ESE 502. PS 1

Dr. Tony E. Smith

Zixi Liu

Explaining the space-time diffusion effect of Burkitt's Lymphoma cases in Uganda

Introduction

The objective of this epidemiological study is to explain the space-time diffusion effect of Burkitt's Lymphoma cases from the West Nile region in Uganda. Uganda, a landlocked country in East-Central Africa as shown in figure 1, is the world's second youngest country with a median age of 16.7 years and is rich with natural resources and fertile land. Our study area, West Nile region, is located at Northern Uganda where Burkitt's Lymphoma has been reported and analyzed. Relatively rare in Western countries, Burkitt's Lymphoma is a form of lymphoma common in Central Africa and most common among children. It has been recognized as a fast-growing human tumor associated with impaired immunity and could be fatal if left untreated.

This study will involve a regression analysis by nearest neighbor based on a dataset of 188 Burkitt's Lymphoma cases from the West Nile region. This data contains the locations and time of occurrence of Burkitt's Lymphoma, patient age, together with the population density in West Nile region. In order to study the diffusion effect of these cancer cases, patients were categorized by age into 5 to 10-year-old children (aged 5 to 10; n=137) and all other cases (all other cases; n = 51). The maps of Burkitt's Lymphoma cases by age group is shown in figure 2, in which there are significant concentrations of cases in both groups.

Figure 1: West Nile District in Uganda

As we can see in the figure, the concentrations of these cancer cases are mainly in densely populated areas, which indicates that the pattern of Burkitt's Lymphoma cases is influenced by population density and spatial clustering of patients of Burkitt's Lymphoma. However, there are other factors to consider before we determine there is clustering; we must evaluate the spatial autocorrelation as spatial dependence may exist between regions. In view of the key assumption when testing on the hypothesis of Complete Spatial Randomness, we must ensure the locations of these cases have no influence on each other. That is, under Complete Spatial Randomness, the random variables are assumed to be statistically independent in the region.

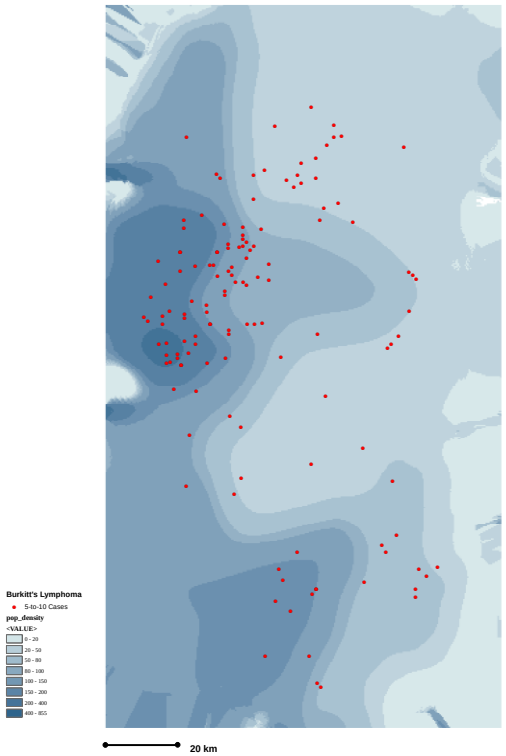

Figure 2a: Burkitt's Lymphoma cases aged 5 to 10.

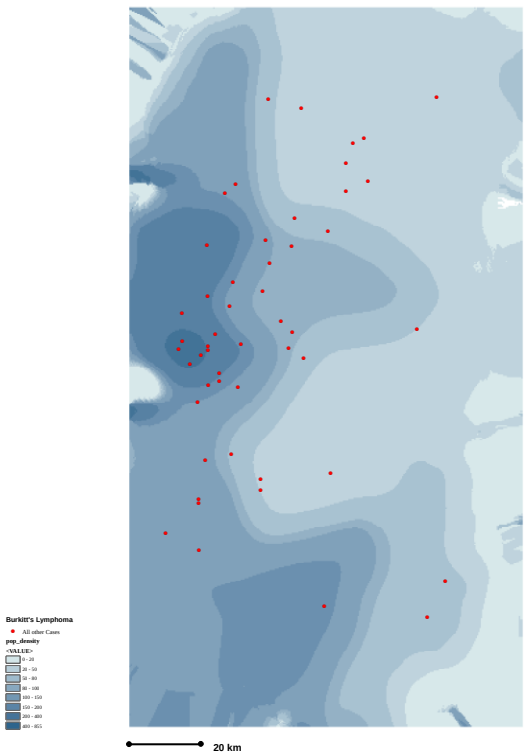

Figure 2b: Burkitt's Lymphoma cases of all other ages.

Methodologies and Outcomes

In this study, we think that the onset time for each Burkitt's Lymphoma case is associated with the onset time for its nearest neighbor. If there is similarity between each case's and its nearest neighbor's onset time, then we would recognize this as the evidence of a possible spatial diffusion effect. Therefore, we used nearest neighbor onset time to predict onset time by linear regression; we regressed the onset time for person i on the onset time for person j. Here is the formula we used for regression: $Time = \beta_0 + \beta_1 NeighTime + \varepsilon$.

In this formula, β_0 and β_1 are coefficients of regression and ε is the unobserved residual, which is the vertical distance between a data point and the fit line. If there is no contagion effect, then the coefficients β_0 and β_1 would both be zero. Then we constructed nearest neighbors using MATLAB and ran regression using JMP. From the regression outcome as shown in figure 3, we can see a point scatter of the two variables, which shows little monotonic relation. We finally arrived at a positive β_1 with significance level of 0.05, which means there is a little contagion effect.

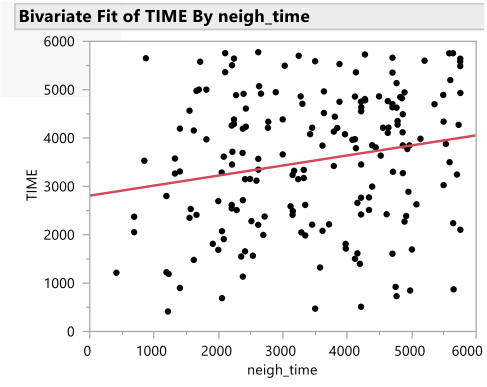

Figure 3: Scatterplot of Time by Neighbor Time, and the Best Fit Line.

The regression result we reached is as follows: $Time = 2808.931 + 0.2078091 * NeighTime$. In order to interpret our regression result, it is important to first examine R-squared value as it explains the goodness of fit. In this case, we arrived at a regression model with very low R-squared ($R^2 = 0.041$), low adjusted R-Square ($R^2adj = 0.036$) and low P-values (<0.0001 for β_0 and 0.0052 for β_1). As can be seen from figure 3, the slope coefficient is 0.2078091, which means if we move left or right along the x-axis by an amount that represents one unit change in time, the fitted line rises or falls by 0.2078091 units. Then we have a very low R-squared value of 0.041 and very low P-values, which indicates that even noisy, high-variability data can have a significant trend. The trend indicates that neighbor onset time still provides information about the onset time even though there are data points fall further from the regression line. In terms of P-values, the P-value for each term tests the null hypothesis that the coefficient is equal to zero. In this case, we have low values for both β_0 and β_1 (<0.05), which indicates that we can reject the null hypothesis and conclude that changes in the neighbor onset time is related to changes in Burkitt's Lymphoma's onset time.

Discussion

The reason there is a positive slope is that there is an agent transmitted between individuals that enables one Burkitt's Lymphoma case to induce the case of its neighbor. That is, we think having Burkitt's Lymphoma is associated with some other disease that was able to transmit between people. Williams et al. pointed out in 1978 in his article *Space-time Clustering of Burkitt's Lymphoma in the West Nile District of Uganda: 1961-1975* that chronic infection of malaria is the cofactor of EBV infection, which leads to the development of Burkitt's Lymphoma, and it is the former which primarily determined the onset time of Burkitt's Lymphoma. This suggests that the diffusion effect of Burkitt's Lymphoma in West Nile region is not caused by Epstein-Barr virus itself but induced by the malaria infection and has a relation to the exposure to mosquitos.