International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-15
Available Online at http://acascipub.com/Journals.php

Research article

Regressively prioritizing sociodemographic and landscape covariates for iteratively cartographically quantitating vulnerability for oral cancer in a county, grid-stratified, georeferenced zip code polygon

Karan Talreja[a], Toni Panaou[a], Benjamin G Jacob[a]*

[a]College of Public Health, University of South Florida, Tampa, United States
E-mail: karantalreja@mail.usf.edu, bjacob1@health.usf.edu, apanaou@mail.usf.edu

Abstract

An estimated 49,670 people in the United States will get oral cavity or oropharyngeal cancer with 9,700 deaths related to these cancers (American Cancer Society 2017). Oral cancers are named cancers of lips, tongue, cheeks, floor of the mouth, hard and soft palate, sinuses and throat. There are various factors influencing oral cancer frequency such asdemographics like age, race, gender, and ethnicity. Behavioral factors like tobacco and alcohol consumption, hereditary qualities and exposure to the human papillomavirus (HPV) can accelerate the dysplasia. As reported by the Centers for Disease Control and Prevention (CDC) (2017) oral cancer incidence has been increasing in recent years. The possible reason for this would be the lack of a robust predictive models to comprehend the factors influencing oral cancer.

This research examines significant factors employing a multiple regression model and uses Geographic Information System (GIS) software to cartographically illustrate the vulnerable areas employing these influencing factors. Albeit oral cancers are thoroughly researched and documented in the literature, the incidence rates show no sign of diminishing. A conceivable explanation for this would be a dearth of models clarifying the influence of different factors. This model, if applied effectively, would not only help in distinguishing the specific hazards but also aid in predicting the occurrence of future cases on the basis of vulnerability. Thus, accordingly, resource allocation decisions could be made and prevention strategies be planned.Here a predictive, grid-stratified, GIS regression model found that race, gender, health insurance, and poverty to be significant for oral cancer. The p value reported for these variables was less than 0.05. Land use land cover (LULC) was found to be significant at 90% confidence with a p value less than 0.10. Using the significant variables as a model,predictive regressionmaps may be created to show areas of vulnerability to oral cancer ranging from none to very high in Hillsborough County, Florida.

Keywords: Oral Cancer; Regression; sociodemographic; GIS, Florida.

---

1. Introduction

Shockingly, there have been 529,000 cases of Oral cancer worldwide; resulting in 292,300 deaths (Walkansuriya 2008). These account for 3.8% of all cancer cases and 3.6% of deaths related to cancers (Ferlay et al. 2012). This figure is expected to rise 62% to 856,000 cases by 2035 (Shield et al. 2012). Tragically, an estimated 49,670 people in the United States will get oral cavity or oropharyngeal cancer in the year 2017 with 9,700 death related to these cancers (Siegel et al. 2017). There has been an increased incidence of oral cancers cases in Florida, especially in males during the last 5 years (McGorray et al. 2012). Florida ranks 5th in the country for oral cancer incidence with a rate of 13 per 100,000 people (United States Cancer Statistics: Data Visualizations 2017).The 5-year survival rate for oral cancer is 61%, one of the least among all major cancer types making it a serious public health issue (American Cancer Society 2015).

Johnson et al. (2010) determined that socioeconomic status has a significant influence on oral cancer outcomes. Their research showed that even after adjusting for factors like tobacco and alcohol consumption, there was an increased incidence of oral cancer in lower socioeconomic households. Further, Auluck et al. (2016) examined neighborhoods which were classified on basis of socioeconomic status and determined that there was increased incidence in the poorer neighborhoods. Peters et al. (2017) conducted a regression analysis consisting of covariates like age, gender, sex, race, income level to determine the effect of these factors on knowledge of oral cancer in African Americans. This research employed a regression analysis to determine statistically significant variables related to oral cancers.

Conway et al. (2008) and Johnson et al. (2010) both revealed that education level influences oral cancer outcomes. Their research revealed that individuals who have less than high school degree and are at increased risk to develop oral cancer. There exists a definite relation between education status and oral cancer rates and one that requires more investigation (Pororski et al. 2014). The education level also affects mortality trends pertaining to oral cancer (Borges et al. 2015, Chen 2011). Thus, educational attainment is a key influencing factor for oral cancer incidence.

Peters et al. (2015) research showed a relationship between race and increased oral cancer rates, specifically in the African American community. This was further supported by Daraei and Moore (2014) where they included adjustments for factors like age, gender and poverty level. One of the major reason that these racial disparities exist and continue growing is due lack of oral cancer screenings among African Americans (Shepperd et al. 2013). Another reason is the lack of knowledge of oral cancer compounded by low education levels in the community (Riley et al. 2013).

Another important factor that influences oral cancer incidence is gender (Javadi et al. 2016). Males have an increased risk to develop oral cancer as compared to females (Walkansuriya 2008, Weatherspoon et al. 2015). Males also show an increase mortality rate as compared to women. The possible explanation for this would be increased alcohol and tobacco consumption (Peters et al. 2015). In the United States, there is an increased incidence of oral cancer especially in African American males with a rate differential of 2.1(Walkansuriya 2008). Lack of health insurance adversely affects outcome related to oral cancer; due to lack of insurance people cannot get routine screening and available preventive measures thus increasing their risk for oral cancer(Shepperd et al. 2013).

What deems this research novel is that regressioncan quantitatevarioussociodemographic,explanatory, variables like gender, income, race, education level, health insurance, and LULCfor determining georeferenceablelocations of county-level, vulnerability,grid-stratified at the zip-code level. Our assumption was that multiple regression model renderings and cartographic algorithmic techniques along with sociodemographic statisticsmay help predict geospatial locations of higher oral cancer risk to design and implement effective strategies. Once the significant factors have been recognized and vulnerable populations identified; resources could be allocated effectively to reduce the cases of oral cancers.We examined a multiple regression analysis with sociodemographic and other covariates to optimally cartographically delineate and forecast vulnerable populations to oral cancer in Hillsborough County, Florida.

2. Study Site

Florida is a state located in the southeastern region of the United States. The state is bordered to the west by the Gulf of Mexico, to the north by Alabama and Georgia, to the east by the Atlantic Ocean, and to the south by the Straits of Florida and Cuba. Florida is the 22nd most extensive, the 3rd most populous and the 8th most densely populated of the United States(Geography of Florida - World Atlas 2017). Jacksonville is the most populous municipality in Florida and the largest city by area in the contiguous United States. The Miami metropolitan area is Florida's most populous urban area (US Census Bureau 2017). The city of Tallahassee is the state capital.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(Geography of Florida - World Atlas 2017). 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 (Geography of Florida - World Atlas 2017).

For this research, emphasis was placed on Broward County since it had the highest number of cases of oral cancer in Florida (356). Broward County is located in the south east Florida and is the second most populous county in Florida and 17th most in the United States (U.S. Census Bureau 2017). Its county seat is Fort Lauderdale. It has a total area of 3,430 km^2 of which 3,100 km^2 is land.

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-15
Available Online at http://acascipub.com/Journals.php

The computed regression model obtained after analysis of Broward was applied to Hillsborough County to delineate vulnerable populations. Hillsborough County is a part of the Tampa metropolitan area. It is the fourth most populous county in Florida (Hillsborough County Community Atlas 2017). It has a total of 3,280 km^2 of which 2,600 km^2 is land(Hillsborough County Community Atlas 2017).

3. Data Collection and Analysis

The initial data pertaining to the incidence of oral cancer was collected from the Florida Cancer Data system (www.flhealthcharts.com). Employingthis data, a GIS map was created showing the incidence count by county. Counties shown as 0 either had no cases or data for oral cancers was not collected.After cartographically illustrating the data, it was seen that Broward County had the highest incidence of oral cancers in Floridawith 356 oral cancer cases being reported in 2014 as illustrated in Figure 1.

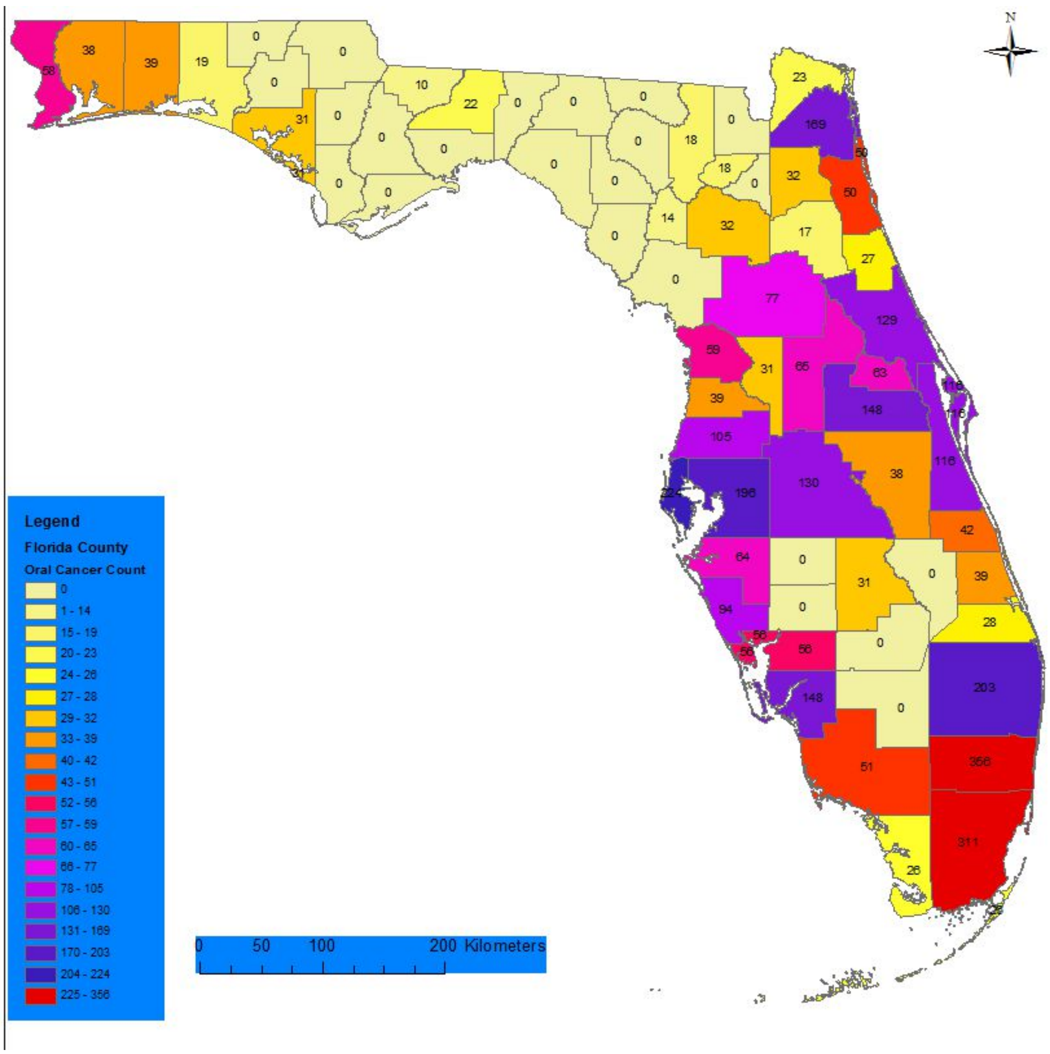

Figure 1. Oral cancer cases by county

A univariate statistics and regression models was generated by employing zip code level data for geospatially regressively summarizing an empirical dataset of grid-stratified, multiple socio-demographic epidemiological, covariate coefficients. A misspecification term for constructing an explanative model in PROC REG was produced. Percentage values for the different regressionexplanatorswere obtained from the 2010 Census and the 2011-2015 American Community Survey 5 Year Estimates (US Census Bureau 2017). Land use land cover map for Broward County was created. The zip codepolygon georeferenceableboundaries map was obtained from ArcGIS (Campbell 2013). The LULC data for Broward County in the form was obtained from the United States geological survey (USGS) land cover institute (https://landcover.usgs.gov/landcoverdata.php#regional). These data were used to create a map depicting LULC by zip code. Using the map the LULC was classified into 3 categories, urban residential, urban commercial and pastureland, depending which one was predominant in the zip code (Figure 2).

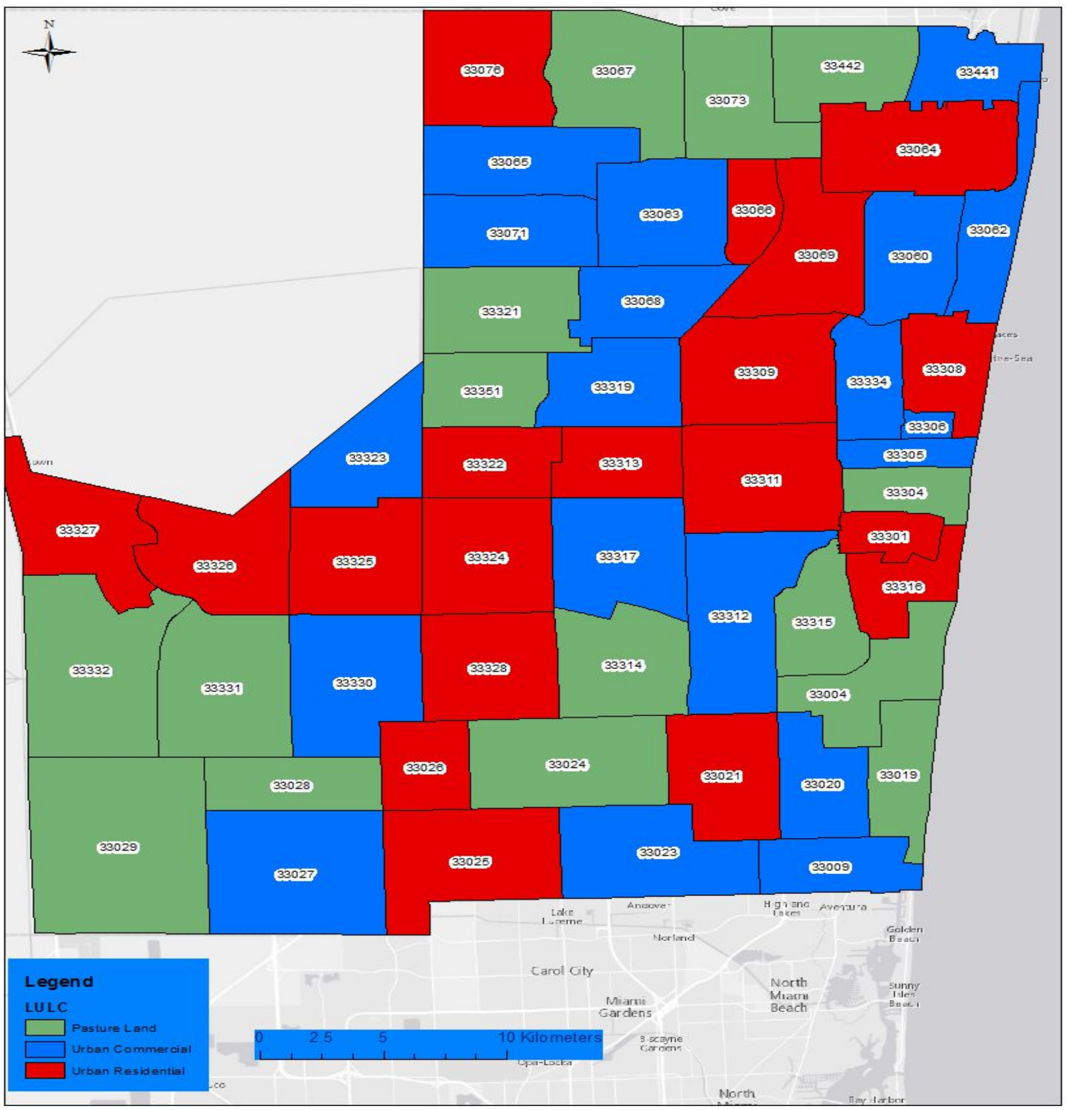

Figure 2.LULC map for Broward County

Since zip code incidence data for oral cancer was not available, it was calculated by dividing the population living in the particular zip code by the total number of people living in the county (1,748,245) and multiplying that by the number of cases in Broward County which was 356 (Equation 1). This is how a dependent variable for each zip code was created.

The equation used was

$$Zip\ Code\ Cancer\ Rate = \frac{\text{Population at zip code}}{\text{Total Population in Broward}} \text{xBroward Cancer Rate} \qquad (1.1)\ .$$

Table 1. Parameterizable sociodemographic samples within 53 zip code polygons of Broward County as entered in SAS®

Variable	Description	Units
AA	African American	Percentage
HSE	High School Education	Percentage
Males	Males	Percentage
PIP	People in Poverty	Percentage
No_HI	No Health Insurance	Percentage
DV	Incidence of Oral Cancer	Numeric Value
LULC	Land Use Land Cover	Numeric Value

4. Regression Analysis

The relationship between the geosampled, zip-code level, sociodemographic, covariates was investigated by single variable regression analysis in PROC REG. Since incidence data are binomial fractions, a regression model was employed, as is standard practice for the analysis of the zip code-level, polygon data. Multiple linear regression analyses were employed to infer the relationship between the geosampled, zip code data variables (i.e., independent variables) and incidence of oral cancer (i.e. dependent variable).

Multiple linear regression is a standard statistical tool that regresses p independent variables against a single dependent variable. The objective of our model construction was to find a linear oral cancer forecast vulnerability model,unbiased estimator that best predicted the dependent variable () from the sociodemographic, LULC and other independent variables. Information criteria uses the covariance matrix in PROC REG and the number of parameters in the model to calculate a statistic that summarizes the information represented by the model by balancing a trade-off between a lack of fit term and a penalty term. SAS models can determine the best subset of variables that minimizes the information criteria among all possible subsets (https://support.sas.com/).

Simulated multivariate data was used to compare the performance of AIC to select the true model with standard statistical techniques such as minimizing RMSE, forward selection, backward elimination, and stepwise regression.The regression analyses assumed independent counts (i.e., Ni), taken at multiple, geosampled, georeferenced, county-level , zip code sub-geolocations $i = 1, 2, \ldots, n$. The geo-spatiotemporal-related county-level count data

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-15
Available Online at http://acascipub.com/Journals.php

were then described by a set of discrete integers denoted by matrix $\mathbf{X}_i$, where a $1 \times p$ was a vector of covariate coefficient indicator measurement values for an interpretively geosampled, endemic transmission-oriented, explanative foci. The expected value of these data was given by $\mu_i(\mathbf{X}_i) = n_i(\mathbf{X}_i)\exp(\mathbf{X}_i\beta)$, where β was the vector of parameterizable non-redundant covariates in the interpretively interpolative, operationalizable, epidemiological, prognosticative, zip code model.

5. Results

The relationship between county-level incidence and each explanatory, sociodemographic regressorselected was investigated by employing a single variable regression analysis in PROC REG. The first line of the code began the PROC REG command. The second line specified the fixed portion of each epidemiological, sociodemographic, zip-code, risk model, [i.e., the model without the random intercept, value (i.e., *xb*)]. The model statement specified that the parameterizable, explanative, covariate, estimators that were distributed (~) normally with a mean of *x*b and variance *s*2.

We employed the regression line $(\mathbf{y}_i - \bar{\mathbf{y}}) = (\hat{\mathbf{y}}_i - \bar{\mathbf{y}}) + (\mathbf{y}_i - \hat{\mathbf{y}}_i)$ to generate a pseudo R^2 value where the first term was the total variation in the response *y* (e.g., zip code level oral cancer incidence) and the second term was the variation in mean response based on the asymptotical, normalized, parameterizable, sociodemographic, zip-code-level,and estimators. Squaring each of these terms and adding over all the, zip-code level polygon, the oral cancer incidence, model observations generated the equation $\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$. This equation was written as SST = SSM + SSE, where SS was notation for sum of squares and *T*, *M*, and *E* were the notation for total quantized model error estimates. The square of the sample correlation was then equal to the ratio of the estimates while the sum of squares was related to the total sum of squares: R^2 = SSM/SST. This formalized the interpretation of R^2 for explaining the fraction of variability in the county-level, epidemiological, zip code data explained by the regression model. The sample variance s_y^2 was equal to $\sum\frac{(y_i - \bar{y})^2}{n-1}$, which in turn was equal to the SST/df, the total sum of squares divided by the total df.

A regression equation was constructed by employing the mean square model (i.e., MSM) = $\sum\frac{(\hat{y}_i - \bar{y})^2}{l}$, which was equal to the SSM/df. The corresponding mean square error (MSE) was $\sum\frac{(y_i - \hat{y}_i)^2}{n-2}$ which was determined to be equal to SSE/df and the quantitated, operationalizable, zip code level polygonised, oral cancer, explicatory estimate of the variance about the regression line (i.e., σ^2). The MSE is an estimate of σ^2 for determining whether or not the null hypothesis is true (Schluchter 2014). The pseudo R^2 value obtained in this analysis was 0.6564.

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-15
Available Online at http://acascipub.com/Journals.php

For robustly, parsimoniously, quantizing, operationalizable, explanatory,iteratively interpolative, county-level incidence count, zip code, prognosticators (p)a DFM, was generated which we noted was equal to p and the error degrees of freedom (DFE). This product was also equal to $(n - p - 1)$, and the total degrees of freedom (DFT) which was subsequently equal to $(n - 1)$. The sum of DFM and DFE was determined. The relationship between the mean of the response variable (i.e., zip code incidence count) and the level of the explanatorily, parameterizable, zip code, polygon covariate coefficients in the regression equation were assumed to be approximately linear (i.e., straight line). The corresponding table generated classified each clinical, field and remote, asymptotical, unbiased, covariate estimator in SAS®. (see Table 2).

Table 2. Oral Cancer incidence regression-based model parameter estimates

Source	Sun of Squares	Formula
Model	$\sum(\hat{\mathbf{y}}_i - \overline{\mathbf{y}})^2$	SSM/DFM
Error	$\sum(y_i - \hat{\mathbf{y}}_i)^2$	SSE/DFE
Total	$\sum(y_i - \overline{\mathbf{y}})^2$	SST/DFT

In the multiple regression analyses, the test statistic MSM/MSE had a $F(p, n - p - 1)$ distribution. The null hypothesis was $\beta_1 = \beta_2 = \ldots = \beta_p = 0$, and the alternative hypothesis was evaluated by encompassing the zip code-level, predictive, epidemiological, sociodemographic risk-related parameters in $\beta_j \neq 0$, $j = 1, 2, \ldots, p$. The F test did not indicate which of the parameters $\beta_j \neq 0$ nor, which was not equal to zero only that at least one of them was linearly related to the response variable. The ratio SSM/SST $= R^2$ (i.e., squared multiple correlation coefficient) was thereafter the proportion of the variation in the response variable that was explained by the zip code incidence data. The square root of R^2 (i.e., the multiple correlation coefficient) was the correlation between the explanatorial, time-series, empirical sampled observations (i.e., y_i) and the fitted values (i.e., $\hat{\mathbf{y}}_i$).

Additionally, from the sampling distribution generated from the *t* parameters, the probability of obtaining an *F* was calculated. There were only two means to compare, the *t*-test and the *F*-test, which coincidentally were equivalent. The relation between ANOVA and *t* was then given by $F = t^2$. Thereafter, significant differences by ANOVA were noted for the quantitated mean numbers of the explicative operationalized,iteratively interpolative, asymptotically normalized, datageoreferenceablefeature attributes captured throughout the sampling frame (F Value=14.65 DF=6) (see Figure 3).

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	358.70640	59.78440	14.65	<.0001
Error	46	187.75565	4.08164		
Corrected Total	52	546.46205			

Root MSE	2.02031	R-Square	0.6564
Dependent Mean	6.84906	Adj R-Sq	0.6116
Coeff Var	29.49761		

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	26.45034	12.51981	2.11	0.0401
AA	AA	1	0.06741	0.02494	2.70	0.0096
HSE	HSE	1	-0.11075	0.12824	-0.86	0.3923
LULC	LULC	1	-0.64237	0.36231	-1.77	0.0829
No_HI	No_HI	1	0.30310	0.08242	3.68	0.0006
Males	Males	1	-0.22249	0.09743	-2.28	0.0271
PIP	PIP	1	-0.29328	0.09624	-3.05	0.0038

Figure 3. Output obtained from SAS®

The residual diagnostic plots obtained for the explanatorialregressors were examined to make sure that they meet the regression assumptions (e.g., homoscedasticity of variance). By examining the residual plots, thelinearity and constant variance assumptions were metin the oral cancer regression model (seeFigure 4).

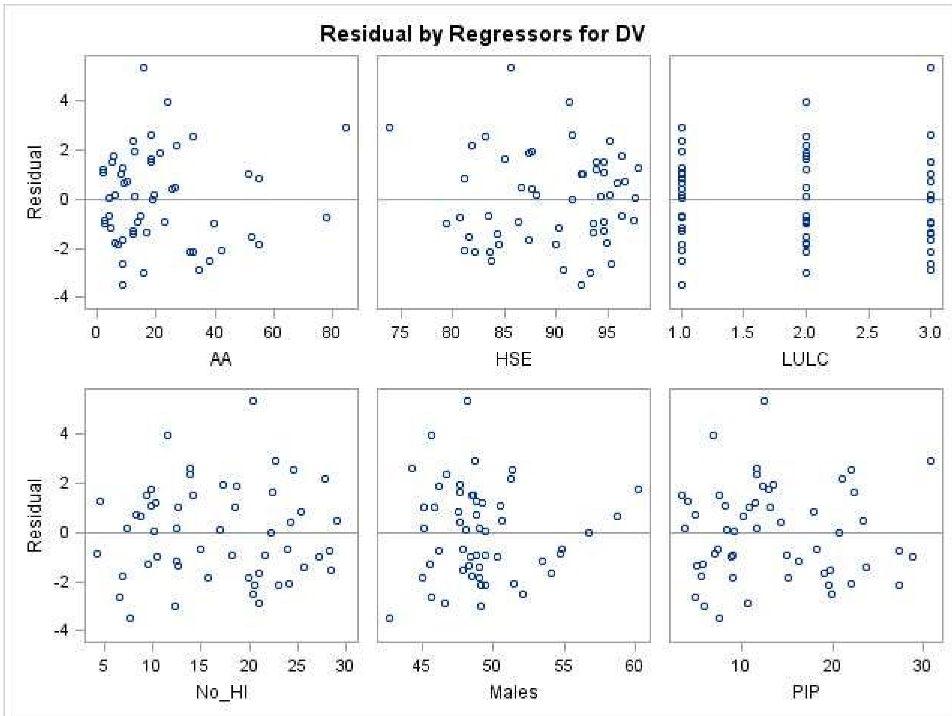

Figure 4. Residual plots for all covariant regressorsas obtained from SAS®

We then generated a stepwise backward regression model to tease out any propagational,probabilistic, uncertainties measurements in the dataset of zip code-level, sociodemographic, epidemiological, vulnerability forecasts. In statistics, stepwise regression is a method of fitting regression models in which the choice of predictive variables is carried out by an automatic procedure (Johnsson 1992). In each step, a county-level, zip code-level, explicative, sociodemographic, prognosticators were considered for addition to or subtraction from a dataset of diagnostic variables based on some pre-specified criterion. According to Jacob et al. (2012) the frequent practice of fitting the final selected model followed by reporting estimates and confidence intervals in an epidemiological model can occur without adjusting the model building process for accounting for model uncertainty.

Running the regression model in SAS rendered the significant variables which were African American, males, no health insurance, poverty which had a p value less than 0.05 and LULC specified variables which had a p value of less than 0.10. Out of the five dependent variables in the regression model, one was found to be non-significant which was having a high school education (see Figure 5).

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-15
Available Online at http://acascipub.com/Journals.php

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	355.66222	71.13244	17.52	<.0001
Error	47	190.79983	4.05957		
Corrected Total	52	546.46205			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	16.42240	4.66870	50.22970	12.37	0.0010
AA	0.07485	0.02335	41.73001	10.28	0.0024
LULC	-0.65885	0.36083	13.53433	3.33	0.0742
No_HI	0.32682	0.07750	72.19491	17.78	0.0001
Males	-0.24411	0.09390	27.43466	6.76	0.0124
PIP	-0.24541	0.07846	39.71688	9.78	0.0030

Figure 5. Backward selection process output as obtained from SAS®

This was confirmed using the backward selection option in PROC REG which resulted in the elimination of High School Education from the analysis (see Table 3).

Table 3. Oral Cancer incidence regression-based model parameter estimates for Backward Elimination

Step	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr> F
1	HSE	HSE	5	0.0056	0.6508	5.7458	0.75	0.3923

Once the significant variables had been established they were georeferenced against the zip code in Broward County contributing to the most number of cases. The final variables used to determine risk in Hillsborough County were Males, African American, Persons in poverty and No health insurance. If all statistics for the zip code were higher than the reference markers then that zip code was assigned a risk of 'very high'. If three geolocations were higher, then a risk of 'high' was assigned. If two geolocations were higher then 'medium' and if one was higher then 'low' risk was assigned. If all statistics were lower than the georeferenced markers then 'none' risk was assigned as illustrated in Figure 6. Zip codes for which the sociodemographic data were not available were not considered in this analysis.

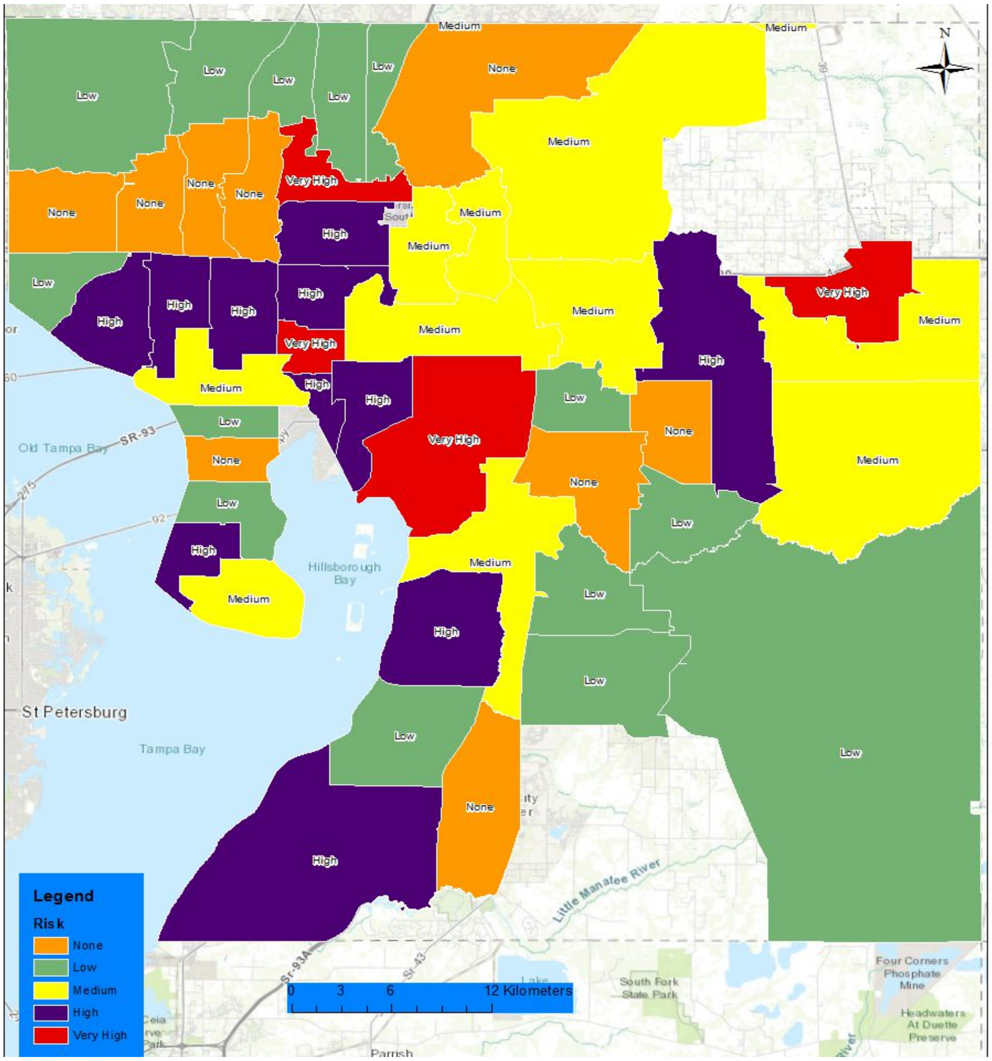

Figure 6. Risk map for Hillsborough County

6. Conclusion

The research corroborates the effect of sociodemographics and LULC covariates on oral cancer prevalence. Our model revealed the influence of race, gender, socioeconomic status and health insurance on outcomes related to oral cancer.

The LULCregressors were significant at 90% confidence suggesting that this predictor variable influences oral cancer incidence at the county-level. This finding warrants further research into the association between oral cancer and land cover. For example in future research an LULC change may be geoclassified in an ArcGIS cyberenvironment to determine if georeferenceable transitioning landscape-relatedregressors such as rural agro-pasturelands to urban residentialor urban commercialmay be a statistically significant predictors of oral cancer prevalence at the county-level.

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-15
Available Online at http://acascipub.com/Journals.php

This research also provides a contradictory result pertaining to the association of high school education and oral cancer incidence. Despite literature suggesting that high school education influences oral cancer, this research found it to be non-significant. The data results obtained from the study could be applied to different county-level geolocations to identify vulnerable population using sociodemographicdata.

The major limitation of the study is the lack of oral cancer incidence data by zip code. Due to the unavailability of this data a weighted method was utilized to quantitate the dependent variable for each georeferenced zip code polygon. This statistic was heavily reliant on the population at the particular zip code which could possibly reduce the reliability of the data.

An iterative Bayesian in PROC MCMC may be usable to identify spatial autoregressive regressors for further targeting of vulnerable populations at the zip code, grid-stratified level. A product moment correlation coefficient (e.g., Moran's*I*) may tease out error autocorrelation in PROC AUTOREG while revealing clustering propensities in county-level sociodemographic and landscape covariates associated with oral cancers.

References

1. American Cancer Society, 2015. *Cancer Facts & Figures 2015* [online]. Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2015/cancer-facts-and-figures-2015.pdf [Accessed 8th 2017].
2. American Cancer Society. *Cancer Facts & Figures 2017* [online] Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2017/cancer-facts-and-figures-2017.pdf [Accessed 8th 2017]
3. Auluck, A., Walker, B.B., Hislop, G., Lear, S.A., Schuurman, N., and Rosin, M., 2016. Socio-economic deprivation: a significant determinant affecting stage of oral cancer diagnosis and survival. *BMC Cancer*, 16 (1).
4. Borges, C.P., FéLix, L.C., Viana, T.S.A., Benevides, B.S., Brizeno, L.A.C., Mota, M.R.L., and Sousa, F.B., 2015. Influence of Education Level on Oral Cancer Survival in a Brazilian Population: A Retrospective Study of 10 Years in a Tertiary Institution. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology,* 120 (2), e90.
5. Briggs, D.J., Collins, S., Elliott, P., Fischer, P., Kingham, S., Lebret, E., Pryl, K., Van Reeuwijk, H., Smallbone, K., and Van Der Veen, A., 1997. Mapping urban air pollution using GIS: a regression-based approach. *International Journal of Geographical Information Science*, 11 (7), 699–718.
6. Campbell M, 2013. [online]. Available from: https://www.arcgis.com/home/item.html?id=80aba8feca084c8c80ca93aa25bd8966 [Accessed 8 July 2017].
7. Chen, A.Y., 2011. US Mortality Rates for Oral Cavity and Pharyngeal Cancer by Educational Attainment. *Archives of Otolaryngology–Head & Neck Surgery*, 137 (11), 1094.
8. Conway, D.I., Petticrew, M., Marlborough, H., Berthiller, J., Hashibe, M., and Macpherson, L.M.D., 2008. Socioeconomic inequalities and oral cancer risk: A systematic review and meta-analysis of case-control studies. *International Journal of Cancer*, 122 (12), 2811–2819.

9. Daraei, P. and Moore, C.E., 2015. Racial Disparity Among the Head and Neck Cancer Population. *Journal of Cancer Education*, 30 (3), 546–551.
10. Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F., 2015. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012: Globocan 2012. *International Journal of Cancer*, 136 (5), E359–E386.
11. Florida Cancer Data System 2014 [online] Available from: http://www.flhealthcharts.com/charts/OtherIndicators/NonVitalIndNoGrpDataViewer.aspx?cid=0444[Accessed 25 June 2017].
12. Geography of Florida - World Atlas 2017 [online]. Available from: http://www.worldatlas.com/webimage/countrys/namerica/usstates/flland.htm [Accessed 8 Jul 2017].
13. Hillsborough County:Demographics: Hillsborough County Community Atlas [online], 2017. Available from: http://hillsborough.communityatlas.usf.edu/demographics/default.asp?ID=12057&level=cnty [Accessed 8 Jul 2017].
14. Jacob, B.G., Novak, R.J., Toe, L., Sanfo, M.S., Afriyie, A.N., Ibrahim, M.A., Griffith, D.A., and Unnasch, T.R., 2012. Quasi-likelihood techniques in a logistic regression equation for identifying Simulium damnosum s.l . larval habitats intra-cluster covariates in Togo. *Geo-spatial Information Science*, 15 (2), 117–133.
15. Javadi, P., Sharma, A., Zahnd, W.E., and Jenkins, W.D., 2017. Evolving disparities in the epidemiology of oral cavity and oropharyngeal cancers. *Cancer Causes & Control*, 28 (6), 635–645.
16. Johnson, S., McDonald, J.T., Corsten, M., and Rourke, R., 2010. Socio-economic status and head and neck cancer incidence in Canada: A case-control study. *Oral Oncology*, 46 (3), 200–203.
17. Johnsson, T., 1992. A procedure for stepwise regression analysis. Statistical Papers, 33 (1), 21–29.
18. Land Cover Institute (LCI) [online], 2011. Available from: https://landcover.usgs.gov/landcoverdata.php#regional [Accessed 11 Aug 2017].
19. McGorray, S.P., Guo, Y., and Logan, H., 2012. Trends in incidence of oral and pharyngeal carcinoma in Florida: 1981-2008: Oral and pharyngeal carcinoma trends. *Journal of Public Health Dentistry*, 72 (1), 68–74.
20. Osazuwa-Peters, N., Adjei Boakye, E., Hussaini, A.S., Sujijantarat, N., Ganesh, R.N., Snider, M., Thompson, D., and Varvares, M.A., 2017. Characteristics and predictors of oral cancer knowledge in a predominantly African American community. *PLOS ONE*, 12 (5), e0177787.
21. Osazuwa-Peters, N., Massa, S.T., Christopher, K.M., Walker, R.J., and Varvares, M.A., 2016. Race and sex disparities in long-term survival of oral and oropharyngeal cancer in the United States. *Journal of Cancer Research and Clinical Oncology*, 142 (2), 521–528.
22. Posorski, E., Boyd, L., Giblin, L.J., and Welch, L., 2014. Oral Cancer Awareness Among Community-Dwelling Senior Citizens in Illinois. *Journal of Community Health*, 39 (6), 1109–1116.
23. Riley, J.L., Pomery, E.A., Dodd, V.J., Muller, K.E., Guo, Y., and Logan, H.L., 2013. Disparities in Knowledge of Mouth or Throat Cancer Among Rural Floridians: Disparities in Knowledge of Mouth or Throat Cancer. *The Journal of Rural Health*, 29 (3), 294–303.

24. Schluchter, M.D., 2014. Mean Square Error. *In*: N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, and J.L. Teugels, Wiley StatsRef: Statistics Reference [Online]. Chichester, UK: John Wiley & Sons, Ltd.
25. Shepperd, J.A., Howell, J.L., and Logan, H., 2014. A survey of barriers to screening for oral cancer among rural Black Americans: Barriers to oral cancer screening. *Psycho-Oncology*, 23 (3), 276–282.
26. Shield, K.D., Ferlay, J., Jemal, A., Sankaranarayanan, R., Chaturvedi, A.K., Bray, F., and Soerjomataram, I., 2017. The global incidence of lip, oral cavity, and pharyngeal cancers by subsite in 2012: Lip, Oral Cavity, and Pharyngeal Cancers. CA: *A Cancer Journal for Clinicians*, 67 (1), 51–64.
27. Siegel, R.L., Miller, K.D., and Jemal, A., 2017. Cancer statistics, 2017. CA*: A Cancer Journal for Clinicians*, 67 (1), 7–30.
28. TIGER/Line with Selected Demographic and Economic Data [online]. The *United States Census Bureau*. Available from:https://www.census.gov/geo/maps-data/data/tiger-data.html>[Accessed 8 July 2017].
29. United States Cancer Statistics: Data Visualizations [online], 2017. Available from: https://nccd.cdc.gov/USCSDataViz/rdPage.aspx [Accessed 8 Jul 2017].
30. U.S Census Bureau, 2010 American Fact Finder [online]. Available from: https://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml [Accessed 25 June 2017].
31. Warnakulasuriya, S., 2009. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncology*, 45 (4–5), 309–316.
32. Weatherspoon, D.J., Chattopadhyay, A., Boroumand, S., and Garcia, I., 2015. Oral cavity and oropharyngeal cancer incidence trends and disparities in the United States: 2000–2010. *Cancer Epidemiology*, 39 (4), 497–504.