Heat Maps of Hypertension, Diabetes Mellitus, and Smoking in the Continental United States

Matthew Shane Loop, PhD; George Howard, DrPH; Gustavo de los Campos, PhD; Mohammad Z. Al-Hamdan, PhD; Monika M. Safford, MD; Emily B. Levitan, ScD; Leslie A. McClure, PhD

Background—Geographic variations in cardiovascular mortality are substantial, but descriptions of geographic variations in major cardiovascular risk factors have relied on data aggregated to counties. Herein, we provide the first description of geographic variation in the prevalence of hypertension, diabetes mellitus, and smoking within and across US counties.

Methods and Results—We conducted a cross-sectional analysis of baseline risk factor measurements and latitude/longitude of participant residence collected from 2003 to 2007 in the REGARDS study (Reasons for Geographic and Racial Differences in Stroke). Of the 30239 participants, all risk factor measurements and location data were available for 28 887 (96%). The mean (±SD) age of these participants was 64.8(±9.4) years; 41% were black; 55% were female; 59% were hypertensive; 22% were diabetic; and 15% were current smokers. In logistic regression models stratified by race, the median(range) predicted prevalence of the risk factors were as follows: for hypertension, 49% (45%–58%) among whites and 72% (68%–78%) among blacks; for diabetes mellitus, 14% (10%–20%) among whites and 31% (28%–41%) among blacks; and for current smoking, 12% (7%–16%) among whites and 18% (11%–22%) among blacks. Hypertension was most prevalent in the central Southeast among whites, but in the west Southeast among blacks. Diabetes mellitus was most prevalent in the west and central Southeast among whites but in south Florida among blacks. Current smoking was most prevalent in the west Southeast and Midwest among whites and in the north among blacks.

Conclusions—Geographic disparities in prevalent hypertension, diabetes mellitus, and smoking exist within states and within counties in the continental United States, and the patterns differ by race.

Key Words: cardiovascular diseases ■ diabetes mellitus ■ epidemiology ■ hypertension ■ risk factors

Geographic disparities in risk factors for cardiovascular disease (CVD) have been documented by county for hypertension, diabetes mellitus, and obesity. Identifying variation of key risk factors between regions can help direct prevention efforts. County-level maps are often created because data are available only on a per-county basis. However, county-level maps of risk factor prevalences cannot identify disparities within a county and can distort the true patterns of disparities by using somewhat arbitrary administrative boundaries. Therefore, we used data from a national cohort study of geographic disparities in stroke mortality, which collected the street addresses of participants’ residences to describe geographic disparities in the prevalences of hypertension, diabetes mellitus, and smoking by creating heat maps for the continental United States that showed geographic disparities below the county level.

Methods

Study Sample
For our analysis, we used the REGARDS study (Reasons for Geographic and Racial Differences in Stroke), which is a longitudinal, population-based cohort study designed to identify factors associated with the excess stroke mortality among blacks and residents of the southeast United States. REGARDS was designed to be a near-national cohort of individuals aged 45 years, randomly selected with approximately equal representation of whites and blacks, men and women, with oversampling from the Stroke Belt (8 southeastern states of North Carolina, South Carolina, Georgia, Tennessee, Alabama, Mississippi, Louisiana, and Arkansas) and Stroke Buckle (coastal plains of North Carolina, South Carolina, and Georgia). The study enrolled 30239 participants from across the continental United States from 2003 to 2007 and has continued to follow them prospectively. Participants were recruited by mail and telephone, and a baseline computer-assisted telephone interview was used to obtain demographic variables and self-reported health information.
WHAT IS KNOWN

• The prevalence of self-reported hypertension, diabetes mellitus, and smoking vary by county in the United States.

WHAT THE STUDY ADDS

• A description of the geographic heterogeneity of hypertension, including directly measured blood pressure, and diabetes mellitus, including glucose assessment.

• The prevalence of hypertension, diabetes mellitus, and self-reported smoking vary on a scale across the continental United States.

• The prevalence of hypertension, diabetes mellitus, and smoking are not uniformly high across the Southeast United States, and regions of high prevalence of the risk factors differ.

• Public health interventions for cardiovascular disease prevention should be mindful of geographic heterogeneity and be planned according to local data and factors.

(eg, preexisting history of stroke). The computer-assisted telephone interview was followed by an in-home visit from a trained health professional, who collected anthropometric measurements (eg, height, weight, and waist circumference), blood pressure, an ECG, and blood and urine samples. All participants provided written informed consent, and the study protocol was approved by the Institutional Review Boards of all participating institutions.

Participants were considered to have hypertension if (1) they had a systolic blood pressure ≥140 mm Hg; (2) they had a diastolic blood pressure ≥90 mm Hg; or (3) they self-reported currently taking antihypertensive medications. Blood pressure was estimated as the average of 2 blood pressure measurements after the participant had been seated for 5 minutes. Participants were considered to have diabetes mellitus if (1) they had a fasting glucose level ≥126 mg/dL (or a glucose level ≥200 mg/dL if they failed to fast); or (2) they self-reported currently taking medication for diabetes mellitus management. Participants were considered current smokers if they had positive responses to the questions: (1) “Have you smoked at least 100 cigarettes in your lifetime?”; and (2) “Do you smoke cigarettes now, even occasionally?” which is similar to the definition used by the Behavioral Risk Factors Surveillance System. Past smokers were grouped with never smokers. We initially considered dyslipidemia as well, but we found no evidence of spatial clustering during preliminary testing (see Methods in the Data Supplement). Participant residential locations were geo-coded using SAS (SAS Institute, Cary, NC).

Statistical Methods

We calculated descriptive statistics of the sample by race. To generate heat maps, we used logistic regression models stratified by race with smooth functions of latitude and longitude of residence, adjusting for age and sex as standard potential confounders, which were chosen a priori. In other words, separate smooth maps were created for blacks and whites. We did not include other variables such as lifestyle factors that were available from the REGARDS study in our models because our goal was to identify potential geographic variation in risk factor prevalence below state and county levels, not to explain its potential causes. Spline-based techniques were used to model the smooth functions of location, as well as smooth functions of age. We chose the spline for age to be a restricted cubic spline with 3 knots, and the spline for location to be a second-order thin plate spline, fit using thin plate regression splines. The functions of location were essentially 1000 smooth planes joined together to provide estimates of prevalence that were seamless from one location to another. Model fit was assessed using percent of deviance of null model (intercept only) explained by fitted model. We then used the estimated model to provide predicted probabilities of the risk factor at the intersections of a 10 km x 10 km grid across the continental United States. Maps were created for typical REGARDS participants for each race by assuming populations of the average age and percent of women of each race in the REGARDS sample. Model fitting and prediction were performed using the mgcv package (version 1.8–14) in R (version 3.3.1), and map creation was done using ggplot2 (version 2.1.0) in R. Technical details of the analyses are presented in the Methods in the Data Supplement.

Results

After excluding participants with anomalous data (n=56), missing hypertension status (n=74), missing smoking status (n=116), missing diabetes mellitus status (n=1095), and missing location (n=11), our final sample size for this study was 28,887 (96% of the original sample).

The mean (SD) age of the sample was 64.8 (9.4) years. The sample was 41% black and 55% female. The sampling frame of the REGARDS study is reflected in the geographic distribution of participants: 35% of the participants lived in the Stroke Belt, 21% lived in the Stroke Buckle, and 45% lived in the rest of the continental United States. Fifty-nine percent of the participants had hypertension, 22% had diabetes mellitus, and 15% were current smokers. Summary statistics by race are presented in Table 1. Of the 17,107 participants with hypertension, 61% did not have an elevated systolic blood pressure or diastolic blood pressure but reported taking medication to lower blood pressure, and 11% had an elevated systolic blood pressure or diastolic blood pressure but did not report taking medication to lower blood pressure. Of the 6362 participants with diabetes mellitus, 45% did not have elevated blood glucose but reported using pills or insulin to treat diabetes mellitus, and 11% had elevated blood glucose but did not report using pills or insulin to treat diabetes mellitus. Figure 1 shows the participant residential locations by race.

Table 1. Characteristics of the Study Sample by Race

<table>
<thead>
<tr>
<th>Variable</th>
<th>Level</th>
<th>Blacks</th>
<th>Whites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>11,896 (41.2%)</td>
<td>16,991 (58.8%)</td>
<td></td>
</tr>
<tr>
<td>Age,* y</td>
<td>64.0 (9.3)</td>
<td>65.4 (9.5)</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>7,360 (61.9%)</td>
<td>8,472 (49.9%)</td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>5821 (48.9%)</td>
<td>7032 (41.4%)</td>
<td></td>
</tr>
<tr>
<td>Stroke Belt</td>
<td>3,947 (33.2%)</td>
<td>6,054 (35.6%)</td>
<td></td>
</tr>
<tr>
<td>Stroke Buckle</td>
<td>2,128 (17.9%)</td>
<td>3,905 (23.0%)</td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>8,497 (71.4%)</td>
<td>8,610 (50.7%)</td>
<td></td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>3,671 (30.9%)</td>
<td>2,691 (15.8%)</td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>2,064 (17.4%)</td>
<td>2,140 (12.6%)</td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>5,414 (45.5%)</td>
<td>7,639 (45.0%)</td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>4,418 (37.1%)</td>
<td>7,212 (42.4%)</td>
<td></td>
</tr>
</tbody>
</table>

All statistics are numbers (percentages) unless otherwise noted. *Mean (standard deviation).
High-Resolution Maps of Hypertension, Diabetes Mellitus, and Smoking Prevalence

Table 2 shows the age- and location-adjusted odds ratios (95% confidence intervals) for women versus men for each risk factor and race. Age was significantly associated ($P<0.001$) with odds of hypertension, diabetes mellitus, and current smoking in both blacks and whites, after adjustment for sex. The smoothed maps for age- and sex-adjusted hypertension prevalence, diabetes mellitus prevalence, and smoking prevalence, stratified by race, are presented in Figure 2. The percent deviance explained by the fitted models for whites was 3.9% for hypertension, 1.5% for diabetes mellitus, and 3.8% for smoking. The percent deviance explained by the fitted models for blacks was 2.5% for hypertension, 1.3% for diabetes mellitus, and 3.7% for smoking. Hypertension, diabetes mellitus, and current smoking proved to have significant geographic variation in both blacks and whites ($P<0.01$). Predicted prevalences for whites assumed a population aged 65 years and 50% female (mean age and proportion of women among white REGARDS participants), and predicted prevalences for blacks assumed a population aged 64 years and 62% female (mean age and proportion of women among black REGARDS participants). Thus, predicted prevalences for each risk factor represent the typical black and white participant in the REGARDS study. The median (range) predicted prevalence of hypertension was 49% (45%–58%) among whites and 72% (68%–78%) among blacks. The median (range) predicted prevalence of diabetes mellitus was 14% (10%–20%) among whites and 31% (28%–41%) among blacks. The median (range) predicted prevalence of current smoking was 12% (7%–16%) among whites and 18% (11%–22%) among blacks. Hypertension was most prevalent in the central Southeast among whites (Mississippi, Alabama, and Georgia), but in the west Southeast among blacks (Louisiana, Arkansas, and Mississippi). Diabetes mellitus was most prevalent in the west and central Southeast among whites (Louisiana, Arkansas, Mississippi, Alabama, Tennessee, and south Kentucky, as well as parts of North Carolina and South Carolina), but in south Florida among blacks. Current smoking was most prevalent in the west Southeast and east Midwest among whites (Arkansas, Missouri, Illinois, Kentucky, Tennessee, and Mississippi) and in the north among blacks (north Minnesota, north Wisconsin, and central and north Michigan).

Discussion

We found that the age- and sex-adjusted prevalence of hypertension, diabetes mellitus, and current smoking among both black and white REGARDS participants varied across the continental United States. The areas with the highest prevalence of hypertension were fairly concordant among whites and blacks, with the highest prevalence for blacks slightly more concentrated around northeast Louisiana, southeast Arkansas, and central Mississippi. Higher than average prevalence of hypertension was found not only in the Southeast, but also in the Midwest and parts of the Northeast. Higher than average prevalence of diabetes mellitus was present throughout the Southeast in whites, but was limited to the coastal regions of the Southeast among blacks. Although portions of the Southeast had higher than average prevalence of current smokers, current smokers were most prevalent in the Midwest among whites. These maps revealed geographic disparities for major CVD risk factors not only within states, but also within counties.

Compared with previous studies, we provided smooth heat maps of prevalence as opposed to maps of prevalence by county. Because of the high resolution, we were able to identify high hypertension prevalence in north Louisiana among blacks, which was not identified in a previous study. A previous study by the Centers for Disease Control using data from Behavioral Risk Factors Surveillance System identified a Diabetes Belt in the Southeast, which had higher levels of...
prevalence of self-reported diabetes mellitus compared with the rest of the United States. Although our maps showed high prevalences of diabetes mellitus in the same region, we also identified an area in the west and central Southeast that had higher prevalences of diabetes mellitus than the rest of the Diabetes Belt. The high smoking prevalence identified among whites in the midwest was similar to previous maps for smoking prevalence among women in 2012.9 The map for smoking for men in 2012 in that study included the same counties as for women, but also extended throughout much of the Southeast. By creating smooth maps, we were able to identify subregions of increased prevalence for hypertension and diabetes mellitus that were previously unidentified using county-level maps.

Geographic disparities in the estimated prevalence of hypertension and diabetes mellitus in our study could have arisen from 3 main sources: (1) variation in the prevalence of factors that led to the development of hypertension or diabetes mellitus (eg, diet, physical activity, or chronic kidney disease); (2) variation in the treatment of hypertension or diabetes mellitus, given that our definition of these risk factors included self-reported use of medications for these conditions; or (3) variation in the degree of self-report bias of use of medications for these conditions. Geographic disparities in socioeconomic status could explain the geographic disparities in these risk factors, but socioeconomic status explains <16% of the excess stroke mortality in the Stroke Belt compared with the rest of the United States.10 Because hypertension, diabetes mellitus, and smoking are key risk factors for stroke,11 it is unclear whether disparities in socioeconomic status were a major contributor to the geographic disparities in these risk factors observed in our study. The contribution of socioeconomic status is an area we hope to pursue in future work. Given that we included measured blood pressures and blood glucose levels, variation in degree of self-report bias for medication use would affect our results only if, among participants with normal blood pressures or blood glucose levels, there was large-scale over-reporting of use of these medications. Geographic variation in smoking prevalence could have been because of geographic variation in other lifestyle risk factors, but could also have been influenced by state-, county-, or city-level political and economic policies regarding tobacco. The

![Maps of estimated hypertension, diabetes mellitus, and current smoking prevalence among whites and blacks, adjusted for age and sex.](http://circoutcomes.ahajournals.org/)

Figure 2. Maps of estimated hypertension, diabetes mellitus, and current smoking prevalence among whites and blacks, adjusted for age and sex. High prevalence is indicated by red, while low prevalence is indicated by blue. Predicted prevalences assumed a population with the same proportion of women for each race and the same age as the mean age of each race. Thus, the prevalences reflect the sex and age composition of REGARDS participants of each race. REGARDS indicates Reasons for Geographic and Racial Differences in Stroke.
multiple potential causes of geographic variation require further investigation to determine the contribution of each cause.

Our study had some notable limitations. Because the study used simple random sampling within race-sex-region strata, there was a lack of uniform data coverage across the entire continental United States, which increased the uncertainty of our predictions in Western states (where the US population is more sparse), particularly for blacks (see Figure II in the Data Supplement). However, when we restricted the sample to participants living in the east United States, where the majority of REGARDS participants were sampled, there were not any meaningful differences in the predicted geographic patterns of each risk factor (see Figure III in the Data Supplement). Other spatial prediction methods could have potentially identified smaller scale variation than the spline-based approach we used. Spline-based models are known to capture long-range variation well, but often fail to effectively capture short-range variation. We attempted to overcome this limitation by using the largest number of planes joined together that balanced goodness of fit for the statistical model and computation time. However, we could not completely overcome the limited ability of spline-based methods to capture short-range trends. The data presented in this analysis were collected ≈10 years ago, and the prevalences of these risk factors might have changed. However, the patterns we observed in our study would only be affected if there were important differences in the temporal trends of these risk factors among regions. Such a trend was not obvious in a previous analysis of hypertension prevalence from 2001 to 2009. Finally, the REGARDS study oversampled blacks and residents of the Stroke Belt and Stroke Buckle, which likely affected our estimates of prevalence of these conditions. However, the patterns of geographic disparities depicted in the heat maps should not have depended on the estimated prevalences. Although we could have used sampling weights to produce estimates of prevalence that reflected the estimated prevalences. Although we could not completely overcome the limited ability of spline-based methods to capture short-range trends. The data presented in this analysis were collected ≈10 years ago, and the prevalences of these risk factors might have changed. However, the patterns we observed in our study would only be affected if there were important differences in the temporal trends of these risk factors among regions. Such a trend was not obvious in a previous analysis of hypertension prevalence from 2001 to 2009. Finally, the REGARDS study oversampled blacks and residents of the Stroke Belt and Stroke Buckle, which likely affected our estimates of prevalence of these conditions. However, the patterns of geographic disparities depicted in the heat maps should not have depended on the estimated prevalences. Although we could have used sampling weights to produce estimates of prevalence that reflected the US population, this would have required the derivation of new statistical methods that correctly estimate standard errors for a thin plate regression splines, which was outside the scope of this article. The estimated prevalences were also not representative of the general US population because we excluded adults aged <45. This exclusion criterion may not have been overly negative, given that hypertension and diabetes mellitus tend to affect the older US population more than the younger population. However, most current smokers are <45 years of age.

Despite its limitations, our study also had important strengths. Because of the sampling strategy of the REGARDS study, we were able to obtain dense coverage in areas currently purported to have high prevalences of CVD risk factors. Although previous studies have generally relied on only self-reported risk factors (eg, Behavioral Risk Factors Surveillance System), we were able to use both measured blood pressures and blood glucose levels, in addition to self-reported medication use, to identify those with hypertension or diabetes mellitus. The sensitivity of self-reported antihypertensive medication use was 94% in a sample of ≈36000 men and women from the Monitoring Project on Cardiovascular Disease risk factors in the Netherlands. The sensitivity and specificity of self-reported hypertension have been reported to be 71% and 90%, respectively, and the sensitivity and specificity of self-reported diabetes mellitus have been reported to be 58% to 71% and 96% to 97%, respectively. By using self-reported medication use instead of self-reported risk factor status, we were able to more reliably identify participants with these risk factors. In addition to potentially missing undiagnosed hypertension and diabetes mellitus, self-report bias can also vary geographically and prevent valid comparisons of estimated prevalences between regions, as noted in a previous study of geographic variation in diabetes mellitus. Although we do not have direct evidence of such geographic variation in misreporting for hypertension, diabetes mellitus, or smoking status, self-report of obesity has been shown to have such a bias in the continental United States. Finally, the heat maps produced in this analysis were on an 10 km×10 km grid, which allows for public distribution of maps of risk factor prevalence on a fine scale without compromising participant confidentiality. These maps could then be combined with heat maps of other risk factors for CVD already available, such as ambient air pollution.

Conclusions
We have shown evidence that the prevalences of hypertension, diabetes mellitus, and smoking varied on a finer scale than the state or county level. We have also shown that the patterns of these geographic disparities differed between blacks and whites. This information could inform prevention efforts targeted at particular populations within the United States. A clear future direction of this research is to determine the degree to which geographic disparities in these risk factors contribute to geographic disparities in CVD, including coronary heart disease and stroke, which are the first and fifth leading causes of death in the United States, respectively. Coronary heart disease mortality rates are known to be higher in the Mississippi and Ohio River valleys compared with those at the rest of the country, and stroke mortality rates are highest in the Stroke Belt and the Stroke Buckle. Knowledge of the degree to which geographic disparities in risk factors contribute to geographic disparities in CVD could lead to focused efforts that reduce the burden of CVD in the most affected regions of the United States.

Acknowledgments
We thank the anonymous reviewers for their helpful comments, as well as the other investigators, the staff, and the participants of the REGARDS study for their valuable contributions.

Sources of Funding
This research project is supported by a cooperative agreement U01 NS041588 from the National Institute of Neurological Disorders and Stroke, National Institutes of Health, Department of Health and Human Service. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Neurological Disorders and Stroke or the National Institutes of Health. Representatives of the funding agency have been involved in the review of the article but not directly involved in the collection, management, analysis, or interpretation of the data. A full list of participating REGARDS investigators and institutions can be found at http://www.regardsstudy.org. This research project was also partially funded by an American Heart Association 2014 Predoctoral Fellowship (14PRE18830073) and 5T32HL00745734 from the
Disclosures
Dr Loop has received salary support from Amgen Inc, and Dr Levitan receives salary support from Amgen, Inc. Dr Safford has received salary support from Amgen, Inc for an investigator-initiated project and has served as a consultant for diaDexus. Dr Levitan has received research grants from Amgen, Inc, serves on the Amgen, Inc advisory board, and served as a consultant for diaDexus. Dr Levitan has received research support from Amgen, Inc for an investigator-initiated project and has receives salary support from Amgen, Inc. Dr Safford has received salary support from Amgen Inc, and Dr Levitan has received salary support from Amgen, Inc.

References
Heat Maps of Hypertension, Diabetes Mellitus, and Smoking in the Continental United States
Matthew Shane Loop, George Howard, Gustavo de los Campos, Mohammad Z. Al-Hamdan, Monika M. Safford, Emily B. Levitan and Leslie A. McClure

Circ Cardiovasc Qual Outcomes. 2017;10:
doi: 10.1161/CIRCOUTCOMES.116.003350
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2017 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/10/1/e003350

Data Supplement (unedited) at:
http://circoutcomes.ahajournals.org/content/suppl/2017/01/09/CIRCOUTCOMES.116.003350.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org//subscriptions/
Supplemental Methods

Geocoding

Participant residential locations were geocoded using SAS (SAS Institute, Cary, NC), which uses the North American Datum 1983 (NAD83) as the geographic coordinate system. We projected the participant locations to the USA Contiguous Albers Equal Area Conic projected coordinate system in meters, in order to ease interpretation of the distances.

Preliminary test for spatial clustering

We tested for spatial variation in the prevalence of hypertension, diabetes, smoking, and dyslipidemia. Participants were considered to have dyslipidemia if: (1) they met the criteria for recommended treatment by lipid-lowering medications according to the Adult Treatment Panel III (APT III) guidelines;¹ or (2) they self-reported taking lipid lowering medications. In order to test for spatial variation in disease prevalence, we used the difference in Ripley’s functions test,²,³ which is the most commonly used test of disease clustering among epidemiologists when locations of diseased and nondiseased participants are available.⁴ This method tests whether the risk for the disease is constant across the region of interest.

In order to account for the overlapping points in the REGARDS dataset due to geocoding error, we excluded overlapping points. There is little precedent for “best practices” with a percentage of overlapping points this large. We tested for clustering of each risk factor in turn, using 1,166 equidistant ranges, which spanned from 0 to approximately 725 km. The maximum range (725 km) was chosen as one quarter the height of the enclosing rectangle for the continental US, and the interval between ranges was the median distance to the nearest neighbor (approximately 0.5 km). The polygonal window used for this study was the 1:20,000,000
resolution boundary shapefile of the US, which we modified to remove polygons not representing the continental US. We performed 200 Monte Carlo simulations and tested for clustering at an alpha level of 0.05. The null hypothesis was that the difference between Ripley’s $K$ function for the cases and Ripley’s $K$ function for the controls was 0 (i.e., $D(h) = K_{\text{cases}}(h) - K_{\text{controls}}(h) = 0$). Simultaneous critical regions for $D(h)$ over all ranges tested (i.e., all values of $h$) were constructed in order to prevent inflation of the family-wise error rate (FEW) caused by performing over 1,000 hypothesis tests. The null hypothesis was rejected if $D(h)$ was ever outside the simultaneous critical regions, leading to the conclusion of evidence of clustering for that particular risk factor. The test was performed using the spatstat package (v. 1.43-0) in the R statistical environment (v. 3.2.3).

**Results**

Seven participants were excluded from each test of disease clustering, given that their locations lay outside polygonal window used for the study. The exclusion of these participants, as well as participants that were missing dyslipidemia status, resulted in a sample size of 27,780 for the tests of disease clustering. The value of $D(h)$ for all values of $h$, as well as the 95% simultaneous critical envelopes for all $h$, for hypertension, diabetes, smoking, and dyslipidemia are presented in Supplemental Figure 1.

As shown in Supplemental Figure 1, there is evidence of clustering of hypertension up to a range of approximately 500 km, clustering of diabetes up to 600 km and 700 km, and clustering of current smoking up to 550 – 600 km. In other words, within these distances you would find more people around someone with the risk factor of interest, who also had the risk factor, than you would expect by chance. For comparison, the driving distance from Boston, MA
to Philadelphia, PA is about 500 km. On the other hand, we found no evidence of clustering of
dyslipidemia despite a large sample size (black curve never exited the grey critical envelopes).

Supplemental Figure 2 shows the widths of the 95% confidence intervals of the predicted
prevalences across the US, which conveys the uncertainty in our predictions. Mean prevalences
whose confidence intervals do not overlap can be considered statistically significantly different.
However, no adjustment was made to the interval widths for multiple comparisons, so such
results should be interpreted with caution. The widest confidence intervals among blacks tended
to be twice as wide as the widest confidence intervals in whites, reflecting both the fewer number
of blacks and the low proportion of blacks in large sections of the continental US. Supplemental
Figure 3 shows the estimated prevalences when only participants living in the eastern US were
used in the prediction models.
Supplemental Figure 1. Estimated difference in K functions, D(h), as a function of range for each risk factor of interest. The grey lines are the critical values for each corresponding range, the black curve is D(h), and the grey dashed line is the estimated mean value for D(h) under the null hypothesis of constant risk. Hypertension, diabetes, and smoking show evidence of clustering, while dyslipidemia does not.
Supplemental Figure 2. Maps of widths for 95% confidence intervals for predicted mean hypertension, diabetes, and smoking prevalence among whites and blacks, adjusted for age and gender. The widest intervals (most uncertainty) are indicated by red, while the shortest intervals (least uncertainty) are indicated by blue.
Supplemental Figure 3. Maps of estimated hypertension, diabetes, and current smoking prevalence among whites and blacks, adjusted for age and gender, using only REGARDS participants living in the east half of the US. High prevalence is indicated by red, while low prevalence is indicated by blue. Predicted prevalences assumed a population with the same proportion of women for each race and the same age as the mean age of each race. Thus, the prevalences reflect the gender and age composition of REGARDS participants of each race.
Supplemental references


