



# Examining spatial clusters of high & low proportions of late stage cervical cancer in the U.S.: a look at geographic disparities & associated risk factors

Yamisha Rutherford<sup>1</sup>, Lee R. Mobley<sup>1,2</sup>

<sup>1</sup>School of Public Health, Georgia State University, Atlanta, GA, USA; <sup>2</sup>Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA, USA

*Contributions:* (I) Conception and design: All authors; (II) Administrative support: LR Mobley; (III) Provision of study materials or patients: LR Mobley; (IV) Collection and assembly of data: Y Rutherford; (V) Data analysis and interpretation: Y Rutherford; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*Correspondence to:* Yamisha Rutherford, PhD. School of Public Health, Georgia State University, Atlanta, GA 30302, USA.

Email: yrutherford1@student.gsu.edu.

**Background:** Despite known effectiveness of Pap screening in detecting precancerous cells before developing into cancer, 51% of women with cervical cancer (CVC) are diagnosed at a late stage (regional or distant). Previous epidemiological surveillance indicates that the burden of late stage CVC varies widely between states in the US. However, little is known about the spatial clustering of county-level late stage CVC rates across the total US. Examining county as opposed to state-level data will help us to identify highly burdened places that are potentially masked when using aggregated state-level data. In addition, examining spatial clusters as opposed to simple geographic distributions of high and low proportions of late stage CVC will allow us to generate and test hypotheses regarding underlying risk factors that may be common to counties within adjacent states.

**Methods:** This cross-sectional study includes CVC cases diagnosed during the ten-year period from 2005–2014 from the United States Cancer Statistics (USCS) database. Using CVC data among 43 states and their 2,357 constituent counties in the US, we employ Empirical Bayes (EB) LISA tests to identify clusters of counties considered to be high risk “hotspots” and clusters of counties considered to be low risk “coolspots,” during two time periods, pre and post the implementation of the Patient Protection and Affordable Care Act (ACA) in 2010. Using a series of t-tests and co-location mapping, we also assess whether the hotspots identified in both time periods are associated with various contextual and compositional factors and whether geographic hotspots persist over both time periods.

**Results:** The proportion of CVC cases diagnosed at a late stage has increased from 47% to 54% over time. There were also substantial changes in the number and distribution of clusters over time and the distribution of county-specific hotspots were not consistent with the state-level burden of late stage CVC identified in previous literature. Over time, higher concentrations of Black women were associated with hotspot clusters and access to care barriers became primary drivers of clusters of higher proportions of late stage CVC diagnoses.

**Conclusions:** Study results demonstrate that there are both geographic and demographic disparities in late stage CVC. Without further investigation into these relationships we cannot adequately inform late stage CVC interventions in the US. As a result, the overall proportion of late stage CVC in the US is likely to remain high and at a disproportionately higher rate among African American and Hispanic women and women in the identified hotspot clusters across the US.

**Keywords:** Cervical cancer (CVC); late stage; spatial clusters; hotspots; associated factors

Received: 23 October 2019; Accepted: 07 May 2020; Published: 30 June 2020.

doi: 10.21037/ace-19-36

View this article at: <http://dx.doi.org/10.21037/ace-19-36>

## Introduction

Prior to the introduction of Pap testing in the early 1950s, cervical cancer (CVC) was the most common cancer among women in the United States (1). The ability to detect precancerous cervical lesions through Pap test screening has made CVC one of the most preventable of all cancers (2). In fact, due to screening, the incidence of CVC decreased by more than 50% during 1975–2014 (3,4). Since the end of a sharp decline in 2001, however, CVC incidence rates have remained stable (4).

Despite known effectiveness of Pap screening in detecting precancerous cells before developing into cancer, recently available data from the Surveillance, Epidemiology, and End Results (SEER) registry indicate that 51% of women with CVC were diagnosed at a late stage (regional or distant) (5). Stage at diagnosis is a significant public health concern as it has been found to play a leading role in CVC treatment, prognosis and survival (6). Specifically, late stage CVC is associated with increased morbidity and lower 5-year survival rates (6). The 5-year survival rates for distant and regional CVC are as low as 17% and 56%, respectively, compared to a 5-year survival of 91.7% for localized CVC (7).

Previous epidemiological surveillance using 2001–2003 county-level data from SEER and the North American Association of Central Cancer registries indicated that the percent of late stage CVC was highest in Iowa, Connecticut, California, New Jersey and Missouri (8). A later report using 2004–2006 data from cancer registries affiliated with CDC's National Program of Cancer Registries (NPCR) and SEER indicated that late stage CVC incidence rates were highest in Arkansas, the District of Columbia, Illinois, Kentucky, Louisiana, Mississippi, Nevada, New Mexico, and Oklahoma (9). These studies suggest that the incidence and thus the burden of late stage CVC varies widely between states in the US. However, little is known about how the incidence and burden of late stage CVC differs within specific counties across each of the different states. A few studies have examined the county-level distribution of late-stage CVC incidence within a single state (10–13), however there is a lack of literature describing the distribution of late stage CVC at the county-level across the entire or majority

of the US.

There are no current studies specifically assessing the spatial clustering of county-level late stage CVC rates across the entire US. Identification of county-level clusters of higher than average rates of late stage CVC “hotspots” would provide important new information to inform our approaches to reducing the burden of late stage CVC. Examining county as opposed to state-level data will help us to identify highly burdened places that are potentially masked when using aggregated state-level data. In addition, examining spatial clusters as opposed to simple geographic distributions of late stage CVC rates will allow us to generate and test hypotheses regarding underlying risk factors that may be common to counties within adjacent states.

Hotspot analysis is an essential technique used within geographic information system (GIS) studies. However, sophisticated spatial analytic methods such as these have not been frequently applied to identify late stage CVC hotspots or generate hypotheses regarding the factors underlying late stage diagnoses. This is due in part to a lack of knowledge regarding the utility of spatial data, lack of appropriate spatial databases, and previously insufficient spatial analytic software (14). As a result, previous research on this topic largely underestimates the contribution of place; a “social context deeply connected to larger patterns of social advantage and disadvantage.” (15). To our knowledge, this is the first time this approach has been used to understand late stage CVC.

Using spatial cluster methods, our first aim is to examine geographic disparities in late stage CVC diagnosis rates in the US by robustly identifying and characterizing clusters of counties considered to be high risk “hotspots” and clusters of counties considered to be low risk “coolspots,” during two different time periods, both before and after implementation of the preventive services provisions of the Patient Protection and Affordable Care Act (ACA). Our second aim is to determine whether the hotspots identified in both time periods are associated with various contextual and compositional factors. Finally, our third aim is to determine whether there are geographic hotspots that persist over both time periods, as these places may represent those counties in more urgent need of intervention. It is

our hypothesis that there will be spatial clusters of higher than expected late stage CVC diagnosis rates in the US and that clusters will differ across the two time periods. We also believe these clusters will contain geographic factors that can help better explain why late stage CVC rates varies across counties in the US and how risk changes over time.

Ultimately, identifying patterns in late stage CVC rates will allow us to pinpoint the areas in greatest need of intervention, to better allocate intervention resources and evaluate performance of existing prevention and early detection programs. These findings can also be used to inform further research aimed at gaining a better understanding of the underlying causes of late stage CVC. Finally, comparing spatial patterns and risk factors pre- and post-ACA implementation will allow researchers to generate hypotheses regarding the policy's impact and effectiveness as it relates to cancer prevention and control.

We present the following article in accordance with the STROBE reporting checklist (available at <http://dx.doi.org/10.21037/ace-19-36>).

## Methods

### *Study population*

This cross-sectional study includes CVC cases diagnosed during the 10-year period from 2005–2014 from the United States Cancer Statistics (USCS) database, available at the National Center for Health Statistics Research Data Center. The USCS database is a population-based surveillance system of cancer registries with data representing 98% of the US population (16). This database has information on demographics (age, gender, race, and ethnicity), tumor characteristics, and geographic location (county of residence) at time of diagnosis (16). The confidentiality of data with geographic identifiers for county of residence is preserved by restricting access to researchers with approved research plans with analyses conducted inside secure federal Research Data Centers (RDCs) (16). There is no access to the Internet from inside the RDC, and all results must be reviewed before they can be released from the RDC and published (16).

All states participate in the USCS registry system, however, five did not allow use of county of residence information (Kansas, Minnesota, Illinois, Michigan, and Missouri) (16). Therefore, we excluded these five states and two additional states, Alaska and Hawaii, because of missing contextual data. Our final analysis includes 43 states and

their 2,357 constituent counties. The study sample was further restricted to include all diagnosed cases for women whose primary cancer was cervical, and less than 1% of these were excluded due to lack of staging information. This restriction resulted in 120,325 individuals diagnosed with CVC in the US during 2005–2014. Cases were then divided into two 5-year time periods: those diagnosed during 2005–2009 (pre ACA), and those diagnosed during 2010–2014 (post ACA). We further categorized cases into late stage (regional and distant) or early stage (localized, including *in situ*) diagnosis. We then created a county-level late stage diagnosis rate variable for both time periods, which was used to address study aim 1. This variable aggregated the total number of late stage cases within each county by Federal Information Processing Standard (FIPS) code and divided this by the total number of CVC cases.

In addition to the USCS's geographic location, case identification and stage variables, we also extracted two additional variables for use in study aim 2, race or ethnicity and age. We created the race or ethnicity variable by combining USCS's race and Hispanic variables. Race or ethnicity is a recoded variable categorized into six race or ethnicity groups representing the proportion of the total population that was non-Hispanic White, non-Hispanic Black, Hispanic, Asian Pacific, American Indian and other. We also created five age variables representing the percent of the total population that was either 0–40, 40–49, 50–64, 65–74, or 75 years or older. Additional county-level contextual variables needed for aim 2 were extracted from a number of external data sources. Data describing the percentage of the county population living in poverty (2005 and 2010), the percentage of individuals under age 65 with no health insurance (2005 and 2010) and the percent of individuals unemployed (2005 and 2010) were obtained from the U.S. Census Bureau, SAHIE data (17). The proportion of the population age 18–64 that speaks English poorly (2007–2011) and the percent of population who came to the US from a different country in the prior year variables were extracted from the American Community Survey (18). Data describing county level population density (i.e., urbanicity) was extracted from the Economic Resource Services (ERS) agency (19). This measure was calculated by dividing the total population in 2010 by the total square miles of land area. Higher values of this measure indicate more urban places.

The percent of the state population insured by employers in private self-insured health plans in 2010 was obtained from the Agency for Healthcare Research and

Quality (AHRQ) (20). Data describing the percent of HMO penetration in 2010 was extracted from Kaiser (21). Finally, the number of federally qualified health centers (FQHCs) in 2005 and in 2010 were extracted from the Guttmacher Institute's State- and County-level Family Planning Clinic dataset. We recoded the 2005 and 2010 FQHC count variables into rates per capita by multiplying each of the count variables by 100,000 and dividing the product by the total US population in 2005 and 2010 (22).

### *Statistical analysis*

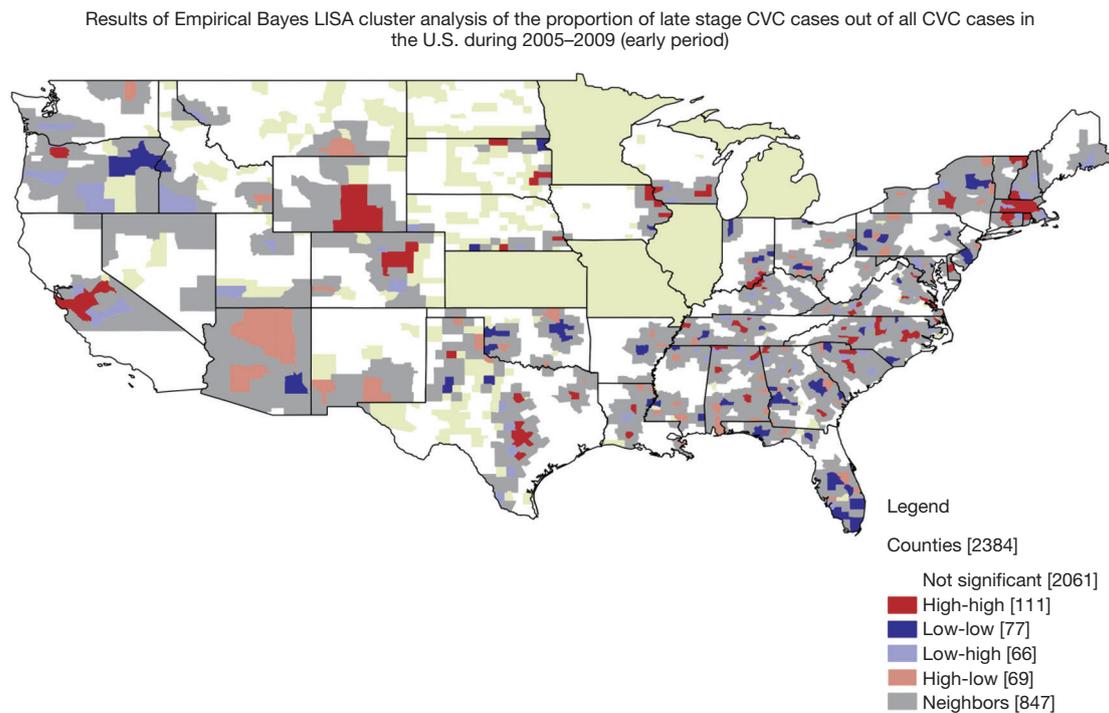
Using Moran's I statistics computed using GeoDa software, we robustly identify clusters of counties considered to be high risk "hotspots" and clusters of counties considered to be low risk "coolspots" across the US during two time periods (pre- and post-ACA implementation), aim 1. Moran's I statistics are computed based on an underlying assumption of constant variance among rates, which can be violated when county population sizes significantly vary. Therefore, we first assess whether this assumption holds or is violated by developing histograms of the distribution of county populations used as the denominators in constructing the rates during both time periods and comparing them to a normal curve. Histograms show skewness in the distribution of the late stage CVC rates during both time periods, suggesting potential for variance instability due to the fact that the underlying populations at risk (all women with CVC) vary in size across counties. Such variance instability in the rates can lead to spurious inferences for global and local Moran's I (23,24). To correct for variance instability among late stage CVC rates, we use Empirical Bayes (EB) standardization techniques to compute global and local spatial autocorrelation statistics. This method is known for adjusting these statistics for small sample sizes, reducing the variability of estimates, removing erroneously suggested spatial outliers and thus computing robust and reliable clusters (23,24). To accomplish EB standardization GeoDa computes spatial autocorrelation for transformed standardized random variables (23). To get these transformed standardized random variables GeoDa replaces the original crude rates with new standardized rates that have a mean of zero and standard deviation of one (23). Thus, the EB standardization method directly standardizes or rescales crude rates to account for instability in variance (23).

Global spatial autocorrelation was determined by performing the EB-adjusted global Moran's I spatial

clustering test, which produces a EB Moran's I coefficient test statistic. Given a statistically significant EB Moran's I coefficient, we reject the null hypothesis of spatial randomness and conclude that there is global clustering in the patterns of late stage CVC rates across counties. After confirming that there was global clustering we calculated EB-adjusted Local Indicators of Spatial Association (LISA) to pinpoint the specific locations of the statistically significant clusters within both time periods. At significance level  $<0.01$ , the EB LISA test first calculates a test statistic for each county representing whether the county has a statistically significant higher or lower than the national average rate of late stage CVC.

To determine statistical significance of EB LISA test statistics, GeoDA uses a permutations approach called bootstrapping. This approach compares the actual correlation between late stage CVC measures among a county and its neighbors with 1,000 or more correlations between the county in question and groups of randomly chosen neighbors. Queen contingency matrix weights are used to define neighboring counties. A statistical distribution is generated by the more than 1,000 permuted repetitions with the random neighbors and is assessed to determine where along the distribution the actual correlation falls. If the actual correlation with neighbors falls in the tail of the distribution then we reject the null hypothesis of local spatial randomness and conclude that the county's correlation with actual neighbors is statistically significantly unlikely to have occurred by chance. This assessment is repeated independently for each county in the dataset, and the collection of test statistic findings for all counties are mapped together in a single LISA clustering map.

Using statistically significant EB LISA test statistics, four distinct cluster types are formed in both time periods: high-high, low-low, low-high and high-low. High-high clusters include counties with higher than average rates surrounded by other counties with higher than average rates. Similarly, low-low clusters include counties with lower than average rates surrounded by other counties with lower than average rates. Low-high and high-low clusters are developed in a similar fashion. Among all cluster types, those that were statistically significant were presented in two separate maps, one for each time period, using QGIS software (25). To represent the entire cluster both maps included the counties at the center of the cluster and their surrounding neighbors. To determine whether the hotspot clusters identified during both time periods were associated with various contextual



**Figure 1** Results of Empirical Bayes LISA cluster analysis of the proportion of late stage CVC cases out of all CVC cases in the U.S. during 2005–2009 (early period). LISA, Local Indicators of Spatial Association; CVC, cervical cancer.

and compositional factors, aim 2, we employed two sets of independent sample  $t$ -tests in SAS version 9.0. Specifically, for both time periods, we grouped all high-high clusters together and all low-low clusters together and treated them as two independent groups. We then tested for statistically significant differences in the means of the underlying biological and contextual factors between the two cluster groups, at significance level 0.05. We also carried out the Benjamini-Hochberg procedure to address the issue of multiple comparisons. However, at a false discovery rate of 0.10, we found that the results were consistent with the results of the  $t$ -test.

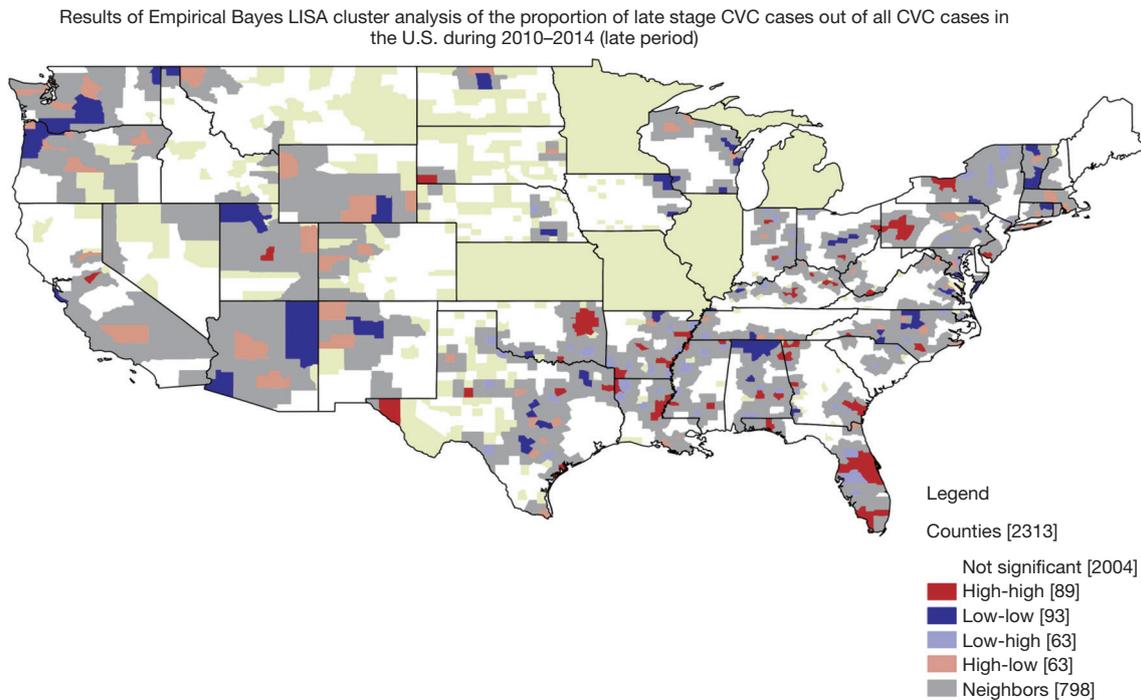
Finally, to determine whether there were geographic hotspots that persisted over both time periods, aim 3, we developed a colocation map. To develop the colocation map we grouped counties into three categories: those that belonged to significant hotspot clusters in both periods (persistently hotspots), those that did not belong to significant hotspot clusters in either period (persistently non-hotspots), and those that transitioned into or out of a hotspot clusters (transitional hotspots). The colocation of these three categories across the two time periods was then mapped using QGIS software.

## Results

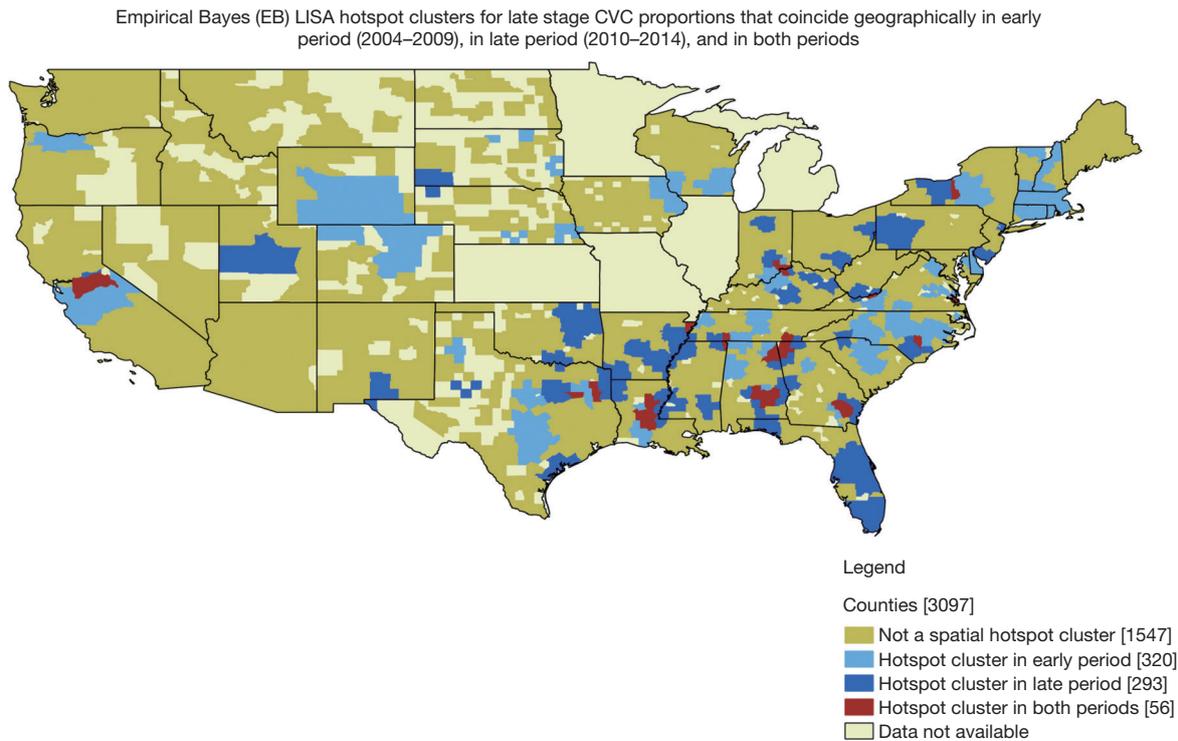
### *LISA cluster and persistent hotspot results (Figures 1-3)*

EB adjusted Global Moran's  $I$  tests indicate that there is significant positive spatial autocorrelation among the proportions of late stage CVC during both time periods, (significance level  $\alpha=0.01$ ). Thus, for both time periods, we reject the null hypothesis of spatial randomness and conclude that the proportions of late stage CVC across neighboring counties were too similar in some local areas to have occurred by chance.

Using EB adjusted Local Moran's  $I$  tests, we further determined which local areas were statistically significantly spatially correlated with one another with regards to late stage CVC proportions- location of local clusters. During both time periods, we found several statistically significant local high and low rate cluster centers. High-rate clusters centers are areas where counties and their neighbors have statistically significantly higher proportions of late stage CVC than would be observed by chance, using a 5% level of significance. These clusters will be referred to as "hotspots" going forward. Low-rate clusters centers are areas where counties and their neighbors have statistically significantly



**Figure 2** Results of Empirical Bayes LISA cluster analysis of the proportion of late stage CVC cases out of all CVC cases in the U.S. during 2010–2014 (late period). LISA, Local Indicators of Spatial Association; CVC, cervical cancer.



**Figure 3** Empirical Bayes (EB) LISA hotspot clusters for late stage CVC proportions that coincide geographically in early period (2004–2009), in late period (2010–2014), and in both periods. LISA, Local Indicators of Spatial Association; CVC, cervical cancer.

**Table 1** *T*-test comparing the mean of contextual and demographic variables between EB adjusted hotspot and coolspot clusters during 2005–2009

Variable description	Mean in hotspots (N=385)	Mean in coolspots (N=339)	P value, for <i>t</i> -test of differences in means
Contextual characteristics of counties of residence			
Percent underserved by a primary care provider, 2005	46.9216	45.0018	<0.0001
Poor English speaking among 18–64 years old (proportion)	0.1291	0.1448	0.0347
Percent of people of all Ages in poverty for Income year 2005	3.58	3.36	0.5678
Percent insured by employers in self-insured plans exempt from state regulations 2006	26.78	27.53	0.5979
Percent unemployed 2005	9.32	12.01	0.0075
Percent HMO Penetration 2005	5.0935	5.7215	<0.0001
Percent of total pop <65 uninsured 2005	9.38	10.72	0.1462
Percent of population that moved from different country last year	0.3209	0.2984	0.0030
Population Density 2005 (urbanicity)	349.4	228.5	0.1051
Sample population demographic characteristics			
Percent under age 50	55.93	50.14	0.0003
Percent American Indian	9.74	9.62	0.2675
Percent Black	12.16	17.80	0.0005
Percent Asian	1.58	0.751	0.0025
Percent White	76.84	71.81	0.0082
Percent Hispanic	2.29	2.62	0.4992

lower proportions of late stage CVC than would be observed by chance, using a 5% level of significance. These clusters will be referred to as “coolspots” going forward.

During 2005–2009, we found 111 statistically significant hotspots (colored red) and 77 statistically significant coolspots (colored blue). Hotspots were observed in 24 of 43 states but were most apparent throughout the Eastern and Southern regions of the US, as well as California, Colorado, Connecticut, Massachusetts and Wyoming. Coolspots were observed in 23 of 43 states but were most apparent throughout the Eastern and Southern regions of the US, as well as Oregon, Florida, Georgia and Oklahoma (*Figure 1*).

During 2010–2014, we found 89 statistically significant hotspots (colored red) and 93 statistically significant coolspots (colored blue). Hotspots were observed in 19 of 43 states but were most apparent in Florida, Pennsylvania, Oklahoma, New York and other states within the Southern and Eastern regions of the US. Coolspots were observed in 26 of 43 states but were most apparent in Georgia, Arizona,

Utah, Oregon, Washington and other states within the Eastern region of the US (*Figure 2*).

Over time the number of statistically significant hotspot clusters decreased while the number of statistically significant coolspot clusters increased. However, colocation mapping shows that there were 56 hotspot clusters that persisted over time. Persistent hotspot clusters were observed in 13 of 43 states and were most apparent in California, Louisiana, Alabama and Georgia (*Figure 3*).

#### ***Comparison of high and low rate clusters: t-test results (Tables 1,2)***

To determine what factors were associated with hotspot clusters, we tested for significant differences in the means of the underlying compositional and contextual variables between the hotspot and coolspot clusters observed in both time periods. During the early period, we found that hotspot clusters had a statistically significantly higher proportion of women who were White, Asian and individuals less than

**Table 2** T-test comparing the mean of contextual and demographic variables between EB adjusted hotspot and coolspot clusters during 2010–2014

Variable description	Mean in Hotspots (N=350)	Mean in coolspots (N=335)	P value, for t-test of differences in means
Contextual characteristics of counties of residence			
Percent underserved by a primary care provider, 2012	13.0312	10.9681	<0.0001
Poor English speaking among 18–64 years old (proportion)	0.0191	0.0216	0.5614
Percent of people of all Ages in poverty for income year 2010	18.61	14.50	<0.0001
Percent insured by employers in self-insured plans exempt from state regulations 2013	62.3106	60.5290	0.0003
Percent unemployed 2010	10.4811	9.2087	<0.0001
Percent HMO penetration 2010	13.5544	16.5991	<0.0001
Percent of total pop <65 uninsured 2010	20.1954	17.3681	<0.0001
Percent of the population that moved from a different country	0.297	0.379	0.0098
Population density 2010 (urbanicity)	451.3	336.7	0.5989
Sample population demographic characteristics			
Percent under age 50	45.55	53.51	<0.0001
Percent American Indian	1.77	2.67	0.2345
Percent Black	16.78	8.38	<0.0001
Percent Asian	1.12	1.91	0.0150
Percent White	72.59	76.76	0.0394
Percent Hispanic	7.04	9.45	0.0504

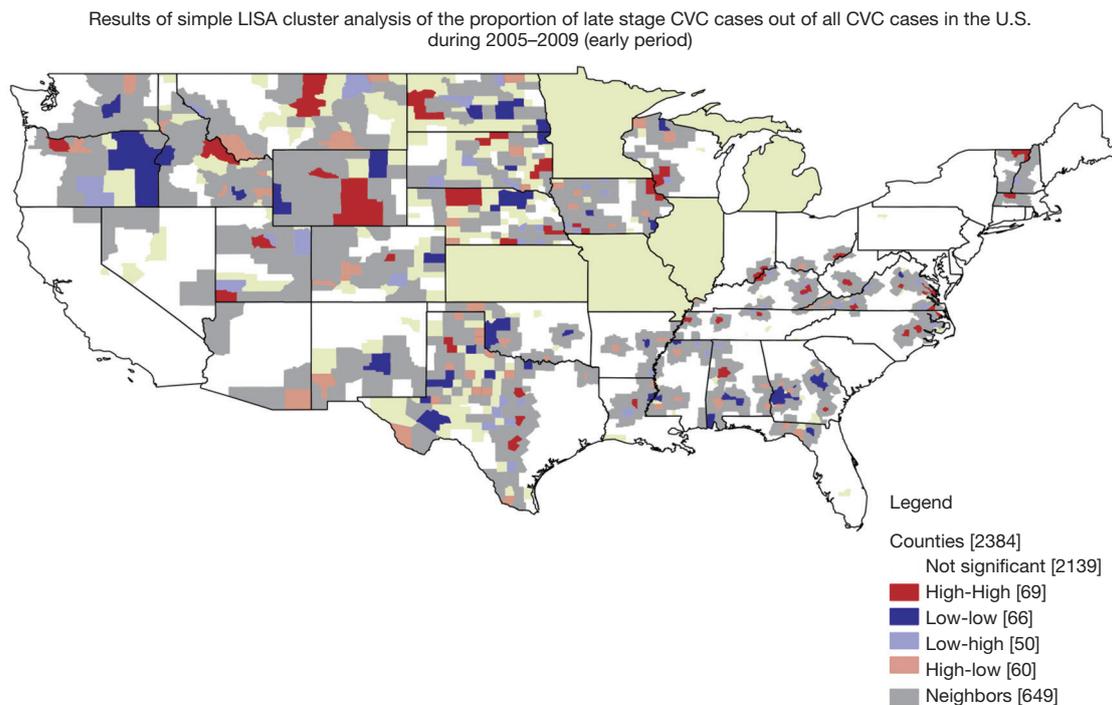
age 50 among the CVC sample population, compared to coolspot clusters. Hotspot clusters in the early period also had a statistically significantly higher proportion of counties that were underserved by a primary care provider, compared to coolspot clusters. In addition, compared to coolspot clusters, hotspot clusters had statistically significantly lower proportions of African American women among the CVC sample population, and lower proportions of unemployed persons and 18–64 years old that spoke English poorly in the general population. Finally, we found that hotspot clusters in the early period had a statistically significantly lower proportion of HMO penetration (*Table 1*).

When comparing hotspot and coolspot clusters observed during the later period we found drastically different associations than what was found when comparing cluster groups observed during the early period. During the later period, we found that hotspot clusters had a statistically significantly higher proportion of people that were in poverty, unemployed, and in self-insured insurance plans

that were exempt from state regulations. Compared to coolspot clusters, hotspot clusters also had a statistically significantly higher proportion of African American women in the CVC sample women and a higher proportion of counties that were underserved by a primary care provider. We also found that hotspot clusters had a statistically significantly lower proportion of White and Asian women as well as individuals less than age 50 in the CVC sample population, compared to coolspot clusters. During the later period, the percent of HMO penetration was also statistically significantly lower in hotspot clusters compared to coolspot clusters (*Table 2*).

## Discussion

Although the number of CVC cases has decreased over time, the proportion of CVC cases diagnosed at a late stage has increased from 47% to 54% overall. This highlights the need to identify former, existing and persisting clusters



**Figure 4** Results of simple LISA cluster analysis of the proportion of late stage CVC cases out of all CVC cases in the U.S. during 2005–2009 (early period). LISA, Local Indicators of Spatial Association; CVC, cervical cancer.

of high proportions of late stage CVC and to determine what factors are associated with these clusters. The results of this study are essential for pinpointing areas in need of intervention and generating hypothesis regarding the causes of late stage CVC.

The LISA clustering method is a sophisticated spatial method used to identify areas in need of intervention by pinpointing areas of local clustering of rates. This method has been used to identify high risk areas in a number of studies (26,27). However, this method assumes constant variance in the rates across the areas, which was not the case for our study measure. This was due to there being small counts in both the numerator (CVC cases diagnosed at a late stage) and denominator (all CVC cases) of the proportion of late stage CVC variable. To ensure that this did not bias the clustering results we employed a more robust LISA technique called EB adjusted LISA. This method has never been used to identify clusters of high proportions of late stage CVC. However, there were significant differences in the location of clusters when EB and traditional LISA methods were used, which emphasize the importance of adjusting for variance instability in order to properly identify clusters. For example, using traditional LISA

methods there were hotspot clusters observed in Montana and North Dakota during the early period (*Figure 4*). However, using EB adjusted LISA methods there were no hotspot clusters observed in either Montana or North Dakota during the early period (*Figure 1*). This suggests that the late stage CVC LISA results were overestimated when traditional LISA techniques were used. Therefore, we take clusters observed using EB adjusted LISA to be most reliable and robust.

Using EB adjusted LISA, we found that there were substantial changes in the number and distribution of clusters over time and that the distribution of county-level hotspots were not consistent with the state-level burden of late stage CVC identified in previous literature. Maps of EB adjusted LISA clusters show that the overall number of hotspots decreased from 111 to 89 over time while the number of coolspots increased from 77 to 93 over time. Maps also show that hotspots observed in Massachusetts, Connecticut, Wyoming and Colorado during the early period were no longer observed in the later period. We also found that over time local areas in both Utah and Arizona developed coolspots. On the other hand, some places such as Florida, Oklahoma and Pennsylvania developed hotspots

over time. There are also areas in California, Louisiana, Alabama and Georgia that presented hotspots during both time periods.

There are several implications that can be drawn from the changes in the distribution of clusters over time. Local clustering of proportions into significant hot and coolspots during both the early and late period suggests that there were and still are geographic disparities in the proportion of late stage CVC across counties and states. It can also be implied that places that developed coolspots or lost hotspots over time may have implemented effective CVC interventions or early detection programs that worked to attenuate the geographic disparities that were once present. On the other hand, it can be implied that places with newly developed hotspots likely represent those places where there was a release in the pent-up demand for Pap testing services over time. During the early period (2005–2009), there was a pent-up demand for Pap testing across the US due to a number of women having limited or no health insurance coverage. However, during the later period (2010–2014), millions gained access to Pap testing services via ACA provisions that mandated full coverage for preventative services in 2010 (28) and expanded Medicaid in 2014 (29). Thus, newly developed hotspots are likely driven by a higher number of women, who had never been screened and whose CVC was predominately asymptomatic, being screened for CVC and in turn a higher number of late-stage diagnoses over time. Lastly, places displaying persist hotspots over time such as those observed in California, Texas and Southeast regions of the US represent those places in greatest need of interventions.

Implications can also be drawn from the comparisons of the differences in the means of contextual and compositional variables between hotspots and coolspots. The early period *t*-test results indicate that the percent of sample women with CVC who were Asian and White and women under the age of 50 was significantly higher in hotspot clusters compared to coolspot clusters. However, this association changed over time. During the later period, the percent of sample women with CVC who were African American and women over the age of 50 became significantly higher and the percent of sample women with CVC who were White or Asian became significantly lower in hotspot clusters compared to coolspot clusters. These findings are consistent with the current literature which suggests that African Americans, Hispanics and women over 50 are now among those disproportionately burdened by late stage CVC (9,30). Additional research is needed to understand why African

American women and women over 50 are at greatest risk for late stage CVC. Similarly, additional research is also needed to understand what preventive programs, policies of behavior changes are associated with decreased risk of late stage CVC among Asian and White women over time. This information could be useful for developing intervention strategies among other race or ethnic populations.

In addition to the associated compositional factors, we also found that several factors that were not associated with hotspot clusters in the early period showed a significant relationship with hotspot clusters over time. Specifically, we found that the percent uninsured, unemployed, in poverty and insured by employers in self-insured plans exempt from state regulations became statistically significantly higher in hotspot clusters compared to coolspot clusters over time. These results suggest that over time the proportion of late stage CVC became more strongly influenced by barriers to access to care, as a higher percent of unemployed, uninsured, those in poverty, individuals insured by employers in self-insured plans exempt from state regulations, and individuals underserved by a primary care provider, each represent access to care barriers.

Although coverage and cost sharing provisions were implemented under ACA between 2012–2014, we were not surprised to find that several access to care barriers were associated with higher proportions of late stage CVC during the 2010–2014 time period. The ACA has made significant strides toward improving access to Pap testing, however, it has not eliminated access barriers for all women. Following the implementation of the ACA, there were still 44.4 million individuals left uninsured and thus potentially facing issues regarding access and affordability of care (31). Furthermore, the shortage of primary care physicians (PCP) is projected to increase to as many as 49,300 PCPs by the year 2030 (32). This presents a potential barrier to access to healthcare for women in the US, including those with insurance, as the US is already experiencing a PCP shortage. Together these unresolved issues of affordability and availability of care can significantly shape access to Pap testing and thus stage at diagnosis. This demonstrates the need to further develop strategies to combat the issue of access to care, as the protective effects of the ACA are limited.

Although the current study offers a significant contribution to the literature on CVC diagnoses, study results are limited in two major ways. First, each of the LISA tests were performed using a limited sample of the U.S. population ( $n=43$  states). Carrying out LISA cluster

analyses using data that excludes seven states imposes limitations on all study results as the clustering of counties into distinct cluster groups (i.e., hotspots or coolspots) is based on whether or not a county is surrounded by neighboring counties with similar rates. Therefore, it is likely that the distribution of clusters across the US, using all summary measures, would be different if the counties of the remaining seven states were included and assessed relative to their neighboring counties. Second, although the current study applied the spatial autocorrelation-based approach there are several other spatial clustering methods that could have been applied to identify spatial clusters. The spatial autocorrelation-based approach was applied over other spatial clustering methods because of its many advantages (24,32). However, the largest benefit of using spatial autocorrelation methods is that clustering results are derived using several underlying statistical techniques (32). These statistical techniques are computationally efficient and equipped for solving large statistical problems (32) such as variance instability using a sophisticated standardization process (24). Despite these advantages, this method is limited by the issue of “multiple comparisons” (32). This issue occurs when testing more than one local statistic for significant (32). When testing more than one local statistic for significant, the correlation among tests that are near one another in space can bias both results and interpretations (32). Like the spatial autocorrelation method, other spatial clustering methods such as non-hierarchical, hierarchical and scan-based spatial approaches also demonstrate several advantages and drawbacks (32). However, it is important to note that each approach seeks to find clusters in a different way and can therefore result in different cluster patterns (32).

## Conclusions

Together these results demonstrate that there are both geographic and demographic disparities in late stage CVC. Study results also suggest that late stage CVC incidence and geographic disparities are likely influenced by county- and state-level factors, as clusters vary across counties and states. Results further demonstrate that the county-level factors associated with the current burden of late stage CVC are all indicators of access to care. These indicators include employment status, insurance coverage, poverty level, primary care shortage, HMO penetration and insurance plan exemptions from state-based health

regulations. Advanced inferential statistics are needed to further investigate the relationships between various county- and state-level access to care barriers and late stage CVC incidence and disparities. More specifically, these relationships should be further investigated using mixed modeling methods which consider the hierarchical structure of the data. Using this approach the researcher can simultaneously examine the effects of county and state level variables and the interactions within and between them.

Without further investigation into these relationships we cannot adequately inform late stage CVC interventions in the US. As a result, the overall proportion of late stage CVC in the US is likely to remain high and at a disproportionately higher rate among African American and Hispanic women and women in the identified hotspot clusters across the US. Stage at diagnosis is of significant public health concern as it plays a leading role in CVC treatment, prognosis and survival (6). In fact, the 5-year survival rate for distant and regional CVC is as low as 17% and 56% respectively, compared to a 5-year survival of 91.7% for localized CVC (7). The results of the current study will help to reduce the number of late CVC cases and associated mortality by informing further research aiming to gain a better understanding of the underlying causes of late stage CVC. This study also pinpoints areas in greatest need of late stage CVC interventions, by identifying geographic hotspots that persist over both time periods, as seen in states including California, Louisiana, Alabama and Georgia.

## Acknowledgments

*Funding:* This study was funded by a National Cancer Institute grant (2R01CA126858). CDC's National Program of Cancer Registries also contributed funds to cover the standard RDC fees for researchers conducting analyses under approved research projects. The content is solely the responsibility of the authors and does not necessarily represent the official views of Georgia State University, the National Center for Health Statistics, the National Cancer Institute, or the National Institutes of Health.

## Footnote

*Provenance and Peer Review:* This article was commissioned by the Guest Editors (Peter Baade and Susanna Cramb) for the series “Spatial Patterns in Cancer Epidemiology” published in *Annals of Cancer Epidemiology*. The article has undergone external peer review.

*Reporting Checklist:* The authors have completed the STROBE reporting checklist. Available at <http://dx.doi.org/10.21037/ace-19-36>

*Conflicts of Interest:* Both authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/ace-19-36>). The series “Spatial Patterns in Cancer Epidemiology” was commissioned by the editorial office without any funding or sponsorship. The authors have no other conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

- Shaw P. The History of Cervical Screening I: The Pap Test. *J Soc Obstet Gynaecol Can* 2000;22:110-14
- Sawaya GF, McConnell KJ, Kulasingam SL, et al. Risk of cervical cancer associated with extending the interval between cervical-cancer screenings. *N Engl J Med* 2003;349:1501-9.
- Adegoke O, Kulasingam S, Virnig B. Cervical Cancer Trends in the United States: A 35-Year Population-Based Analysis. *J Womens Health (Larchmt)* 2012;21:1031-7.
- American Cancer Society Cancer Action Network. Screening Leads to Cervical Cancer Decline in the United States. Accessed December 15, 2018. Available online: <https://www.fightcancer.org/sites/default/files/FINAL%20-%20Cervical%20Cancer%20General%20Factsheet%2003.06.18.pdf>
- SEER. <https://seer.cancer.gov/statfacts/html/cervix.html> SEER Stat Fact Sheets: Cervix Uteri Cancer. Cancer Stat 2014. Available online: <https://seer.cancer.gov/statfacts/html/cervix.html>
- Survival Rates for Cervical Cancer, by Stage. American Cancer Society Website. Available online: <https://www.cancer.org/cancer/cervical-cancer/detection-diagnosis-staging/survival.html>. Updated November 16, 2016. Accessed October 14, 2017.
- SEER. SEER Stat Fact Sheets: Cervix Uteri Cancer. Cancer Stat 2014. Available online: <https://seer.cancer.gov/statfacts/html/cervix.html>
- Horner MJ, Altekruse SF, Zou Z, et al. U.S. Geographic Distribution of Prevalence Era Cervical Cancer Screening, Incidence, Stage, and Mortality. *Cancer Epidemiol Biomarkers Prev* 2011;20:591-9.
- Henley SJ, King JB, German RR, et al. Surveillance of screening-detected cancers (colon and rectum, breast, and cervix) - United States, 2004-2006. *MMWR Surveill Summ* 2010;59:1-25.
- Saghari S, Soret S, Ghamsary M, et al. Geographic Distribution of Cervical Cancer in California: A Population Based Study. *JSM Women's Health* 2016;1:1001.
- Roche LM, Niu X, Henry KA. Invasive Cervical Cancer Incidence Disparities in New Jersey—a Spatial Analysis in a High Incidence State. *J Health Care Poor Underserved* 2015;26:1173-85.
- Zhan FB, Lin Y. Racial/Ethnic, Socioeconomic, and Geographic Disparities of Cervical Cancer Advanced-Stage Diagnosis in Texas. *Womens Health Issues* 2014;24:519-27.
- Polednak AP. Trends in Late-Stage Breast and Cervical Cancer Incidence Rates in Connecticut (United States). *Cancer Causes Control* 2003;14:361-5.
- Lacey L. Cancer prevention and early detection strategies for reaching underserved urban, low-income black women. Barriers and objectives. *Cancer* 1993;72:1078-83.
- Janelle D, Hodge D, editors. Information, place, and cyberspace: Issues in accessibility. Berlin: Springer-Verlag, 2000.
- Centers for Disease Control and Prevention (CDC). National Center For Health Statistics. U.S. Cancer Statistics; 2018. Available online October 2018. Available online: <http://www.cdc.gov/rdc/b1datatype/dt131.htm>. Accessed 1.9.2019. US Census.
- U.S. Census Bureau. Data. Accessed October 2018. Updated n.d. Available online: <https://www.census.gov/data.html>
- United States Census Bureau. American Community Survey Data. Accessed November 2018. Updated October 11, 2018 Available online: <https://www.census.gov/programs-surveys/acs/data.html>
- U.S. Department of Agriculture Economic Research

- Service. Data Download. Accessed October 2018. Updated May 18, 2017. Available online: <https://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data/>
20. Agency for Health Research and Quality. Data. Accessed October 2018. Updated n.d. Available online: <https://www.ahrq.gov/data/index.html>
  21. Kaiser Family Foundation. State HMO penetration Rates. Accessed May 2017.
  22. Frost J, Frohwirth L, Blades N, et al. Contraceptive Needs and Services, 2010, New York: Guttmacher Institute, 2013. Available online: <https://www.guttmacher.org/report/publicly-funded-contraceptive-services-us-clinics-2015>
  23. Anselin L. An Introduction to Spatial Data Analysis-Global Spatial Autocorrelation (2). GeoDa Available online: [http://geodacenter.github.io/workbook/5b\\_global\\_adv/lab5b.html](http://geodacenter.github.io/workbook/5b_global_adv/lab5b.html). Updated March 6, 2016. Accessed February 10, 2019.
  24. Anselin L. Exploring Spatial Data with GeoDA: A Workbook 2005 Center for Spatially Integrated Social Science. Available online: <https://s3.amazonaws.com/geoda/software/docs/geodaworkbook.pdf>. Accessed February 10, 2019.
  25. QGIS Development Team (YEAR). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available online: <http://qgis.osgeo.org>
  26. Sasson C, Cudnik MT, Nassel A, et al. Identifying High-risk Geographic Areas for Cardiac Arrest Using Three Methods for Cluster Analysis. *Acad Emerg Med* 2012;19:139-46.
  27. Schieb LJ, Mobley LR, George M, et al. Tracking Stroke Hospitalization Clusters Over Time and Associations With County-Level Socioeconomic and Healthcare Characteristics. *Stroke* 2013;44:146-52.
  28. Uberoi N, Finegold K, Gee E. Health Insurance Coverage and the Affordable Care Act, 2010-2016. Department of Health and Human Services, ASPE Issue Brief 2016. Available online: <https://aspe.hhs.gov/system/files/pdf/187551/ACA2010-2016.pdf>
  29. Henry J Kaiser Family Foundation. Preventive Services Covered by Private Health Plans under the Affordable Care Act. (2015). Available online: <https://www.kff.org/health-reform/fact-sheet/preventive-services-covered-by-private-health-plans/>
  30. Virnig BA, Baxter NN, Habermann EB, et al. A matter of race: early-versus late-stage cancer diagnosis: African Americans receive their cancer diagnoses at more advanced stages of the disease than whites do. *Health Aff (Millwood)* 2009;28:160-8.
  31. Kaiser Family Foundation analysis of 2017 American Community Survey (ACS), 1-Year Estimates. Available online: <http://files.kff.org/attachment//fact-sheet-key-facts-about-the-uninsured-population>.
  32. Grubestic T, Wei R, Murray A. Spatial Clustering Overview and Comparison: Accuracy, Sensitivity, and Computational Expense. *Annals of the Association of American Geographers* 2014;104:1134-55.

doi: 10.21037/ace-19-36

**Cite this article as:** Rutherford Y, Mobley LR. Examining spatial clusters of high & low proportions of late stage cervical cancer in the U.S.: a look at geographic disparities & associated risk factors. *Ann Cancer Epidemiol* 2020;4:5.