Geospatial Analysis Of High-Grade Cervical Lesions To Address Health Disparities

Savanah M. Russ
savanah.russ@yale.edu

Follow this and additional works at: https://elischolar.library.yale.edu/ysphtdl

Recommended Citation
https://elischolar.library.yale.edu/ysphtdl/1841

This Open Access Thesis is brought to you for free and open access by the School of Public Health at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Public Health Theses by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
Geospatial Analysis of High-Grade Cervical Lesions to Address Health Disparities

Savanah Russ
Master of Public Health Thesis
May 2019

Epidemiology of Microbial Disease Department
Yale School of Public Health
Primary Reader: Dr. Linda Niccolai, PhD
Secondary Reader: Dr. Lynn Sosa, MD
**Introduction:** Incidence of high-grade cervical lesions (HGCL) caused by human papillomavirus (HPV) has declined in the U.S following the introduction of the HPV vaccine in 2006. However, disparities in HPV cervical infection and subsequent sequelae by race, ethnicity and income continue to persist. The purpose of this analysis was to identify spatial areas with significantly elevated HGCL burden in Connecticut, and determine socioeconomic characteristics associated with high incidence clusters.

**Methods:** Data from statewide surveillance in Connecticut for cervical intraepithelial neoplasia grades 2, 2/3, 3 and adenocarcinoma in situ (CIN 2+) from 2008-2016 were used for this analysis. Spatial analyses were performed using SaTScan v9.6 to identify significant clusters of HGCLs by census tract among women age 21-39 years across aggregated year groups from 2008–2010, 2011–2013, and 2014–2016. Four separate mixed effects models with varying sociodemographic covariates were constructed to assess the fit of the model to the number of HGCL cases per census tract. The likelihood of the predicted incidence for each model was used to calculate the Deviance Information Criterion (DIC), which balances goodness of fit and model complexity.

**Results:** From 2008–2016, incidence of HGCLs declined, particularly in women aged 20–29. Spatial analyses identified four significant clusters of HGCLs over time. These clusters varied in time, number of census tracts, as well as racial, ethnic and economic composition. The most recent cluster of HGCL, located in south central Connecticut among women aged 30–39, displayed significantly higher proportions Black, Hispanic and below poverty populations compared to the rest of the state. The proportion of Hispanic individuals per census tract was the most significant predictor of the number of HGCL cases.
Conclusions: The decline in HGCL incidence overall and among young women may suggest HPV vaccine impact in CT. The differences in location and socioeconomic composition of clusters suggest there are differences in incidence of HGCLs between neighborhoods of varying socioeconomic statuses. The relevant nature of ethnicity in predicting HGCL incidence may be indicative of a cultural disparity in the health coverage of Hispanic women in CT. The disparities in racial, ethnic and income characteristics in census tracts with high incidence of HGCLS indicate a need for continued surveillance and targeted interventions.
Table of Contents:

1. Abstract ..................................................................................................................2

2. Thesis Body
   a. Introduction .............................................................................................................4
   b. Materials & Methods ..............................................................................................6
   c. Results ....................................................................................................................9
   d. Discussion .............................................................................................................11
   e. Conclusion ...........................................................................................................13

3. List of Tables
   a. Table 1 ..................................................................................................................15
   b. Table 2 ..................................................................................................................16
   c. Table 3 ..................................................................................................................17
   d. Table 4 ..................................................................................................................18

4. List of Figures
   a. Figure 1 ..............................................................................................................19
   b. Figure 2 ..............................................................................................................20
   c. Figure 3 ..............................................................................................................21
   d. Figure 4 ..............................................................................................................22
   e. Figure 5 ..............................................................................................................23
   f. Figure 6 ..............................................................................................................24
   g. Figure 7 ..............................................................................................................25

5. References ............................................................................................................26
Introduction:

Human papillomavirus (HPV) is the most common sexually transmitted infection in the US with an estimated 79 million prevalent cases prior to vaccine introduction in 2006\textsuperscript{1}. Many infections are asymptomatic and transient, but persistent infection with a high-risk HPV type is a well-established necessary cause of cervical cancer\textsuperscript{2}. Two high-risk types, HPV 16 and 18, are responsible for 70\% of cervical cancers. Along the continuum to carcinoma, HPV infections also cause high-grade cervical lesions (HGCL) that affect nearly 500,000 women in US annually, of which ~50\% are caused by high-risk types HPV 16/18\textsuperscript{3}. In addition to being a precursor to invasive disease, HGCL are an important public health consideration for several reasons including high disease burden, costs, and associated health care utilization and psychological distress\textsuperscript{4, 5, 6}.

Differences by race, ethnicity, and income exist throughout the natural history of cervical HPV infections. National prevalence studies show that HPV infections are most common in black and low-income women\textsuperscript{7, 8, 9}. Previous work has demonstrated disparities by area-based measures of race, ethnicity, and poverty in HGCL\textsuperscript{10}. Disparities also pervade the continuum of the ‘cancer disparities grid’ for cervical cancer including diagnosis, treatment, and survival\textsuperscript{11}. Specifically, women living in poverty and racial/ethnic minorities continue to bear a disproportionate burden of incidence and mortality despite the decrease in rates that has resulted from widespread cervical cancer screening\textsuperscript{12, 13, 14, 15, 16, 17}.

Differences by race, ethnicity, and income that are described as health disparities may reflect more complex processes and social inequities\textsuperscript{18}. Observable individual characteristics, such as race, may signal contextual or geographic factors that are more difficult to measure\textsuperscript{11}. Contextual factors are social, cultural, political, and economic characteristics such as poverty,
discrimination, racism, sex ratios, interpersonal networks, and residential segregation that differ between racial and ethnic groups and likely contribute to persistent health disparities\textsuperscript{19}. Other factors may reflect the geographic context, that is, characteristics that operate at the level of place such as neighborhoods. These factors may include the built environment, access to health services, and community characteristics with known impacts on health outcomes. Previous research had displayed geographic distribution of cervical cancer at the county level in the United States, with spatial clustering of cervical cancer mortality related to high proportion black population, low socioeconomic status and low health care coverage\textsuperscript{20,21,22}.

Health disparities observed across incidence of HPV-related disease, specifically HGCLs, may be further understood through geospatial analyses that can provide insight into factors relating to location that may be driving health disparities. This, in turn, can identify areas at the local level for targeted interventions. These types of analyses are particularly salient for HPV-related diseases as identifying particular geographic areas where there may be a need for additional resources, such as access to more timely screening or HPV vaccination programs. To evaluate disparities in HGCL incidence, we sought to identify spatial areas with significantly elevated HGCL burden in Connecticut, and to determine the salient demographic & socioeconomic characteristics associated with high HGCL incidence.

\textbf{Materials & Methods:}

In 2008, the Centers for Disease Control and Prevention began to monitor the impact of HPV vaccination through population-based surveillance of cervical intraepithelial neoplasia grades 2, 2/3 or 3 and adenocarcinoma in situ (CIN2+) conducted by five of the Emerging Infections Program sites. At the Connecticut (CT) HPV-IMPACT site, the CT Department of Public Health added CIN2+ to the list of reportable diseases statewide, effective January 1, 2008.
All 34 pathology laboratories that have served CT residents are in compliance with the reporting requirement. These laboratories are regularly contacted to ensure ongoing, complete, and timely reporting, and quality assurance protocols are routinely implemented. Pathology laboratories that collectively report >80% of cases are routinely audited for completeness. This work has been deemed public health surveillance by university, state, and federal institutional review boards and thus exempt from the need for human subject approval.

All reported cases were individually geocoded by residential address to the census tract level using the Federal Financial Institutions Examination Council (FFIEC) Geocoding/Mapping System database. Previous research has shown that census tract-based measures that describe the neighborhoods in which people live reflect important aspects of social context and are important for determinants for health\(^\text{23}\). These geocodes were then matched to American Community Survey data for census tract-based measures of race (% black), ethnicity (% Hispanic) and poverty (% living below the federal poverty line) for that year of HGCL diagnosis.

Data for this analysis included the period January 1, 2008–December 31, 2016. The number of HGCL cases were tabulated for the 833 census tracts in the state of Connecticut, as well as grouped by year and by age at first diagnosis. For women with multiple reports, only the first CIN 2+ diagnosis was included in this analysis. Cases during the study time period were aggregated by year of diagnosis into three time periods (2008–2010, 2011–2013, 2014–2016) to analyze temporal changes in geographic distribution of HGCL cases.

Spatial analyses were performed using SaTScan v9.6 to identify clusters of census tracts in Connecticut having significantly elevated incidence of HGCL compared to the expected number of cases for that census tract. A purely spatial analysis using a discrete Poisson model and a Monte Carlo simulation using 999 permutations was used to compare the aggregated year
groups of case data to the null hypothesis that women diagnosed with HGCLs were equally geographically distributed across all census tracts in Connecticut. A cluster was considered to be significant if the Monte Carlo simulation yielded a p-value of ≤0.05. Overlapping clusters were not permitted for this analysis, and the percent of the population at risk was set to 50%. The input parameter for the number of cases was the observed number of HGCL cases per census tract, and the population denominator was the predicted number of cases per census tract following adjustment for age and year of diagnosis, percent Black population, percent Hispanic population, and percent below poverty population using a generalized linear model, assuming the number of cases followed a Poisson distribution. Nine different runs of the spatial analysis were performed: one for each aggregated year group across all ages, one for each aggregated year group for women 21–29 years of age at first diagnosis, and one for each aggregated year group for women 30–39 years of age at first diagnosis. Geospatial distributions of significant clusters were mapped using ArcMap v10.4.1 and shapefiles obtained from the University of Connecticut GIS Data Library.

To evaluate the relationship between the number of women in each census tract with a HGCL and fixed covariates, including age and year of primary HGCL diagnosis, and area-based measures of proportion Black, Hispanic and below poverty populations per census tract, a mixed effects model was constructed. In this model, census tracts were treated as spatially independent. A Markov Chain Monte Carlo simulation was used to sample from the posterior distributions defined in the model for random effects values of the model, as well as for the coefficients for each fixed covariate, to estimate the predicted number of HGCL cases by census tract. Four separate mixed effects models were run to assess the success of the model in predicting the number of HGCL cases per census tract, given the fixed covariates included in the model. Model
1 included age and year at diagnosis, while models 2–4 included age, year and one of the three area-based measures of race, ethnicity or poverty. Age was aggregated into two-year groups, and area-based measures of race, ethnicity and poverty were scaled by the mean and standard deviation of each measure. The likelihood of the predicted incidence values was used to calculate goodness-of-fit criteria for each model, including Deviance Information Criterion (DIC) and the effective number of parameters ($P_D$), which balances goodness of fit and model complexity. All DIC and $P_D$ values were compared to the reference model including only age and year. R software version 3.4.3 was used for this analysis.

**Results:**

A total of 16,038 women were diagnosed with a high-grade cervical lesion between January 1st, 2008–December 31st, 2016 and reported to the Connecticut Emerging Infections Program and the Connecticut Department of Public Health. Of these women, 9567 (59.6%) were 20–29 years of age at their first HGCL diagnosis and 6471 (40.3%) were 30–39 years of age (Table 1). Of the 16,038 HGCL cases reported between 2008–2016, 97% were successfully matched to a census tract. HGCL cases were reported from 823 of 833 (98%) census tracts with annual incidence rates from 2008–2016 ranging from 0–2840 cases per 100,000 (median 1330 per 100,000).

Figures 1–3 show the incidence rates by census tract of the observed and adjusted number of HGCL cases. After adjusting for age and year of HGCL diagnosis, proportion Black, Hispanic and below poverty by census tract, the incidence rates by census tract is more equally spatially distributed and no longer show aggregation around urban areas.
There were four significant clusters of HGCL detected using the SaTScan method across the aggregated year and age groupings (Table 2). Two significant clusters were detected across all age groups: one among the 2008–2010 data (1.1) (Figure 4) and one among the 2014–2016 data (3.1) (Figure 6). The number of census tracts included in these clusters range from 78 to 91. A third significant cluster was detected among the 21–29 year olds in 2008–2010, with 27 census tracts included in the cluster (2.1) (Figure 5). The fourth significant cluster was detected among the 30–39 year olds in 2014–2016, with 11 census tracts included in the cluster (4.1) (Figure 7). All of the census tracts included in cluster 4.1 were also included in cluster 3.1.

Compared to state estimates for area-based measures of socioeconomic status, the census tracts included in clusters 1.1, 2.1 and 3.1 had a lower median % Black population, median % Hispanic population and median % population below poverty. The census tracts included in cluster 4.1, however, had higher median % Black population, median % Hispanic population and median % population below poverty than the respective estimates for the state of Connecticut (Table 3). The racial, ethnic and income composition of cluster 3.1 and 4.1 varied significantly, even with the clusters overlapping on census tracts and in year group. Clusters 1.1 and 2.1 were similar in their area-based measures of socioeconomic status composition, with both falling under the state median estimates for all three categories.

After assessing goodness-of-fit criteria for each mixed effects model, the fixed covariates that, when included, yielded the lowest DIC score were age, year and proportion Hispanic per census tract, and were therefore the most predictive of the number of HGCL cases per census tract (Table 4). The model with the highest DIC score, and was therefore least predictive of the number of HGCL cases, was the reference model including only age and year of first HGCL diagnosis.
Discussion:

Declines in overall incidence of HGCLs, as well as in the age group 21–29 year olds, from 2008–2016 is suggestive of an impact from the introduction of the HPV vaccine, as previously shown for this age group\(^7\). The unchanging incidence of HGCLs in the 30–39 year old women in Connecticut is suggestive that women in this group having lagging HPV vaccine coverage, were ineligible to receive the vaccine due to age, or received the HPV vaccine following HPV infection that caused them to develop a HGCL. With the coverage of the nonavalent HPV vaccine for all children under age 18 in Connecticut through the CT Vaccine Program, along with proposed legislation for a mandate of HPV vaccination for seventh graders in Connecticut, the HPV vaccination coverage in CT should increase to the Healthy People 2020 goal of 80% coverage.

The spatial clustering of HGCL cases after adjustment for age, year and area-based measures indicates an underlying disparity or phenomenon occurring in those census tracts that cannot be explained by the fixed covariates included in the analysis, such as individual level sociodemographic characteristics, and abnormally low vaccine coverage. While these clusters only span three-year aggregated year groups and do not persist beyond these time periods, the significant cluster in central Southern Connecticut, composed largely of 30–39 year old women diagnosed with HGCL, indicates a need for continued surveillance of this population of women at a state level to assess persistence and socioeconomic composition of these clusters. The high median proportions Black, Hispanic and below poverty, as compared to the state medians, in this most recent cluster may lead to greater insights into the differences in vaccination or screening for cervical cancers that may be driving higher than expected incidence.
The high and geographically widespread incidence of high-grade cervical lesions in 823/833 census tracts in Connecticut between 2008–2016 supports the continued need for surveillance. Incidence, prior to adjustment for age and year at diagnosis, and area-based measures of race, ethnicity and poverty, is centered largely around the urban areas of Connecticut. Following the adjustment for age, year and area-based measures of race, ethnicity and poverty, incidence is more evenly distributed across census tracts which is suggestive of the salient influence nature that neighborhood level characteristics have on HGCL burden distribution through this time period. Potential drivers behind the observed disparate incidence of HGCLs across Connecticut include variability in vaccine-uptake by neighborhood level sociodemographic characteristics, as well as availability of health insurance and access to screening prior to the progression of disease to HGCL.

The effect of these area-based measures on predicting the number of HGCL cases is greatest when using proportion Hispanic, followed by proportion Black and then proportion below poverty. While previous studies have indicated the strength of the associations between poverty and CIN2+ incidence, the role of race and ethnicity in predicting incidence of HGCL has been less clear previously as women who are Black and Hispanic have been shown to have a lower risk for the diagnosis of CIN3 when compared to White, non-Hispanic women\textsuperscript{24}. While Hispanic populations in the United States have not experienced large declines in the number of uninsured individuals since the expansion of Medicaid eligibility, Hispanic women as a sub-demographic population have seen a greater decline in the number of uninsured and an increase in cervical cancer screening rates since Medicaid expansion in 2014\textsuperscript{25,26}. This may result in a future increase in the number of Pap smears and catching of HPV infection and HGCLs earlier.
during the natural history of HPV infection, resulting in faster treatment and decreased incidence of HGCL beyond 2016.

Some limitations of this analysis should be noted. The only fixed covariates included in the adjustment for the population denominator for SaTScan analysis were those that were available, which excluded individual level race, ethnicity and poverty, HPV vaccine coverage by census tract, and screening history prior to HGCL diagnosis. Without the availability of HPV vaccination, differences in local healthcare systems, or cervical cancer screening data at the census tract level, we cannot truly understand what may be driving HGCL incidence at a local level. Within the mixed effects, a full model containing all fixed covariates could not be performed due to inability of the model to converge from collinearity between area-based measures of race and poverty, which may have resulted in an under or overestimate the DIC goodness of fit values produced for this analysis with a single area-based measure included.

Findings for this study may not be generalizable beyond Connecticut. Important strengths of this study include robust data collection performed through mandated population-based surveillance case ascertainment of women diagnosed with HPV-related precancerous lesions, as well as census-tract level information about HGCL incidence.

Conclusion:

High HPV vaccination coverage, along with consistent and timely cervical cancer screening across all sociodemographic groups of women in CT will best lower incidence of HGCLs. The continuation of Medicaid expansion eligibility will be critical for ensuring that women of all races, ethnicities and income-levels receive the proper preventive and therapeutic care necessary to limit differences in HGCL incidence seen across various sociodemographic groups of women. Vaccine impact on age-specific incidence of HGCL may be observed in these findings, however,
continued surveillance of HGCL incidence by age and location will be necessary as incidence in later age groups is unchanging. Future work in monitoring vaccination at a finer geographic level, such as census tract, will be necessary to understand drivers behind differences in HGCL incidence at a local level.
Table 1. Total Number of High-Grade Cervical Lesion Cases, 2008–2016

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of HGCL Cases</th>
<th>Annual Incidence Rate(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008–2010</td>
<td>6189</td>
<td>9100</td>
</tr>
<tr>
<td>2011–2013</td>
<td>5403</td>
<td>7900</td>
</tr>
<tr>
<td>2014–2016</td>
<td>4446</td>
<td>6500</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>21-29 Years of Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008–2010</td>
</tr>
<tr>
<td>2011–2013</td>
</tr>
<tr>
<td>2014–2016</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>30-39 Years of Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008–2010</td>
</tr>
<tr>
<td>2011–2013</td>
</tr>
<tr>
<td>2014–2016</td>
</tr>
</tbody>
</table>

\(^1\)Across a three-year period per 100,000
Table 2. Relative Risk in Clusters of High-Grade Cervical Lesions, 2008–2016

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Year/Age Group</th>
<th>Relative Risk¹</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>2008–2010 All Ages</td>
<td>1.24</td>
<td>0.0016</td>
</tr>
<tr>
<td>2.1</td>
<td>2008–2010 21–29</td>
<td>1.48</td>
<td>0.0260</td>
</tr>
<tr>
<td>3.1</td>
<td>2014–2016 All Ages</td>
<td>1.25</td>
<td>0.0077</td>
</tr>
<tr>
<td>4.1</td>
<td>2014–2016 30–39</td>
<td>2.22</td>
<td>0.0130</td>
</tr>
</tbody>
</table>

¹Observed number of cases per census tract relative to the expected number
Table 3. Median Area-Based Measures of Socioeconomic Status by Cluster, 2008–2016

<table>
<thead>
<tr>
<th>Cluster Number</th>
<th>Year/Age Group</th>
<th>Median % Black</th>
<th>Median % Hispanic</th>
<th>Median % Below Poverty</th>
</tr>
</thead>
<tbody>
<tr>
<td>State</td>
<td>2008–2010</td>
<td>3.3</td>
<td>6.4</td>
<td>6.2</td>
</tr>
<tr>
<td>Cluster 1.1</td>
<td>2008–2010 All Ages</td>
<td>2.7</td>
<td>5.2</td>
<td>4.7</td>
</tr>
<tr>
<td>Cluster 2.1</td>
<td>2008–2010 21–29</td>
<td>2.3</td>
<td>4.1</td>
<td>5.0</td>
</tr>
<tr>
<td>State</td>
<td>2011–2013</td>
<td>3.8</td>
<td>7.0</td>
<td>6.4</td>
</tr>
<tr>
<td>State</td>
<td>2014–2016</td>
<td>4.0</td>
<td>8.2</td>
<td>6.8</td>
</tr>
<tr>
<td>Cluster 3.1</td>
<td>2014–2016 All Ages</td>
<td>1.2</td>
<td>4.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Cluster 4.1</td>
<td>2014–2016 30–39</td>
<td>6.5</td>
<td>12.6</td>
<td>21.4</td>
</tr>
</tbody>
</table>
Table 4. Associations Between HGCL and Area-Based Measures of Race, Ethnicity & Poverty, Age and Year

<table>
<thead>
<tr>
<th>Model Number</th>
<th>Fixed Covariates Included</th>
<th>DIC</th>
<th>Effective Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Age, Year</td>
<td>98867.8</td>
<td>456.9</td>
</tr>
<tr>
<td>2</td>
<td>Age, Year, % Below Poverty</td>
<td>98685.6</td>
<td>382.2</td>
</tr>
<tr>
<td>3</td>
<td>Age, Year, % Black</td>
<td>98680.2</td>
<td>358.4</td>
</tr>
<tr>
<td>4</td>
<td>Age, Year, % Hispanic</td>
<td>98564.2</td>
<td>278.1</td>
</tr>
</tbody>
</table>
Figure 1. Observed and adjusted annual incidence rate per 100,000 of high-grade cervical lesions by census tract, 2008-2010. A, Observed IR of high-grade cervical lesions by census tract, 2008-2010. B, Adjusted incidence rate, which adjusts for age and year at HGCL diagnosis, and area-based measures of race, ethnicity and poverty by census tract.
Figure 2. Observed and adjusted annual incidence rate per 100,000 of high-grade cervical lesions by census tract, 2011-2013. A, Observed IR of high-grade cervical lesions by census tract, 2011-2013. B, Adjusted incidence rate, which adjusts for age and year at HGCL diagnosis, and area-based measures of race, ethnicity and poverty by census tract.
Figure 3. Observed and adjusted annual incidence rate per 100,000 of high-grade cervical lesions by census tract, 2014-2016. A, Observed IR of high-grade cervical lesions by census tract, 2014-2016. B, Adjusted incidence rate, which adjusts for age and year at HGCL diagnosis, and area-based measures of race, ethnicity and poverty by census tract.
Figure 4. Cluster 1.1, State All Ages, 2008-2010
Figure 5. Cluster 2.1, State 20-29 Years of Age, 2008-2010
Figure 6. Cluster 3.1, State All Ages, 2014-20
Figure 7. Cluster 4.1, State 30-39 Years of Age, 2014-2016
References:


24 Socioeconomic status and the risk of cervical intraepithelial neoplasia grade 3 among oncogenic human papillomavirus DNA-positive women with equivocal or mildly abnormal cytology. https://doi.org/10.1002/cncr.21129

