



# HHS Public Access

Author manuscript

*Sex Transm Dis.* Author manuscript; available in PMC 2017 June 08.

Published in final edited form as:

*Sex Transm Dis.* 2016 April ; 43(4): 216–221. doi:10.1097/OLQ.0000000000000421.

## Influence of detection method and study area scale on syphilis cluster identification in North Carolina

Veronica Escamilla, PhD<sup>1</sup>, Kristen H. Hampton, MSPH<sup>2</sup>, Dionne C. Gesink, PhD<sup>3</sup>, Marc L. Serre, PhD<sup>4</sup>, Michael Emch, PhD<sup>2,5</sup>, Peter A. Leone, MD, MPH<sup>2,6</sup>, Erika Samoff, PhD, MPH<sup>7</sup>, and William C. Miller, MD, PhD, MPH<sup>2,6</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, University of Chicago; Chicago, IL USA

<sup>2</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina; Chapel Hill, NC USA

<sup>3</sup>Dalla Lana School of Public Health, University of Toronto; Toronto, ON, Canada

<sup>4</sup>Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina; Chapel Hill, NC USA

<sup>5</sup>Department of Geography, University of North Carolina; Chapel Hill, NC USA

<sup>6</sup>Division of Infectious Diseases, Department of Medicine, School of Medicine, University of North Carolina; Chapel Hill, NC, USA

<sup>7</sup>Division of Public Health, North Carolina Department of Health & Human Services; Raleigh, NC USA

### Abstract

**Background**—Identifying geographical clusters of sexually transmitted infections can aid in targeting prevention and control efforts. However, detectable clusters can vary between detection methods because of different underlying assumptions. Furthermore, because disease burden is not geographically homogenous, the reference population is sensitive to the study area scale, affecting cluster outcomes. We investigated the influence of cluster detection method and geographical scale on syphilis cluster detection in Mecklenburg County, North Carolina.

**Methods**—We analyzed primary and secondary (P&S) syphilis cases reported in North Carolina (2003–2010). Primary and secondary syphilis incidence rates were estimated using census tract-level population estimates. We used two cluster detection methods: local Moran's I using an areal adjacency matrix, and Kulldorff's spatial scan statistic using a variable size moving circular window. We evaluated three study area scales: North Carolina, Piedmont region, and Mecklenburg County. We focused our investigation on Mecklenburg, an urban county with historically high syphilis rates.

---

Correspondence to: Veronica Escamilla, University of Chicago, 5841 S. Maryland Ave, MC 2050, Chicago, IL 60637, Telephone: 773-834-8534, Fax: 773-702-5161, vescamilla@bsd.uchicago.edu.

**Conflict of interest statement:** None declared.

**Results**—Syphilis clusters detected using local Moran’s I and Kulldorff’s scan statistic overlapped but varied in size and composition. Because we reduced the scale to a high incidence urban area, the reference syphilis rate increased, leading to the identification of smaller clusters with higher incidence. Cluster demographic characteristics differed when the study area was reduced to a high incidence urban county.

**Conclusion**—Our results underscore the importance of selecting the correct scale for analysis to more precisely identify areas with high disease burden. A more complete understanding of high burden cluster location can inform resource allocation for geographically targeted sexually transmitted infection interventions.

### Keywords

syphilis; cluster detection; study area scale; local Moran’s I; Kulldorff’s scan statistic

---

## INTRODUCTION

High rates of sexually transmitted infections (STIs) have been observed in small, definable geographical regions<sup>1–3</sup>. Urban STI patients tend to select sexual partners near the patients’ residential locations<sup>4, 5</sup>. Geographical clusters of high STI rates are hypothesized to correspond to areas of high transmission risk, due in part to the selection of sexual partners locally<sup>6</sup>. Identifying geographical clusters of high STI rates has many potential uses, from providing insight into transmission patterns, to the more efficient allocation of resources through the targeting of STI prevention and control programs<sup>1, 7–9</sup>. However, multiple methods for cluster detection have been proposed<sup>10–13</sup>, and STI cluster studies that have similar objectives often use different methods to identify clusters<sup>1, 8, 14–15</sup>. Better understanding of the performance and assumptions inherent in different cluster detection methods is needed to help public health researchers interpret results and accurately target areas for control measures.

Cluster detection depends on the approach taken to identify the cluster. By definition, different cluster detection methods identify different aspects of the spatial pattern, such that different methods may yield varied disease clusters using the same dataset<sup>10, 16–19</sup>. The applicability of findings from STI cluster investigations depends significantly on the selections of cluster detection method and the geographical scale of the study area, such as a state or a county boundary<sup>20</sup>.

Cluster detection methods vary in their underlying assumptions and sensitivity to different aspects of the spatial pattern<sup>17</sup>. For example, two of the most commonly used methods in STI investigations, due in part to freely available implementation software, are Kulldorff’s scan statistic using the SatScan software<sup>1, 7</sup> and the local Moran’s I test for spatial dependency<sup>8</sup>. The scan statistic employs a circular or elliptical scanning window to identify local clusters of exceedingly high STI rates. The local Moran’s I identifies clusters of adjacent areal units (e.g. census tracts) with high STI rates. Both methods are powerful tools that identify high incidence areas relative to the population, and affect the size and shape of the identified clusters using different definitions of clustering. Consideration of method

assumptions is critical when making inferences from cluster detection studies, as the resulting cluster maps can be used to inform public health efforts<sup>9,18</sup>.

The geographical scale of the selected study area also has important implications in spatial investigations. Just as study population selection must be considered in traditional epidemiological studies, the study area scale in cluster analyses affects the reference population against which the null hypothesis of the statistical test is compared<sup>17</sup>. For example, within a county boundary, the null hypothesis for the scan statistic is that rates are geographically uniform across the county, while within a state boundary, the null hypothesis is that rates are uniform across the state. In North Carolina, STI clusters identified in a low prevalence mountain region were no longer detectable when the study area comprised the entire state including high prevalence regions (resulting in a higher reference rate)<sup>7</sup>. The observation that clusters from high prevalence regions dominate the cluster detection process led us to examine more formally the influence of varying study area scale and cluster detection method.

The objective of this study was to assess the impact of cluster detection method and study area scale on the identification of high syphilis burden areas in Mecklenburg County, an urban county in North Carolina. We compared the performance of Kulldorff's spatial scan statistic and the local Moran's I in identifying census tract clusters of high syphilis rates over an 8-year period. We also compared the influence of geographical scale on cluster size and composition using three study area scales: the state of North Carolina, Piedmont region, and Mecklenburg County (Figure 1). We examined how the use of these varied scales influenced clusters identified within Mecklenburg County. We focused our investigation on Mecklenburg County, the most populated county in North Carolina with historically high STI rates, because STI geographical clusters have predominantly been identified in urban areas<sup>21</sup>.

## MATERIALS AND METHODS

Our study area, Mecklenburg County, houses the largest metropolitan area (Charlotte) in North Carolina. The urban county has approximately 970,000 persons with 75 percent of the population over 18 years of age<sup>22</sup>. Mecklenburg County reported the highest number of new syphilis cases in North Carolina in 2012 and has historically high syphilis rates<sup>23</sup>.

### Data

We analyzed new primary and secondary (P&S) syphilis cases reported in North Carolina from January 1, 2003 to December 31, 2010. In North Carolina, physicians, other health care providers, and laboratories are required to report suspected and identified P&S syphilis cases to the local health department. Data recorded for each case of reportable infection included disease stage, report date, date of symptom onset, and residence. We obtained de-identified and geomasked (donut method)<sup>24</sup> syphilis case data from the Communicable Disease Branch of the North Carolina Division of Public Health.

Incident P&S syphilis cases were aggregated over an 8-year period. Primary and secondary syphilis incidence rates were calculated using census tract-level population estimates.

Census tract-level P&S syphilis incidence rates were the unit of analysis. Smoothing temporal variations and aggregating cases to the census tract provided a more stable metric for identifying areas with a high P&S syphilis burden. We calculated the mean P&S syphilis incidence rate for the state of North Carolina, Piedmont region, and Mecklenburg County. The mean P&S syphilis incidence rate is the reference rate against which observed rates are compared to define high incidence census tract clusters using both the local Moran's I and the scan statistic.

North Carolina census tract boundary files were obtained from the 2000 US Census<sup>25</sup>. Census tract population estimates were obtained from the 2000 census and the 2009 American Community Survey<sup>25,26</sup>. We calculated population estimates for each year between 2000 and 2009 using linear interpolation, and extrapolated 2010 population estimates. Demographic data were obtained from the 2000 census block group files and aggregated to the census tract.

### Local Moran's I statistic

We identified clusters of census tracts with high P&S syphilis burden in Mecklenburg County using the local Moran's I statistic. The local Moran's I statistic is a test of spatial association that identifies census tracts with high P&S syphilis incidence rates that are close together. The local Moran's I yields a measure of spatial association for each observation (i.e. census tract) based on a weighted average of P&S syphilis rates among adjacent neighboring observations<sup>27</sup>. The rook adjacency matrix defines 1<sup>st</sup> order neighbors as areas with a shared border, while the queen adjacency matrix defines 1<sup>st</sup> order neighbors as areas with shared borders and vertices. We selected a 2<sup>nd</sup> order queen areal adjacency matrix to account for the potential influence of non-adjacent neighbors as previous studies have identified high transmission clusters consisting of multiple order census tracts<sup>1,7</sup>, Census tracts sharing a common border with 1<sup>st</sup> order neighbors are defined as 2<sup>nd</sup> order neighbors in the weights matrix and are assigned less weight. An areal adjacency matrix was preferred over a distance threshold matrix given the variation in tract size. Distance matrices define neighbors using census tract centroids resulting in very few neighbors for larger rural tracts.

We systematically calculated multiple local Moran's I statistics using empirical Bayes (EB) standardized P&S syphilis rates in GeoDa 1.2.0<sup>27</sup>. We selected the empirical Bayes (EB) standardized rate to account for the varying population across tracts. The local Moran's I employs a randomization process to determine whether observed clusters of high burden census tracts are statistically significant. Statistical significance was determined using a Markov Chain Monte Carlo (MCMC) simulation. Observed values are randomly re-assigned to different tracts (999 permutations). The value in a given tract is compared to the randomly permuted values under the null hypothesis of no spatial association among neighboring tracts with P&S syphilis incidence rates exceeding the reference rate (mean P&S syphilis incidence rate for a given study area). The local Moran's I statistic was calculated using North Carolina, Piedmont region, and Mecklenburg County study area boundaries.

### Kulldorff's scan statistic

We also analyzed Mecklenburg P&S syphilis incidence rate data aggregated by census tract using Kulldorff's spatial scan statistic in SaTScan v9.1.1<sup>28</sup> to identify cluster locations and compare to the clusters detected using local Moran's I. An adaptive circular scanning window is employed to identify high incidence clusters in the study area. Data are aggregated to the census tract centroid, and centroids that fall within a given scan window represent the cluster. The adaptive window increases incrementally to encompass 0.0 to 5.0 percent of the study population. We selected the 5.0 percent maximum window size based on existing research<sup>7</sup> and to maintain comparability with the average neighborhood size (5% of the study population) defined using the local Moran's I adjacency matrix.

A discrete Poisson model was used to identify high P&S syphilis incidence rate clusters, where clusters were defined as windows where the number of cases was greater than the expected number of cases given the underlying population. The window with the maximum likelihood was defined as the most likely cluster, followed by secondary clusters. The significance of each cluster was determined using Monte Carlo simulation (999 permutations)<sup>28</sup>. We identified clusters using the North Carolina, Piedmont region, and Mecklenburg County boundaries.

### Comparison of cluster demographic characteristics by method and study area scale

The spatial distribution of syphilis transmission is likely influenced by the geographical variation of neighborhood level determinants (e.g. poverty, education, and sex ratios) that may place persons at increased risk of STI transmission<sup>29-31</sup>. The clusters identified and the demographic characteristics of the area covered by the cluster can vary by detection method<sup>19</sup> and by study area scale. Therefore, the observed neighborhood level (e.g. census tract) determinants of high transmission areas may change as we examine and interpret associations at a different scale. For example, differences in P&S syphilis burden across socio-demographic groups and areas accounting for most of the burden may be more notable at a smaller scale. When the process scale of the study area is increased, the study area becomes more diverse and demographic information associated with high risk can be muted.

We examined all clusters identified within Mecklenburg County using the state, region, and county scales, and local Moran's I and the scan statistic, for agreement in size and demographic composition. The selected demographic characteristics were previously identified as STI risk factors<sup>30,32-34</sup>. The following numerical values were calculated for each cluster: male to female ratio, percent female headed households, percent renting, percent less than a high school diploma, percent income below \$30,000, and percent unemployed. Numerical values were qualitatively compared by detection method and study area scale.

## RESULTS

Of the 2,572 P&S syphilis cases reported in North Carolina over the 8 year study period, 621 (24%) were from Mecklenburg County. 2,267 (88%) cases in the state and 572 (92%) cases in Mecklenburg could be geocoded and geomasked to a census tract. The highest P&S

syphilis incidence rates were reported in central Charlotte and declined at further distances (Figure 2).

### **Impact of method on cluster detection**

Several descriptive differences were found when comparing clusters. The local Moran's I clusters were more discontinuous due to excluding census tracts with a rate of zero (Figure 3:A–C). As a result, the local Moran's I clusters more closely matched the pattern of high burden tracts (Figure 2). The local Moran's I statistic identified a single cluster within Mecklenburg County while the scan statistic detected multiple clusters as the study area scale decreased (Figure 3:C,F). The scan statistic's inclusion of zero rate tracts increased the underlying population in the scan clusters resulting in lower incidence compared to the local Moran's I clusters (Table 1). Cluster characteristics were similar between the two detection methods, including the proportion of tracts in a cluster and cluster incidence (Table 1).

### **Impact of study area scale on cluster detection**

The reference P&S syphilis incidence rates varied substantially across study area scales. Reducing the study area scale from state to region to county increased the reference P&S syphilis incidence rate from 3.2 per 100,000 person-years for the state, to 4.1 per 100,000 person-years for the region, and to 8.9 cases per 100,000 person-years for the county. Peripheral tracts were lost as the study area decreased (Figure 3) and the P&S syphilis incidence reference rate increased (Table 1).

Within Mecklenburg County, reducing the scale from state to region resulted in the loss of few peripheral tracts using both local Moran's I and the scan statistic (Figure 3:A–B,D–E). The P&S syphilis incidence rates in lost peripheral tracts ranged from 0.0 to 13.1 cases per 100,000. The local Moran's I identified a single cluster with both state and region study boundaries, whereas the scan statistic identified two distinct clusters at the regional level (Figure 3:B,E). The minimum P&S syphilis incidence rate for census tracts included in the local Moran's I region cluster (5.7 cases per 100,000) was 39% higher than the study area reference rate (4.1 cases per 100,000 person years; Table 1). Nine percent of tracts (N=7) included in the scan statistic clusters had P&S syphilis incidence rates below the study area reference rate.

Cluster size decreased significantly when the scale was restricted to Mecklenburg County (Figure 3:C,E). Peripheral tracts that remained within the study area but were lost due to the increased reference rate had P&S syphilis incidence rates as high as 41.8 per 100,000. The minimum P&S syphilis incidence rate for census tracts included in the local Moran's I county cluster was 54% higher than the study area reference rate (Table 1). The scan statistic detected three clusters and 12.5% (N=3) of tracts had zero cases (Table 1).

### **Impact of method and study area scale on cluster demographic characteristics**

Demographic differences between the local Moran's I and scan statistic clusters were most pronounced at the most local scale, Mecklenburg County, where the local Moran's I captured a higher percentage of female headed households, and the scan statistic captured a higher percentage of households renting (Table 2).

Variation of cluster demographic characteristics was most pronounced between state and county scales (Table 2). For example, the percent of female headed households in the county cluster (15.9%) was 1.5 times the percent in the state cluster (10.4%) using local Moran's I. Similarly, the percent of female headed households in the county cluster (13.6%) was 1.4 times the state percent (9.5%) using the scan statistic. In addition, the percent of households with less than a high school diploma in the county cluster (23.6%) was 1.3 times greater than the state cluster (17.8%) using local Moran's I. Using the scan statistic, the percent of households with less than a high school diploma in the county cluster (22.4%) was 1.4 times greater than the state cluster (16.5%).

## DISCUSSION

Clusters detected using local Moran's I and Kulldorff's scan statistic overlapped but varied in size and composition. The local Moran's I statistic measures spatial association between adjacent neighbors and identifies high incidence census tracts surrounded by other high incidence census tracts; census tracts with a rate of zero are excluded from the cluster. In contrast, the adaptive circular scanning window of SatScan identified multiple clusters that included tracts with a rate of zero. Demographic differences between the local Moran's I and scan statistic clusters were most pronounced at the county scale.

Our results are similar to previous studies that found differences in chronic and infectious disease cluster locations and composition using multiple detection methods<sup>16,18,19</sup>. However, the difference in results indicate that using different methods are not as important as selecting the correct scale for analysis. One can be confident using either method but should consider the strengths and weaknesses when selecting a method. Local Moran's I may be more efficient in identifying only census tracts where Disease Intervention Specialists (DIS) might be sent, while Kulldorff's scan statistic may capture a higher number of cases and better identify areas that are more likely to have new cases (i.e. areas adjacent to areas with cases).

We selected the local Moran's I and Kulldorff's scan statistic because they are two of the most commonly used methods in STI research and are currently implemented in publicly available software programs<sup>35,36</sup>. However, additional methods exist including Bayesian disease mapping, generalized additive models (GAM), and the Getis Ord  $G_i^*$  statistic<sup>10,20</sup>. Further analysis with additional methods is necessary to gain a better understanding of the most appropriate approaches for STI cluster detection.

Study area scale selection affects the reference population, thereby influencing cluster detection. In our study, identified clusters included over 50 percent of Mecklenburg census tracts when using the state as the reference population. Using the county as the reference population yielded a higher reference rate, resulting in the identification of smaller clusters with higher incidence and the inclusion of less than 20 percent of Mecklenburg census tracts in identified clusters. Our findings correspond to previous research that identified differences in cluster detection after restricting the reference population to geographical regions with similar incidence rates<sup>7,17</sup>.

Although we cannot be certain of the generalizability of our findings to other STIs, our results highlight research issues to consider for practical STI cluster detection studies. Although the clusters detected were similar in location and size, census tracts identified as high burden were not identical between methods, especially around the peripheral areas, due to differences in method assumptions. Second, selecting a reference population with a different average rate influenced cluster outcomes. The data structure must also be considered for cluster interpretation. We analyzed syphilis data aggregated to the census tract to protect patient confidentiality. However, point data are also sensitive to cluster detection methods<sup>16</sup> and could yield different results. It is important to note that surveillance data are subject to the availability of, and participation in, testing; more cases will be diagnosed in areas with more availability of testing. Geographical analyses are susceptible to this potential bias.

Our results have implications for public health programs, particularly resource allocation and the targeting of STI interventions. Limited resources may require the targeting of control programs toward areas with the highest disease burden. Inconsistency around cluster edges also suggests that control efforts may need to consider peripheral and other nearby, non-cluster areas<sup>14</sup>. A state-level analysis would identify large clusters in high incidence regions. However, disease burden within these high incidence regions is not homogeneous, and additional analysis at the county-level would be needed to identify census tracts with the highest burden. Disease burden is also not homogeneous across low incidence regions, and additional county-level analysis may be needed to identify localized clusters that are undetectable when using the state average rate. In this situation, relying on too coarse of a study area scale for the spatial process of interest would not identify areas with increased cases in low-burden regions.

Our findings underscore the need for an exploratory and integrative approach to examine spatial patterns of high disease burden. The outcome of interest and study objectives should guide the selection of study area scale, and a discussion of strengths and weaknesses of different methods may be warranted. Analysts and decision makers should be mindful of the method and scale at which they identify clusters as this will impact their interpretation of spatial patterns and ultimately resource allocation and intervention decisions.

## Acknowledgments

**Sources of support:** VE received trainee support from the National Institutes of Health (NIH 5 T32AI07001–36, R24 HD050924). Data collection was supported by NIH R01 AI067913.

## References

1. Jennings JM, Curriero FC, Celentano D, et al. Geographic identification of high gonorrhea transmission areas in Baltimore, Maryland. *Am J Epidemiol.* 2005; 161:73–80. [PubMed: 15615917]
2. Becker KM, Glass GE, Brathwaite W, et al. Geographic epidemiology of gonorrhea in Baltimore, Maryland, using a geographic information system. *Am J Epidemiol.* 1998; 147:709–716. [PubMed: 9554611]
3. Bernstein KT, Curriero FC, Jennings JM, et al. Defining core gonorrhea transmission utilizing spatial data. *Am J Epidemiol.* 2004; 160:51–58. [PubMed: 15229117]



4. Rothenberg RB. The geography of gonorrhea. Empirical demonstration of core group transmission. *Am J Epidemiol.* 1983; 117:688–694. [PubMed: 6859024]
5. Zenilman JM, Ellish N, Fresia A, et al. The geography of sexual partnerships in Baltimore: Applications of core theory dynamics using a geographic information system. *Sex Transm Dis.* 1999; 26
6. Rothenberg R. Maintenance of endemicity in urban environments: A hypothesis linking risk, network structure and geography. *Sex Transm Infect.* 2007; 83:10–15. [PubMed: 17283360]
7. Gesink DC, Sullivan AB, Norwood TA, et al. Does core area theory apply to sexually transmitted diseases in rural environments? *Sex Transm Dis.* 2013; 40:32–40. [PubMed: 23254115]
8. Wu X, Tucker JD, Hong F, et al. Multilevel and spatial analysis of syphilis in Shenzhen, China, to inform spatially targeted control measures. *Sex Transm Infect.* 2012; 88:325–329. [PubMed: 22378936]
9. Goswami ND, Hecker EJ, Vickery C, et al. Geographic information system-based screening for TB, HIV, and syphilis (GIS-THIS): A cross-sectional study. *PLoS One.* 2012; 7:e46029. [PubMed: 23056227]
10. Torabi M, Rosychuk RJ. An examination of five spatial disease clustering methodologies for the identification of childhood cancer clusters in Alberta, Canada. *Spat Spatiotemporal Epidemiol.* 2011; 2:321–330. [PubMed: 22748230]
11. Kulldorff M, Song C, Gregorio D, et al. Cancer map patterns: Are they random or not? *Am J Prev Med.* 2006; 30:S37–S49. [PubMed: 16458789]
12. Goujon-Bellec S, Demoury C, Guyot-Goubin A, et al. Detection of clusters of a rare disease over a large territory: Performance of cluster detection methods. *Intl J Health Geogr.* 2011; 10:53.
13. Schmiedel S, Blettner M, Schüz J. Statistical power of disease cluster and clustering tests for rare diseases: A simulation study of point sources. *Spat Spatiotemporal Epidemiol.* 2012; 3:235–242. [PubMed: 22749209]
14. Gesink D, Wang S, Norwood T, et al. Spatial epidemiology of the syphilis epidemic in Toronto, Canada. *Sex Transm Dis.* 2014; 41:637–648. [PubMed: 25299409]
15. Nicolai LM, Stephens N, Jenkins H, et al. Early syphilis among men in Connecticut: Epidemiologic and spatial patterns. *Sex Transm Dis.* 2007; 34:183–187. [PubMed: 16906124]
16. Wheeler DC. A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996–2003. *Int J Health Geogr.* 2007; 6:13. [PubMed: 17389045]
17. Jacquez GM, Greiling DA. Local clustering in breast, lung and colorectal cancer in Long Island, New York. *Int J Health Geogr.* 2003; 2:3. [PubMed: 12633503]
18. White PS, Graham FF, Harte DJ, et al. Epidemiological investigation of a Legionnaires' disease outbreak in Christchurch, New Zealand: the value of spatial methods for practical public health. *Epidemiol Infect.* 2013; 141:789–799. [PubMed: 22697112]
19. Ozdenerol E, Williams BL, Kang SY, et al. Comparison of spatial scan statistic and spatial filtering in estimating low birth weight clusters. *Int J Health Geogr.* 2005; 4:19. [PubMed: 16076402]
20. Fritz CE, Schuurman N, Robertson C, et al. A scoping review of spatial cluster analysis techniques for point-event data. *Geospat Health.* 2013; 7:183–198. [PubMed: 23733283]
21. North Carolina Office of State Budget and Management. Certified County Population Estimates—Ranked by Size. 2012. Available at: [http://www.osbm.state.nc.us/ncosbm/facts\\_and\\_figures/socioeconomic\\_data/population\\_estimates/county\\_estimates.shtm](http://www.osbm.state.nc.us/ncosbm/facts_and_figures/socioeconomic_data/population_estimates/county_estimates.shtm)
22. State & County Quick Facts: Mecklenburg County, North Carolina. Available at: <http://quickfacts.census.gov/qfd/states/37/37119.html>
23. North Carolina: Epidemiologic Profile for HIV/STD Prevention & Care Planning. Available at: [http://epi.publichealth.nc.gov/cd/stds/figures/Epi\\_Profile\\_2012.pdf](http://epi.publichealth.nc.gov/cd/stds/figures/Epi_Profile_2012.pdf)
24. Hampton KH, Fitch MK, Allshouse WB, et al. Mapping health data: Improved privacy protection with donut method geomasking. *Am J Epidemiol.* 2010; 172:1062–1069. [PubMed: 20817785]
25. 2010 TIGER/Line Shapefiles [machine-readable data files]/prepared by the U.S. Census Bureau. 2010.
26. U.S. Census Bureau. American Community Survey. 2009

27. Anselin, L. An introduction to spatial autocorrelation analysis with GeoDa. Champagne–Urbana, IL: Spatial Analysis Laboratory, University of Illinois; 2003.
28. Kulldorff M. A spatial scan statistic. *Communications Stat Theory Methods*. 1997; 26:1481–1496.
29. Law DC, Serre ML, Christakos G, et al. Spatial analysis and mapping of sexually transmitted diseases to optimise intervention and prevention strategies. *Sex Transm Infect*. 2004; 80:294–299. [PubMed: 15295129]
30. Sullivan AB, Gesink DC, Brown P, et al. Are neighborhood sociocultural factors influencing the spatial pattern of gonorrhea in North Carolina? *Ann Epidemiol*. 2011; 21:245–252. [PubMed: 21376271]
31. Le Polain DeWaroux O, Harris R, Hughes G, et al. The epidemiology of gonorrhoea in London: A Bayesian spatial modelling approach. *Epidemiol Infect*. 2014; 142:211–220. [PubMed: 23561246]
32. Bush KR, Henderson EA, Dunn J, et al. Mapping the core: Chlamydia and gonorrhea infections in Calgary, Alberta. *Sex Transm Dis*. 2008; 35:291–297. [PubMed: 18490871]
33. Holtgrave DR, Crosby RA. Social capital, poverty, and income inequality as predictors of gonorrhoea, syphilis, chlamydia and AIDS case rates in the United States. *Sex Transm Infect*. 2003; 79:62–64. [PubMed: 12576618]
34. Semaan S, Sternberg M, Zaidi A, et al. Social capital and rates of gonorrhea and syphilis in the United States: Spatial regression analyses of state-level associations. *Soc Sci Med*. 2007; 64:2324–2341. [PubMed: 17400352]
35. Anselin L, Syabri I, Kho Y. *GeoDa: An introduction to spatial data analysis*. *Geogr Anal*. 2006; 38:5–22.
36. Software: Kulldorff M. and Information Management Services, Inc. SaTScan™ v8.0: Software for the spatial and space-time scan statistics. Available at: <http://www.satscan.org/>. 2009

**SUMMARY**

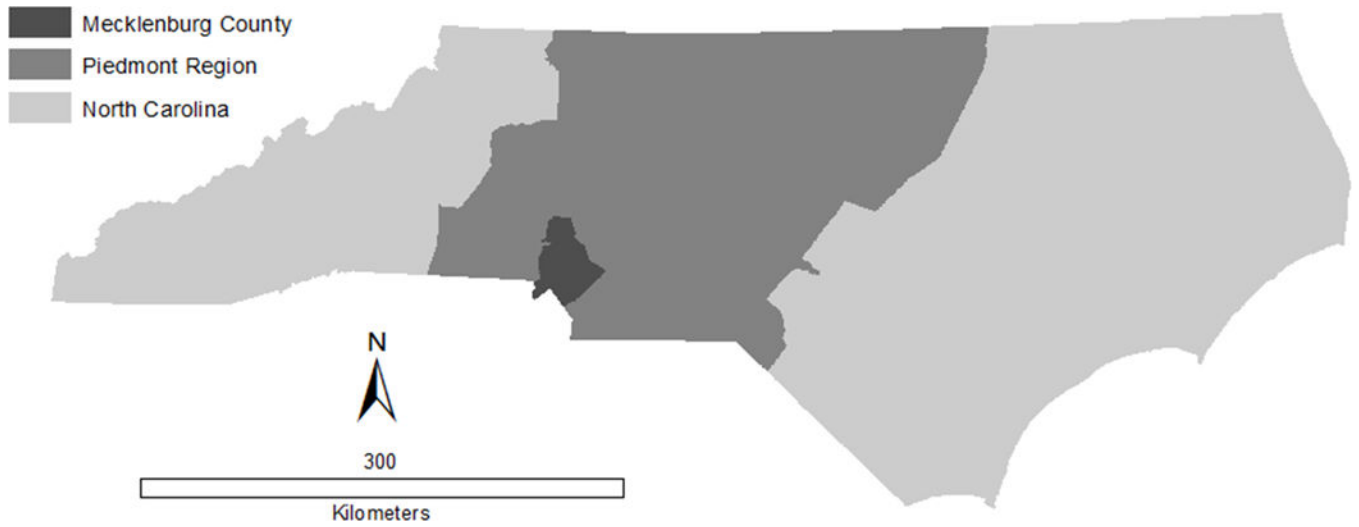
Clusters of high primary and secondary syphilis incidence rates in North Carolina varied by selected cluster detection method and study area scale.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



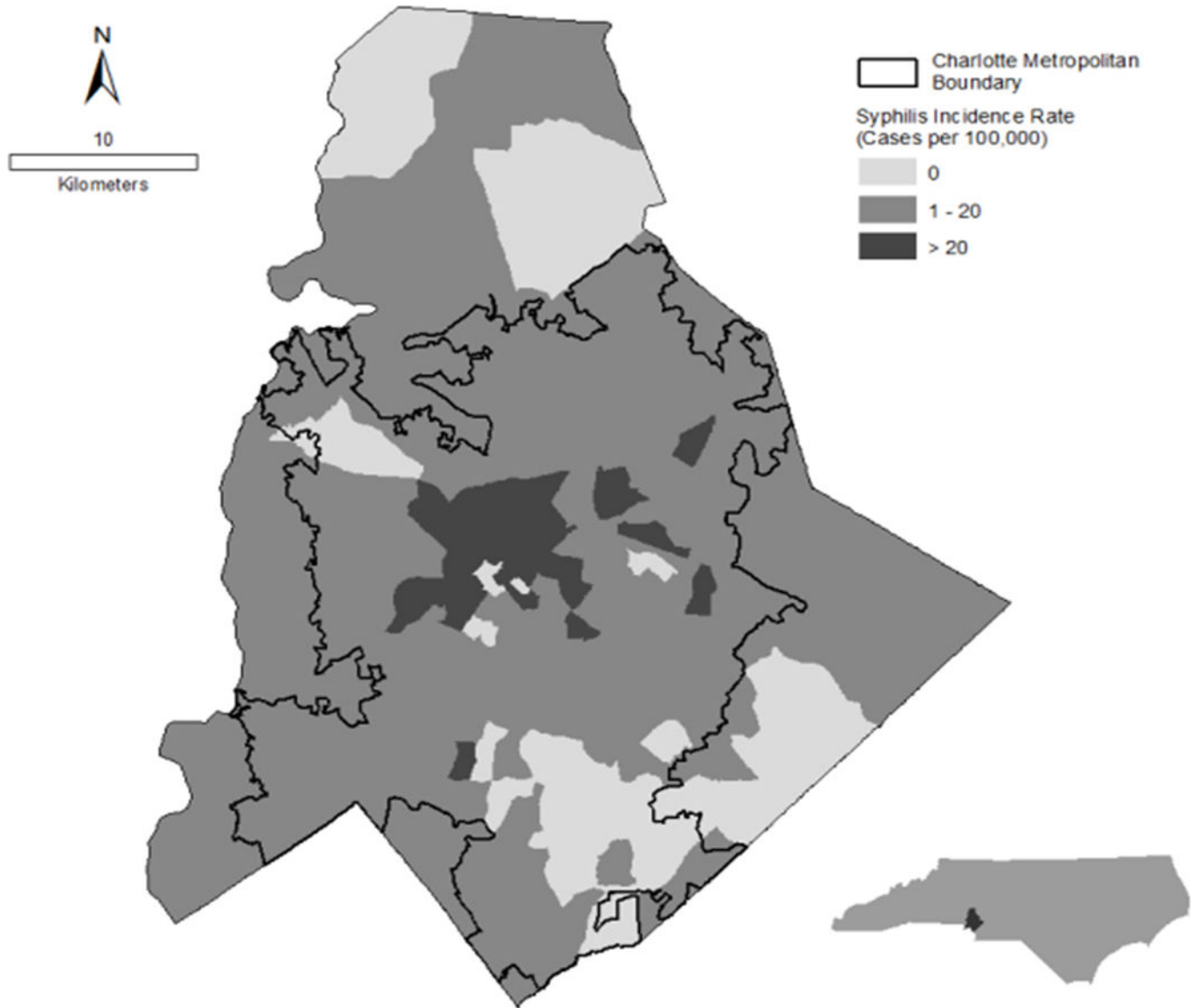
**Figure 1.**  
North Carolina, Piedmont region, and Mecklenburg County study area boundaries

Author Manuscript

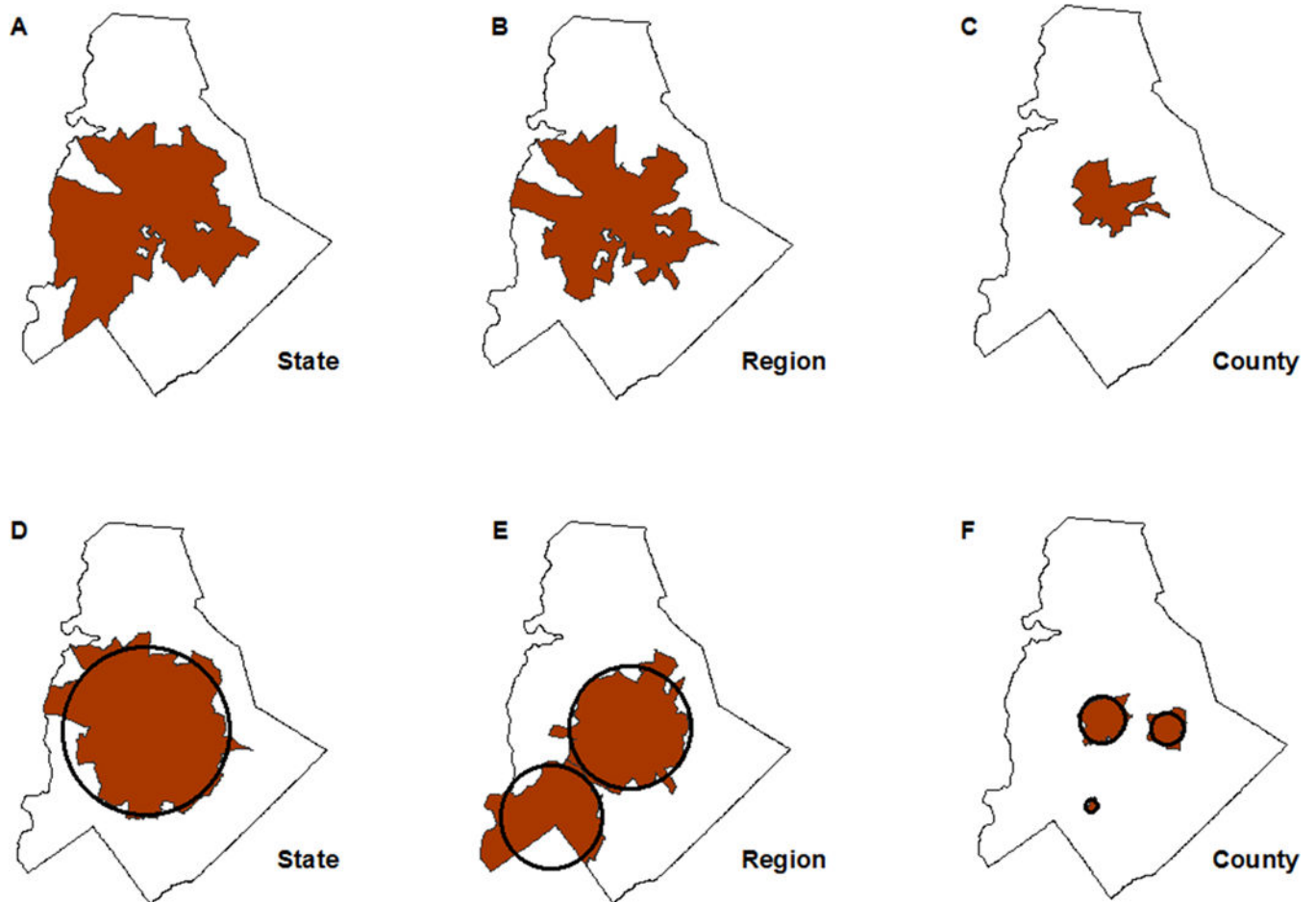
Author Manuscript

Author Manuscript

Author Manuscript



**Figure 2.**  
Mecklenburg County P&S syphilis incidence rates (January 2003 – December 2010)



**Figure 3.**

Local Moran's I clusters (A-C) and Kulldorff's scan statistic clusters (D-F). Clusters detected using North Carolina state scale (A,D), Piedmont region scale (B, E), and Mecklenburg County scale (C, F).

Note: Identifiable clusters displayed in red. Circles placed on panels D-F represent the scan windows that encompassed centroids to identify a tract as part of a cluster.

**Table 1**

Cluster characteristics by detection method and study area scale

	<i>Local Moran's I Statistic</i>				<i>Kulldorff's Scan Statistic</i>				
	<i>State</i>	<i>Region</i>	<i>County</i>	<i>State</i>	<i>Region</i>	<i>County</i>	<i>State</i>	<i>Region</i>	<i>County</i>
<b>Reference syphilis incidence rate by study area scale, cases per 100,000 person years</b>	3.2	4.1	8.9	3.2	4.1	4.1	3.2	4.1	8.9
<b>Cluster characteristics</b>									
Proportion of Mecklenburg tracts included in cluster, <i>n</i> (%)	78 (54.2)	67 (46.5)	26 (18.1)	90 (62.5)	77 (53.5)	24 (16.7)			
Cluster incidence, cases per 100,000 person years	14.8	17.0	30.3	14.1	15.9	28.6			
Total population in cluster	359734	290895	77662	393656	330076	85507			
Total cases in cluster	454	410	179	465	434	190			
Minimum rate for tracts included in cluster	4.3	5.7	13.7	0.0	0.0	0.0			

**Table 2**

Cluster demographic characteristics by cluster detection method and study area scale

Cluster demographic characteristics	Local Moran's I Statistic			Kulldorff's Scan Statistic		
	State	Region	County	State	Region	County
Population under 18 years, %	75.2	75.5	73.1	76.4	76.2	74.7
Male to female ratio	0.97	0.97	0.96	0.96	0.96	0.96
Female headed households, %	10.4	11.2	15.9	9.5	10.0	13.6
Households renting, %	47.3	51.3	57.4	48.3	51.4	59.1
Households with less than high school diploma, %	17.8	19.0	23.6	16.5	17.5	22.4
Households with income < 30,000, %	33.9	37.3	45.0	33.6	34.3	43.0
Unemployed, %	7.0	7.4	8.8	6.6	6.8	7.8