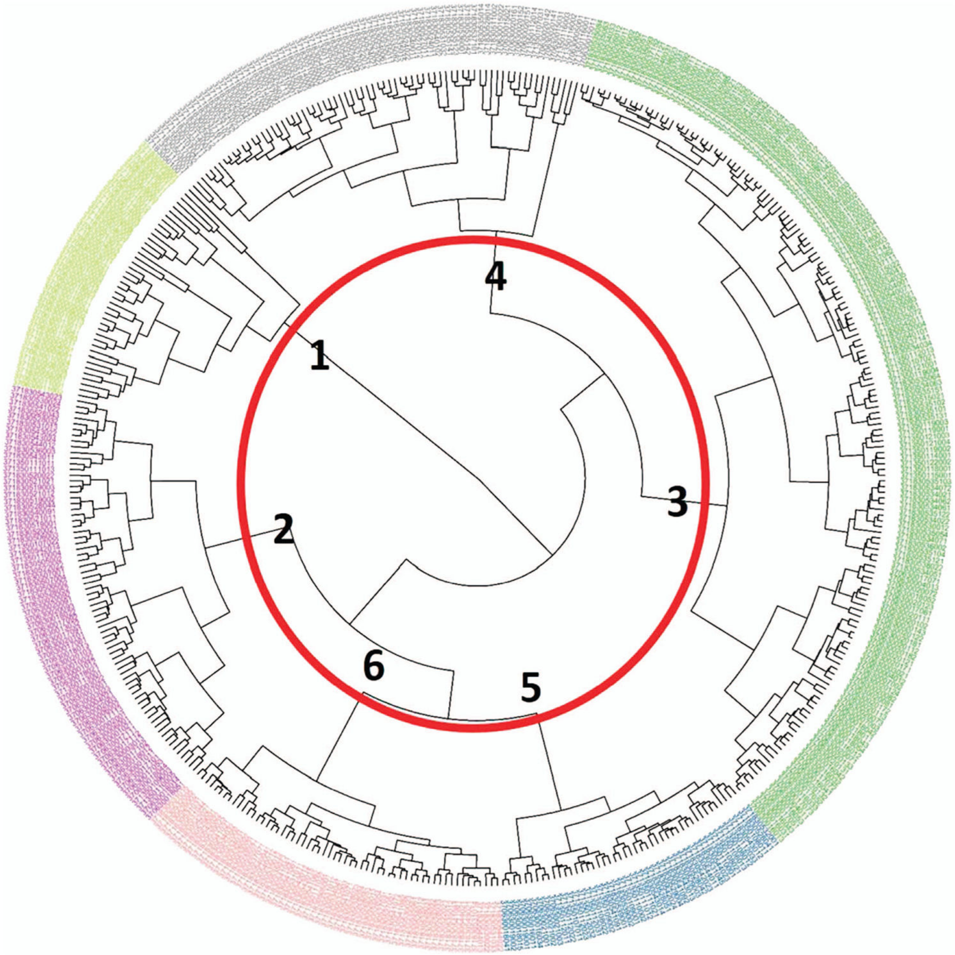

Figure 1.

Circular dendrogram of six clusters comprised of 444 Census block groups in Orange, Wake, and Durham counties, North Carolina. Numbers in bold represent the cluster numbers. Four block groups (BGs) having a population of zero or containing a prison/ correctional facility were removed. The bolded numbers represent each of the clusters.

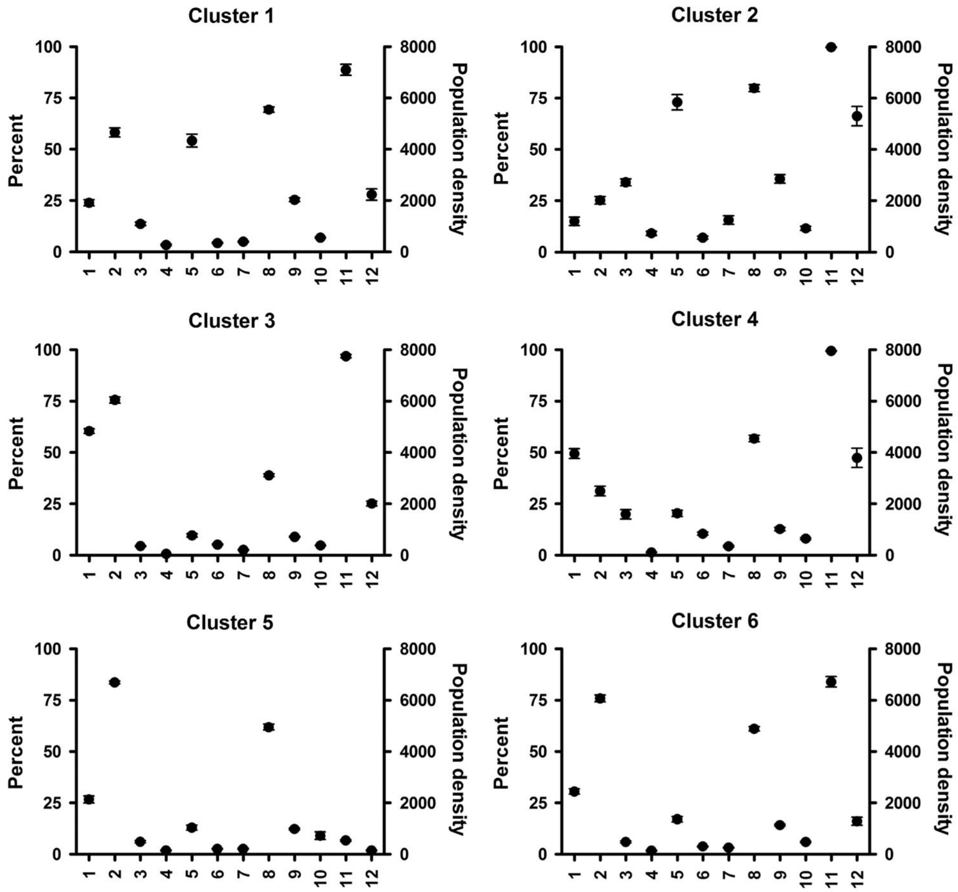

Figure 2.

Residential clusters by R-SES factors. Four block groups (BGs) having a population of zero or containing a prison/correctional facility were removed. Values represent mean ±SEM.

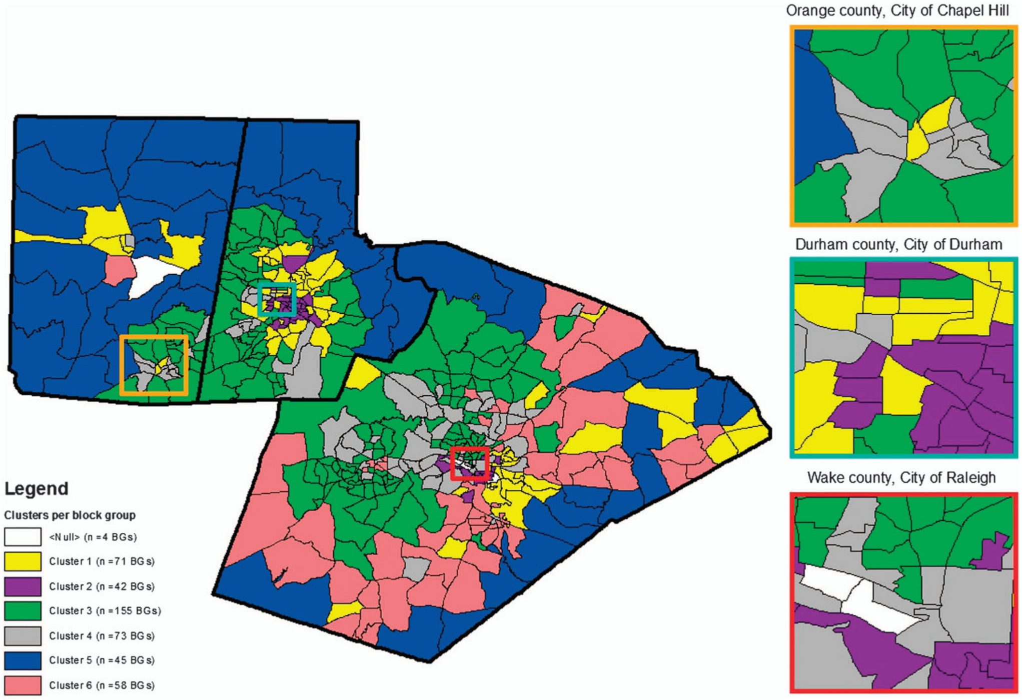

Figure 3.

Six clusters comprised of 444 Census block groups in Orange, Wake, and Durham Counties, North Carolina. Four block groups (BGs) having a population of zero or containing a prison/ correctional facility were removed.

Table 1

Input variables used to form the clusters for analysis.

Cluster input categories	Cluster input variables
Education	Population with at least a Bachelor's degree
Wealth	Population in owner-occupied housing
Income	Population with income below the poverty level
	Households on public assistance income
Race	Black
	Other race
Employment	Population unemployed
	Population in non-managerial positions
Housing	Single parent housing
	Vacant housing
Land-use	Urban environment
	Population density (persons per mile^3)

Table 2

Summary of subject characteristics.

	Total (*n* = 2254)	Cluster 1 (*n* = 392)	Cluster 2 (*n* = 210)	Cluster 3[a] (*n* = 961)	Cluster 4 (*n* = 235)	Cluster 5 (*n* = 361)	Cluster 6 (*n* = 95)
Subject characteristics							
Age (years)	61 (23–93)	60 (27–69)	58 (26–84)	63 (26–93)	61 (23–90)	60 (33–91)	61 (35–84)
Female (%)	39.1	41.3	57.1	37.6	37.4	31.3	40.0
Smoking (%)	43.8	47.7	49.0	40.8	42.6	46.5	40.0
Race							
White (%)	69.1	45.9	16.2	81.2	73.6	85.3	87.4
Black (%)	28.1	53.1	81.9	14.8	23.8	12.7	10.5
Other (%)	2.8	1.0	1.9	4.1	2.6	1.9	2.1
Health outcomes							
BMI (kg/m^2)	30.2 (11.6–83.1)	31.1 (12.5–83.1)	32.7 (11.6–63.3)	29.4 (15.9–69.5)	30.0 (16.2–71.9)	30.0 (15.2–58.6)	29.5 (16.4–52.5)
Fasting glucose (mg/dl)	117 (1–477)	123 (49–416)	130 (56–461)	113 (1–415)	112 (30–339)	115 (59–477)	129 (9–459)
Diabetes (*n* =2254) (%)	26.5	28.1	40.5	22.2	26.0	26.3	35.8
Hypertension (*n* = 2254) (%)	68.7	74.2	82.4	66.0	65.1	65.9	63.2
CHF (*n* = 2140) (%)	25.5	29.4	34.3	21.7	29.0	23.2	28.7
CVD (*n* =2254) (%)	7.1	7.4	7.6	7.4	6.8	6.6	5.3
CAD index >23 (*n* = 2062) (%)	58.6	55.5	47.2	60.8	56.7	63.5	60.5

Abbreviations: BMI, body mass index; CAD, coronary artery disease; CHF, congestive health failure; CVD, cardiovascular disease. Values represent average (range) unless where otherwise noted.

[a] Set to reference cluster. Values represent mean (range) or percent.

Table 3

Logistic and linear regressions of health end points by cluster.

	Cluster 1		Cluster 2		Cluster 4		Cluster 5		Cluster 6	
	Crude	**Adjusted**[a]	**Crude**	**Adjusted**[a]	**Crude**	**Adjusted**[a]	**Crude**	**Adjusted**[a]	**Crude**	**Adjusted**[a]
Logistic regression models										
BMI >25 (kg/m^2)	1.33	1.29 (0.96, 1.74)	1.69	1.65 (1.10, 2.54)	0.98	0.94 (0.68, 1.32)	1.35	1.25 (0.93, 1.70)	1.24	1.23 (0.74, 2.12)
BMI >30 (kg/m^2)	1.53	1.40 (1.09, 1.79)	2.36	1.92 (1.39, 2.67)	1.01	0.96 (0.71, 1.29)	1.08	1.03 (0.80, 1.33)	1.12	1.11 (0.71, 1.70)
Diabetes	1.37	1.31 (1.00, 1.72)	2.39	2.19 (1.57, 3.04)	1.23	1.23 (0.88, 1.71)	1.25	1.30 (0.98, 1.73)	1.96	2.04 (1.29, 3.17)
CHF	1.50	1.56 (1.17, 2.07)	1.88	1.99 (1.39, 2.83)	1.47	1.54 (1.10, 2.14)	1.09	1.17 (0.86, 1.57)	1.45	1.50 (0.90, 2.43)
Hypertension	1.49	1.38 (1.05, 1.82)	2.41	2.05 (1.38, 3.11)	0.96	1.00 (0.73, 1.36)	1.00	1.12 (0.86, 1.46)	0.88	0.94 (0.60, 1.48)
CVD	1.00	0.98 (0.61, 1.53)	1.03	0.97 (0.52, 1.71)	0.92	0.95 (0.52, 1.64)	0.89	1.00 (0.60, 1.61)	0.70	0.75 (0.26, 1.74)
CAD index >23	0.83	0.98 (0.75, 1.28)	0.55	0.83 (0.58, 1.18)	0.74	0.76 (0.55, 1.05)	1.06	1.09 (0.83, 1.42)	0.64	0.65 (0.40, 1.06)
Linear regression models										
Fasting glucose (mg/dl)	9.91	8.25 (2.34, 14.15)	16.98	14.23 (6.58, 21.88)	− 1.40	−2.04 (−9.06, 4.98)	2.18	1.83 (−4.13, 7.79)	16.33	16.54 (5.86, 27.22)

Abbreviations: BMI, body mass index; CAD, coronary artery disease; CHF, congestive health failure; CVD, cardiovascular disease. Cluster 3 was selected as the reference cluster for comparison. Values represent the odds ratio or relative differences (95% CI).

[a] Models adjusted for age, sex, smoking, and race.