International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-18
Available Online at http://acascipub.com/Journals.php

Research article

Employing ArcGIS Spatial Analyst extensions and linear regression statistics to prioritize vulnerability to cervical cancer in a georeferenced grid-stratified zip code polygon geographically classified in Hillsborough County, Florida

Yessica C. Martinez[a*], ToniPanaou[b], Samuel Alao[c], Benjamin G. Jacob[c]

[a] Department of Epidemiology and Biostatistics, University of South Florida, Tampa, FL, USA

[b] Department of Civil and Environmental Engineering, University of South Florida, Tampa, FL, USA

[c] Department of Global Health, University of South Florida, Tampa, FL, USA

*Corresponding author: martinezy@health.usf.edu

Abstract

Cervical cancer disparities continue to exist even with the overall decline in cervical cancer incidence and mortality rates. Although many studies have been conducted to assess the factors which may be disproportionately affecting cervical cancer incidence rates among women, there seems to be a lack of geospatial, cartographic, analysis data which corroborates and identifies statistically significant predictive variables which may pinpoint georeferenceablevulnerable populations. Current literature and research allude to the fact that socioeconomic factors play a major role in the disproportionality of cervical cancer incidence rates. As socioeconomic status decreases, so does health literacy, follow-up rates for abnormal cytology, and the ability to access quality medical care and screening services (Akers *et al.* 2007). By exploring spatial influences which may contribute to the unequal distribution of cervical cancer, we can begin to guide current policies, strategies, and interventions in order to find population specific alternatives. In this study, we employed Spatial Analyst extensions in ArcGIS and a multiple linear regression model to find exploratory predictive variables at the zip code level within our Miami-Dade epidemiological, study site. Three robust explicatory predictors with a 95% confidence level were identified in this study: median income, being uninsured, and the number of medical facilities in a county georeferenced polygon (healthcare access). We then interpolated these variables to create a stratified vulnerabilityindex so as to develop strategies to optimize the usage of allocated resources and targeted funding in Hillsborough County, Florida.

Keywords: cervical cancer; disparities; vulnerability index; incidence; GIS/SAS

1. Introduction

Although cervical cancer was once the number one cause of cancer for women in the United States, it is now considered a rare type of cancer affecting about 12,000 women annually (CDC, 2017b). In 1976 the German virologist HaraldzurHausenintroduced the hypothesis that human papillomavirus played a vital role in the cause of cervical cancer. Today, not only is cervical cancer the only type of cancer preventable with a vaccine, but incidence rates in the United States have seen steady declines (CDC 2017a). However, even though declining rates have been witnessed in all women regardless of race and/or ethnicity, there still remains a stable disproportionality among incident cases (CDC 2017a). The American Cancer Society (2017) reported that Hispanic women, Black women, and Asian women continue to have the highest incidence rates among all races; although Blacks and Whites have been noted to have similar screening rates(Akers *et al.* 2007). The state of Florida ranks third in the country for its cervical cancer rates, which may be attributable to a diverse and multicultural population comprising of more than 40% minorities (United States Census Bureau 2016).

Regression models have the capacity to generate robust predictors at a 95% confidence level (Hosmer and Lemeshow 2000)which have been used for numerous diseases (Jacob *et al.* 2003, Alao *et al.* 2016). By determining regression covariates of significance to an explanatory dependent variable such as incidence of cervical cancer at the county level, clinicians and other epidemiologists may be able to better target vulnerable populations at the zip code level within a county georeferenceablepolygon. Endogenous diagnosticexplanatorialregressors can reveal areas of hyperendemic foci (Griffith 2005, Jacob *et al.*2009, Alao *et al.* 2016). Remote sensing, GIS, and statistical remote algorithms can develop predictive, grid-stratifiablecartographic maps of regression models (Wood *et al.*1991, Hay et al. 1998, Kitron 1998). GIS epidemiological model outputs have been employed to support exploratory analysis, statistical and computational testing of hypotheses, policy decision making, dissemination of information in a variety of forms, resource allocation & prioritization, and program evaluation (GIS for Government 2017).

Currently in literature, a regression model exists for determining proxy correlation of HPV vaccination and screening for cervical cancer at the national level (Kish *et al.* 2016). Likewise, GIS has been used for spatial

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-18
Available Online at http://acascipub.com/Journals.php

analysis of access to health care as well as identification of geostatistically significant characteristics among invasive cervical cancer clusters in New Jersey (Wan*et al.*2011, Roche*et al.*2015). However, our research differs in that it combines the use of a regression model and GIS to create a vulnerability index which can be used to prioritize direct allocation of funding and resources at the zip code level.

We began by employing a grid-stratification algorithmic process at the county level using incidence as the dependent variable. We found that Miami-Dade County had the largest incidence rate among all Florida counties. We then constructed a linear regression model to determine which covariates had statistical significance at a 95% confidence level which were associated with county level incidence. These were used to determine cervical cancer vulnerability in Hillsborough County. Our objectives were:

(1) To create a georefernced, GIS spatial cluster model to determine the county with the highest incidence rate for the disease.
(2) To construct aregression model using covariates to determine predictors of the disease in Miami-Dade County.
(3) To interpolate the regression estimate from the high spatial cluster covariate to deduce vulnerable regions to cervical cancer in Hillsborough County.

2. Materials and Methods

2.1. Study Site

The state ofFlorida is located in the southeastern region of the United States and covers approximately 2.59 square kilometers (km). The state is bordered by the Gulf of Mexico to its West, by Alabama and Georgia to its North, by the Atlantic Ocean to its East, and by the Straits of Florida and Cuba to its South. A peninsula between the Gulf of Mexico, the Atlantic Ocean, and the Straits of Florida, it has the longest coastline in the contiguous United States, approximately 2,170 km, and is the only state that borders both the Gulf of Mexico and the Atlantic Ocean. Much of the state is at or near sea level and is characterized by sedimentary soil. The climate varies from subtropical in the north to tropical in the south(Florida Department of State 2017).

Jacksonville is the most populous municipality in Florida and the largest city by area in the contiguous United States(Factual Facts 2017). Florida's most populous county is Miami-Dade with an estimated 2,600,900 inhabitants from many diverse backgrounds(Florida Department of State 2017). About half of Miami's residents were born outside of the United States (US) and two-thirds speak a language other than English (The Beacon Council 2017). Florida's state capital is the city of Tallahassee.

2.2. Data Attainment

Data of incident casesof cervical cancer wasattained from the National Cancer Institute's State Cancer Profiles; cases included women of all ages, races, and ethnicities from 2010 – 2014 (National Cancer Institute 2017). Geographic features and attributes such as zip code polygons, county boundaries, and base maps were providedby (Campbell 2013,USPS and Hillsborough County 2017, Florida Geographic Data Library 2012,Esri 2017). Demographic data including female population by zip code, median income, number of uninsured women, and

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-18
Available Online at http://acascipub.com/Journals.php

number of medical facilities were acquired through (United States Census Bureau 2016,Florida Geographic Data Library 2014).

2.3. Distribution Maps

Several maps were created in order to gain a spatial perspective of cervical cancer distribution cases by county and/or zip code. Figures1 through 5 were obtained by adding either a zip code shapefile or a county boundary shapefileto a base map. ArcGIS software was then used to transfer demographic data into the base map; joining tables by feature attribute, ArcGIS is able to showcase the data. Symbologies were then altered with distinctive colors for better visual representation. The healthcare access map, Figure 4(a), was linked by using a spatial join of polygons to points; in this manner, we were able to get an exact number of facilities in each zip code polygon.

2.4. Regression Analysis

We used a multi-regression technique to estimate a cervical cancer, county level, diagnostic, forecast, vulnerability model which had more than one outcome variable. When there is more than one independent, explanatory, predictor variable in a multivariate regression model, the model would be a multivariate multiple regression model (Hosmer and Lemeshow 2000). We employed the r^2 (coefficient of determination) to evaluate the model fit.

In the model the r^2 was tabulated 1 minus the ratio of residual variability. When the variability of the values around the regression line is relative to the overall variability, the prediction from the regression equation is robust (Brunsdon*et al.* 2002). For example, if there is no relationship between the *X* and *Y* variables in a cervical cancer, county level, epidemiological, forecast model, then the ratio of the residual variability of the *Y* variable to the original variantsis equal to 1 and the r^2 would be 0. If *X* and *Y* are perfectly related, then there is no residual variance and the original variance is equal to 1, then r^2would be -1. In most regression predictive equations, the statisticalsignificance of the explanatory variable would fall between 0 and 1(Fox 1997). In this analysis, we optimally quantitated the county level sociodemographic and other variables at a 95% confidencelevel.

2.4.1. Regression error quantification

In this analysis, we quantitated Type I and Type II regression error in the cervical cancer probabilistic predictive model. If the negative log likelihood is greater than the chi^2 distribution at 95%, then there is a disregard of a true hypothesis. Alternatively, if the 95% chi^2 distribution was greater than the negative log likelihood, then there is an acceptance of a false hypothesis (Hosmer and Lemeshow 2000). We validated our cervical cancer county level model for Type I and Type II errors.

3. Results

Our foremost result was the development of a vulnerability index with the capability to predict populations most at risk for cervical cancer at the zip code level. Figure 1 displays a cartographic representation of county level crude incidence counts in the State of Florida. A total of 67 counties were included in the map and 25 counties with less than 3 counts were suppressed. The counties with the highest counts were Miami-Dade (n = 145), Broward (n = 94), Hillsborough (n = 70), Palm Beach (n = 60), Orange (n = 52), Pinellas (n = 45), and Duval (n = 44).

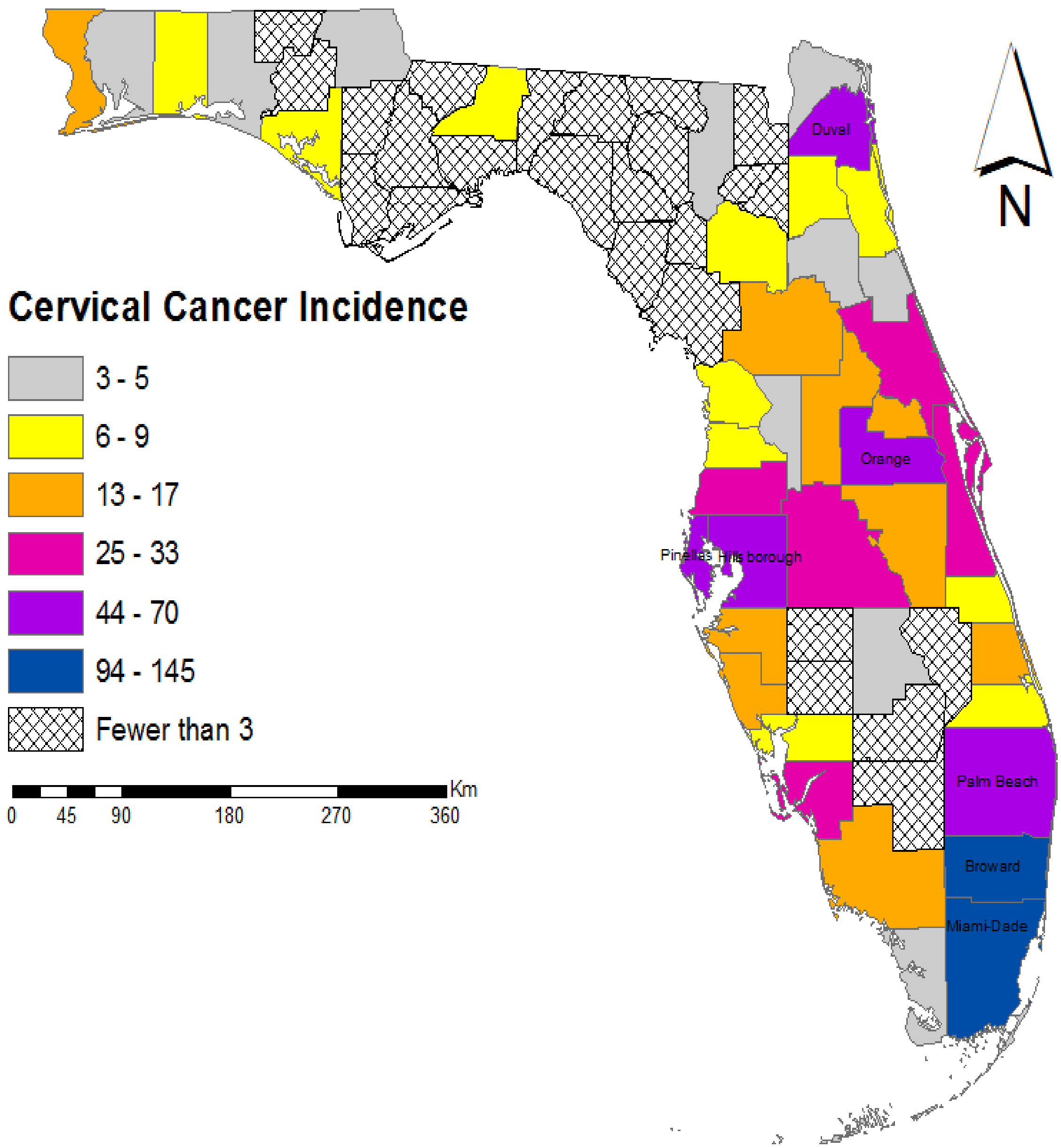


Figure 1. County level crude cervical cancer incidence counts

Note: Study area consisting of 67 total counties.

Figure 2.demonstrated incident counts represented as epidemiological,capture points per county, with 1 point representing one case. A spatial analysis of this figure confirmed the directional distribution of case counts which appear to have a northwest to southeast pattern with the majority of the cases coming from the most southeastern region of the state. Mapping point data also allows for better visualization of clustering as can be seen in Figure 2. Miami-Dade was chosen as our main study site given that it offered the largest contribution of cervical cancer incident cases.

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-18
Available Online at http://acascipub.com/Journals.php

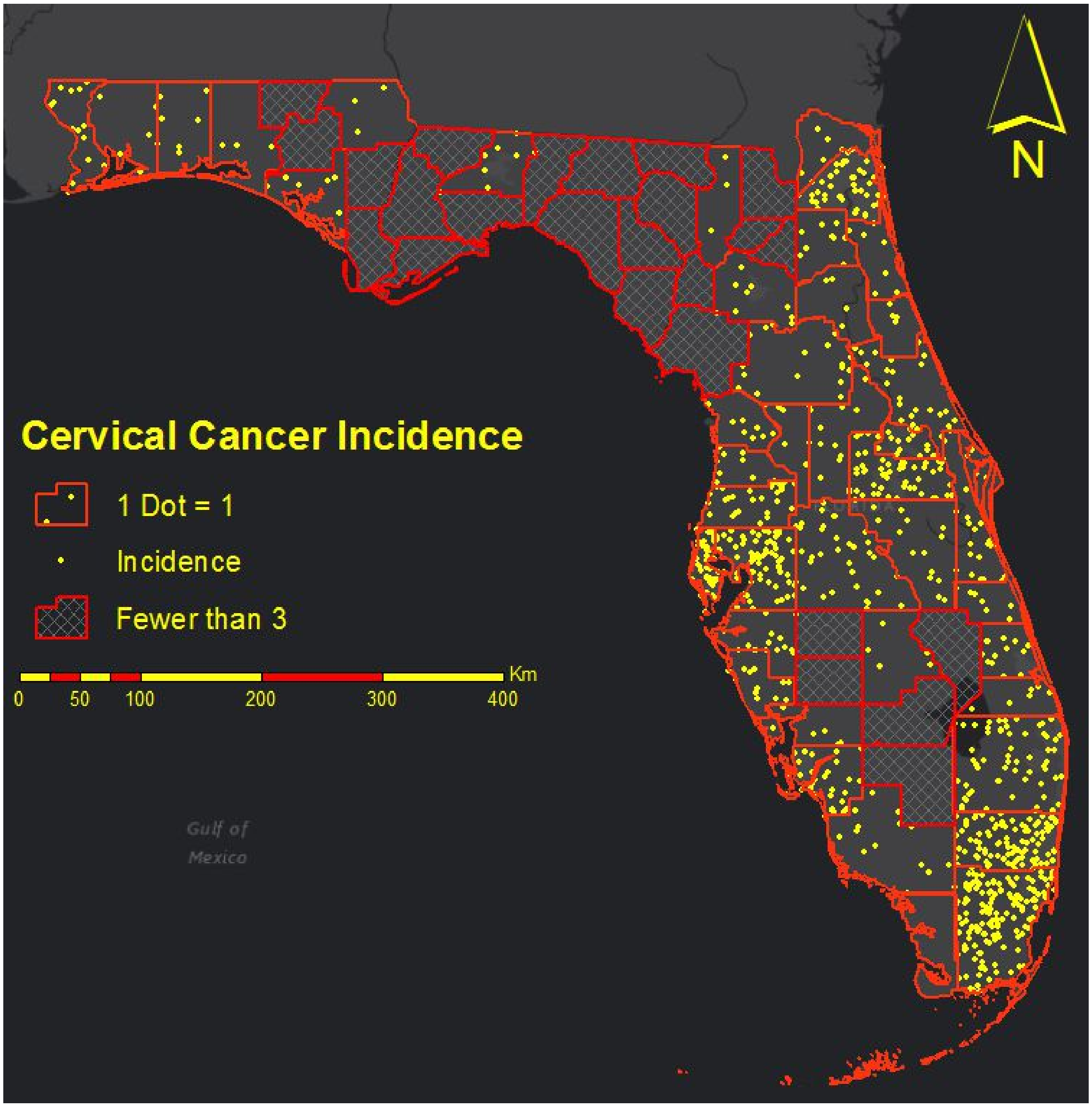


Figure 2. County level crude cervical cancer incidence counts

Note: Study area consisting of 67 total counties; each point representing one case.

In Figure 3 we charted the proportion of county level incident cases at the zip code level taking into account female population for each specific zip code polygon. Larger points represent high contribution, while smaller points represent lower contribution. A visual inspection of the map illustrates a concentration of high contribution zip code polygons in the central and northeastern region of the county. Partitioning grid stratified hyperendemic focal points based on anthropogenic population at the county level can render optimal forecastable values (Griffith 2003).

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-18
Available Online at http://acascipub.com/Journals.php

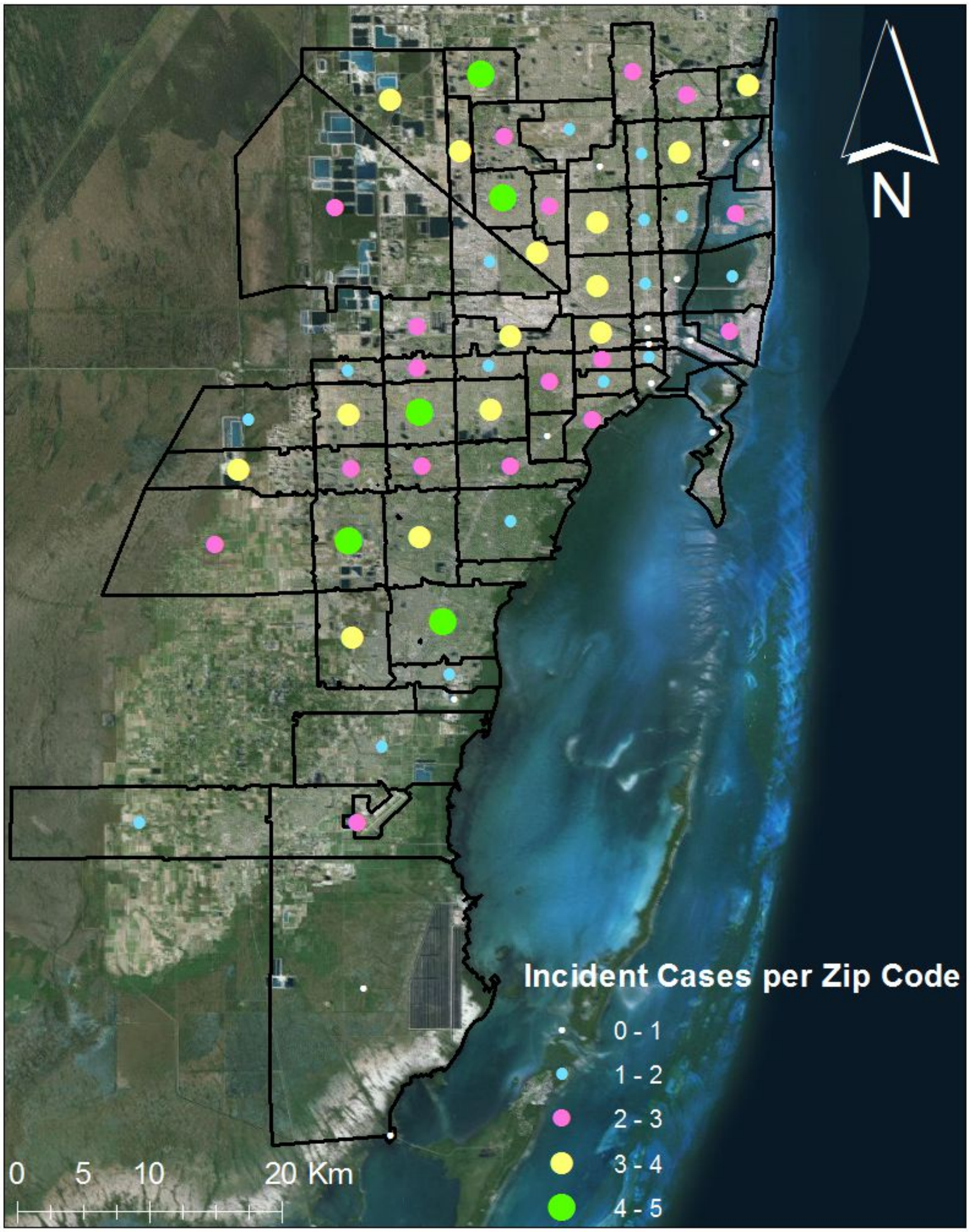


Figure 3. Proportion of county level incident cases of cervical cancer at the zip code level

Note: Female population taken into account

A dataset with zip code-level demographic information was created, stored in excel files, and analyzed using SAS. Multiple regression analysis was used to find explanatory predictor variables showing statistical significance at a 95% confidence level in our Miami-Dade study site. The dependent variable was the proportion of zip code level cases based on county level counts employing an anthropogenic population stratification. The initial explanatory variables included the following sociodemographic characteristics: median income, having less than a high school diploma, having no health insurance, and the number of medical facilities in a zip code polygon (healthcare access). SAS PROC REG was used to obtain Table 1 which reports the predictive strength of our independent variables and the amount of variation explained in the demographics dataset. Overall, our model explains 79.6% ($F_{4, 60} = 58.54$, $p < 0.001$) of the variation in cervical cancer cases from the Miami-Dade study site. The socioeconomic predictors which showed a statistical significance at the $p < 0.05$ level were median income, being uninsured, and the number of medical facilities in that zip code. The remaining coefficient was not significant.

Table 1. Multiple regression model SAS output

Regression Statistics					
Multiple R		0.8922			
R Square		0.7960			
Adjusted R Square		0.7824			
Standard Error		0.4653			
Observations		65			
	df	*SS*	*MS*	*F value*	*Pr> F*
Regression	4	50.6998	12.6749	58.5387	< 0.0001
Residual	60	12.9913	0.2165		
Total	64	63.6911			
	Coefficients	*Standard Error*	*t Value*	*Pr>|t|*	
Intercept	-0.311498	0.231235	-1.347104	0.1830	
Median Income	0.000015	0.000003	4.645590	< 0.0001	
< HS Diploma	0.000135	0.000597	0.226609	0.8215	
Healthcare Coverage	0.000379	0.000039	9.830951	< 0.0001	
Access to Care	0.005244	0.001311	3.999039	0.0002	

Figure 4 was constructed to allow for visual representation of all statistically significant explanatory variables within Miami-Dade County. Median income at the zip code level varies from approximately $15,000 to $106,000, with the majority of higher income zip codes residing in the center of the county. A spatial analysis of the number of healthcare facilities per zip code indicate a clustering pattern which could be indicative of a lack of access to care for certain regions. Finally, the percentage of uninsured women, our last explanatory variable, appears to be greatest at the northeastern region of the county. However, there are a few zip codes at the southwestern region and one in the central west region, showing an elevated percentage of uninsured women.

Figure 4. Cartographic representations of explanatory variables

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-18
Available Online at http://acascipub.com/Journals.php

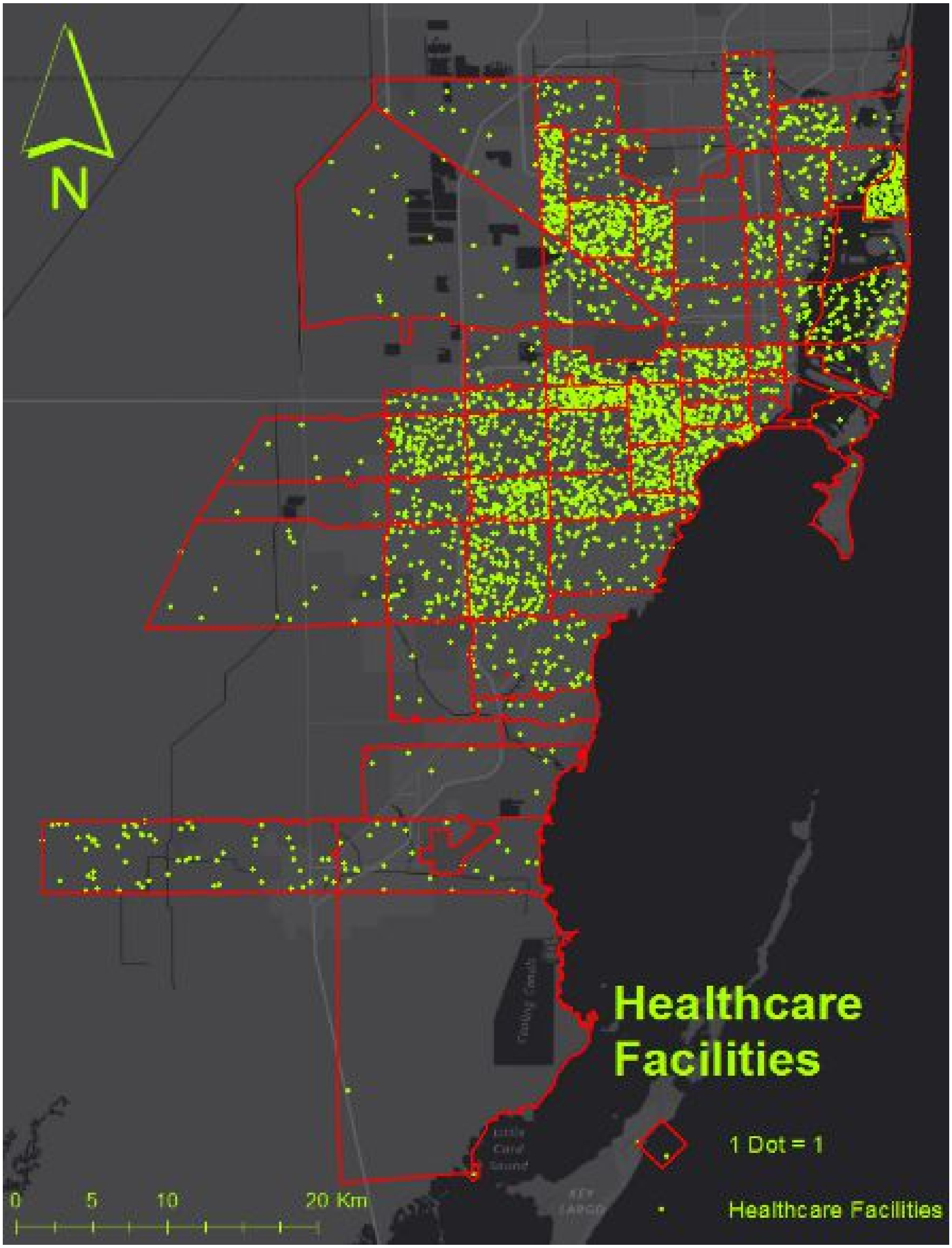


Figure 4(a) Healthcare access

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-18
Available Online at http://acascipub.com/Journals.php

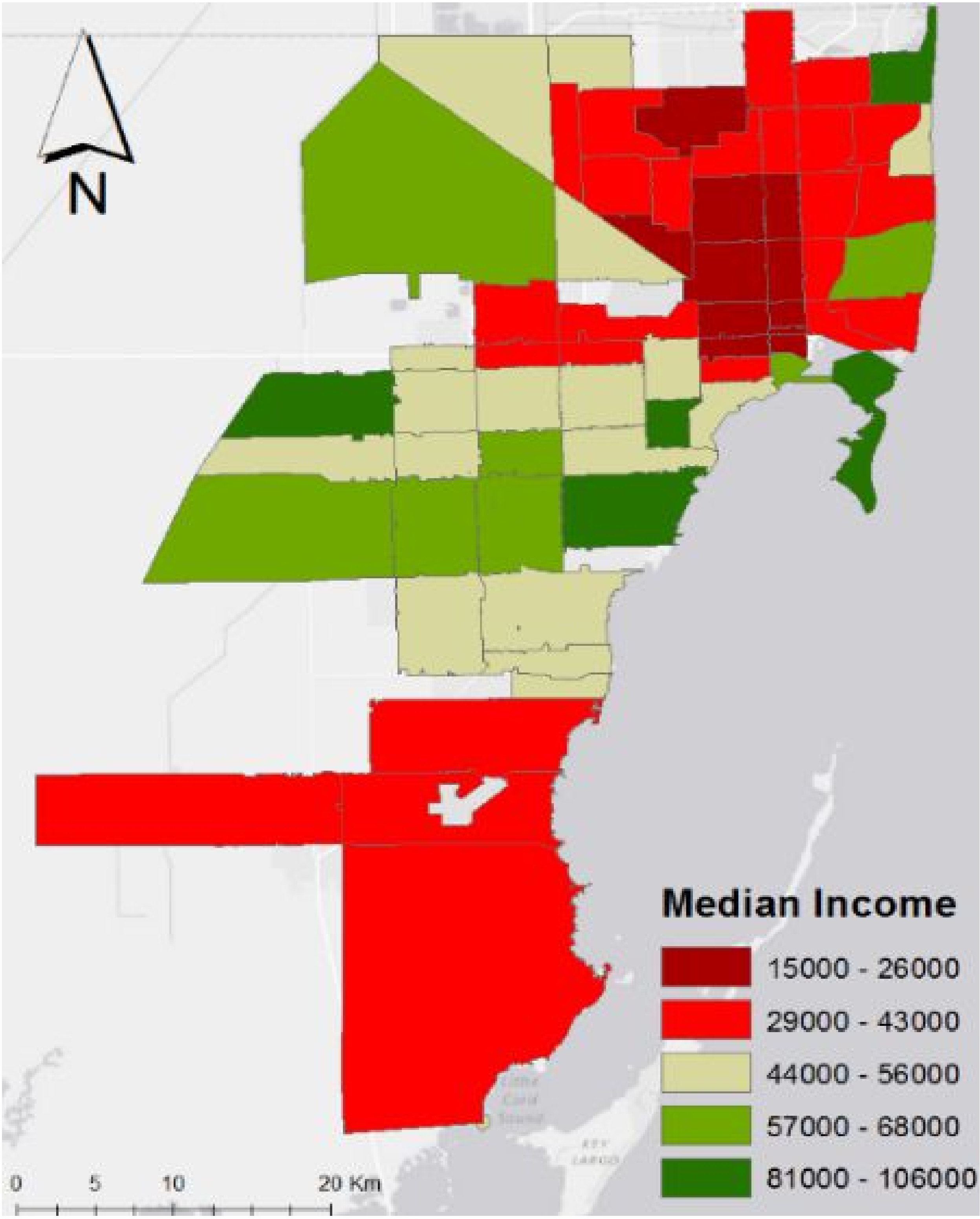


Figure 4(b) Median income

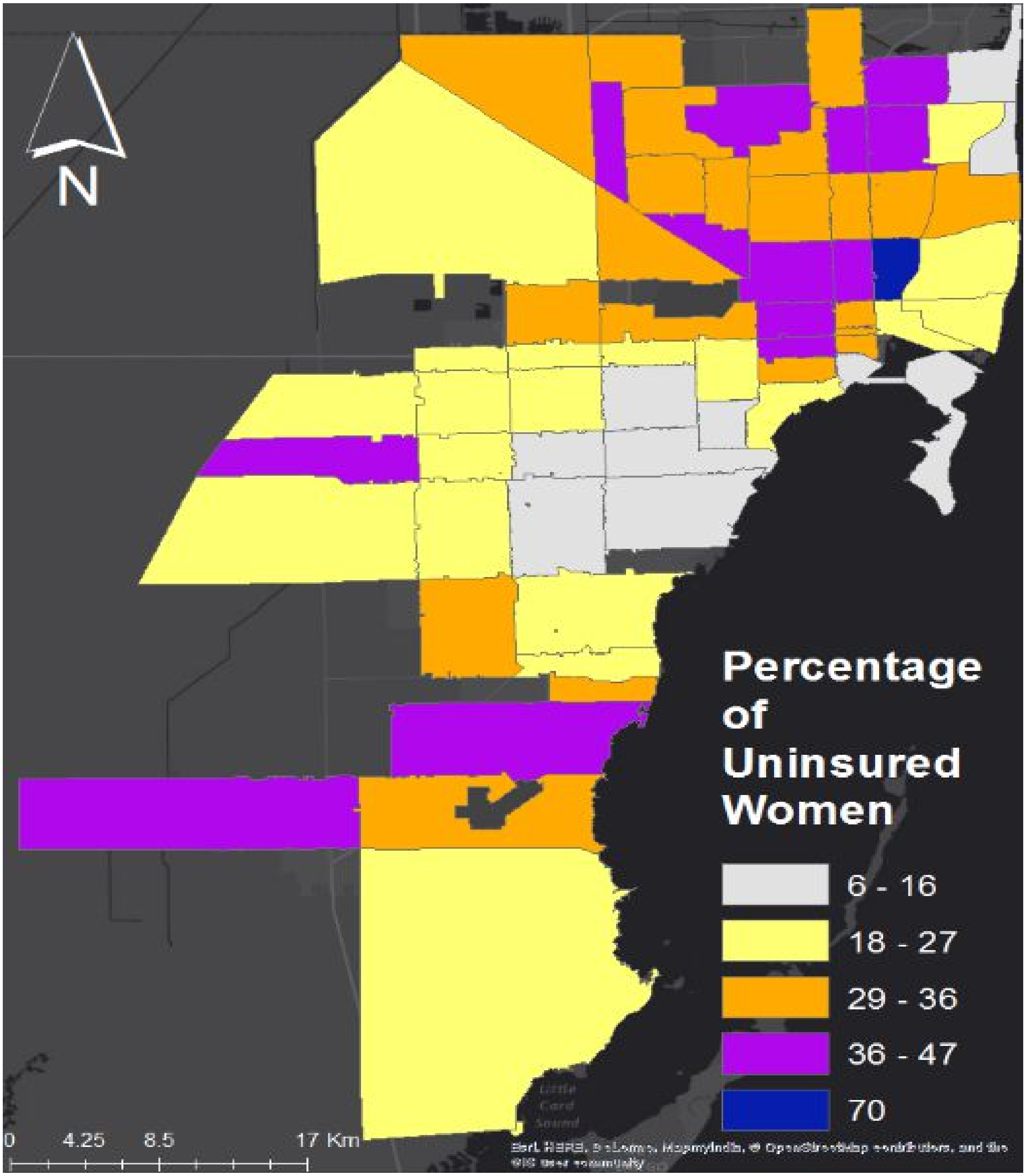


Figure 4(c) Percent of uninsured women

Note: Healthcare access is represented by the number of medical facilities within a zip code polygon.

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-18
Available Online at http://acascipub.com/Journals.php

Our final epidemiological map was created by interpolating statistically significant explanatory variables into Hillsborough County in order to locate vulnerable regions to cervical cancer. Given that three out of our four explanatory variables showed statistical significance at a 95% confidence level, we were able to construct a stratified index which prioritized vulnerability by low risk, medium risk, and high risk zip code polygons. Interpolation of explanatory variables withinHillsborough County can be seen in Figure 5. Suppressed counties did not have enough available data to be included in the index.

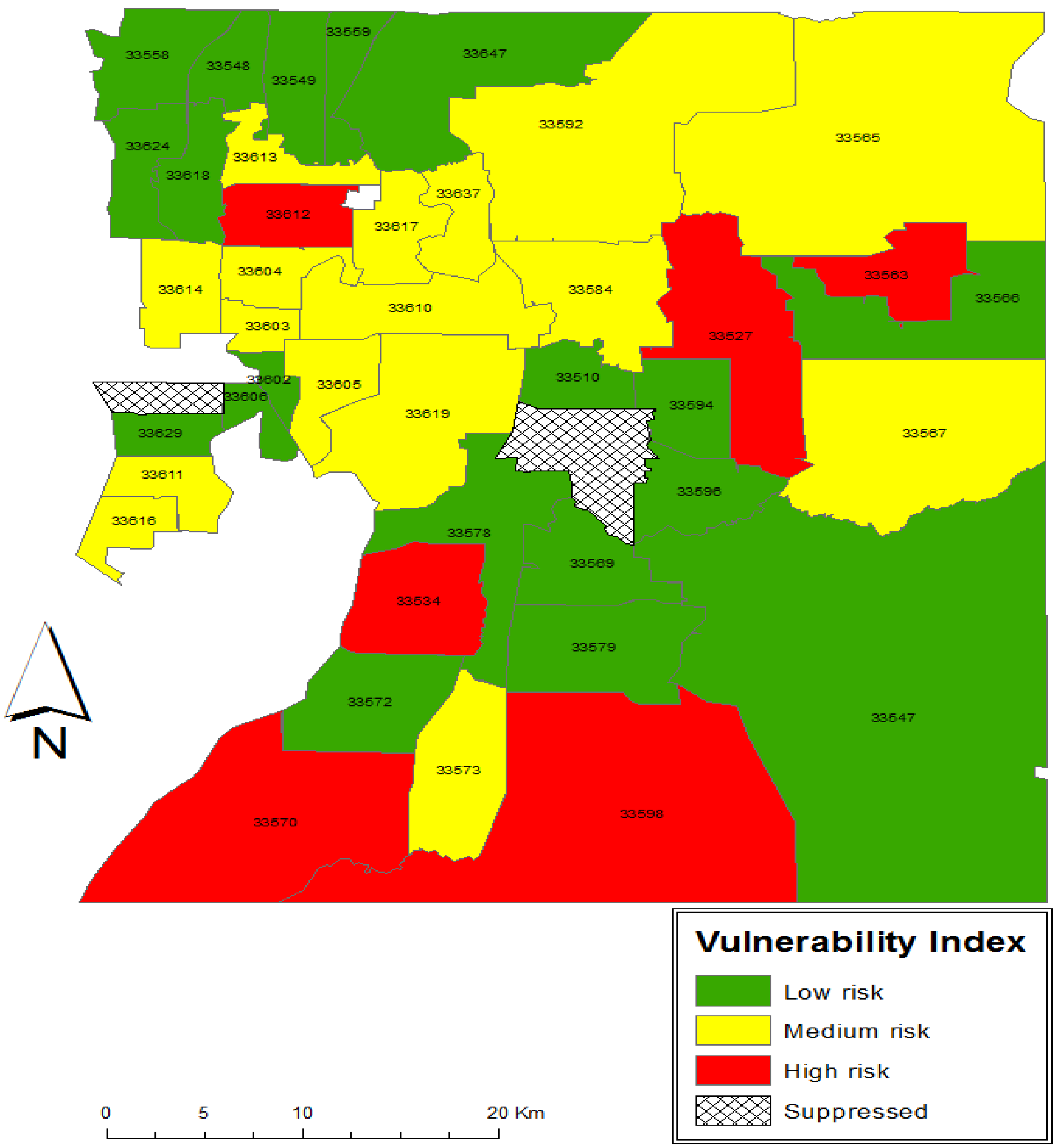


Figure 5. Hillsborough County vulnerability index

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-18
Available Online at http://acascipub.com/Journals.php

4. Conclusion

We were able to construct a robust cervical cancer multivariate linear regression distribution model. In our model, a single response measurement Y (incident cases of cervical cancer) was related to three socioeconomic prognosticators, where X represented each county levelgeosampled observation. The critical assumption of this county level cervical cancer endemic model is that the conditional mean function was linear $E\ (Y / X = \alpha + \beta X)$; this led to the following multiple regression mean function $E\ (Y / X = \alpha + \beta_1 X_1 + \ldots X_i Y_i)$ where α was the intercept and β was the slope or coefficients. Based on the mean function we were able to determine levels of statistical difference in the cervical cancer covariates.

Our final epidemiological model forecast was Gaussian. The residual plots located in Appendix 1 all revealed that the regression model did not violate any assumptions (non-multicollinear). We were able to assess the normality of the regression plots by looking at the residual deviations rendered by PROC REG. All measures of uncertainty were within norm limitations. Hence, our realizations rendered from the probabilistic cervical cancer linear paradigm was deemed to be appropriate. Multivariate regression can have inconspicuous propagational uncertainties which may be teased out in PROC REG employing a stepwise backward elimination method (SAS Institute Inc. 2016).

Ultimately, the ability to produce a stratified vulnerability index ensures optimal use of allocated resources and funding. Unfortunately, federal funding for prevention and health promotion programs in the United States declined by approximately $580 million in 2017 alone; and the Presidents proposed FY 2018 budget would include a nearly 18% cut to the Department of Health and Human Services (Trust for America's Health 2017). This outstanding deficit calls for novel and innovative techniques in which we can specifically target populations most in need. By using targeted interventions, we would not only be able to bridge the gap in cervical cancer disparities, but we would also be contributing to the reduction in economic burden associated with this disease.

This study did not consider sociodemographic factors such as race, foreign born vs. native born women, language spoken, health literacy, and other factors which may also be contributing variables for cervical cancer incidence rates. Therefore, generalizability of these results are limited to comparable populations. However, the proposed method, can be combined with these and other more specific sociodemographic factors in future applications.

A generalizable, hierarchical, county-level, Bayesian, cervical cancer model may illustrate clustering tendencies in a county-level georeferenced, grid-stratified zip code polygon. Bayesian theory calls for the use of the posterior predictive distribution to do predictive inference (i.e., to predict the distribution of a new, unobserved cervical cancer vulnerability population data point). That is, instead of a fixed point as a predictor, a distribution over possible county-level population points may be optimally retrieved.By comparison, prediction in frequentist statistics often involves finding an optimum point estimate of the parameter(s)—e.g., by maximum likelihood or maximum a posteriori estimation (MAP)—and then plugging this estimate into the formula for the distribution of a sampled data point (e.g., hyperendemic cervical cancer capture point). This has the disadvantage that it does not account for any propagational probabilistic uncertainty (e.g., leptokurtoticdistributions) in the sample value of the parameter, and hence could underestimate the variance of the forecasted cervical cancer regression probability distribution.

In some instances, frequentist statistics may quantitate linear cervical cancer endemicity within a grid-stratified county georeferenced zip code polygon in ArcGIS For example, confidence intervals and prediction intervals in frequentist statistics generated in Statistical Analyst ™ may be robust when constructed from a normal distribution with unknown mean and variance employing a Student's t-distribution. This method would correctly estimate the variance in a county-level,ArcGIS, cervical cancer, forecast,vulnerability model, due to the fact that (1) the average of normally distributedrandom variables would also be normally distributed; (2) the predictive distribution of a normally distributed county-level grid-stratified zip code population data point with unknown mean and variance, using conjugate or uninformative priors, has a student's t-distributionin Bayesian statistics.In ArcGIS however, the posterior predictive distribution can always be determined exactly—or at least, to an arbitrary level of precision, especially when numerical methods are employed for parameter estimation.

Note that both types of predictive distributions have the form of a compound probability distribution (as does the marginal likelihood) for quantitating an empirical dataset of diagnostic, clinical cervical cancer, county-level, georeferenced, grid-stratified, polygonised, zip code explanatory, time series covariates. In fact, if the prior distribution is a conjugate prior, in a county-level cervical cancer risk model, the prior and posterior distributions would come from the same family. Hence it can easily be understood that both prior and posterior, predictive, county-level,cervical cancer probability distributions come from the same family of compound distributions. The only difference in the cervical cancer county prognosticative models would be that the tabulated posterior predictive distribution would employ the updated values of the sampled hyperparameters by applying the Bayesian inferences as rendered by the conjugate prior. The prior predictive distribution uses the values of the hyperparameters that appear in the prior distribution (Cressie 1993).

On a long-term basis, this pilot study seeks to expand a cervical regression model surveillance system throughout each county in Florida using clinical field operational and remote satellite data. The proposed methods and resulting information will provide public health officials in Florida with the tools to accurately identify factors regulating outbreaks of county level cervical cancer. We also envision a web based GIS interface for use by county and state level public health officials in Florida which will include:

(1) Generating real-time system based reporting tolls.
(2) Automated immediate alerts to Public Health officials.
(3) Health alerts to doctors and hospitals.
(4) A web based tool for data entry and communication.
(5) Geographical mapping of cervical cancer outbreaks at the zip code level.And
(6) Informative predictive maps determining county locations based on risk of affected populations and hyperendemic related regression explanatory covariates of cervical cancer.

References

Akers, A.Y., Newmann, S.J. and Smith, J.S., 2007. Factors underlying disparities in cervical cancer incidence, screening, and treatment in the United States.*Current Problems in Cancer,* 31(3), 157-181.

Alao, S., Mati K. and Jacob, B., 2016. Differentiating non-homoscedasticity and geospatially extreme outliers for urban and rural landscape dataset using Pearson's product moment correlation coefficients for quantitating clustering tendencies in nonvaccinated measles populations in Nigeria. *Journal of Remote Sensing & GIS,* 5(4), 2469-4134.

American Cancer Society, 2016. *Cervical cancer prevention and early detection* [online]. Available from: http://www.cancer.org/acs/groups/cid/documents/webcontent/003167-pdf.pdf [Accessed 19 July 2017].

Brunsdon, C., Fotheringham, A.S. and Charlton, M., 2002. Geographically weighted summary statistics — A framework for localised exploratory data analysis. *Computers, Environment and Urban Systems,* 26, 501-524.

Centers for Disease Control and Prevention, 2017a. *Cervical cancer statistics* [online]. Available from: https://www.cdc.gov/cancer/cervical/statistics/index.htm [Accessed 7 July 2017].

Centers for Disease Control and Prevention, 2017b. *Cervical cancer rates by race and ethnicity* [online]. Available from: http://www.cdc.gov/cancer/cervical/statistics/race.htm [Accessed 4 July 2017].

Cressie, N., VerHoef J.M., 1993. Multivariable spatial prediction. *International Association for Mathematical Geology*, 25, 219-220.

Factual Facts, 2017. 10 Largest Cities in the US (by Land Area) [online]. Available from: https://factualfacts.com/us-largest-cities/ [Accessed 7 July 2017].

Florida Department of State, 2017. Quick facts [online]. State of Florida. Available from: http://dos.myflorida.com/florida-facts/quick-facts/ [Accessed 10 July 2017].

Florida Geographic Data Library, 2012. Zipbnd_2012 [online]. Available from: http://www.fgdl.org/metadata/fgdl_html/ [Accessed 7 July 2017].

Florida Geographic Data Library, 2014. GC_HEALTH_AUG14 [online]. Available from: http://www.fgdl.org/metadataexplorer/explorer.jsp [Accessed 7 July 2017].

GIS for Government, 2017. [online]. Esri. Available from: https://www.esri.com/industries/government [Accessed 7 July 2017].

Griffith, D.A., 2003. Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization [online]. Springer. Available from: http://www.springer.com/us/book/9783540009320 [Accessed 11 July 2017].

Griffith, D.A., 2005. A comparison of six analytical disease mapping techniques as applied to West Nile Virus in the coterminous United States. *International Journal of Health Geographics,* 4(18), 1-14.

Hay, S.I., Snow, R.W., and Rogers, D.J., 1998. From predicting mosquito habitat to malaria seasons using remotely sensed data: Practice, problems and perspectives. *Parasitology Today,* 14(8), 306-313.

Hosmer, D.W. and Lemeshow, S., 2000. Applied logistic regression. 2nd ed. [online]. Wiley. Available from: http://onlinelibrary.wiley.com/book/10.1002/0471722146 [Accessed 11 July 2017].

Jacob, B.G., *et al*., 2003. Occurrence and distribution of Anopheles (Diptera: Culicidae) larval habitats on land cover change sites in urban Kisumu and urban Malindi, Kenya. *Journal of Medical Entomology,* 40(6), 777-784.

Jacob, B.G., *et al*.,2009. Developing operational algorithms using linear and non-linear squares estimation in Python® for the identification of Culexpipiens and Culexrestuans in a mosquito abatement district (Cook County, Illinois, USA). *Geospatial Health,* 3(2), 157-176.

Kish, J.K., *et al*., 2016. Prioritizing US cervical cancer prevention with results from a geospatial model. *Journal of Global Oncology,* 2(5), 275-283.

Kitron, U., 1998. Landscape ecology and epidemiology of vector-borne diseases: Tools for spatial analysis. *Journal of Medical Entomology,* 35(4), 435-445.

Campbell, M., 2013. County boundaries of Florida [online]. ArcGIS. Available from: http://services.arcgis.com/JMAJrTsHNLrSsWf5/arcgis/rest/services/FloridaCounties/FeatureServer [Accessed 11 July 2017].

National Cancer Institute, 2017. Incidence rate report for Florida by county. State Cancer Profiles. Available from: https://statecancerprofiles.cancer.gov/incidencerates/index.php?stateFIPS=12&cancer=057&race=00&age=001&type=incd&sortVariableName=rate&sortOrder=default [Accessed 11 July 2017].

Roche, L.M., Niu, X. and Henry, K.A., 2015. Invasive cervical cancer incidence disparities in New Jersey-Aspatial analysis in a high incidence state. *Journal of Health Care for the Poor and Underserved,* 26(4), 1173-1185.

SAS Institute Inc., 2016. Model-selection methods [online]. Available from: http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_reg_sect030.htm [Accessed 10 July 2017].

The Beacon Council, 2017. Facts & figures [online]. The Beacon Council. Available from: http://www.beaconcouncil.com/facts-figures/demographics [Accessed 10 July 2017].

Trust for America's Health, 2017. A funding crisis for public health and safety: State-by-state public health funding and key facts, 2017 [online]. Available from: http://healthyamericans.org/report/136/ [Accessed 10 July 2017].

United States Census Bureau, 2016. *American Fact Finder* [online]. Available from: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_14_5YR_DP02&src=pt [Accessed 10 July 2017].

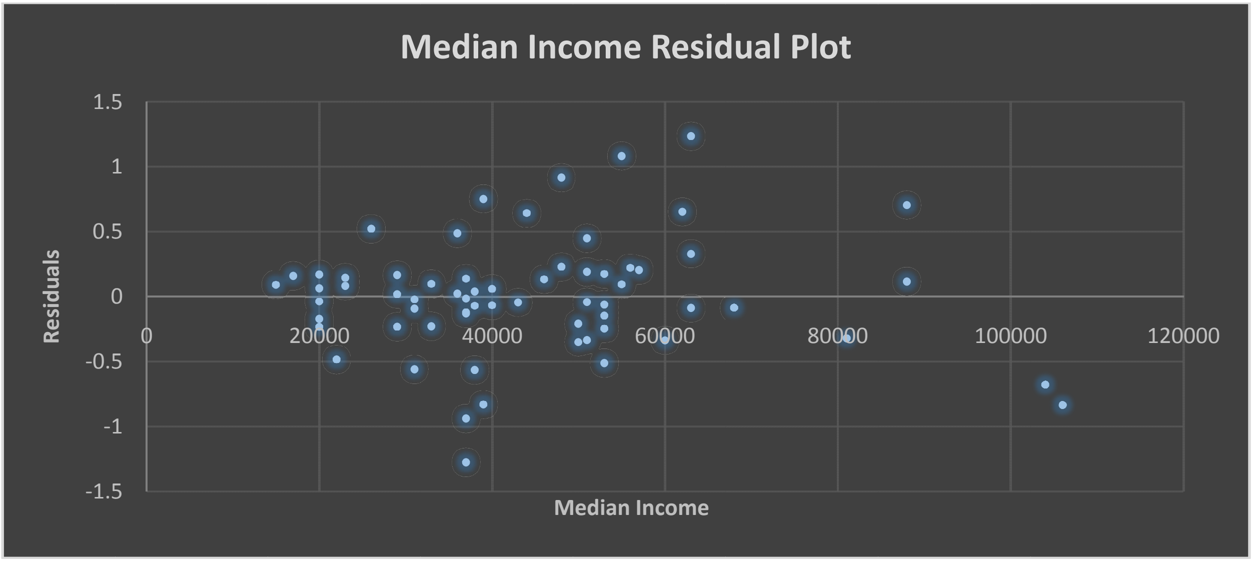
United States Postal Service and Hillsborough County, 2017. Zip codes [online]. ArcGIS. Available from: https://services.arcgis.com/apTfC6SUmnNfnxuF/arcgis/rest/services/Zip_Codes/FeatureServer [Accessed 10 July 2017].

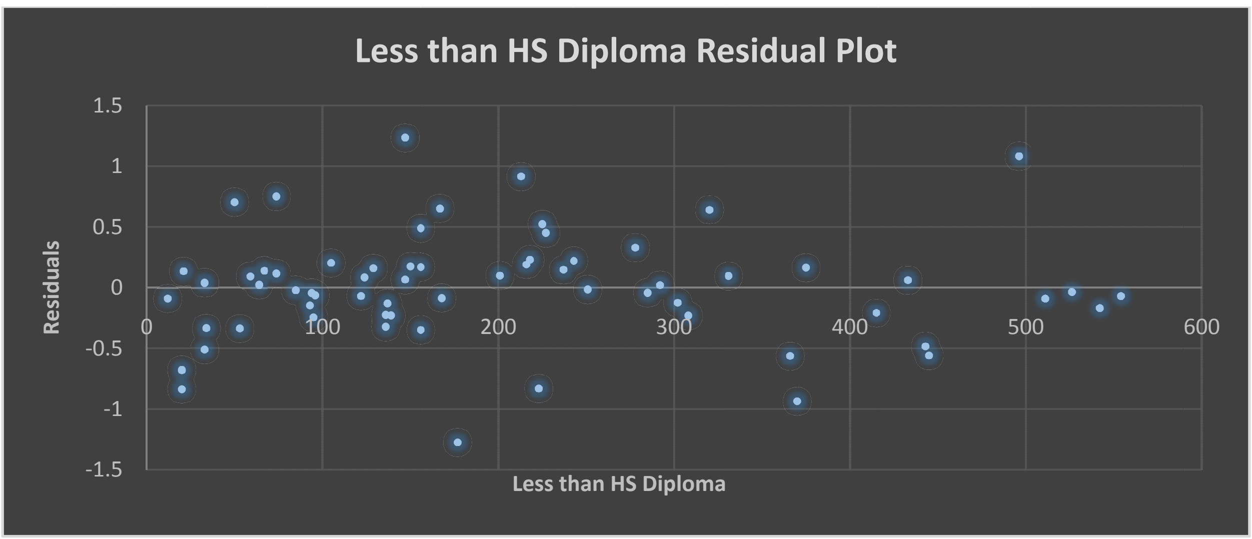
Wan, N., Zou, B. and Sternberg, T., 2012. A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science,* 26(6), 1073-1089.

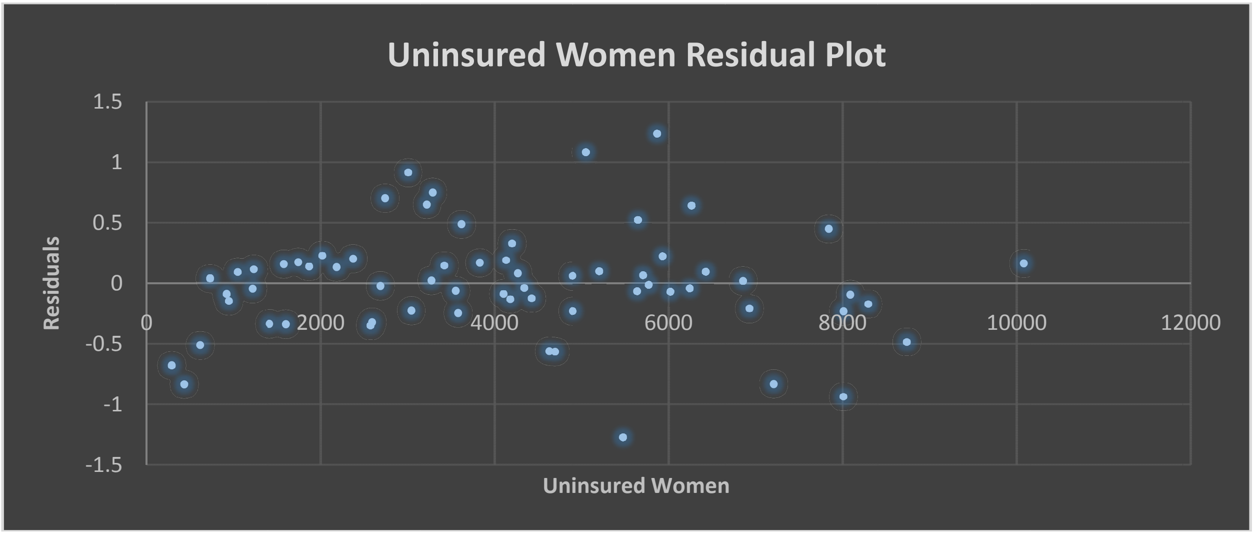
Wood, R.W., *et al*., 1991. Distinguishing high and low anopheline-producing rice fields using remote sensing and GIS technologies. *Preventive Veterinary Medicine*, 11, 277-288.

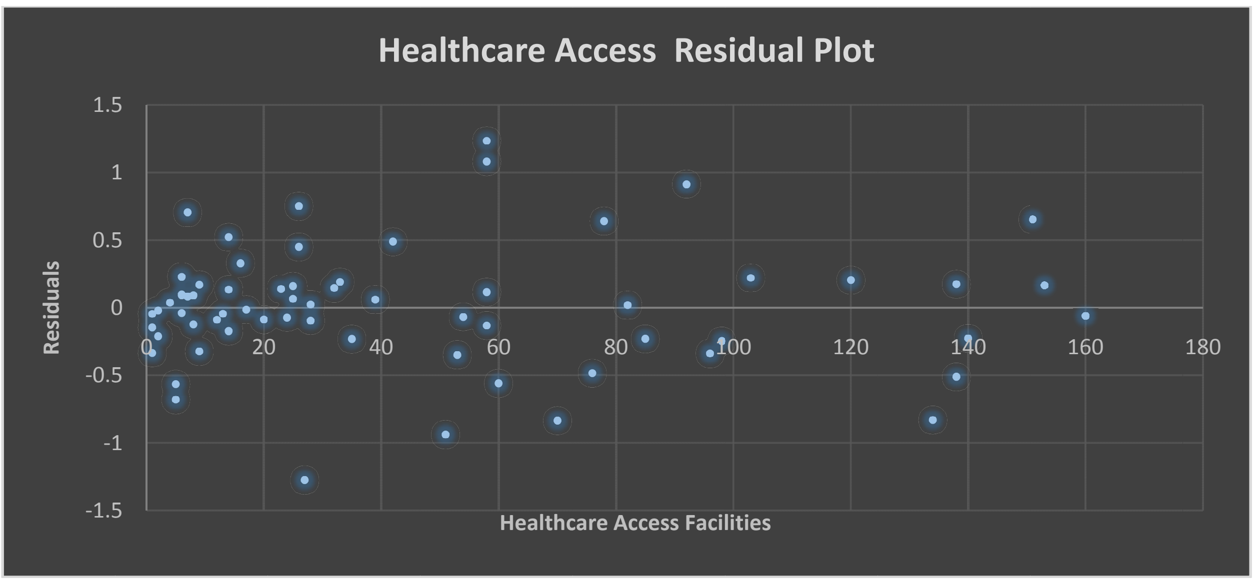
International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-18
Available Online at http://acascipub.com/Journals.php

Appendix 1. Residual plots for explanatory predictor variables






Uninsured Women Residual Plot
Residuals
Uninsured Women
1.5
1
0.5
0
-0.5
-1
-1.5
0
2000
4000
6000
8000
10000
12000


Healthcare Access Residual Plot
Residuals
Healthcare Access Facilities
1.5
1
0.5
0
-0.5
-1
-1.5
0
20
40
60
80
100
120
140
160
180