January 2015

Spatial Distribution And Sociodemographic Composition Of High Grade Cervical Lesion Clusters In Connecticut 2008-2013

Aref Senno
Yale University, tareksenno78@gmail.com

Spatial Distribution and Sociodemographic Composition of High Grade Cervical Lesion Clusters in Connecticut 2008-2013

A Masters of Public Health Thesis

Presented to the Department of Epidemiology of Microbial Diseases

Yale School of Public Health

Primary Reader: Linda M. Niccolai, PhD, ScM

Secondary Reader: Daniel M. Weinberger, PhD

By:

Aref T. Senno

April 2015

Introduction: HPV is necessary for the development of invasive cervical cancer. There were 530,000 cases and 275,000 deaths attributable to cervical cancer in 2008. Since the introduction of HPV vaccines in 2006, the incidence of high grade cervical lesions (HGCL)—precancerous lesions which may develop into cervical cancer—has decreased significantly. Some changing sociodemographic trends in disease burden have been previously reported since 2008.

Objectives: The objectives of this analysis were to identify regions of Connecticut with high rates of HGCLs and to describe how trends of disease changed over time in the post-vaccine era.

Methods: This is a descriptive analysis of surveillance data on HGCLs that were collected from all 34 state pathology laboratories between 2008 and 2013 through the HPV-IMPACT surveillance project. Spatial analyses were performed to identify clusters of high rates of HGCLs in 20-39 and 20-24 year old groups. Census sociodemographic data were linked to clusters to analyze trends in disease burden. A Poisson regression model was fit to describe the changing statewide patterns of disease in different racial, ethnic, socioeconomic, and age groups.

Results: Spatial analyses identified several transient clusters of HGCLs from 2008 to 2013. These clusters varied in size, racial, ethnic, and socioeconomic composition. A persistent cluster of HGCLs was observed in the northwestern region of the state across age groups. The number of cases declined significantly from 2009 to 2013 in the 20-24 and 25-29 year old groups, but did not change significantly in the 30-39 year old groups. Census tracts with ≥20% in poverty had a significantly lower rate of cases compared to census tracts with <5% in poverty. These same census tracts had a higher rate of cases in the 30-34 and 35-39 year old group compared to census tracts with <5% in poverty. No significant interactions between poverty level and time were observed in predicting the number of cases in any age strata between 20 and 39 years old.

Discussion: Local public health collaborators should further investigate the persistent cluster of HGCLs in the northwestern region of the state. Cases have been declining consistently since the introduction of the vaccine in younger age groups (20-29) which may be affected by the vaccine. This decline was not seen in older groups (30-39) that likely did not receive the vaccine. This method of linking surveillance data to census data with spatio-temporal analysis can be used in other areas to assess changing trends in precancerous lesions in the post-vaccine era.

Table of Contents

2.1 Introduction

HPV and Cervical Cancer

Human Papillomavirus (HPV) is a sexually transmitted infection implicated in the development of genital warts and several types of cancer including anal, vaginal, penile, oral, and cervical.[1, 2] The virus is ubiquitous globally. Approximately 11-12% of all women in the world carry an HPV infection at any time.[3] The highest rates of infection occur in Sub-Saharan Africa (24%), Eastern Europe (21%), and Latin America (16%).[3] Over 100 variants of HPV have been isolated and data suggest that there are well over 200 types present worldwide.[4] While there is an abundance of variation in types of HPV, only two types are responsible for the vast majority of cases of cervical cancer. High risk types, 16 and 18, have are present in 60%-70% of cervical tumors.[5-7] Approximately 11-16 other "high-risk" types are responsible for nearly all other cases of cervical cancer.[8]

There is a high global burden of cervical cancer, especially in low-income regions. HPV is a necessary cause for cervical cancer and thus must be present for the malignant transformation to occur.[9] The worldwide burden of cervical cancer was 530,000 cases and 275,000 deaths in 2008.[10] The burden of cervical cancer disproportionately affects lower income countries as 85.4% of cervical cancer cases occurred in less developed regions of the world.[3] Due to decreased access to screening and preventative care, cervical cancer is the leading cause of cancer-related death in women in Eastern, Western and Middle Africa, Central America, South-Central Asia and Melanesia.[10]

Women in the United States are at lower risk for cervical cancer, but show a high degree of disparity between races, ethnicities, and socioeconomic groups. 12,109 new cases of cervical cancer were diagnosed and 4,092 women died from cervical cancer in the United States in 2011.[11] The incidence of cervical cancer was 43% higher in Black women and 51% higher in Hispanic women compared to White women in 2010.[12] Mortality rates were 100% higher in Black women and 38% higher in Hispanic women compared to White women in 2010.[12] Socioeconomic status has also been shown to play a key role in cervical cancer survival. Uninsured and Medicare/Medicaid insured women were 1.4 more likely to die from cervical cancer than privately insured women.[13] Mortality rates also declined at a faster rate for highly educated women compared to uneducated women between 1993 and 2007.[13]

Vaccine Production

As a result of the high burden of HPV-related diseases, three preventative vaccines were developed to specifically target "high risk" HPV types. Gardasil and Cervarix, which protects against HPV 16 and 18 were approved by the FDA in June 2006 and October 2009 respectively.[14, 15] Gardasil 9, a 9-valent vaccine which protects against an additional five types of "high risk" HPV variants, was approved by the FDA in December 2014.[16] Vaccination is currently recommended for girls aged 13-26 years and boys aged 13-21 years.[17] Nationally, HPV vaccination rates are quite low. Only 37.6% of female and 13.9% of male adolescents completed at least three doses of a HPV vaccine.[18] Connecticut has slightly higher

rates of vaccination where 40.1% of female and 23.4% of male adolescents received at least three doses of a HPV vaccine.[18]

Evaluation of Trends in the Post-Vaccine Era

A number of studies have examined how HPV and HPV-related disease incidence changed since the introduction of HPV vaccines. Prevalence of vaccine-type HPV deceased from 11.5% between 2003 and 2006 to 5.1% between 2007 and 2010 among females 14-19 years old.[19] A meta-analysis of HPV studies showed that countries with ≥50% vaccination coverage in females 13-19 years old had a 68% drop in type 16 and 18 infections compared to pre-vaccine years.[20] High-risk types 31, 33, and 45 also declined significantly in this same age group due to suspected cross protection.[20] Australia, which has coverage above 80% for target females, has had a decline in vaccine-type HPV prevalence from 28.7% to 6.7%.[21]

The effect of vaccine use on cervical cancer incidence is not yet available. Cervical carcinomas take years of persistent infection to develop. The median time from infection to carcinoma in situ is 7-12 years.[22] High grade cervical lesions (HGCL) can be used as an indicator for evaluation of vaccine impact until cervical cancer data is available. HGCLs (cervical intraepithelial neoplasia grades 2 and 3 and adenocarcinoma in situ) are precursors to cervical cancer which form years prior to cervical carcinomas. The annual incidence of HGCLs declined from 834 per 100,000 females age 21-24 to 688 per 100,000 females age 21-24 between 2008 and 2011 in Connecticut.[23]

Spatial analysis of trends in the distribution of HGCLs and the composition of clusters is a currently unutilized method of assessing vaccine impact. No studies have examined how spatial disease trends have changed since the introduction of the HPV vaccine in 2008.

Cluster Analysis and Spatial Trends

Spatial analyses using SaTScan have previously been used to identify clusters of high rates of infectious diseases and cancers. Descriptive studies previously identified areas with high incidence of sexually transmitted infections using census and surveillance data.[24] This method has similarly been used for HPV- and pathogen-related cancers. [25-27] Through the identification of areas with high rates of pathogen-associated cancers, local public health collaborators can more efficiently direct resources toward transmission prevention and treatment.

Analysis of the composition of these clusters provides information on potentially at-risk populations. Community-wide differences in sociodemographic composition (race, ethnicity, poverty, age, etc.) may highlight groups which have not benefitted fully from the HPV vaccine. Spatio-temporal data can also be used to assess changes in disease burden in these groups.

Objectives

The primary objective of this analysis was to identify regions of Connecticut with high rates of HGCLs that persist or change over time. The secondary objective was to describe the composition of these clusters—as described by area-based sociodemographic measures—and how they changed over time in the post-vaccine era.

2.2 Materials and Methods

HPV-IMPACT

The Emerging Infections Program (EIP) is a network of 10 state health departments, local collaborators, and academic institutions under the direction of the CDC which provides surveillance support for several infectious diseases. In 2008, the EIP instituted the Human Papillomavirus Vaccine Impact Monitoring Project (HPV-IMPACT) in five of the ten sites, including Connecticut. The Connecticut EIP collects surveillance data on HGCLs which were added to the list of mandatory reportable diseases in Connecticut on January 1, 2008. The Connecticut EIP additionally collects basic demographic information including date of birth, home address, HGCL diagnosis, race, and ethnicity. This surveillance program is ongoing.

Population

The original study population (n=17,299) includes all cases of HGCL between January 1, 2008 and November 11, 2014 that were reported by all 34 Connecticut pathology laboratories. A case is defined as a laboratory-confirmed case of HGCL: cervical intraepithelial neoplasia grade 2 or 2/3 or 3 and adenocarcinoma in situ (CIN2/3/AIS). Cases are not contingent on the presence of HPV. Information on date of birth, address, race, and ethnicity was collected when available.

Connecticut was made up of 829 census tracts as of the 2010 census. Census tract-level data on gender, age, race, ethnicity, and poverty was retrieved from the American Community Survey (ACS).[28] Black race and Hispanic ethnicity data were retrieved as a proportion of the census tract population which identified as that race or ethnicity. Poverty data was retrieved as a proportion of the census tract that lives below the federal poverty line as determined by family income, size, and composition. All census tract-level data retrieved from the ACS were five-year estimates for the 2008-2012 period. Shapefiles for the state of Connecticut and its corresponding census tracts were retrieved from the University of Connecticut GIS Data Library.[29]

Data Analysis

All cases were first geocoded to X and Y coordinates and then linked to specific census tracts using ArcGIS version 10.2.2 and the Federal Financial Institutions Examination Council Website.[30] The dataset was cleaned to remove all cases without ages, without addresses, with addresses incapable of being geocoded, and any cases with diagnosis outside of the period January 1, 2008 to December 31, 2013. The final case number for analysis was n=15,584. Analyses were controlled for age by restriction to the 20-24 (n=3,897), 25-29 (n=3,835), 30-34 (n=2,633), 35-39 (n=1,527), and 20-39 (n=11,892) year old groups.

Spatial representations of HGCL distribution were produced using ArcMap version 10.1. Analyses were age-restricted to the 20-24 and 20-39 year old groups. Spatial cluster analyses were performed in SaTScan, version 9.3.1. Cluster analyses was also disaggregated by two-year periods (2008-2009, 2010-2011, and 2012-2013) to observe temporal trends. Two-year periods were used as opposed to one-year periods in order to provide a larger sample of cases and to decrease noise across time.

A Poisson regression model was fit to age-restricted groups using SAS, version 9.3 to predict case rate by differences in area-based sociodemographic measures. Three ordinal variables of race, ethnicity, and poverty were considered in the Poisson regression where proportions were classified as <5.0%, 5.0% to 9.9%, 10.0% to 19.9%, ≥20.0%. Time was also considered as an ordinal variable from 2008 to 2013. We additionally modeled a linear spline with a knot at 2010. An additional Poisson regression model was fit with an interaction term between time and poverty. Supplementary preliminary models and bivariate analyses are presented in the Appendix.

2.3 Results

Census tract-aggregated cases (n=15,584) were normalized to the total number of females in each census tract between 2008 and 2013 (Figure 1). The case rate was stratified into 20th percentile quantiles. ACS population data was not available for four census tracts. The annual case rate ranged from 0 to 460 cases per 100,000 females. The lowest quintile case rates were observed in the eastern and western perimeter census tracts. The highest quintile case rates were observed in the central, south-central, and the southeastern regions of the state.

This analysis was repeated in the age-restricted 20-39 year old group (Figure 2). The annual case rate ranged from 0 to 1,515 cases per 100,000 20-39 year old females. The lowest quintile case rates of HGCLs were observed in the northeastern and northwestern perimeter census tracts. The highest quintile case rates of HGCLs were observed diffusely across the state.

A spatial cluster analysis was performed to find clusters of high rates of HGCLs in the 20-39 year old group (Figure 3). Four statistically significant ($p<0.05$) clusters were observed. Cluster size ranged from 30 to 126 census tracts (Table 1). The median proportion that are Black, Hispanic, or in poverty in Non-Cluster census tracts is 3.80%, 6.47%, and 5.50% respectively. The median proportion that are Black ranges from 1.30% to 6.55% in Cluster tracts. The median proportion that are Hispanic ranges from 4.24% to 11.53 % in Cluster tracts. The median proportion that are in poverty ranges from 4.50% to 10.65% in Cluster tracts.

This analysis was repeated in the age-restricted 20-24 year old group (Figure 4). Seven statistically significant ($p<0.05$) clusters were observed. Cluster size ranged from 3 to 85 census tracts (Table 1). The median proportion that are Black, Hispanic, or in poverty in Non-Cluster census tracts is 5.10%, 8.07%, and 6.70% respectively. The median proportion that are Black ranges from 0.75% to 32.90% in Cluster tracts. The median proportion that are Hispanic ranges from 2.68% to 32.49 % in Cluster tracts. The median proportion that are in poverty ranges from 1.75% to 21.20% in Cluster tracts. Cluster 4.4 is a small cluster (3 census tracts) with a high median proportion of the population that is Black (32.90%), Hispanic (32.49%), and in poverty (21.20%).

Cluster analyses were performed on the age-restricted 20-39 year old group in two year periods from 2008 to 2013 (Figure 5). Clusters presented all have significantly ($p<0.05$) high rates of HGCLs for their respective populations. Between 2008 and 2009, two clusters were

observed. Cluster 5.1 is located in the northwestern region. Cluster 5.2 is located in the southeastern region. Between 2010 and 2011, three clusters were observed. Cluster 5.3 is located in the northwestern region. Cluster 5.4 is located in the southeastern region. Cluster 5.5 is located in the south-central region. Between 2012 and 2013, one cluster was observed. Cluster 5.6 is located in the south-central region. Clusters ranged in size from 7 to 360 census tracts (Table 2). Cluster 5.4 is a small cluster (7 census tracts) with a high median proportion of the population that is Black (9.30%), Hispanic (20.37%), and in poverty (21.90%).

This analysis was repeated in the age-restricted 20-24 year old group (Figure 6). Between 2008 and 2009, two clusters were observed. Cluster 6.1 is located in the northwestern region. Cluster 6.2 is located in the southeastern region and spans across the south-central region. Between 2010 and 2011, three clusters were observed. Cluster 6.3 is located in the northwestern region. Cluster 6.4 is located in the south-central region. Cluster 6.5 is located in the southwestern region. Between 2012 and 2013, three clusters were observed. Cluster 6.6 is located in the northwestern region. Clusters 6.7 and 6.8 are located in the south-central region. Clusters ranged in size from 1 to 156 census tracts (Table 2). Cluster 6.7 is a small cluster (1 census tract) with a high proportion of the population that is Black (31.10%), Hispanic (32.15%), and in poverty (16.00%).

A Poisson model was fit to predict the census tract case rate in age-restricted 20-24, 25-29, 30-34, 35-39, and 20-39 year old groups (Table 3). Poverty was negatively associated with the case rate ratio in the 20-24 year old group in a dose-dependent manner. Census tracts that were 10% to ≤20% below the federal poverty line had 30.8% [95% CI 39.7%-20.7%] lower case rates than those that are <5% in poverty in the 20-24 year old group ($p<0.001$). Census tracts that were ≥20% in poverty had 43.9% [95% CI 51.7%-34.8%] lower case rates than census tracts that were <5% below the federal poverty line ($p<0.001$). The case rate has been decreasing annually on average by 14.6% [95% CI 17.0%-12.2%] since 2009 in the 20-24 year old group ($p<0.001$). The case rate decreased on average 2.8% [95% CI 5.5%-0.1%] annually in the 25-29 year old group ($p=0.039$). However, no changes were seen temporally in the 30-34 or 35-39 year old group. The 35-39 year old group had significantly higher case rates in census tracts with 5% to <10% and ≥20% of the population in poverty.

A Poisson model was fit to observe the interaction of poverty and time in age-restricted 20-24, 25-29, 30-34, 35-39, and 20-39 year old groups (Table 4). No significant differences in case rates were observed for any interaction term across any age group.

2.4 Discussion

Burden and Overall Distribution

The distribution of HGCLs in Connecticut is widespread. HGCLs occurred in nearly all of the census tracts in Connecticut (825/829) over the 6 year surveillance period. The spatial distribution of HGCLs (Figure 1) provides a visual representation of their overall burden in Connecticut. High burden census tracts are easily identifiable in particular regions of the state,

namely the center. Low burden areas, likewise, can be seen with relative ease. However, the incidence of HGCLs is largely dependent on the age composition of the underlying population. The age range 20-39 years is considered the "at risk" population as 69% of cases occurred in this age group. Extending the age range to 65 years—when national guidelines no longer recommend pap smears—only adds an additional 20% of cases. This group is largely composed of individuals that did not benefit from the introduction of the HPV vaccine. They were too old to be vaccinated or were likely exposed to HPV before being vaccinated. They may additionally not benefit from herd immunity from the HPV vaccine due to age-specific sexual practices because over half of women aged 15-44 had sex with a partner aged within 2 years of them.[31]

The case rate of HGCLs must be normalized to the age of the underlying population to account for age-specific differences. To understand where cases are occurring with unusually high rates, cases and the underlying population were age-restricted. When age-restricted to the 20-39 year old group (Figure 2), the burden of disease becomes much more dispersed. The highest quintile rates of HGCLs were observed throughout much of the state. Lowest quintile census tracts are seen in some parts of the northwestern and northeastern region of the state. The range of case rates varies widely from 0 to 1,515 cases per 100,000 female-years, suggesting large discrepancies in the burden of disease.

Spatial Clusters and Sociodemographic Measures (2008-2013)

Cluster analyses were performed on the 20-39 year old group to identify regions with significantly higher rates of HGCLs in the at-risk population in Figure 3. The four clusters identify regions where HGCLs are occurring at unusually high rates over the entire surveillance period. Clusters 3.1-3.3 are composed of census tracts that have generally higher median proportions Black, Hispanic, and in poverty as compared to Non-Cluster census tracts. However, Cluster 3.4 is composed of census tracts with a lower median proportion Black, Hispanic, and in poverty compared to Non-Cluster census tracts. This suggests that the at-risk population with high rates of HGCLs may largely come from higher proportion Hispanic and impoverished areas. However, the identification of a cluster that has a lower proportion Black, Hispanic, or in poverty (Cluster 3.4) suggests that unusually high rates may occur in higher income, non-Black, non-Hispanic areas also.

Cluster analysis was further age-restricted to the 20-24 year old group. This age group is also within the at-risk 20-39 year old group. However, this group falls within the HPV vaccination guidelines and may also have sex with partners that are vaccinated. Seven clusters over the entire surveillance period (2008-2013) were identified in Figure 4. These clusters had high rates of HGCLs despite potential benefits of the HPV vaccine in this age group. Clusters 4.1-4.3 and 4.6-4.7 all have lower median proportions Black, Hispanic, and in poverty. Cluster 4.4 and 4.5, however, generally have higher median proportions Black, Hispanic, and in poverty than Non-Cluster census tracts.

Temporal Trends

Cluster analyses were performed in two year periods for age-restricted 20-39 and 20-24 year old groups to identify sociodemographic and spatial trends over time (Figure 5 & 6). Most

clusters identified were diverse in location, size, and were mostly transient. However, there is one persistent cluster which overlaps with Clusters 5.1, 5.3, 6.1, 6.3, and 6.6. The consistent identification of this cluster across both 20-24 and 20-39 year old groups suggests that this area may be subject to unusually high rates of HGCLs. This may be due to higher rates of HPV 16 or 18, more prolific screening, or some other unknown reason. The identification of this persistent cluster calls for further investigation at the local level.

The composition of these overlapping clusters is not entirely apparent (Table 2). The clusters characteristically have a lower median proportion Black (0.80%-2.60%) compared to Non-Cluster census tracts in the same years and age groups (3.80%-5.30%). Median proportion Hispanic and in poverty seem to have no distinct association with these Clusters.

The composition of clusters identified over the surveillance period does not have any discernible temporal pattern. Clusters 5.4 and 6.7 are both composed of census tracts with unusually high median proportions Black, Hispanic, and in poverty. However, the majority of clusters (11/14) identified were composed of census tracts with <10% Black, Hispanic, or in poverty. These clusters are similar in composition to the Non-Cluster census tracts which all had median proportions Black, Hispanic, and in poverty under 10%.

Poisson Regression

The Poisson model highlights differences in case rates due to their sociodemographic composition across age groups. Statewide, overall incidence of HGCLs significantly declined in the 20-24 and 25-29 year old groups from 2009 to 2013 (Table 3). However, the 30-34 and 35-39 age groups did not see any significant decline in incidence across the same time period. This result is to be expected if HPV vaccination is effective in preventing HPV infection and subsequent HGCLs. This trend has been reported previously in Connecticut between 2008 and 2011 for 21 year olds, but not older groups.[23]

Poverty was commonly associated with significant differences in case rates. The 20-24 year old group had a lower incidence of cases in census tracts with ≥20% in poverty compared to census tracts with <5% in poverty. There was no significant difference across poverty strata in the 25-29 year old group. The 30-34 and 35-39 year old groups showed increasingly higher numbers of cases in census tracts with ≥20% in poverty compared to census tracts with <5% in poverty.

An interaction model was fit to observe how the number of cases in poverty strata changed over time in different age groups (Table 4). No significant associations were found in the interaction term. This suggests that the incidence of HGCLs did not change differentially by poverty strata across any age group. The trend of increasing rate ratios in the ≥20% poverty strata across age groups may be explained by a "harvesting effect" in which very low-income individuals may not be screened for HGCLs until they are much older. Thus, cases may appear to be lower in younger age groups and higher in older age groups. Screening data is not presently available to validate this hypothesis.

Methodology

The methodology of this project provides insights for other groups interested in HGCL surveillance and vaccine impact studies. This methodology links surveillance data to independent census data and thus provides a simple way to identify trends and compare areas of interest. There is no need to collect individual-level demographic data outside of the geospatial location of the individual. Low and middle income countries rarely have the resources to collect large amounts of individual-level data. Simple, robust methodologies such as the one described here are necessary for use in resource-limited settings.

Surveillance of HGCLs is an efficient way to collect data that can be used to observe changes in burden after HPV vaccine introduction. Regions where the burden of HPV and HPV-related diseases is highest—largely low-income countries—can especially benefit from this methodology. The costs of HPV vaccination in low- and middle-income countries varies widely and can be very expensive ($1.49 to $18.94 per female).[32] In order to justify these high startup costs, programs must have high efficacy at the local level.

In November 2011, the GAVI Alliance added the HPV vaccine to their list of funding-eligible vaccines for low- and middle-income countries.[32] However, vaccination must be properly supported at the national level to receive funding. The success of these immunization campaigns is dependent upon the continuous identification of at-risk, hidden, and missed populations. Spatial mapping of clusters—like the methodology described here—provides a way to identify regions that may require additional resources. Linkage to sociodemographic data can also identify additional covariates that may influence disease transmission.

Strengths and Limitations

This analysis has several strengths. Firstly, HPV-IMPACT provides the most robust dataset for the presence of HGCLs in Connecticut. Since it is mandatory to report HGCLs in Connecticut, reported cases represent essentially all known cases of HGCLs in the state. Second, spatial analyses of HGCL data have never been performed in the post-vaccine era, providing a novel understanding of how HPV-related diseases may be changing spatially and temporally. The added benefit of census-linked data provides a brief understanding of populations which may disproportionately benefit from the introduction of the HPV vaccine.

While this dataset is the most robust information on HGCLs in Connecticut, there are limitations to this analysis. The first limitation is the number of census tracts with missing or incomplete ACS data. While the actual number of census tracts with missing data or suspected missing data is low (10 out of 829), any systematic differences in the underlying population due to missing data may cause unforeseen changes in spatial analysis. Missing ACS data did not seem to correspond with any spatial pattern. Second, the analysis of 20-24 year olds examines a somewhat mobile population. Many within this age group are attending university or starting jobs out of high school or college. Given that HGCLs may take years to develop, spatial trends in HGCL burden may be disparate from actual local trends. Finally, analyses use population-level demographic data for census tracts because individual level data were sparse. Data on screening and vaccination rates were additionally unavailable. Data on screening and vaccination rates

would allow for a more direct evaluation of HPV vaccination's effect on HGCL incidence. This analysis thus only provides a bird's eye view and does not necessarily draw causative conclusions.

2.5 Conclusion

A persistent cluster of high rates of HGCLs was identified in the northwestern region of the state. Public health collaborators in Connecticut should investigate the cause of unusually high burden of disease that persists over time. The presence of clusters of varying sociodemographic composition suggests that HGCLs are not unique to specific types of ethnic, racial, or sociodemographic populations. A lower case rate was found in the 20-24 year old group with a high proportion in poverty. A higher case rate was found in the 30-34 and 35-39 year old groups with a high proportion in poverty. However, the incidence of HGCLs are not changing differentially across poverty strata across all 5-year age groups from 20 to 39. This finding may be due to a "harvesting effect", which requires further research into screening rates. This methodology should be considered for use in low- and middle-income countries when establishing national HPV vaccination campaigns.

3.1 Table 1—Median proportions of census tracts that are Black, Hispanic, and in poverty amongst Cluster and Non-Cluster census tracts in two age-restricted groups (20-39 and 20-24 year old) between 2008 and 2013. Cluster and Non-Cluster census tracts correspond with those in Figure 3 and Figure 4.

Age-Restricted Sociodemographic Profile of Clusters of HGCLs						
Age	**Year**	**Cluster**	**Median Black**	**Median Hispanic**	**Median Poverty**	**N**
20-39						
	2008-2013					
		Cluster 3.1	2.60%	9.18%	9.10%	71
		Cluster 3.2	5.75%	9.65%	10.65%	30
		Cluster 3.3	6.55%	11.53%	7.25%	126
		Cluster 3.4	1.30%	4.24%	4.50%	54
		Non-Cluster	3.80%	6.47%	5.50%	548
20-24						
	2008-2013					
		Cluster 4.1	1.10%	3.85%	5.70%	85
		Cluster 4.2	1.30%	4.17%	4.50%	65
		Cluster 4.3	4.50%	7.01%	4.15%	36
		Cluster 4.4	32.90%	32.49%	21.20%	3
		Cluster 4.5	11.50%	12.15%	6.65%	82
		Cluster 4.6	0.75%	2.90%	1.75%	4
		Cluster 4.7	3.40%	2.68%	6.30%	3
		Non-Cluster	5.10%	8.07%	6.70%	550

3.2 Table 2— Median proportions of census tracts that are Black, Hispanic, and in poverty amongst Cluster and Non-Cluster census tracts in two age-restricted groups (20-39 and 20-24 year old) for two year intervals (2008-2009, 2010-2011, 2012-2013). Cluster and Non-Cluster census tracts correspond with those in Figure 5 and Figure 6.

Age-Restricted Sociodemographic Profile of Clusters of HGCLs						
Age	**Year**	**Cluster**	**Median Black**	**Median Hispanic**	**Median Poverty**	**N**
20-39						
	2008-2009					
		Cluster 5.1	1.50%	6.73%	7.90%	81
		Cluster 5.2	2.20%	4.39%	4.40%	102
		Non-Cluster	4.60%	7.84%	6.30%	646
	2010-2011					
		Cluster 5.3	2.60%	9.18%	9.10%	71
		Cluster 5.4	9.30%	20.37%	21.90%	7
		Cluster 5.5	3.35%	5.93%	5.00%	8
		Non-Cluster	3.90%	6.94%	5.70%	743
	2012-2013					
		Cluster 5.6	5.45%	8.42%	7.60%	360
		Non-Cluster	2.80%	6.16%	5.50%	548
20-24						
	2008-2009					
		Cluster 6.1	0.90%	3.74%	5.70%	59
		Cluster 6.2	2.20%	5.29%	4.95%	156
		Non-Cluster	5.30%	8.15%	6.60%	614
	2010-2011					
		Cluster 6.3	1.90%	7.09%	7.20%	42
		Cluster 6.4	2.45%	4.85%	4.65%	12
		Cluster 6.5	10.20%	19.76%	5.60%	17
		Non-Cluster	3.80%	6.96%	6.20%	758
	2012-2013					
		Cluster 6.6	0.80%	3.36%	7.10%	33
		Cluster 6.7	31.10%	32.15%	16.00%	1
		Cluster 6.8	4.60%	7.08%	4.20%	35
		Non-Cluster	4.10%	7.33%	6.40%	760

3.3 Table 3—Poisson regression model fit to predict the number of age-restricted cases of HGCLs in a census tract using Black race, Hispanic ethnicity, poverty, and time as predictors. The Poisson regression was run by 5-year age groups: 20-24, 25-29, 30-34, and 35-39 year old. A summary age group, 20-39 year old, was also included. Significant (p<0.05) values are bolded.

Age-Restricted Poisson Regression by 5-Year Age Groups											
Age Group	**N (%)‡**	**20-24†**	**p***	**25-29†**	**p***	**30-34†**	**p***	**35-39†**	**p***	**20-39†**	**p***
Characteristic											
Black											
<5%	452 (54.7)	Reference	---	Reference	---	Reference	---	Reference	---	Reference	---
5% to <10%	118 (14.3)	**0.786 (0.690 – 0.895)**	**<0.001**	0.959 (0.843 – 1.090)	1.000	1.023 (0.876 – 1.194)	0.778	1.029 (0.858 – 1.235)	0.756	0.947 (0.887 - 1.012)	0.107
10% to <20%	108 (13.1)	0.868 (0.746 – 1.011)	0.069	0.968 (0.836 – 1.120)	0.948	1.092 (0.919 – 1.298)	0.317	1.062 (0.854 – 1.319)	0.589	1.007 (0.934 - 1.087)	0.850
≥20%	149 (18.0)	0.990 (0.850 – 1.153)	0.898	0.975 (0.836 – 1.137)	0.630	1.062 (0.885 – 1.273)	0.518	0.953 (0.760 – 1.195)	0.675	1.019 (0.942 - 1.103)	0.636
Hispanic											
<5%	316 (38.2)	Reference	---	Reference	---	Reference	---	Reference	---	Reference	---
5% to <10%	183 (22.1)	**0.891 (0.796 – 0.998)**	**0.046**	0.890 (0.790 – 1.002)	0.138	0.974 (0.846 – 1.122)	0.720	1.078 (0.760 – 1.195)	0.352	0.956 (0.902 - 1.014)	0.138
10% to <20%	124 (15.0)	1.066 (0.919 – 1.237)	0.400	0.980 (0.845 – 1.135)	0.109	1.118 (0.939 – 1.331)	0.211	1.080 (0.872 – 1.264)	0.48	1.064 (0.986 - 1.147)	0.109
≥20%	204 (24.7)	**1.270 (1.077 – 1.498)**	**0.004**	0.873 (0.740 – 1.030)	0.335	0.976 (0.802 – 1.187)	0.810	1.160 (0.910 – 1.338)	0.231	1.042 (0.958 - 1.134)	0.335
Poverty											
<5%	355 (43.0)	Reference	---	Reference	---	Reference	---	Reference	---	Reference	---
5% to <10%	184 (22.3)	0.918 (0.822 – 1.025)	0.127	0.921 (0.820 – 1.035)	0.433	1.093 (0.956 – 1.250)	0.191	**1.208 (1.039 – 1.403)**	**0.014**	1.023 (0.967 - 1.083)	0.433
10% to <20%	140 (17.0)	**0.692 (0.603 – 0.793)**	**<0.001**	0.914 (0.802 – 1.042)	0.258	1.150 (0.980 – 1.348)	0.086	1.117 (0.916 – 1.362)	0.274	0.961 (0.898 - 1.029)	0.258
≥20%	146 (17.7)	**0.561 (0.483 – 0.652)**	**<0.001**	1.025 (0.875 – 1.201)	0.959	**1.360 (1.122 – 1.649)**	**0.002**	**1.403 (1.107 – 1.780)**	**0.005**	1.002 (0.925 - 1.085)	0.959
Year											
SPLINE		**0.854 (0.830 – 0.878)**	**<0.001**	**0.972 (0.945 – 0.999)**	**0.039**	0.995 (0.965 – 1.027)	0.775	0.967 (0.931 – 1.004)	0.079	**0.936 (0.923 - 0.949)**	**<0.001**

† Values are rate ratios

‡ Number of census tracts

* P-value is for χ^2 test.

3.4 Table 4—Poisson regression model fit to predict the number of age-restricted cases of HGCLs in a census tract using poverty, time, and a poverty-time interaction term as predictors. The Poisson regression was run by 5-year age groups: 20-24, 25-29, 30-34, and 35-39 year old. A summary age group, 20-39 year old, was also included. Significant (p<0.05) values are bolded.

Age-Restricted Poisson Regression by 5-Year Age Groups with Interaction Term

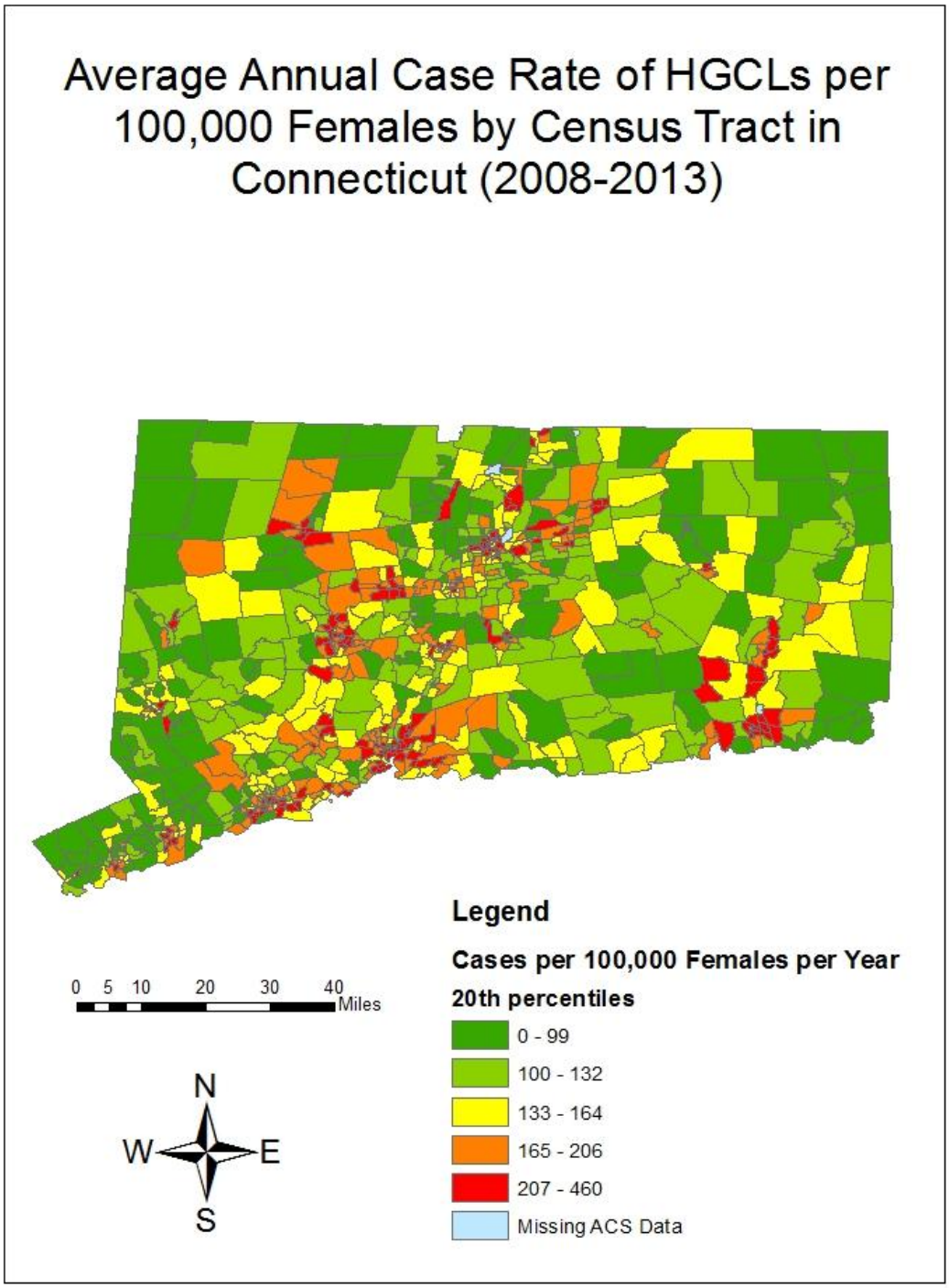
Age Group	N (%)‡	20-24†	p*	25-29†	p*	30-34†	p*	35-39†	p*	20-39†	p*
Characteristic											
Poverty											
<5%	355 (43.0)	Reference	---	Reference	---	Reference	---	Reference	---	Reference	---
5% to <10%	184 (22.3)	0.954 (0.822 - 1.106)	0.523	0.861 (0.726 - 1.022)	0.087	1.166 (0.961 - 1.415)	0.119	**1.274 (1.030 - 1.574)**	**0.025**	1.045 (0.965 - 1.131)	0.281
10% to <20%	140 (17.0)	**0.800 (0.686 - 0.934)**	**0.005**	0.954 (0.813 - 1.119)	0.561	**1.272 (1.051 - 1.539)**	**0.014**	1.235 (0.984 - 1.55)	0.069	1.044 (0.964 - 1.131)	0.289
≥20%	146 (17.7)	**0.641 (0.550 - 0.745)**	**<.0001**	0.972 (0.824 - 1.146)	0.734	**1.394 (1.149 - 1.692)**	**<0.001**	**1.312 (1.037 - 1.658)**	**0.023**	1.026 (0.946 - 1.112)	0.539
Year											
SPLINE		**0.865 (0.825 - 0.906)**	**<0.001**	0.987 (0.940 - 1.036)	0.588	1.006 (0.95 - 1.064)	0.846	0.956 (0.897 - 1.018)	0.158	**0.939 (0.917 - 0.962)**	**<0.001**
SPLINE*Poverty											
<5%		Reference	---	Reference	---	Reference	---	Reference	---	Reference	---
5% to <10%		0.966 (0.894 - 1.043)	0.374	1.025 (0.950 - 1.106)	0.528	0.976 (0.895 - 1.065)	0.591	0.982 (0.889 - 1.085)	0.726	0.990 (0.953 - 1.028)	0.601
10% to <20%		0.965 (0.891 - 1.046)	0.387	0.938 (0.870 - 1.010)	0.091	0.969 (0.889 - 1.057)	0.479	0.986 (0.886 - 1.096)	0.790	0.974 (0.937 - 1.012)	0.176
≥20%		1.000 (0.925 - 1.080)	0.813	0.976 (0.905 - 1.052)	0.524	1.010 (0.927 - 1.101)	0.813	1.096 (0.988 - 1.215)	0.084	1.021 (0.983 - 1.060)	0.282

† Values are rate ratios

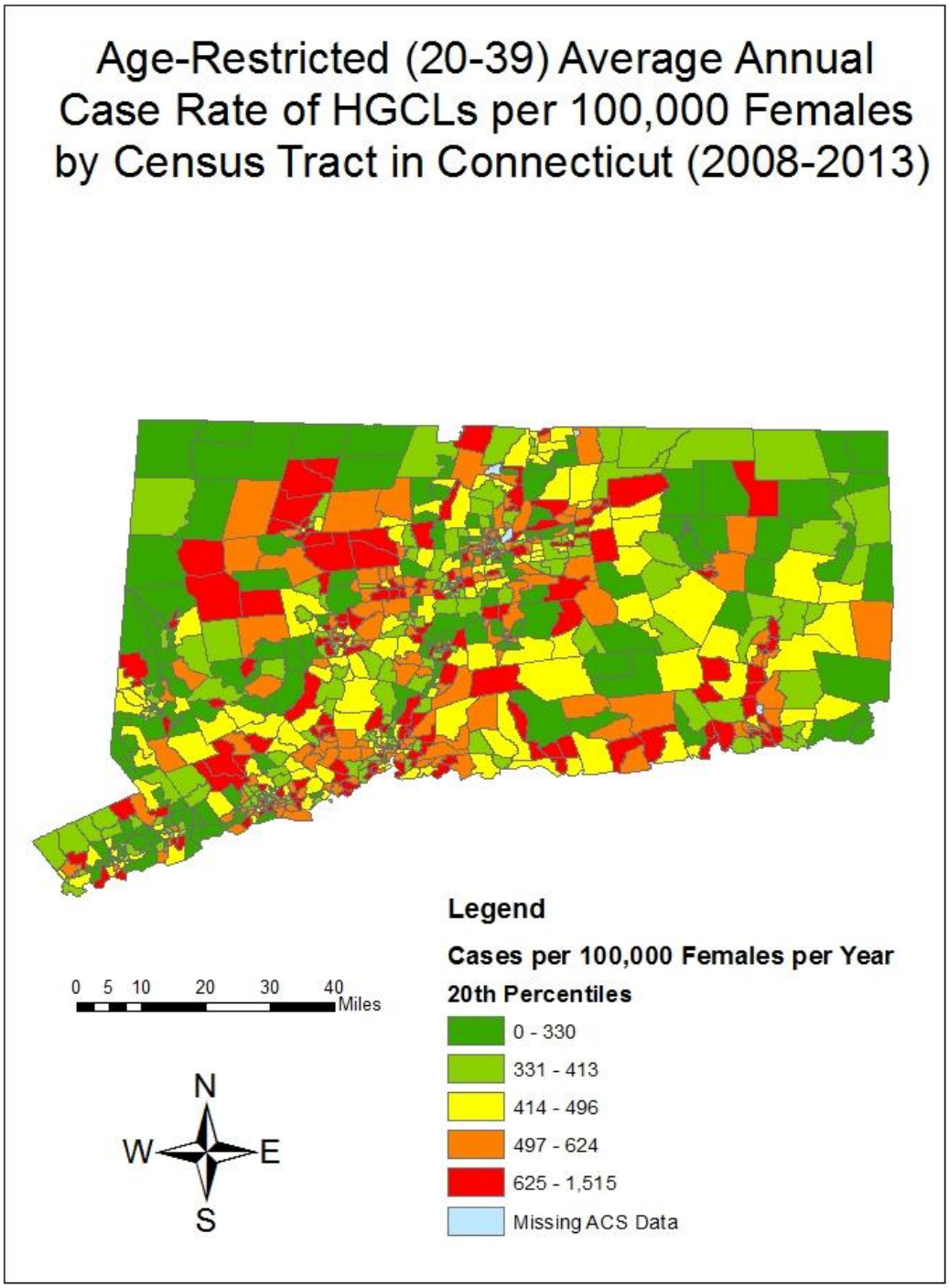
‡ Number of census tracts

* P-value is for χ^2 test

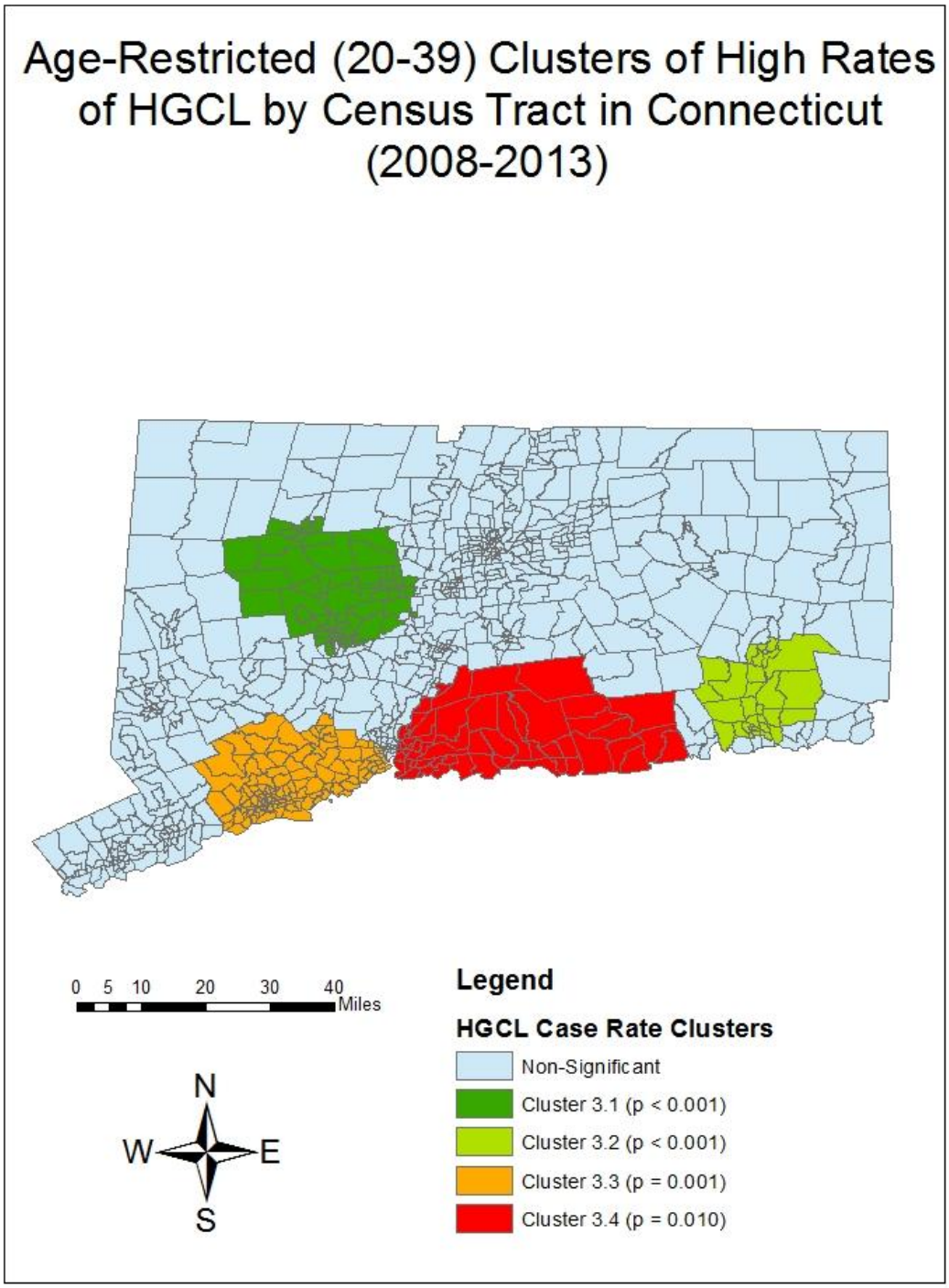
4.1 Figure 1—The average annual case rate of HGCLs per 100,000 females by census tract in Connecticut (2008-2013).



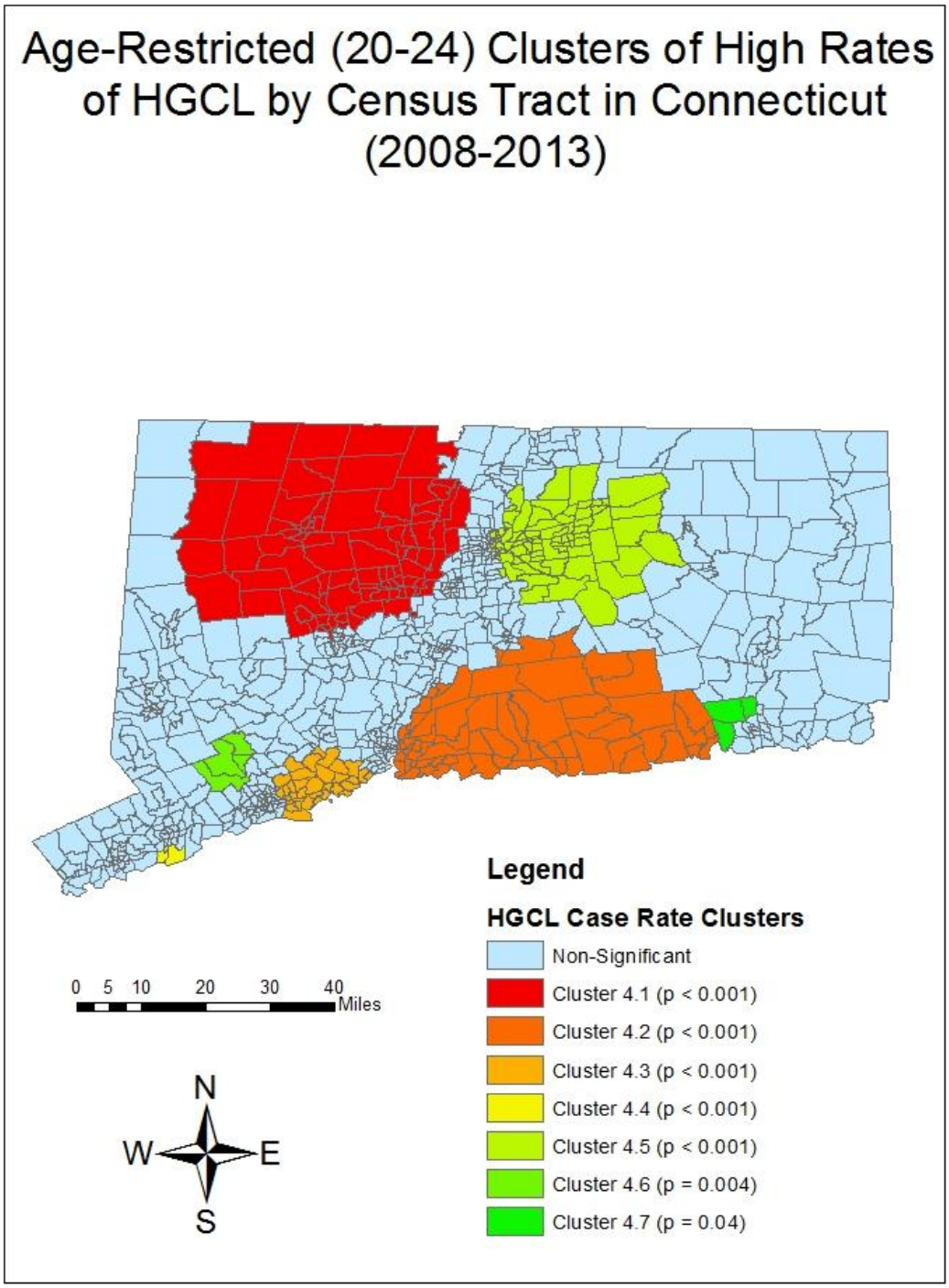
4.2 Figure 2—The age-restricted (20-39) average annual case rate of HGCLs per 100,000 females by census tract in Connecticut (2008-2013).



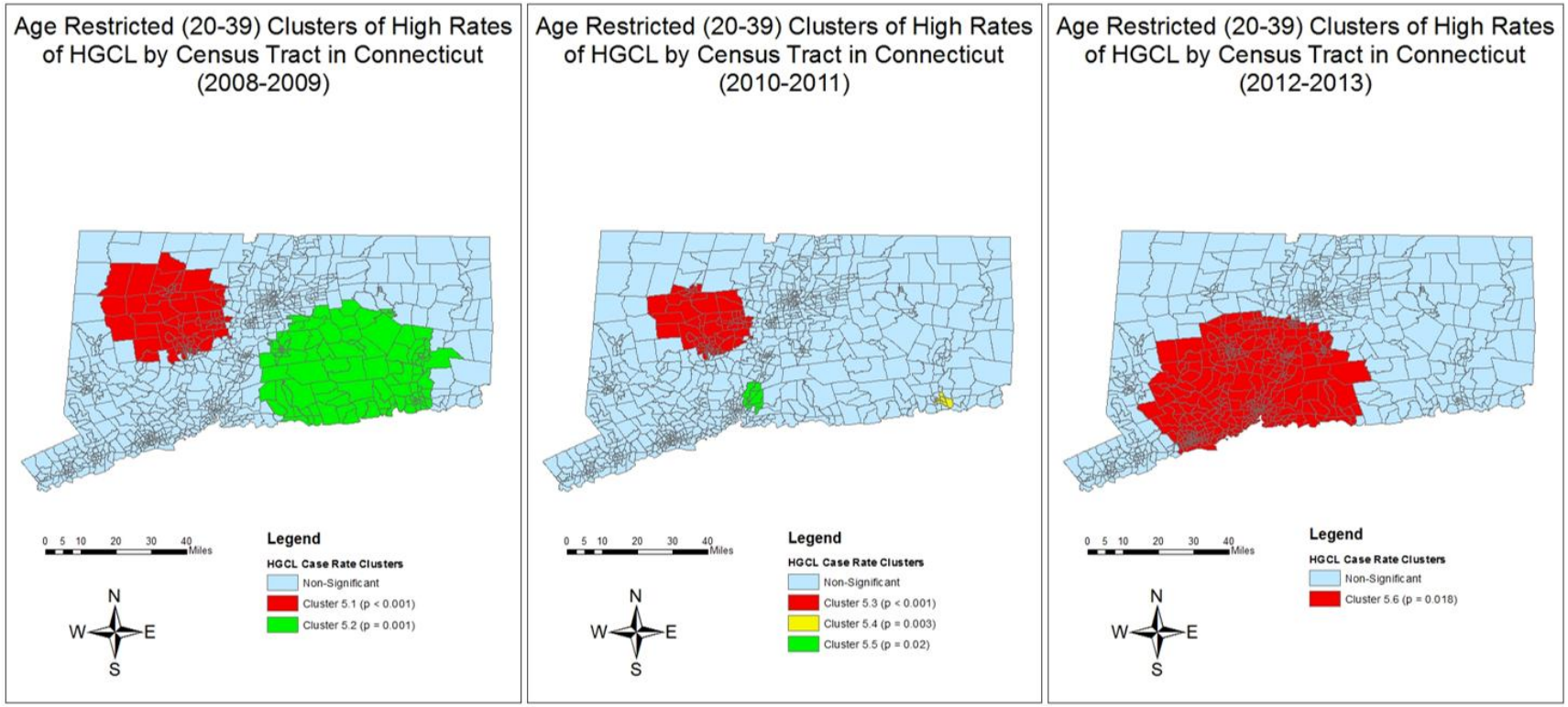
4.3 Figure 3—The age-restricted (20-39) clusters of high rates of HGCL by census tract in Connecticut (2008-2013). All clusters are statistically significant (p<0.05).



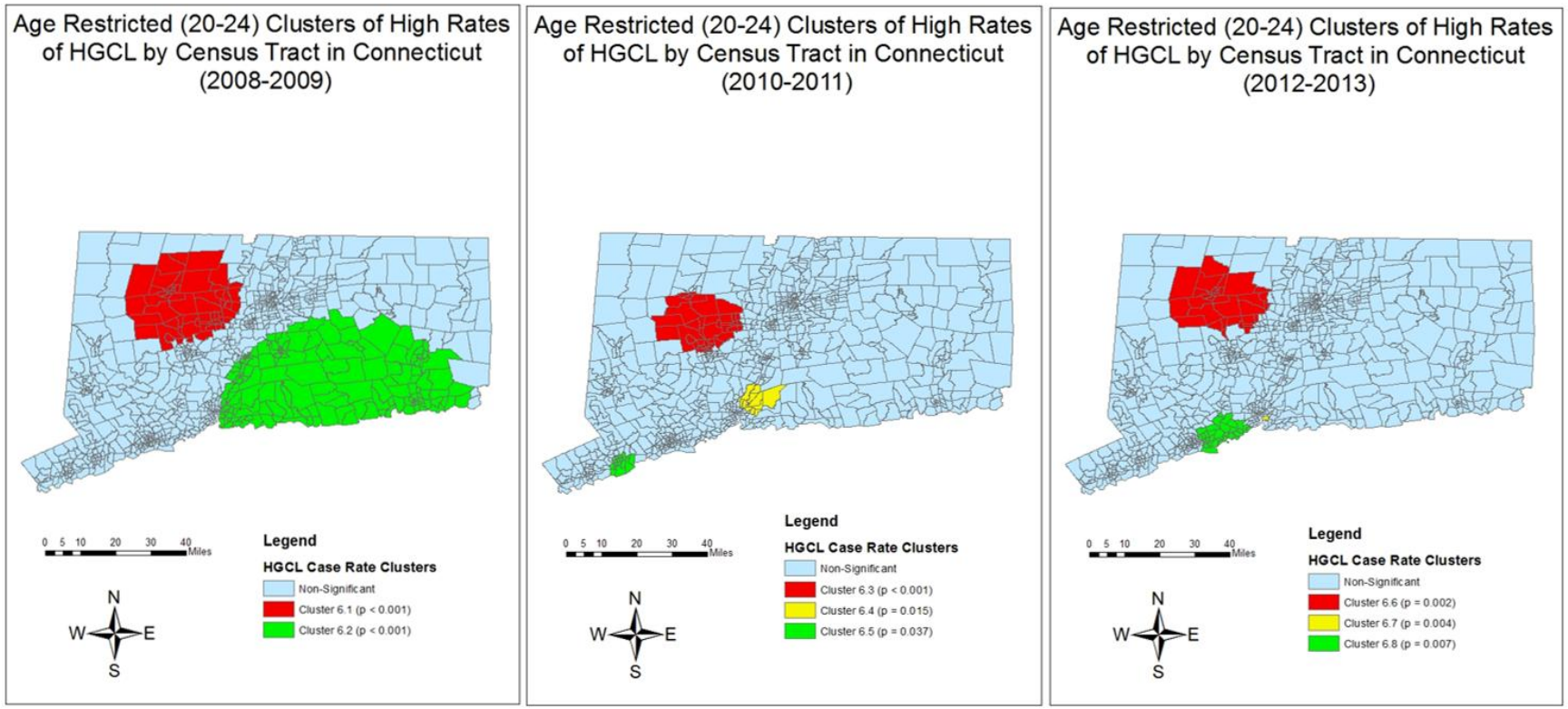
4.4 Figure 4—The age-restricted (20-24) clusters of high rates of HGCL by census tract in Connecticut (2008-2013). All clusters are statistically significant (p<0.05).



4.5 Figure 5—The age-restricted (20-39) clusters of high rates of HGCL by census tract in Connecticut across three consecutive two-year periods (2008-2009, 2010-2011, and 2012-2013). All clusters are statistically significant (p<0.05).



4.6 Figure 6—The age-restricted (20-24) clusters of high rates of HGCL by census tract in Connecticut across three consecutive two-year periods (2008-2009, 2010-2011, and 2012-2013). All clusters are statistically significant (p<0.05).



5. References

1. Schiffman, M.H., et al., *Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia.* J Natl Cancer Inst, 1993. **85**(12): p. 958-64.
2. Munoz, N., et al., *Chapter 1: HPV in the etiology of human cancer.* Vaccine, 2006. **24 Suppl 3**: p. S3/1-10.
3. Forman, D., et al., *Global burden of human papillomavirus and related diseases.* Vaccine, 2012. **30 Suppl 5**: p. F12-23.
4. Bernard, H.U., *The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses.* J Clin Virol, 2005. **32 Suppl 1**: p. S1-6.
5. Bosch, F.X., et al., *Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group.* J Natl Cancer Inst, 1995. **87**(11): p. 796-802.
6. Boshart, M., et al., *A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer.* EMBO J, 1984. **3**(5): p. 1151-7.
7. Durst, M., et al., *A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions.* Proc Natl Acad Sci U S A, 1983. **80**(12): p. 3812-5.
8. Trottier, H. and E.L. Franco, *The epidemiology of genital human papillomavirus infection.* Vaccine, 2006. **24**: p. S4-S15.
9. Walboomers, J.M., et al., *Human papillomavirus is a necessary cause of invasive cervical cancer worldwide.* J Pathol, 1999. **189**(1): p. 12-9.
10. Arbyn, M., et al., *Worldwide burden of cervical cancer in 2008.* Ann Oncol, 2011. **22**(12): p. 2675-86.
11. Group, U.C.S.W., *United States cancer statistics: 1999–2010 incidence and mortality web-based report.* Atlanta (GA): Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute, 2014.
12. Siegel, R., et al., *Cancer statistics, 2014.* CA: a cancer journal for clinicians, 2014. **64**(1): p. 9-29.
13. Simard, E.P., et al., *Widening socioeconomic disparities in cervical cancer mortality among women in 26 states, 1993-2007.* Cancer, 2012. **118**(20): p. 5110-6.
14. Food, U. and D. Administration, *FDA licenses new vaccine for prevention of cervical cancer and other diseases in females caused by human papillomavirus.* FDA News, 2006. **10**.
15. Food, U. and D. Administration, *Approval Letter-Cervarix.* 2009.
16. Kirby, T., *FDA approves new upgraded Gardasil 9.* The Lancet. Oncology, 2014.
17. Petrosky, E., et al., *Use of 9-Valent Human Papillomavirus (HPV) Vaccine: Updated HPV Vaccination Recommendations of the Advisory Committee on Immunization Practices.* MMWR Morb Mortal Wkly Rep, 2015. **64**(11): p. 300-4.
18. Elam-Evans, L.D., et al., *National, state, and selected local area vaccination coverage among children aged 19-35 months - United States, 2013.* MMWR Morb Mortal Wkly Rep, 2014. **63**(34): p. 741-8.
19. Markowitz, L.E., et al., *Reduction in human papillomavirus (HPV) prevalence among young women following HPV vaccine introduction in the United States, National Health and Nutrition Examination Surveys, 2003-2010.* J Infect Dis, 2013. **208**(3): p. 385-93.
20. Drolet, M., et al., *Population-level impact and herd effects following human papillomavirus vaccination programmes: a systematic review and meta-analysis.* Lancet Infect Dis, 2015.

21. Tabrizi, S.N., et al., *Fall in human papillomavirus prevalence following a national vaccination program.* J Infect Dis, 2012. **206**(11): p. 1645-51.
22. Ylitalo, N., et al., *A prospective study showing long-term infection with human papillomavirus 16 before the development of cervical carcinoma in situ.* Cancer Res, 2000. **60**(21): p. 6027-32.
23. Niccolai, L.M., et al., *Declining rates of high-grade cervical lesions in young women in Connecticut, 2008-2011.* Cancer Epidemiol Biomarkers Prev, 2013. **22**(8): p. 1446-50.
24. Jennings, J.M., et al., *Geographic identification of high gonorrhea transmission areas in Baltimore, Maryland.* Am J Epidemiol, 2005. **161**(1): p. 73-80.
25. Centers for Disease, C., *A cluster of Kaposi's sarcoma and Pneumocystis carinii pneumonia among homosexual male residents of Los Angeles and Orange Counties, California.* MMWR Morb Mortal Wkly Rep, 1982. **31**(23): p. 305-7.
26. Mohebbi, M., et al., *Geographical spread of gastrointestinal tract cancer incidence in the Caspian Sea region of Iran: spatial analysis of cancer registry data.* BMC Cancer, 2008. **8**: p. 137.
27. Nelson, E.J., J. Hughes, and S.L. Kulasingam, *Spatial patterns of human papillomavirus-associated cancers within the state of Minnesota, 1998-2007.* Spat Spatiotemporal Epidemiol, 2014. **9**: p. 13-21.
28. *American Community Survey*. Available from: http://www.census.gov/acs/www/.
29. *University of Connecticut Magic Library*. Available from: http://magic.lib.uconn.edu/connecticut_data.html.
30. *Federal Financial Institutions Examination Council*. Available from: http://www.ffiec.gov/geocode.
31. Darroch, J.E., D.J. Landry, and S. Oslak, *Age differences between sexual partners in the United States.* Family planning perspectives, 1999: p. 160-167.
32. Levin, A., et al., *Costs of introducing and delivering HPV vaccines in low and lower middle income countries: inputs for GAVI policy on introduction grant support to countries.* PLoS One, 2014. **9**(6): p. e101114.

6. Appendix

6.1 Table 5—Age-restricted (20-39) full Poisson regression model. Significant (p<0.05) values are bolded.

Age Restricted 20-39				
		Full Model (N=829 Census Tracts)		
Characteristic	N (%)	Beta (SE)	Rate Ratio (95% CI)	p
Black				
<5%	452 (54.7)	Reference	Reference	---
5% to <10%	118 (14.3)	-0.054 (0.034)	0.947 (0.887 - 1.012)	0.107
10% to <20%	108 (13.1)	0.007(0.039)	1.007 (0.934 - 1.087)	0.850
≥20%	149 (18.0)	0.019 (0.040)	1.019 (0.942 - 1.103)	0.636
Hispanic				
<5%	316 (38.2)	Reference	Reference	---
5% to <10%	183 (22.1)	-0.045 (0.030)	0.956 (0.902 - 1.014)	0.138
10% to <20%	124 (15.0)	0.062 (0.039)	1.064 (0.986 - 1.147)	0.109
≥20%	204 (24.7)	0.042 (0.043)	1.042 (0.958 - 1.134)	0.335
Poverty				
<5%	355 (43.0)	Reference	Reference	---
5% to <10%	184 (22.3)	0.023 (0.029)	1.023 (0.967 - 1.083)	0.433
10% to <20%	140 (17.0)	-0.040 (0.035)	0.961 (0.898 - 1.029)	0.258
≥20%	146 (17.7)	0.002 (0.041)	1.002 (0.925 - 1.085)	0.959
Year				
2008		Reference	Reference	---
2009		0.000 (0.034)	1.000 (0.936 - 1.069)	1.000
2010		**-0.097 (0.034)**	**0.908 (0.848 - 0.972)**	**0.005**
2011		**-0.111 (0.035)**	**0.895 (0.836 - 0.958)**	**0.001**
2012		**-0.224 (0.036)**	**0.799 (0.745 - 0.858)**	**<0.001**
2013		**-0.260 (0.036)**	**0.771 (0.718 - 0.828)**	**<0.001**

6.2 Table 6—Age-restricted (20-24) full Poisson regression model. Significant (p<0.05) values are bolded.

Age Restricted 20-24				
		Full Model (N=829 Census Tracts)		
Characteristic	N (%)	Beta (SE)	Rate Ratio (95% CI)	p
Black				
<5%	452 (54.7)	Reference	Reference	---
5% to <10%	**118 (14.3)**	**-0.241 (0.067)**	**0.786 (0.690 – 0.895)**	**<0.001**
10% to <20%	108 (13.1)	-0.141 (0.078)	0.868 (0.746 – 1.011)	0.069
≥20%	149 (18.0)	-0.010 (0.078)	0.990 (0.850 – 1.153)	0.898
Hispanic				
<5%	316 (38.2)	Reference	Reference	---
5% to <10%	**183 (22.1)**	**-0.115 (0.058)**	**0.891 (0.796 – 0.998)**	**0.046**
10% to <20%	124 (15.0)	0.064 (0.076)	1.066 (0.919 – 1.237)	0.400
≥20%	**204 (24.7)**	**0.239 (0.084)**	**1.270 (1.077 – 1.498)**	**0.004**
Poverty				
<5%	355 (43.0)	Reference	Reference	---
5% to <10%	184 (22.3)	-0.086 (0.056)	0.918 (0.822 – 1.025)	0.127
10% to <20%	**140 (17.0)**	**-0.368 (0.070)**	**0.692 (0.603 – 0.793)**	**<0.001**
≥20%	**146 (17.7)**	**-0.578 (0.077)**	**0.561 (0.483 – 0.652)**	**<0.001**
Year				
2008		Reference	Reference	---
2009		**0.006 (0.062)**	**1.006 (0.891 – 1.136)**	**0.923**
2010		**-0.260 (0.067)**	**0.771 (0.677 – 0.879)**	**<0.001**
2011		**-0.282 (0.067)**	**0.754 (0.661 – 0.860)**	**<0.001**
2012		**-0.464 (0.071)**	**0.629 (0.548 – 0.722)**	**<0.001**
2013		**-0.665 (0.075)**	**0.514 (0.444 – 0.596)**	**<0.001**

6.3 Table 7—Age-restricted (20-24) bivariate Poisson regression model. Significant (p<0.05) values are bolded.

Age Restricted 20-24 SPLINE Bivariate (Hispanic)				
		Full Model (N=829 Census Tracts)		
Characteristic	N (%)	Beta (SE)	Rate Ratio (95% CI)	p
Hispanic				
<5%	**316 (38.2)**	**Reference**	**Reference**	**---**
5% to <10%	**183 (22.1)**	**-0.248 (0.057)**	**0.781 (0.698 – 0.873)**	**<0.001**
10% to <20%	**124 (15.0)**	**-0.233 (0.063)**	**0.792 (0.699 – 0.897)**	**<0.001**
≥20%	**204 (24.7)**	**-0.224 (0.053)**	**0.800 (0.721 – 0.887)**	**<0.001**
Year				
SPLINE		**-0.158 (0.015)**	**0.854 (0.830 – 0.879)**	**<0.001**

Age Restricted 20-24 SPLINE Bivariate (Black)				
		Full Model (N=829 Census Tracts)		
Characteristic	N (%)	Beta (SE)	Rate Ratio (95% CI)	p
Black				
<5%	**452 (54.7)**	**Reference**	**Reference**	**---**
5% to <10%	**118 (14.3)**	**-0.318 (0.061)**	**0.727 (0.646 – 0.819)**	**<0.001**
10% to <20%	**108 (13.1)**	**-0.231 (0.060)**	**0.794 (0.705 – 0.893)**	**<0.001**
≥20%	**149 (18.0)**	**-0.178 (0.055)**	**0.837 (0.752 – 0.931)**	**0.001**
Year				
SPLINE		**-0.158 (0.015)**	**0.854 (0.830 – 0.879)**	**<0.001**

Age Restricted 20-24 SPLINE Bivariate (Poverty)				
		Full Model (N=829 Census Tracts)		
Characteristic	N (%)	Beta (SE)	Rate Ratio (95% CI)	p
Poverty				
<5%	**355 (43.0)**	**Reference**	**Reference**	**---**
5% to <10%	184 (22.3)	-0.094 (0.055)	0.911 (0.817 – 1.015)	0.090
10% to <20%	**140 (17.0)**	**-0.270 (0.058)**	**0.764 (0.682 – 0.855)**	**<0.001**
≥20%	**146 (17.7)**	**-0.433 (0.056)**	**0.649 (0.581 – 0.724)**	**<0.001**
Year				
SPLINE		**-0.158 (0.015)**	**0.854 (0.830 – 0.879)**	**<0.001**