Reassessing the Stroke Belt
Using Small Area Spatial Statistics to Identify Clusters of High Stroke Mortality in the United States

David N. Karp, MUSA; Catherine S. Wolff, MS; Douglas J. Wiebe, PhD; Charles C. Branas, PhD; Brendan G. Carr, MD, MS; Michael T. Mullen, MD, MS

Background and Purpose—The stroke belt is described as an 8-state region with high stroke mortality across the southeastern United States. Using spatial statistics, we identified clusters of high stroke mortality (hot spots) and adjacent areas of low stroke mortality (cool spots) for US counties and evaluated for regional differences in county-level risk factors.

Methods—A cross-sectional study of stroke mortality was conducted using Multiple Cause of Death data (Centers for Disease Control and Prevention) to compute age-adjusted adult stroke mortality rates for US counties. Local indicators of spatial association statistics were used for hot-spot mapping. County-level variables were compared between hot and cool spots.

Results—Between 2008 and 2010, there were 393 121 stroke-related deaths. Median age-adjusted adult stroke mortality was 61.7 per 100 000 persons (interquartile range=51.4–74.7). We identified 705 hot-spot counties (22.4%) and 234 cool-spot counties (7.5%); 44.5% of hot-spot counties were located outside of the stroke belt. Hot spots had greater proportions of black residents, higher rates of unemployment, chronic disease, and healthcare utilization, and lower median income and educational attainment.

Conclusions—Clusters of high stroke mortality exist beyond the 8-state stroke belt, and variation exists within the stroke belt. Reconsideration of the stroke belt definition and increased attention to local determinants of health underlying small area regional variability could inform targeted healthcare interventions. (Stroke. 2016;47:1939-1942. DOI: 10.1161/STROKEAHA.116.012997.)

Key Words: attention ■ cause of death ■ chronic disease ■ risk factors ■ stroke

The stroke belt, an 8-state region in the southeastern United States, defined for its disproportionately high stroke mortality rates,1,2 has been present since at least 1940 and persists despite recent decreases in stroke mortality, overall.3 Differences in vascular risk factors may explain approximately half of the excess burden,4 yet underlying drivers in this region are not fully understood.1 Access to primary stroke centers (PSCs) is lower within the region, suggesting that differential access to care could be one contributing factor.5,6

Using the existing state-based definition may not provide adequate geographic precision to fully understand the impact of local demographic and healthcare factors as determinants of health. We used county-level data and spatial statistics to empirically identify geographic clusters of high stroke mortality, or hot spots, at a finer geographic resolution to compare with the traditional state-based stroke belt. We compared multiple county-level variables between high stroke mortality hot spots and adjacent low stroke mortality cool spots, to understand local factors contributing to stroke outcomes.

Methods
We used Multiple Cause of Death data (Centers for Disease Control and Prevention) from 2008 to 2010 to calculate a 3-year average age-adjusted adult stroke mortality rate for all US counties. International Classification of Diseases—Tenth Revision codes I60-I69 were used to identify stroke as cause of death.

Local indicators of spatial association statistics were used to describe spatial patterns of mortality rates. This technique categorizes counties as: clustered high-rate counties (High-High); low-rate counties adjacent to High-High counties (Low-High); clustered low-rate counties (Low-Low); and high-rate counties adjacent to Low-Low counties (High-Low). Nonsignificant counties demonstrate spatial randomness among neighbors. We compared High-High and Low-High counties, respectively hot spots and cool spots. Hot spots do not include HL counties and cool spots do not include Low-Low counties.

Received January 31, 2016; final revision received April 28, 2016; accepted May 3, 2016.

From the Department of Biostatistics and Epidemiology (D.N.K., D.J.W., C.C.B.), Leonard Davis Institute of Health Economics (D.J.W., C.C.B., M.T.M.), and Department of Neurology (M.T.M.), University of Pennsylvania, Philadelphia; Department of Emergency Medicine, Thomas Jefferson University, Philadelphia, PA (B.G.C.); and Duke University School of Medicine, Duke University, Durham, NC (C.S.W.).


The online-only Data Supplement is available with this article at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.116.012997/-/DC1.

Correspondence to Michael T. Mullen, MD, MS, Department of Neurology, University of Pennsylvania, 3400 Spruce St, 3W Gates Bldg, Philadelphia, PA 19104. E-mail Michael.mullen@uphs.upenn.edu

© 2016 American Heart Association, Inc.

Stroke is available at http://stroke.ahajournals.org DOI: 10.1161/STROKEAHA.116.012997

1939
We analyzed county-level demographic characteristics (2008–2012 American Community Survey), healthcare resources, and utilization rates (2008, 2010 Area Health Resource File), and prevalence of diabetes mellitus, obesity (2011 Food Environment Atlas), and hypertension (Institute for Health Metrics and Evaluation). Data describing access to PSCs used previously described methods. We

Figure. Stroke mortality rates and hot spots mapped by county. A, The US county map showing stroke mortality rates (2008–2010). Rates are reported per 100,000 persons; binned by quintile. B, Stroke mortality hot-spot classifications. Local indicators of spatial association (LISA) statistics identify counties (significance level $P<0.05$), as high-high (HH), low-high (HL), low-low (LL), high-low (HL), or not significant. In A and B, stroke belt states are outlined in black.
used Wilcoxon rank-sum tests to compare averages. Spatial methods are further described in the online-only Data Supplement.

**Results**

From 2008 to 2010, 393,121 stroke deaths were reported across 3,137 counties. Median county-level age-adjusted stroke mortality was 61.7 per 100,000 persons (interquartile range=51.4–74.6; Figure [A]). The hot-spot analysis identified 705 High-High (22.4%), 234 Low-High (7.5%), 238 Low-Low (7.6%), and 52 HL (1.7%) counties (Figure [B]). Of the 688 counties included in the 8-state stroke belt, there were 391 High-High (57.4%) and 91 Low-High (13.2%). The other 314 High-High counties were located outside of the 8-state region.

Compared with cool spots, hot spots had significantly larger proportions of black residents; higher rates of unemployment, obesity, diabetes mellitus, and hypertension; more hospital admissions and emergency department visits per capita; and lower median income and educational attainment (Table). Sixty-minute access to a PSC was available for 65.2% of people nationally, 31.5% of the population in hot-spot counties, and 50.7% in cool-spot counties. Median county-level PSC access did not differ overall or stratified by urbanicity (Table I in the online-only Data Supplement).

**Discussion**

Our results confirmed past findings that high stroke mortality is geographically clustered in the southeast, and detected clusters of high stroke mortality existing outside of the traditional 8-state stroke belt. A state-based approach misses nearly half of the counties identified empirically as hot spots. We identified cool spots within the stroke belt where mortality is lower than expected. County-level heterogeneity suggests that state-based analyses may limit our understanding of the underlying drivers of survival through misclassification bias. Because things that are geographically close tend to be more similar, identifying areas that have significantly different mortality rates, despite close proximity, may help explain the drivers of disparities in outcomes.

There was a higher rate of 60-minute PSC access for people living in hot spots than cool spots; however, a statistically significant difference for county-level averages was not observed. We found a paradoxical relationship between the number of physicians and mortality, with more physicians on average in hot spots than cool spots. These findings combined may suggest that differential access to health care and specialty stroke care may not be the key factor underlying geographic variability in stroke mortality. It is possible that more granular scales of analysis are needed to detect meaningful differences (the mortality data used restricted us to a county-level analysis).

We found statistically significant differences in prevalence of diabetes mellitus, obesity, and hypertension between hot spots and cool spots. However, differences in disease severity or disease control may be magnifying relatively small differences in prevalence.

This study has limitations. The stability of spatial patterns of mortality over time is not known, but a previous study

---

**Table. Comparison of Hot-Spot and Cool-Spot Counties**

<table>
<thead>
<tr>
<th></th>
<th>All Counties, n=3137</th>
<th>Hot-Spot Counties (HH), n=705</th>
<th>Cool-Spot Counties (LH), n=234</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted stroke mortality rate (per 100k pop)</td>
<td>56.1 (51.4–74.7)</td>
<td>83.6 (76.5–95.5)</td>
<td>46.1 (38.9–51.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Urban population, %</td>
<td>80.6 (12.2–67)</td>
<td>31.4 (10.4–51.5)</td>
<td>26.5 (0–55.1)</td>
<td>0.459</td>
</tr>
<tr>
<td>Median age, y</td>
<td>37.2 (34.4–43.3)</td>
<td>39.9 (37.6–42)</td>
<td>40.4 (38–42.9)</td>
<td>0.033</td>
</tr>
<tr>
<td>Black population, %</td>
<td>12.6 (0.5–10.4)</td>
<td>10.4 (1.9–32.1)</td>
<td>4.8 (1–19)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hispanic population, %</td>
<td>16.4 (1.6–8.3)</td>
<td>2.6 (1.4–5.7)</td>
<td>2.9 (1.5–6.3)</td>
<td>0.221</td>
</tr>
<tr>
<td>Median household income ($, 2012)</td>
<td>53,046 (37,970–50,697)</td>
<td>37,508 (33,333–42,256)</td>
<td>40,682.5 (34,848–48,596)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Unemployment rate, %</td>
<td>9.3 (6.2–10.7)</td>
<td>9.2 (8–12.6)</td>
<td>9 (6.9–10.7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Education high school or less, %</td>
<td>42.5 (43.8–58.7)</td>
<td>58.7 (53.1–63.7)</td>
<td>56 (47–62.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No health insurance, %</td>
<td>14.9 (11.8–18.5)</td>
<td>17.1 (14.7–19.9)</td>
<td>16.7 (13.8–20.2)</td>
<td>0.131</td>
</tr>
<tr>
<td>Adult diabetes mellitus, %</td>
<td>9.0 (8.5–11.3)</td>
<td>11.6 (10.4–12.7)</td>
<td>10.7 (9.5–11.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Adult obesity, %</td>
<td>26.5 (27.2–31)</td>
<td>31.1 (29.5–32.8)</td>
<td>29.9 (27.5–31.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Adult hypertension, %</td>
<td>37.2 (37–41.2)</td>
<td>42.2 (40.4–44.9)</td>
<td>40.4 (38.7–42.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>MDs (per 100k pop)</td>
<td>30.3 (15.5–39.4)</td>
<td>22.7 (14.5–33.3)</td>
<td>20.1 (11.5–31.8)</td>
<td>0.031</td>
</tr>
<tr>
<td>Hospitals (per 100k pop)</td>
<td>2.1 (1–2)</td>
<td>1 (1–1)</td>
<td>1 (0–1)</td>
<td>0.034</td>
</tr>
<tr>
<td>Hospital admissions (per 100k pop)</td>
<td>86.8 (27.2–119.2)</td>
<td>71.6 (21.1–126.7)</td>
<td>48.2 (0–95.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ED visits (per 100k pop)</td>
<td>396.3 (169.7–536.1)</td>
<td>432.1 (110.5–603.3)</td>
<td>331.7 (0–521.9)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Medicare eligible population, %</td>
<td>14.7 (15.3–21.3)</td>
<td>19.1 (16.5–21.5)</td>
<td>18.3 (15.1–20.6)</td>
<td>0.003</td>
</tr>
</tbody>
</table>

The Table provides national totals and county medians for multiple variables and compares hot spots and cool spots using Wilcoxon rank-sum (Mann–Whitney) tests. ED indicates emergency department; HH, High-High; IQR, interquartile range; and LH, Low-High.
found that 75% of stroke hospitalization clusters were stable >10 years. This cross-sectional, population-level analysis cannot assess causality. Population-level associations may not apply at the individual level. Because of the low number of HL counties (n=52), comparisons of HL and Low-Low precluded a meaningful analysis.

Conclusions
Clusters of high stroke mortality exist beyond the 8-state stroke belt, and variation exists within the stroke belt. Reconsideration of the stroke belt definition and increased attention to small area regional variability using spatial methods may allow for better classification of regional disparities and inform targeted healthcare interventions.

Sources of Funding
This work was supported by the Agency for Healthcare Research and Quality (AHRQ-R01-HS018362-01A1).

Disclosures
Dr. Carr spends a portion of his time as Director of the Emergency Care Coordination Center in the Office of the Assistant Secretary for Preparedness and Response. Findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the Department of Health and Human Services.

References
Reassessing the Stroke Belt: Using Small Area Spatial Statistics to Identify Clusters of High Stroke Mortality in the United States

David N. Karp, Catherine S. Wolff, Douglas J. Wiebe, Charles C. Branas, Brendan G. Carr and Michael T. Mullen

Stroke. 2016;47:1939-1942; originally published online May 19, 2016; doi: 10.1161/STROKEAHA.116.012997

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/47/7/1939

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2016/05/19/STROKEAHA.116.012997.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke 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 Stroke is online at:
http://stroke.ahajournals.org//subscriptions/
Online Supplement

METHODS

Spatial Autocorrelation and Geo-statistical Analysis

Geo-statistical methods were used to assess the presence and nature of spatial autocorrelation in the data. That is, whether values of a given variable are random across geography, or alternatively whether a variable is correlated with itself across space, in which case we observe spatial patterns in the data among geographically proximate “neighbors.” A test finding of positive correlation indicates that values are clustered in space, with locations having values that are more similar to values of nearby locations than values of distant locations. A test finding of negative correlation indicates dispersion of values, with locations having high values being nearby locations with low values, and vice versa, thus representing dispersion rather than clustering. A global test of spatial autocorrelation tests for the presence of clustering or dispersion in the dataset overall. A local test of spatial autocorrelation can reveal the precise location of clustering or dispersion. In that way, a local test of spatial autocorrelation can be used to determine the geographic location of outliers, i.e. locations where low values are clustered with high values, and high values are clustered with low values. These tests are evaluated for statistical significance by comparing the observed spatial distributions to what would be expected given spatial randomness.

With both types of tests, the investigator specifies how “neighbor” is defined for the purposes of comparing values of each geographic area to values of neighboring or nearby areas, to test whether values of the index locations are positively or negatively associated with the mean of the values of the index locations’ neighbors. The most straightforward and intuitive method treats the geographic areas that are contiguous to and share a boarder with a given area as the neighbors of that area. Formally, this definition is called first-order contiguity and the comparison is accomplished by building a spatial weights matrix that for each location identifies its contiguous neighbors. That approach is not ideal if the areas being studied differ in size. An alternative is to treat all other geographic locations in the dataset as a neighbor of a given location, and compute the mean of the values of those locations by down-weighting the value of each location by the inverse of its distance from the index location. This weighing scheme is called inverse distance weighting (IDW). This method has the added benefit of enabling an analysis without making the investigator decide arbitrarily the distance or bandwidth within which areas are treated as neighbors.

In our analysis we used the Local Indicator of Spatial Association, or LISA\(^1\) statistic, to test for spatial autocorrelation of stroke mortality rates in counties across the US, and defined neighbors according to inverse distance weighting. We presented the results in the conventional manner, which entails mapping the geographic areas under study and shading those areas where statistically significant
spatial autocorrelation was identified. Different shading colors are used to indicate the nature of the spatial autocorrelation in each local area.

**Supplemental Reference:**

**Supplemental Table I.** Primary Stroke Center (PSC) Access by Urbanicity

<table>
<thead>
<tr>
<th></th>
<th>Hot-spots</th>
<th></th>
<th>Cool-spots</th>
<th></th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Population</td>
<td>PSC Access Median (IQR)</td>
<td>N</td>
<td>Population</td>
</tr>
<tr>
<td>All Counties</td>
<td>705</td>
<td>29,615,396</td>
<td>0% (0-0%)</td>
<td>234</td>
<td>11,494,779</td>
</tr>
<tr>
<td>Rural-Urban:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>141</td>
<td>1,338,768</td>
<td>0% (0-0%)</td>
<td>61</td>
<td>565,715</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.5%)</td>
<td></td>
<td></td>
<td>(4.9%)</td>
</tr>
<tr>
<td>Suburban</td>
<td>361</td>
<td>10,886,902</td>
<td>0% (0-0%)</td>
<td>95</td>
<td>2,363,007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(36.8%)</td>
<td></td>
<td></td>
<td>(20.6%)</td>
</tr>
<tr>
<td>Minor City</td>
<td>145</td>
<td>11,483,276</td>
<td>10.2% (0-65.1%)</td>
<td>47</td>
<td>3,076,269</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(38.8%)</td>
<td></td>
<td></td>
<td>(26.8%)</td>
</tr>
<tr>
<td>Major City</td>
<td>58</td>
<td>5,906,450</td>
<td>13.4% (0-68.2%)</td>
<td>31</td>
<td>5,489,788</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(19.9%)</td>
<td></td>
<td></td>
<td>(47.8%)</td>
</tr>
</tbody>
</table>

Table S1 provides population totals stratified by urbanicity, and median 60-minute access to PSCs. We binned counties into 4 categories, using 2013 Rural-Urban Continuum Codes (USDA): rural (<2,500 urban pop.), suburban (2,500+ urban pop. adjacent to a
metro area), minor city (in metro area of <1M pop.), or major city (in metro areas of 1M+ pop.). Wilcoxon rank-sum (Mann-Whitney) tests compare median PSC access in hot-spots and cool-spots within each of the 4 categories.