



# **Employing Markov Chain Monte Carlo (MCMC) Bayesian Poissonian and a Second-Order Eigenfunction Eigendecomposition Algorithm to Geostatistically Target Landscape Covariates Associated with Leukemia in Hillsborough County, Florida**

**Aarya Satardekar**

ORCID: 0009-0004-3749-8401

Samuel P. Bell III College of Public Health,  
University of South Florida, Tampa, FL, USA

**Jing Liu**

Department of Epidemiology and Biostatistics,  
Samuel P. Bell III College of Public Health,  
University of South Florida, Tampa, FL, USA

**Heather McDonald**

Samuel P. Bell III College of Public Health,  
University of South Florida, Tampa, FL, USA

**Benjamin Jacob**

Department of Global Health,  
Samuel P. Bell III College of Public Health,  
University of South Florida, Tampa, FL, USA

## **ABSTRACT**

Leukemia is a cancer of the blood and bone marrow, that hinders the normal production of healthy blood cells. In exploring mathematical hypotheses for leukemia, three distinct approaches are proposed. First is an over-dispersed Poisson leukemia regression model, with the consideration of outliers being addressed by applying a negative binomial model featuring a non-homogenously distributed mean. Secondly, an eigenfunction, eigendecomposition spatial filter algorithm is introduced, aiming to identify potential leukemia clusters based on hyper/hypo-endemic aggregation/non-aggregation orientations. Lastly, a Bayesian hierarchical model is advocated for determining causation covariates within a non-frequentistic model. This research examined the spatial aggregation of leukemia cases by utilizing sociodemographic data at the zip code level in Hillsborough County, Florida. The investigation involved spatial autocorrelation and Bayesian analyses to pinpoint the covariates linked to the risk of leukemia. The Poissonian regression model revealed a nondispersed paradigm. Hence, there was no need to utilize the negative binomial regression to treat the outliers. A second-order eigenfunction eigendecomposition revealed multiple non-zero autocorrelated clusters throughout various zip codes in Hillsborough County. The hot spots were 33647, 33578, and 33511 and the cold spots were 33621, 33503, and 33530. Our

**proposed approach identifies leukemia hotspots among whites and Asians aged 65+. Urban residential communities in 33647 were most vulnerable to leukemia. By leveraging social media and fostering collaboration between public health and technology sectors, we can work towards achievable leukemia prevention for at-risk populations. Future research should explore the method's applicability at the state level and promote routine blood tests among people.**

**Keywords:** Leukemia, zip-code, Spatial autocorrelation, Bayesian analyses, Hillsborough County, Florida

## INTRODUCTION

Leukemia is a prevalent form of hematologic cancer characterized by abnormal growth of blood cells and bone marrow. The various forms of leukemia are classified based on their pace of progression, categorized as either acute or chronic, and their origin in either lymphocytic or myelogenous cells [1]. According to the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program, leukemia's general five-year relative survival rate is 66.7 percent [2].

There are two main types of leukemia: acute and chronic. Acute leukemia progresses quickly, while chronic leukemia advances more slowly and requires less rapid treatment. Furthermore, leukemia is categorized as myelogenous or lymphocytic. Atypical cell development in bone marrow cells intended to develop into lymphocytes—essential elements of the immune system—occurs in lymphocytic leukemia. On the other hand, aberrant proliferation of marrow cells that mature into red blood cells, white blood cells, and platelets is indicative of myelogenous leukemia. Acute Lymphocytic Leukemia (ALL), Acute Myelogenous Leukemia (AML), Chronic Lymphocytic Leukemia (CLL), and Chronic Myelogenous Leukemia (CML) are the four primary classifications of leukemia. Leukemia affects both adults and children, regardless of age. Acute Lymphocytic leukemia is the most common kind of childhood leukemia, closely followed by Acute Myelogenous leukemia; in adults, AML and CLL are the most common varieties. Leukemia symptoms might vary depending on the specific type and stage of the disease [3].

Leukemia specifically targets the genetic material of white blood cells, affecting both children and adults, with a predominant impact on the body's white blood cells. White blood cells play a vital role in the human immune system, yet individuals with leukemia often generate an excessive quantity of malfunctioning white blood cells. The condition triggers uncontrolled growth and division of cells, disrupting the normal regulation of blood cell production. Consequently, mutated cells overcrowd healthy blood cells in the bone marrow, diminishing the population of functional white blood cells. The treatment for leukemia is highly complex and has not yet proven entirely effective in curing the disease [4].

Common signs and symptoms include fever, chills, and night sweats, along with general weakness and exhaustion. Other signs and symptoms could be bleeding or swollen gums, headaches, enlarged tonsils and liver, bone discomfort, paleness, and tiny red patches on the skin. Weight loss is another important symptom [3].

Chemotherapy, radiation therapy, monoclonal antibodies, and hematopoietic stem cell transplantation are among the possible treatments for acute leukemia. The patient's age, any coexisting medical conditions, the cytogenetic and molecular results, the subtype of leukemia, and other factors all influence the treatment plan. Tyrosine kinase inhibitors brought about a revolutionary change in the way chronic myelogenous leukemia is treated, and they are currently the primary method of treatment. The curative therapeutic option is hematopoietic stem cell transplantation, which is often recommended for younger individuals or in situations in which the disease is not responding to tyrosine kinase inhibitors [5].

Treatment-related complications encompass tumor lysis syndrome and severe infections resulting from immunosuppression. Once a leukemia patient has entered remission, they require vigilant monitoring for potential secondary malignancies, cardiac issues, and endocrine disruptions such as metabolic syndrome, hypothyroidism, and hypogonadism. The highest five-year survival rates are observed among younger patients and those diagnosed with chronic myelogenous leukemia or chronic lymphocytic leukemia [6].

While the exact risk factors for different types of leukemia may vary, some commonly known risk factors include age (as older adults are more susceptible), demographics (white men are more statistically at risk), radiation exposure (especially from atomic bombs), history of cancer treatments (such as chemotherapy or radiation therapy), genetic disorders (including blood disorders and Down syndrome), family history of leukemia (indicating a genetic predisposition), and other environmental factors (like tobacco smoke, pesticides, and industrial solvents) that may contribute to the development of leukemia. Leukemia is a disease that often goes unnoticed in its early stages and can only be detected by routine blood tests [3]. Ionizing radiation is a major risk factor that can contribute to the development of leukemia. As per the American Society for Hematology (2021), there was an elevated risk of leukemia-related death even with minimal accumulated radiation doses (<5 milliGray unit) [7]. This implies that the presumed threshold at which radiation becomes harmless should be set at an exceptionally low level. Furthermore, Leuraud et al. presents compelling proof of links between extended exposure to low doses of radiation and the occurrence of leukemia [8]. Lastly, according to Little et al, the likelihood of developing acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) showed a noteworthy rise with cumulative doses of ionizing radiation below 100 mSv (Sievert unit) during childhood or adolescence. Additionally, an excess risk was observed for cumulative radiation doses below 50 mSv (sievert unit) for certain outcomes [9]. These results indicate an elevated risk of leukemia connected to low-dose radiation exposure, suggesting that the existing radiological protection system is judicious and not excessively cautious [8].

Second-order eigendecomposition autocorrelation has been used to determine hot and cold spots for various diseases [10, 11]. Spatial autocorrelation is the correlation among values of a single variable strictly attributable to their relatively close locational positions on a two-dimensional surface, introducing a deviation from the independent observation's assumption of classical statistics [12]. A Bayesian hierarchical estimation of variables associated with hot and cold spots may identify prognosticators associated with the seasonal cluster [10]. Wang (2021) employed spatial autocorrelation to confirm that there was a correlation among confirmed cases of COVID-19 in China. In the literature contribution, the correlation characteristics were first increased and then decreased. However, considering localized,

residual, unbiased, non-zero, and geo-spatiotemporal autocorrelation, the characteristics tended to stabilize over time. Wang (2021) models revealed positive spatial autocorrelation (PSA) stratified hot spots stabilized over time in the areas bordering Hubei (Henan et al.) [13]. Substituting its modified version devised, we assumed we would be able to accommodate residual PSA in a subcounty, georeferenceable, [GPS indexable] hierarchical, non-zero, autocorrelatable, unbiased, diagnostic, zip-code-level, stratifiable, asymptotical, non-multicollinear, homoscedastic, leukemia vulnerability model. We did so precisely for predictively targeting georeferenceable, potential occurrence, abundance, and distribution sites of leukemia at the zip-code level [14].

We employed the Spatial Autocorrelation Moran's  $I$  tool in ArcGIS Pro™ to measure residual, non-zero autocorrelation in the empirical, georeferenced, eigendecomposed, leukemia-related, estimator determinant dataset. Using a set of sociodemographic zip code stratified, epidemiological, georeferenceable data, capture point, diagnostic, and feature attributes, this tool evaluated whether eigen-decomposable, eigenfunction eigenvectors derived from weighted, aggregation/non-aggregation-oriented, hyper/hypo-endemic model were clustered dispersed, or random. Here the tool calculated the Moran's  $I$  value and both a z-score (i.e., standard deviations) and p-value to evaluate the significance of the eigen-spatial filter, eigen-orthogonalized, georeferenceable, diagnostic, zip code-stratified estimator determinants.

Moran's Indices ( $I$ ) statistics were employed in this analysis to estimate non-zero, probabilistic, latent autocorrelation and spatial distribution in the multivariate, geosampled, zip code-level, stratifiable, sociodemographic, leukemia-related, georeferenced covariates. Moran's  $I$  is a measure of geospatial autocorrelation, which is characterizable by a correlation in a signal among nearby georeferenced locations in eigenvector eigen-geospace [12]. Geo-spatiotemporal autocorrelation exists because real-world phenomena are typified by orderliness, pattern, and systematic concentration rather than randomness. Tobler's First Law of geography encapsulates this situation: "Everything is related to everything else, but near things are more related than distant things." To this maximum, the qualifier should be added: "but not necessarily through the same mechanisms." In other words, autocorrelation means a dependency exists between georeferenced epidemiological, sampled, data capture points [i.e., re-infection leukemia cases in a hyper/hypo-endemic, transmission-oriented, 'hot/ cold spot zip code geolocation] based on an empirical epidemiological proxy dataset of georeferenceable, geosampled, regression-related, prognosticative, diagnostic, estimator determinants [e.g., a high housing socioeconomic landscape index or uninsured migrant farmer neighborhood] in proximal stratifiable geolocations. Autocorrelation can also occur as a systematic pattern in values of other exogenous observational predictors across geolocations on a georeferenced district/county map due to underlying common co-factors [e.g., high leukemia prevalence in a zip code, geospatial cluster] [12]. Latent geo-spatiotemporal autocorrelation in epidemiological models is more complex than one-dimensional (d) autocorrelation since spatial correlation is multi-dimensional (i.e., 2 or 3 dimensions of space) and multi-directional [10]. Hence, we assumed that Moran's  $I$  may help to optimize [i.e., remove residual, non-normal, multicollinear and/or skew heteroscedastic, zero autocorrelated, non-asymptotical parameters) in an eigendecomposed, georeferenced, empirical, epidemiological sample dataset of zip code stratified, hierarchical, non-diffusion-related, hyper/hypo-endemic, leukemia-related sociodemographic estimator determinants] and their distribution patterns in geo-spatiotemporal, eigenvector eigen-geospace.

Maximum likelihood or Bayesian approaches can account for geo-spatiotemporal classifiable dependencies in a parametric framework, whereas recent Bayesian eigenvector spatial filtering approaches focus on semi-parametrically removing autocorrelation [10]. In this experiment, we employ a Bayesian Poissonian semiparametric eigen-spatial filtering approach which allowed us to deal explicitly with (a) lagged autoregressive models and (b) simultaneous autoregressive models for heuristically optimizing forecasting empirically dependent, potential georeferenceable hierarchical, diffusion-oriented, zip code stratifiable, hot/cold spot, quantifiable, leukemia-related, estimator determinants for cartographically and geostatistically delineating, hyper/hypoendemic, clustering/non-clustering propensities. In so doing, a leukemia researcher or epidemiologist may be able to iteratively quantitate trends in an empirical geosampled, geo-referenceable dataset of sociodemographic, leukemia-related epidemiological, and estimator determinants in Bayesian eigenvector eigen-geospace.

Our proposed iterative Bayesian leukemia model recursively updated the posterior distribution of the zip code stratified, georeferenced, epidemiological diagnostic model estimators. While simultaneously aggressively quantitating the process with new empirical, geo-spatiotemporal, predictive samples (i.e., evidential likelihoods) were drawn from a proposal density in highly probable eigenvector eigen-Bayesian eigen-geospace. We assumed that over iterations, the proposal density would progressively become localized near the posterior modes and, in doing so, would allow defining stratifiable, georeferenceable, hot/cold spot, zip code geolocations using asymptotically normalized, (e.g., non-multicollinear, non-skew, non-heteroscedastic, non-zero auto-correlatable, unbiased), geosampled, Gaussian, georeferenced, diagnostic estimator determinants. In our leukemia model, the posterior mean and posterior mode were the mean and mode of the posterior distribution of  $\Theta$ .

Here, the Dirichlet process Gaussian mixture was trained with sparse eigenvalues from the previous iteration to update the proposal density in a multivariate, zip code stratified, prognosticative, leukemia-related, epidemiological, forecast non-frequentistic model framework. We employed the Dirichlet Process Gaussian-mixture model, which is a fully non-parametric Bayesian Poissonian methodology to estimate probability density functions (PDF) with a flexible set of assumptions. The probability density function is a statistical expression that defines a probability distribution (the likelihood of an outcome) for a discrete random variable as opposed to a continuous random variable [15]. This experiment presents a novel algorithm, based upon the dependent Dirichlet process mixture model (DDPMM), for optimally capturing batch-sequential, non-time, series-dependent, epidemiological, empirical, leukemia-related, sociodemographic, zip code stratified, estimator determinants containing an unknown number of evolving georeferenceable, geospatial clusters. The algorithm is derived via a low variance asymptotic analysis of the Markovian sampling algorithm for the DDPMM and provides a hard clustering with convergence guarantees similar to those of the k-means algorithm [i.e., a method of vector quantization, originally from signal processing], that aims to partition  $n$  observations into  $k$  clusters in which each observation [i.e., georeferenced, sociodemographic, zip code stratified, leukemia model estimator determinant] belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

We assumed that the empirical results from a synthetic test with moving Gaussian clusters and a test for unbiasing non-normality in a regressed dataset of zip code, stratifiable, sociodemographic, georeferenced, leukemia-related, geosampled, epidemiological, data,

capture points in Bayesian eigenvector eigen-geospace would demonstrate that the algorithm requires orders of magnitude less computational time than contemporary probabilistic clustering uncertainty algorithms. In so doing, multiple varying erroneous heterogeneities such as non-asymptotical, multicollinear, estimator determinant uncertainties due to violations of regression assumptions would be teased out. We assume the eigen-geospatial eigen-spatial filter algorithm would provide higher estimation accuracy on the examined, semi-parameterizable, generalizable, hierarchical, autocorrelated, leukemia-related, clustering/non-clustering, diagnostic, estimator determinants, residuals, iteratively derived from a Bayesian Poissonian model output.

We assumed that an eigendecomposed, Bayesian Poissonized, asymptotic, estimator model may optimize predictively targeting and prioritizing, hierarchical, diffusion-related, potential hyper/hypoendemic, aggregation/non-aggregation-oriented, leukemia-related, georeferenceable, zip code level geolocations by generating error-free estimator determinants. Asymptotic regression is appropriate when a contiguous region on the right or left (or both) of the input is expected to deviate systematically from the regression model in a finite fashion [12]. Such regions ("non-asymptotic Bayesian regressed regions") often occur in time series and non-time series dependent, vulnerability-oriented, viral infection, risk model analysis [10,12]. We assumed that the local asymptotic normality property may be non-frequentistically quantifiable for an empirical, epidemiological georeferenceable dataset of semi-parameterizable, sociodemographic stratified, leukemia-related estimator determinants, for constricting robust prognosticative, vulnerability-oriented, regression models for heuristically optimizing targeting and prioritizing, sub-county, potential, hyper/hypo endemic, aggregation/non-aggregation-oriented, sites at the zip code level in Hillsborough County.

Regrettably, regression models have not been able to delineate the occurrence, abundance, and distribution of leukemia cases in any county in the United States. Initially, in this research, we created a vulnerability index for every zip code in Hillsborough County using a population stratification model. Then, we constructed a Poissonian count variable model to determine covariates associated with leukemia at the zip code level. Subsequently, we employed an eigendecomposition algorithm to determine hot and cold spots for leukemia in each zip code. Utilizing count variable regression models and eigencorrelation can describe co-variables associated with occurrence, abundance, and distribution at the zip code level [16]. In exploring mathematical hypotheses for leukemia, three distinct approaches are proposed. Firstly, an over-dispersed Poisson leukemia regression model is suggested, with the consideration of outliers being addressed by applying a negative binomial model featuring a non-homogeneously distributed mean. Secondly, an eigenfunction, eigendecomposition spatial filter algorithm is introduced to identify potential leukemia clusters based on hyper/hypo-endemic aggregation/non-aggregation orientations. Lastly, a Bayesian hierarchical model is advocated for determining causation covariates within a non-frequentistic spatial autocorrelation model. This research examined the spatial aggregation of leukemia cases by utilizing sociodemographic and clinical data at the zip code level in Hillsborough County, Florida. The investigation involved spatial autocorrelation and Bayesian analyses to pinpoint the covariates linked to the risk of leukemia in Hillsborough County, Florida. The Poissonian regression model revealed a nondispersed paradigm. Hence, we did not need to utilize the negative binomial regression to treat the outliers. We conducted a second-order eigenfunction

eigendecomposition, which revealed multiple non-zero autocorrelated clusters throughout various zip codes in Hillsborough County.

## MATERIAL AND METHODOLOGY

### Study Site

Hillsborough County is located in the west-central portion of the U.S. state of Florida, as part of the Tampa–St. Petersburg–Clearwater Metropolitan Statistical Area. According to the U.S. Census Bureau [17], the population in this county was 1,459,762, one of the most populous counties in the state. The estimated population in 2022 of Hillsborough County was 1,513,301 people, with a yearly growth rate of 3.7% [17]. The county has a total area of 1,266 square miles (3,279 km<sup>2</sup>), of which 1,020 square miles (2,647 km<sup>2</sup>) are land and 246 square miles (637 km<sup>2</sup>) (19.4%) are covered by water [17]. Hillsborough is home to some important water bodies, such as the Alafia River and Hillsborough River the Hillsborough River, and the Little Manatee River [18]. The county's unincorporated area is around 888 square miles (2,300 km<sup>2</sup>), more than 84% of the total land area. Municipalities account for 163 square miles (420 km<sup>2</sup>). The modern boundaries of the county place it midway along the west coast of Florida. Hillsborough County has 55 standard zip codes [19].

### Zip Code, Topological, Land Cover, Vulnerability Indices

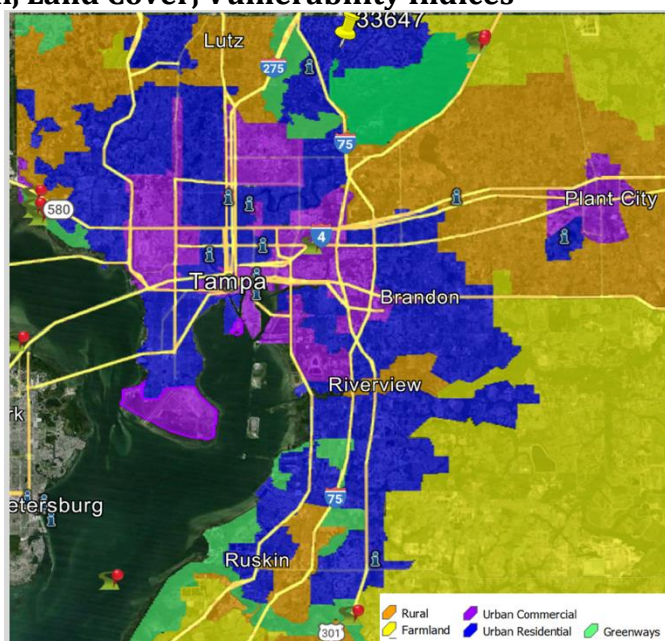


**Figure 1: Land Use Land Cover map for Hillsborough County**

We constructed zip code probabilities from population-stratified leukemia-related age and racial-related covariates such as being American Indian or Alaska Native, Asian, Black, Hispanic, Native Hawaiian or other Pacific Islander which were acquired from the United States Census Bureau website. We employed satellite, sensed, signature, zip code-stratified, land use land cover [i.e., LULC] (Figure 1) capture point, sentinel site surveys of the potential, sociodemographic hyper/hypoendemic, aggregation/non-aggregation-oriented, georeferenced covariates. The digital overlay revealed the georeferenceable, capture point, LULC-classified surface area (m<sup>2</sup>) of each epidemiological, remote, sampled sentinel site, prognosticative variable within a zip code intervention geolocation in Hillsborough County. Land cover types

were classified into urban residential, urban commercial, rural, farmland, and greenways in ArcGIS Pro.

### Poisson Model

Initially, a Poisson regression with statistical significance was calculated by a 95% confidence in ArcGIS Pro using

$$P_p(n|N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} \quad (2.1)$$

We viewed the distribution as a function of the expected number of count variables using the sample size  $N$  for quantifying the fixed  $p$  in equation (2.1), which was then transformed into the equation:

$$\frac{P_v(n|N)}{N} = \frac{N!}{n!(N-n)!} \left(\frac{v}{N}\right)^n (1-p)^{N-n}$$

Based on the vulnerability index computed employing the zip code population as a response variable the model revealed;

$$\begin{aligned} & \lim_{n \rightarrow \infty} P_p(n|N) \\ &= \lim_{N \rightarrow \infty} \frac{N(N-1)\dots(N-n+1)}{n!} \frac{v^n}{N^n} \left(1 - \frac{v}{N}\right)^N \left(1 - \frac{v}{N}\right)^{-n} \\ &= \lim_{N \rightarrow \infty} \frac{N(N-1)\dots(N-n+1)}{n!} \frac{v^n}{n!} \left(1 - \frac{v}{N}\right)^N \left(1 - \frac{v}{N}\right)^{-n} \\ &= 1 \cdot \frac{v^n}{n!} \end{aligned}$$

As expected, the Poisson distribution was normalized so that the sum of probabilities equaled 1. The ratio of probabilities was then quantifiable by

$$\sum_{n=0}^{\infty} P_v(n) = e^{-v} \sum_{n=0}^{\infty} \frac{v^n}{n!} = e^{-v} e^v =$$

which was then subsequently expressed as

$$\frac{P_v(n=i+1)}{P(n=i)} = \frac{\frac{v^{i+1} e^{-v}}{(i+1)!}}{\frac{e^{-v} v^i}{i!}} = \frac{v}{i+1}.$$

### Autocorrelation Model

Spatial autocorrelation was evaluated amongst the sampled leukemia covariates at the Hillsborough County study site using Moran's  $I$ . In statistics, Moran's  $I$  is a measure of spatial autocorrelation [12].

An autocorrelation model specification was used to describe Poisson's random autocorrelated, leukemia-related explanatory variables. The resulting model specification took on the following form:  $\mathbf{Y} = \mu(\mathbf{I} - \rho)\mathbf{1} + \rho\mathbf{W}\mathbf{Y} + \varepsilon$ , (2.1) where  $\mu$  was the scalar conditional mean of  $\mathbf{Y}$ , and  $\varepsilon$  was an  $n$ -by-1 error vector whose elements were statistically independent and identically distributed (i.i.d.) normally random variates. The spatial covariance matrix for equation [2.1] employing the sampled leukemia sociodemographic covariates was  $E[(\mathbf{Y} - \mu\mathbf{1})(\mathbf{Y} - \mu\mathbf{1})'] = \Sigma = [(\mathbf{I} - \rho\mathbf{W})(\mathbf{I} - \rho\mathbf{W})']^{-1}\sigma^2$ , where  $E(\bullet)$  denoted the calculus of expectations,  $\mathbf{I}$  was the  $n$ -by- $n$  identity matrix delineating the matrix transpose operation, and  $\sigma^2$  was the error variance.

We assumed that a mixture of PSA and NSA, latent, spatial autocorrelation may be present in the sociodemographic prognosticative, leukemia model, which in turn would require an even more explicit representation of the Poissonian results. Varying spatial, autoregressive parameters appeared in the covariance matrix which for our model specification was  $\Sigma = [(\mathbf{I} - \langle \rho \rangle_{\text{diag}} \mathbf{W})(\mathbf{I} - \langle \rho \rangle_{\text{diag}} \mathbf{W})']^{-1}\sigma^2$ , (2.2) where the diagonal matrix of autoregressive parameters,  $\langle \rho \rangle_{\text{diag}}$ , contained two sampled parameters:  $\rho_+$  for those sociodemographic covariates displaying positive spatial dependency, and  $\rho_-$  for those displaying negative spatial dependency. For example, by letting  $\sigma^2 = 1$  and employing a 2-by-2 regular square tessellation, revealed

$$\Sigma = \left[ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \rho_+ & 0 & 0 & 0 \\ 0 & \rho_+ & 0 & 0 \\ 0 & 0 & \rho_- & 0 \\ 0 & 0 & 0 & \rho_- \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \right]^{-1} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$$

in the autocorrelation model [i.e.,  $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$ ]. This enabled positing a positive relationship between the sampled leukemia covariates,  $y_1$  and  $y_2$ , a negative relationship between covariates,  $y_3$  and  $y_4$ , and no relationship between covariates  $y_1$  and  $y_3$  and between  $y_2$  and  $y_4$ . This covariance specification yielded:

$$\mathbf{Y} = \mu(\mathbf{I} - \rho_+ \langle \mathbf{I}_+ \rangle_{\text{diag}} - \rho_- \langle \mathbf{I}_- \rangle_{\text{diag}})\mathbf{1} + (\rho_+ \langle \mathbf{I}_+ \rangle_{\text{diag}} + \rho_- \langle \mathbf{I}_- \rangle_{\text{diag}})\mathbf{W}\mathbf{Y} + \varepsilon \quad (2.3)$$

where  $\mathbf{I}_+$  was a binary 0-1 indicator variable which denoted those sociodemographic covariates displaying positive spatial dependency, and  $\mathbf{I}_-$  was a binary 0-1 indicator variable denoting

those sampled covariates displaying negative spatial dependency, using  $\mathbf{I}_+ + \mathbf{I}_- = \mathbf{1}$ . Expressing the preceding 2-by-2 example in terms of equation [2.3] yielded:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \mu \left[ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \rho_+ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} - \rho_- \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \left[ \rho_+ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \rho_- \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right] \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}.$$

If either  $\rho_+ = 0$  (and hence  $\mathbf{I}_+ = \mathbf{0}$  and  $\mathbf{I}_- = \mathbf{I}$ ) or  $\rho_- = 0$  (and hence  $\mathbf{I}_- = \mathbf{0}$  and  $\mathbf{I}_+ = \mathbf{I}$ ), then equation [2.3] is reduced to equation (2.1). This indicator variables classification was made in accordance with the quadrants of the corresponding Moran's scatterplot generated employing the Poissonian sampled socio-demographic and age-related leukemia zip code-stratified covariates sampled in the Hillsborough County intervention study site.

To identify georeferenceable, sociodemographic leukemia stratified clusters of Thiessen polygon surface partitioning were generated in ArcGIS Pro to construct geographic neighbor matrices, which were subsequently employed in the autocorrelation analysis. Entries in the matrix were 1 if two sampled covariates shared a common Thiessen polygon boundary and 0, otherwise. Next, the linkage structure for each surface was edited to remove unlikely geographic neighbors to identify pairs of zip code stratifiable leukemia covariates sharing a common Thiessen polygon boundary. Attention was restricted to those georeferenceable map patterns associated with at least a minimum level of spatial autocorrelation, which, for implementation purposes. In this experiment, we defined by  $|MC_j/MC_{\max}| > 0.25$ , where  $MC_j$  denoted the  $j$ th value and  $MC_{\max}$ , the maximum value of MC. We assumed that this threshold value would allow two candidate sets of leukemia-related sociodemographic stratified, eigendispersed, eigen-spatial filter eigenvectors to be considered for substantial PSA and NSA respectively. Of note is that the ratio of the PRESS (i.e., predicted error sum of squares) statistic to the sum of squared errors from Moran's scatterplot trend line was also evaluated in eigenfunction geospace.

## Bayesian Analyses

Bayesian estimation and MCMC methods were subsequently employed to model the georeferenced, zip code-level, epidemiological, and diagnostic determinants. MCMC methods are primarily used for calculating numerical approximations of multi-dimensional integrals, for instance in Bayesian statistics, computational physics [20], and computational linguistics [21]. In Bayesian statistics, the recent development of MCMC methods has made it possible to compute large, integrative datasets [e.g., georeferenced, district-level sub-county, aggregation/non-aggregation-oriented, hierarchical, diffusion-related, epidemiologically

regressively, forecastable, stratified, time series, dependent, diagnostic, model estimators] that require integrations over hundreds to thousands of unknown parameters [10].

This paper deals with a computational aspect of the Bayesian Poisson analysis of statistical models with intractable normalizing constants. We propose here a general approach to sample from such posterior distributions that bypasses the computation of the normalizing constant. Our method can be considered a Bayesian version of the MCMC estimation approach of [22]. We assumed our approach to the asymptotic behavior of the algorithm could extract normalized, generalizable, inferential model estimator determinants in R. These paradigms were used to quantitate heterogeneity of variances and other propagation non-normalities in the zip code stratified, hierarchical, diffusion-related, diagnostically specified, sociodemographic semi-parameterizable, estimator dataset. The natural logarithms of variances were modeled employing a linearized model for quantifying heterogeneity of the variances (on a logarithmic scale) in terms of the diagnostic and epidemiological geosampled variables.

The model's specific variance parameter was an independent draw from a random sampling distribution. The MCMC sampling began with quantitating conditional (marginal) probability distributions. Subsequently, iterative, space-time, diagnostically stratifiable, leukemia-related, semi-parameterizable estimates were obtained using pseudo-likelihood estimation (i.e., an autoregressive term estimated with a conventional regression procedure). This involved approximating the regression-related coefficients ( $\beta$ ) and  $\rho$  as though the epidemiological diagnostic, estimator determinants were independent. MCMC outputs with sample values for a parameter were drawn from the joint posterior probability distribution.

In the first stage of the inferential Bayesian Poissonian analyses, a likelihood model was specified for the epidemiologic, leukemia-related, geosampled count data. In the second stage, the georeferenced, explanatory, prognosticator variables were analyzed to specify a prior model. MCMC was subsequently employed to recognize conjugate specifications (e.g., Poisson-gamma), in the stratified, diagnostic, epidemiological data. The model assumed that the number of case counts in Hillsborough County,  $i$ ,  $Y_i$ , had a conditional, independent, Poisson distribution with mean  $E_i \exp(\mu_i)$ . The variable  $E_i$  was employed as the expected number of sampling events, which in this experiment was proportional to the corresponding known case count, population,  $n_i$  at the zip code level. The expression  $\exp(\mu_i)$  was the relative risk based on the potentially geosampled, case count, sociodemographic, stratified, leukemia-related, discrete, integer values: regions with  $\exp(\mu_i) > 1$  having greater numbers of observed, leukemia count values than expected, and vice versa for specified, georeferenceable, zip code-level regions with  $\exp(\mu_i) < 1$ . The log-relative term was  $\mu_i$  which modeled the epidemiologically specified, empirical, diagnostically, zip code stratified, vulnerability-oriented, prognosticative, explanatory variables, linearly as:

$$\mu_i = \mathbf{x}_i' \boldsymbol{\beta} + \theta_i + \varphi_i, i = 1, \dots, I \quad (2.4)$$

In this experiment,  $\mathbf{x}_i'$  was the quantifiable propagation-related, aggregation/non-aggregation-oriented, diagnostic, leukemia-related, zip code, stratified regression estimators, and  $\boldsymbol{\beta}$  was a vector of fixed effects in the prognosticated model output. Additionally, the terms  $\theta_i$  and  $\varphi_i$  were employed for capturing georeferenceable, site-specific dependence in the empirical, diagnostic,

subcounty, and district-level estimators. In previous research, [10] employed an MCMC algorithm and an autocovariate uncertainty matrix to geo-spatiotemporally quantitate stochastic error propagation rendered from a Bayesian parametric, variable, estimator dataset approximated from multivariate, time series, regressed, epidemiologically geosampled, Covid-19 explanatory parameters which were grid-stratified by prevalence in the districts of Uganda. An R-based hierarchical agglomerative, polythetic, clustering algorithm was employed to forecast high and low, georeferenced, leukemia-related, zip code-stratified, geospatial clusters. Univariate statistics and Poisson count variable regression models were then constructed in R. An inferential, generalizable probabilistic Bayesian, uncertainty estimation matrix was constructed employing normal priors for each of the error coefficients for revealing both spatially structured random effects (SSRE) and spatially unstructured random effects (SURE).

## RESULTS

The Poisson distribution revealed that the explanatory coefficients reached a maximum Euler-Mascheroni constant and  $H_n$  was a harmonic number, leading to the transcendental equation  $\gamma - H_n + \ln v = 0$ . The regression model also revealed that the Euler-Mascheroni constant arose in the integrals, which were expressible as

$$\begin{aligned}\gamma &= -\int_0^{\infty} e^{-x} \ln x dx = -\int_0^1 \ln \ln\left(\frac{1}{x}\right) dx \\ &= \int_0^{\infty} \left(\frac{1}{1-e^x} - \frac{1}{x}\right) e^{-x} dx = \int_0^{\infty} \frac{1}{x} \left(\frac{1}{1+x} - e^{-x}\right) dx\end{aligned}\quad (3.1)$$

Commonly, integrals that are rendered  $\gamma$  in combination with temporal sampled constants include

$$\int_0^{\infty} e^{-x^2} \ln x dx = -\frac{1}{4} \sqrt{\pi} (\gamma - 2 \ln 2)$$

which is equal to

$$\int_0^{\infty} e^{-x} (\ln x)^2 dx = \gamma^2 + \frac{1}{6} \pi^2 \quad (3.2)$$

in public health epidemiological models [10, 23]. Thereafter, the double integrals in the leukemia county-level prognosticative regression model included

$$\gamma = \int_0^1 \int_0^1 \frac{x-1}{(1-xy) \ln(xy)} dx dy.$$

In this research, an interesting analog of equation (3.2) in the regression-based model was calculated as

$$\ln\left(\frac{4}{\pi}\right) = \sum_{n=1}^{\infty} (-1)^{n-1} \left[ \frac{1}{n} - \ln \frac{n+1}{n} \right] \quad ; \quad \int_0^1 \int_0^1 \frac{x-1}{(1+xy)\ln(xy)} dx dy = 0.241564... \gamma$$

By taking the logarithm of both sides in the county leukemia predictive model, an explicit formula for  $\gamma$  was then derivable employing

$$\gamma = \lim_{x \rightarrow \infty} \left[ \sum_{p \leq x} \ln \left( \frac{1}{1 - \frac{1}{p}} \right) - \ln \ln x \right].$$

This expression was also rendered coincidentally by quantifying the data series employing Euler, and equation (3.2) by first replacing

$$\ln n, \ln(n+1),$$

in the equation

$$\gamma = \sum_{k=1}^{\infty} \left[ \frac{1}{k} - \ln \left( 1 + \frac{1}{k} \right) \right]$$

and then generating

$$\lim_{n \rightarrow \infty} [\ln(n+1) - \ln n] = \lim_{n \rightarrow \infty} \left( 1 + \frac{1}{n} \right) = 0.$$

Thereafter, our product was

$$\lim_{n \rightarrow \infty} \left[ \sum_{k=1}^n \frac{1}{k} - \sum_{k=1}^n \ln \left( 1 + \frac{1}{k} \right) \right] = \lim_{n \rightarrow \infty} \sum_{k=1}^n \left[ \frac{1}{k} - \ln \left( 1 + \frac{1}{k} \right) \right]$$

Additionally, other series in our regression probabilistic model included the equation  $\langle \diamond \rangle$  where

$$\gamma = \sum_{n=2}^{\infty} (-1)^n \frac{\zeta(n)}{n} = \ln\left(\frac{4}{\pi}\right) + \sum_{n=1}^{\infty} \frac{(-1)^{n-1} \zeta(n+1)}{2^n (n+1)}$$

and  $\zeta(z)$  was

$$\gamma = \sum_{n=1}^{\infty} (-1)^n \frac{[\lg n]}{n}$$

plus, the Riemann zeta function. The Riemann zeta function  $\zeta(s)$  is a function of complex variables that analytically continues the sum of the infinite series  $\sum_{n=1}^{\infty} \frac{1}{n^s}$  which converges when the real part of  $s$  is greater than 1 where  $\lg$  is the logarithm to base 2 and the  $[x]$  is the floor function [24].

The outputs revealed that the leukemia model contained a constant term. As such, it was necessary to assume that  $E(e^{\tau_i}) = E(\tau_i) = 1$  to identify the mean of the distribution. We assumed that  $(\tau_i)$  followed a gamma  $(\theta, \theta)$  distribution in the model with  $E(\tau_i) = 1$  and  $V(\tau_i) = 1/\theta$ ,  $g(\tau_i) = \frac{\theta^\theta}{\Gamma(\theta)} \tau_i^{\theta-1} \exp(-\theta \tau_i)$

where  $\Gamma(x) = \int_0^\infty \zeta^{x-1} \exp(-\zeta) d\zeta$  was the gamma function and  $\theta$  was a positive parameter.

Thus, the density of  $y_i$  in the county, prognosticative leukemia, zip code models was  $X_i$  which in this experiment was solved as:

$$\begin{aligned} f(y_i | X_i) &= \int_0^\infty f(y_i | X_i, \tau_i) g(\tau_i) d\tau_i \\ &= \frac{\theta^\theta \mu_i^{y_i}}{y_i! \Gamma(\theta)} \int_0^\infty e^{-(\mu_i + \theta) \tau_i} \tau_i^{\theta + y_i - 1} d\tau_i \\ &= \frac{\Gamma(y_i + \theta)}{y_i! \Gamma(\theta)} \left( \frac{\theta}{\theta + \mu_i} \right)^\theta \left( \frac{\mu_i}{\theta + \mu_i} \right)^{y_i} \\ &= \frac{\theta^\theta \mu_i^{y_i} \Gamma(y_i + \theta)}{y_i! \Gamma(\theta) (\theta + \mu_i)^{\theta + y_i}} \end{aligned}$$

The upper and lower bounds for a spatial matrix generated using Morans indices ( $I$ ) were given by  $\lambda_{\max}(n/1^T W 1)$  and  $\lambda_{\min}(n/1^T W 1)$  where  $\lambda_{\max}$  and  $\lambda_{\min}$ , which were the extreme eigenvalues of  $\Omega = HWH$ . Hence, in this experiment, the eigenvectors of  $\Omega$  were vectors with unit norm maximizing Moran's  $I$ . The eigenvalues of this matrix were equal to Moran's  $I$  coefficients of spatial autocorrelation post-multiplied by a constant. Eigenvectors associated with high positive (or negative) eigenvalues have high positive (or negative) latent autocorrelation [12]. The eigen-orthogonalized eigenfunction eigenvectors associated with the rendered leukemia-related covariate eigenvalues revealed extremely small absolute values corresponding to low spatial autocorrelation for defining spatial structures in the empirical geosampled dataset. The diagonalization of the spatial weighting matrix generated from the covariate coefficients in R consisted of finding the normalized vectors  $u_i$ , stored as columns in the matrix  $U = [u_1 \cdots u_n]$  for satisfying:

$$\Omega = HWH = U\Lambda U^T = \sum_{i=1}^n \lambda_i u_i u_i^T$$

where  $\Lambda = \text{diag}(\lambda_1 \cdots \lambda_n)$ ,  $u_i^T u_i = \|u_i\|^2 = 1$  and  $u_i^T u_j = 0$  for  $i \neq j$  [10]. Note that double centering of  $\Omega$  implied that the eigenvectors  $u_i$  generated from the leukemia stratified covariates were centered and at least one eigenvalue was equal to zero. Introducing these eigenvectors in the original formulation of Moran's index led to:

$$I(x) = \frac{n}{1^T W 1} \frac{x^T H W H x}{x^T H x} = \frac{n}{1^T W 1} \frac{x^T U \Lambda U^T x}{x^T H x} = \frac{n}{1^T W 1} \frac{\sum_{i=1}^n \lambda_i x^T u_i u_i^T x}{x^T H x} \quad (3.3)$$

Considering the centered vector  $z = Hx$  and using the properties of idempotence of  $H$ , equation (3.3) was equivalent to:

$$I(x) = \frac{n}{1^T W 1} \frac{\sum_{i=1}^n \lambda_i z^T u_i u_i^T z}{z^T z} = \frac{n}{1^T W 1} \frac{\sum_{i=1}^n \lambda_i \|u_i^T z\|^2}{\|z\|^2} \quad (3.4)$$

As the eigenvectors  $u_i$  and the vector  $z$  were centered, equation (3.4) was rewritten:

$$I(x) = \frac{n}{1^T W 1} \frac{\sum_{i=1}^n \lambda_i \text{cor}^2(u_i, z) \text{var}(z)n}{\text{var}(z)n} = \frac{n}{1^T W 1} \sum_{i=1}^n \lambda_i \text{cor}^2(u_i, z) \quad [3.5]$$

In this experiment,  $r$  was the number of null eigenvalues of  $\Omega$  ( $r \geq 1$ ). These eigenvalues and corresponding eigenvectors were removed from  $\Lambda$  and  $U$  respectively. Equation 3.6 was then strictly equivalent to:

$$I(x) = \frac{n}{1^T W 1} \sum_{i=1}^{n-r} \lambda_i \text{cor}^2(u_i, z) \quad [3.7]$$

Moreover, it was demonstrated that Moran's index for a given eigenvector  $u_i$  was equal to  $I(u_i) = (n/1^T W 1) \lambda_i$  so the equation was rewritten:  $I(x) = \sum_{i=1}^{n-r} I(u_i) \text{cor}^2(u_i, z)$ . The term  $\text{cor}^2(u_i, z)$  represented the part of the variance of  $z$  that was explainable by  $u_i$  in the leukemia model  $z = \beta_i u_i + e_i$ . This quantity was equal to  $\frac{\beta_i^2}{n} \text{var}(z)$ . By definition, the eigenvectors  $u_i$  were orthogonal, and therefore, regression coefficients of the linearized models  $z = \beta_i u_i + e_i$  were those of the multiple regression model  $z = U\beta + \varepsilon = \beta_1 u_1 + \cdots + \beta_{n-r} u_{n-r} + \varepsilon$ . Subsequently, the distribution of the error residuals in the leukemia model in the autocovariance matrix was quantitated in ArcGIS

The maximum value of  $I$  was obtainable by all of the variations of  $z$ , as explained by the eigenvector  $u_1$ , which corresponded to the highest eigenvalue  $\lambda_1$  in the autocorrelation error matrix. In this experiment,  $cor^2(u_i, z) = 1$  (and  $cor^2(u_i, z) = 0$  for  $i \neq 1$ ), and the maximum value of  $I$ , was deducible for Equation (2.9), which was equal to  $I_{\max} = \lambda_1(n/1^T W 1)$ . The minimum value of  $I$  in the error matrix was parsimoniously obtainable as all the variations of  $z$  which was explained by the eigenvector  $u_{n-r}$  corresponding to the lowest eigenvalue  $\lambda_{n-r}$  generated in the leukemia model. This minimum value was equal to  $I_{\min} = \lambda_{n-r}(n/1^T W 1)$ . If the leukemia sampled explanatory predictor variable was not spatialized, the part of the variance explained by each eigen-orthogonal eigenvector was equal, on average, to  $cor^2(u_i, z) = 1/n-1$ . Because the sampled leukemia-related geosampled, socio-demographic and age-related variables in  $z$  were randomly permuted, it was assumed that we would obtain this result. In this experiment the set of  $n!$  random permutations revealed that

$$E_R(I) = \frac{n}{1^T W 1(n-1)} \sum_{i=1}^n \lambda_i = \frac{n}{1^T W 1(n-1)} \text{trace}(\Omega)$$

It was easily demonstrated that  $\text{trace}(\Omega) = -\frac{1^T W 1}{n}$  and it followed that  $E_R(I) = -\frac{1}{n-1}$  in the prognosticated model output.

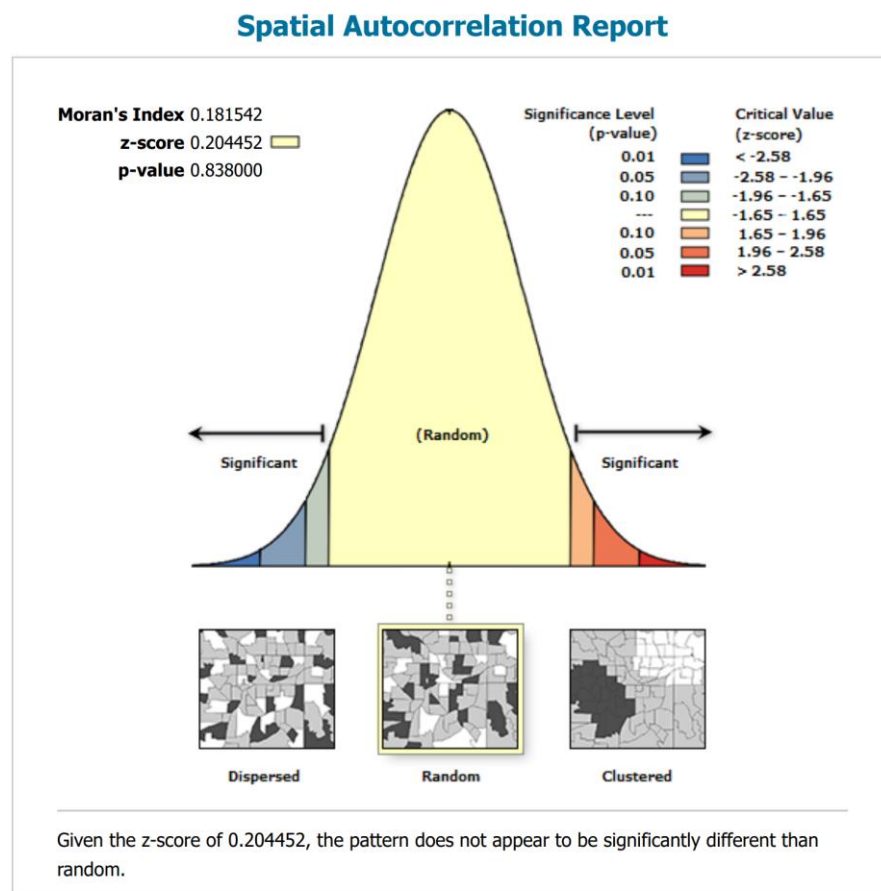


**Figure 2: Spatial Autocorrelation Report depicting moderate positive autocorrelation among leukemia cases**

The Poisson probability model revealed the most important sociodemographic leukemia-related covariate was being White and aged 65+.

The Moran's Index value for the spatial distribution of leukemia cases in Hillsborough County, Florida is 0.182. This indicates a moderate positive spatial autocorrelation among the leukemia cases at the zip code level.

Encountering multicollinearity (dependence of covariates) and heteroscedasticity (unequal variance) in this data presented significant hurdles. Violations of regression assumptions can generate overdispersed data due to outliers [15]. Multicollinearity in the leukemia zip code model was present because the independent variables in the regression model are closely correlated due to race, leading to inflated variances of estimated coefficients. Furthermore, heteroscedasticity emerged because the variance of error terms in the regression model varied across levels of the sociodemographic independent variables, causing erratic spreads of residuals around the regression line. Both of these issues violated the assumption of independence and variance equality, respectively.

The traditional frequentist approach is to make use of the null hypothesis significant test and employ p-values to reject the null hypothesis [25]. Recently, much research has proposed Bayesian versions of parametric and non-parametric frequentist 2 sample t-tests. These proposals include the student 2 sample t-test and its non-parametric counterpart – The main Whitney u test. In this paper, however, we explored using an eigen-Bayesian, semiparametric Markovian Gaussian model. This model allowed us to quantify heteroscedacity, multicollinearity, and zero autocorrelation (geographic chaos) in the output. An extensive similar study may be provided using a semi-parametric eigen-Bayesian Markovian test to control for type 1 and 2 errors. These results are important as balancing type 1 and 2 errors is a crucial goal in leukemia research. Henceforth, we propose shifting towards the eigen-Bayesian sample test to decrease the effects and type 1 and 2 error rates.

An eigen-spatial filter analyses selected eigenvectors as regressors, resulting in spatial autocorrelation being filtered out of the residuals of the leukemia-related data. The spatial filtering analyses transformed all the variables, containing spatial dependence, into covariates free of spatial dependence by partitioning the original georeferenced leukemia sociodemographic attribute variables, within a generalizable linear model framework, into two synthetic variates: (1) an eigen-spatial filter variate capturing latent spatial dependency, that otherwise would have remained in the response residuals, and (2) a nonspatial variate that was free of spatial dependence. The eigenfunction eigen-spatial filter derived from Moran's indices determined the mean, variance, and statistical distribution characterizations and descriptions of the sampled covariates. The spatial autocorrelation residual error analyses employed the estimates from the Monte Carlo simulation suggested slightly positive autocorrelation of the leukemia covariates based on race, age, and insurance. The spatially adjusted models identified the clustering patterns of the sampled leukemia data in the Poisson regressed datasets while accounting for conditional non-Gaussian uncertainty terms in the models. Autocorrelation indices can enable significance testing zip-code stratifiable, georeferenceable, leukemia-related sociodemographic estimator determinants, models using field and remote sampled explanatory variables which can be very useful for model improvement and resource allocation for implementing control strategies [e.g., building a social media platform on **Facebook** for

leukemia prevention messaging, especially in senior citizen homes, the uninsured [i.e., homeless shelter, food refugee camps, etc.] The Bayesian Poisson analyses revealed that the sampled covariate was positively associated with based on the covariate “being White” productivity in the study site.

In this experiment, we complemented the eigen-autocorrelation prognosticative, parameter estimation, empirical model using a Bayesian Poisson matrix in R for formal hypothesis testing. The sampled leukemia-related, sociodemographic parameter estimators were regressed in the Hillsborough County study site at the zip code level. R provides Bayesian analysis including Bayesian zero-inflated Poisson models for zero-inflated count data employing a Markov Chain Monte Carlo (MCMC) algorithm in downloadable, experimental versions of R ([www.R.org](http://www.R.org)).

Applied leukemia researchers interested in Bayesian statistics should be attracted to R or Python because of the ease with which one can code algorithms. These packages can sample from posterior distributions. Additionally, there are a significant number of packages contributed to the Comprehensive R Archive Network (CRAN) that provide tools for Bayesian inference. Markov Chain is a mathematical system that transitions from one state to another, between a finite or countable number of possible states, which can be characterized as a random “memoryless” process [26]. In recent years, MCMC has revolutionized the practicability of inferential Bayesian, allowing for a wide range of posterior distributions to be simulated and their parameters to be quantitated numerically in time series-dependent and non-dependent, leukemia vulnerability, zip-code level prognosticative modeling. In Bayesian statistics, the posterior probability of a random event, or an uncertain proposition in a time series or non-time series dependent leukemia zip code, signature prognosticative vulnerability model would be the conditional probability that is assigned after the relevant evidence [e.g., eigenized Poissonian sociodemographic parameterizable estimators]. Similarly, the posterior probability distribution or leukemia-specified time series covariates would be the zip code distribution of an unknown quantity, treated as a random variable, conditional on the evidence obtained from an experiment or survey.

In this experiment specifically, we employed a Bayesian Poissonian hierarchical model to estimate the risks of leukemia at the Hillsborough County study site. A Bayesian Poissonian vector autoregression model can characterize endogenous dynamic count data with no restrictions on the contemporaneous correlations [27]. Bayesian linear regression techniques may also be applied when the variance is assumed to be a function of the mean in these non-frequentist paradigms. It may be possible in some cases to amend the problem of propagated residual uncertainty in a time series, dependent, zip code stratifiable leukemia model by applying a transformation to the response variable (e.g., fitting the logarithm of the response variable (incidence of leukemia using a linear regression model) which could then imply that the response variable has a log-normal distribution rather than a normal distribution. In probability theory, a log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed [15]. Thus, if  $X$  is a random variable in a geo-spatiotemporal Bayesian generalizable hierarchical, zip code stratifiable prognosticative leukemia model with a normal distribution, then  $Y = \exp(X)$  will have a log-normal distribution; likewise, if  $Y$  is log-normally distributed in the model then  $X = \log(Y)$  has a normal distribution. The log-normal distribution would then be the geo-spatiotemporal leukemia model distribution of the sampled random variables with only positive real values.

Furthermore, the eigendecomposition of the residual forecast uncertainties in the leukemia-regressed covariates was illustratable in the Bayesian Poisson model. The output from the model was able to quantify the effects of exogenous covariate shocks. We then spatially decomposed uncertainty values to quantify the effects of exogenous-sampled explanatory leukemia-related covariate coefficients related to the autocorrelated variables. We assumed that by robustifying Bayesian Poisson model outputs, we would be able to quantify interaction terms between the zip code stratified sampled parameter estimators (i.e., sociodemographic prognosticators) and the leukemia-related zip code vulnerability indices at the Hillsborough County study site. It is well known that leukemia is unevenly distributed and, therefore, morbidity due to the disease can be perceived as problems of local rather than global importance [28]. Unfortunately, the reasons why leukemia may still be so prevalent are largely unknown. Therefore, rectifying estimating vulnerability indices, Poissonian probability estimates, second-order eigen-autocorrelation statistics, and Bayesian hierarchial prognosticative model determinants at a zip code level may have important implications for increased diagnosis, prioritization, and treatment of leukemia at the Hillsborough County study site and elsewhere.

The residuals in the georeferenced stratified clusters revealed three major uncertainty interactions: 1) race, (2) age, and (3) insurance. In this experiment, we assumed that quantitating latent, non-Gaussian (i.e. multicollinearity non-asymptoticalness, zero autocorrelation, heteroscedasticity) error probabilities in an empirical geosampled dataset of regressed, georeferenced, leukemia-related, diagnostic, stratified determinants could be non-heuristically optimizable, by employing a second-order, autoregressive, uncertainty-related, vulnerability-oriented, prognosticative model. Furthermore, a probabilistic, eigen-Bayesian, Poissonian, semi-parametric generalizable, estimation matrix would be usable as a post-treatment of the residual. In this experiment, the georeferenced, sub-county, district-level, diagnostically grid-stratifiable, leukemia-specifiable, geosampled characteristics were robustly and parsimoniously imposed employing the equations:

$$\mu_{\phi_i} = \frac{\sum_{j \neq i} \omega_{ij} \phi_{ij}}{\sum_{j \neq i} \omega_{ij}} \quad \text{and} \quad \sigma^2_{\phi_i} = \frac{1}{\lambda \sum_{j \neq i} \omega_{ij}} \quad (2.5)$$

This paper examined the behaviors of regression methods (the Poisson regression, ordinary least square [OLS], and the inferential Bayesian) employing potential leukemia-related capture points with inherent collinear structure. Datasets with reasonable degrees of multicollinearity at some selected sample sizes were simulated. The regression types were fitted to various, geosampled, zip code stratified, epidemiological data, capture points, and the performances of both the Poisson and OLS estimators were compared with that of the Bayesian model employing normal-Gamma conjugate prior. In probability theory and statistics, the normal-gamma distribution (or Gaussian-gamma distribution) is a bivariate four-parameter family of continuous probability distributions [15]. In this experiment, we used the conjugate prior of a normal distribution with unknown mean and precision. In Bayesian probability theory, if the posterior distribution  $p$  is in the same probability distribution family as the prior probability distribution  $p(\theta)$ , the prior and posterior are then called conjugate distributions, and the prior

is called a conjugate prior for the likelihood function  $p$ , [26]. The goal was to examine the relative efficiency of the aggregation/non-aggregation-oriented, non-zero, auto-correlatable, non-skew, non-heteroscedastic, non-multicollinear, asymptotical, eigen-spatial filter, eigen-estimators integrated with some prior information with the information available from the epidemiological, geosampled, georeferenced, sociodemographic, zip code-level, stratified, diagnostic determinants using varying regression semi-parametrizable, Bayesian, Poissonian, eigen-decomposable, eigen-orthogonal techniques.

Results from Monte Carlo studies were established. Discarding the first set of "burn-in" iterations, this experiment ensured that the chain had reached a steady state and had optimally estimated the Monte Carlo parameters (such as posterior means) from the georeferenced, diagnostic determinants. A posterior mean probability, in Bayesian Poissonian statistics, is the revised or updated probability of an event occurring after taking into consideration current information [26]. The posterior probability here was calculated by updating the prior probability using Bayes' theorem. After the model had converged, samples from the conditional distributions were employed to summarize the posterior distribution of the model. Distributions were approximated for scale parameter estimation in the Bayesian leukemia model.

An uninformative prior or diffuse prior expresses vague or general information about a variable [29]. We constructed conditionally conjugate priors for standard deviation parameters and subsequently considered non-informative and weakly informative priors in this family. An informative prior expresses specific, definite information about a variable [26].

Usage of a conditionally conjugate prior in our epidemiological, geo-spatiotemporal, hierarchical, diffusion-related, district-level, subcounty, leukemia stratified, hyper/hypo-endemic, transmission-oriented, vulnerability model allowed deriving and simulating the marginal posterior density  $\pi(\phi|y)$ . A conjugate prior is an algebraic convenience, giving a closed-form expression for the posterior; otherwise, numerical integration may be necessary [30]. Further, according to chapter 3 of Gelman's Data Bayesian Analysis [DBA], when we generated  $y_i \sim N(\mu, \sigma^2)$ ,  $y_i \sim N(\mu, \sigma^2)$ , and  $p(\mu, \sigma^2) \propto (\sigma^2)^{-1} p(\mu, \sigma^2) \propto (\sigma^2)^{-1}$ , subsequently  $p(\mu, \sigma^2|y) \propto \sigma^{-n-2} \exp(-1/2\sigma^2(n-1)s^2 + n(y - \mu)^2) p(\mu, \sigma^2|y) \propto \sigma^{-n-2} \exp(-1/2\sigma^2(n-1)s^2 + n(y - \mu)^2)$ .

We were interested in  $p(\mu|y) = \int p(\mu, \sigma^2|y) d\sigma^2 p(\mu|y) = \int p(\mu, \sigma^2|y) d\sigma^2$  as we were trying to rectify non-Gaussianism in the Bayesian leukemia model estimates employing the methodology that Gelman employs on page 66 of the third edition of DBA. We assumed that integral may be optimally weighed in the zip code stratified, leukemia-related estimated determinants employing the substitution;  $z = \frac{A}{2\sigma^2}$ , where  $A = (n-1)s^2 + n(\mu - \bar{y})^2$ . We recognized that the result in the leukemia zip code stratified, aggregation/non-aggregation-oriented, epidemiological, forecast-related, vulnerability model used a non-normalized gamma integral; hence we deduced

$$\begin{aligned}
 p(\mu|y) &\propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz \\
 &\propto [(n-1)s^2 + n(\mu - \bar{y})^2]^{-n/2} \\
 &\propto \left[ 1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2} \right]^{-n/2}.
 \end{aligned}$$

Our assumption was conjugate priors may give intuition by more transparently revealing how a likelihood function updates a prior distribution in an epidemiological, hierarchical diffusion-related, district-level, vulnerability, regression model for unbiasedly predictively targeting and prioritizing zip code stratified, hyper/hypoendemic, aggregation/non-aggregation-oriented, prognosticative, leukemia, geo-spatiotemporal, geosampled, epidemiological, exogenous estimators.

The stratified zip code models were compared employing the Deviance Information Criterion (DIC) in R MCMC where  $DIC = \bar{D} + p_D$ , was the sum of the posterior mean of the deviance, ( $\bar{D}$ ), a measure of goodness-of-fit, and the effective number of georeferenced, leukemia geosampled diagnostic parameters ( $p_D$ ). A measure of goodness-of-fit based on the DIC values was applied in R. The  $R^2DIC$  was calculated in line with the standard  $R^2$  measure for the iterated, explanatory residual forecasts (i.e., temporally targeted, zip code -level, hyper/hypoendemic, aggregation / non-aggregation-oriented hot/cold spots). These were optimally definable employing:  $R^2_{DIC} = 1 - \left( \frac{(DIC_k - \bar{D}_{best})}{(DIC_{max} - \bar{D}_{best})} \right)$  when  $DIC_k$  was the DIC value for sub-model k under evaluation and when  $DIC_{max}$  was the DIC value for one-fixed parameter model; and,  $\bar{D}_{best}$  was the posterior deviance as derived iteratively from the model residuals.

Model checking of all residual data input and compilation was conducted in R MCMC. The number of chains had to be specified before compilation, for constructing our vulnerability-oriented, Bayesian uncertainty model. Multiple parallel chains were run. Syntax checking was employed, which involved highlighting the entire model code and then choosing the sequence model specification. The uncertainty-related non-normal quantities in the estimates derived from the MCMC sequence of the random, epidemiological, zip code stratified, diagnostic samples were subsequently determined by  $N_k$  and  $v_k$ . These estimates also revealed a PDF (i.e., a statistical expression that defined a probability distribution and the likelihood) of the georeferenced aggregation / non-aggregation-oriented, hyper/hypoendemic, zip codes being an actual sociodemographic leukemia-related, hot/cold spot based on a regression outcome. Here every individual, exogenously geosampled, hierarchical, diffusion-related, explanatory variable was invasively examined (as opposed to quantitating a continuous random variable) using a scalar quantity  $v$ . The estimated value of  $v$  in the vulnerability-oriented, zip code, model was provided by the sample mean,  $\hat{v} = \frac{1}{N_k} \sum_{K=1}^{N_k} v_k$ .

We then addressed the problem of upper bounding the mean square error [MSE] of the MCMC estimators. Our analysis was asymptotic. We first established a general result valid for all ergodic Markov chains encountered in the Bayesian Poissonian computation and at multiple unbounded target functions. The bound was sharp in the sense that the leading term was

exactly  $\sigma^2(P, f)/n\sigma^2(P, f)/n$ , where  $\sigma^2$  was  $(P, f)$   $\sigma^2(P, f)$  which was the CLT asymptotic variance. In probability theory, the CLT establishes that, when independent random variables are summed up, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed [29].

Next, we specified additional assumptions and generated explicit computable bounds for geometrical and polynomial ergodic Markov chains under quantitative drift conditions. We generated quantitative bounds on the convergence rates of Markov chains under conditions implying polynomial convergence rates. This paper extends an earlier work by [31], which provides quantitative bounds for the total variation norm under conditions implying geometric ergodicity. Explicit bounds for the total variation norm were obtained for the zip code stratified, epidemiological, prognosticative, vulnerability model by evaluating the moments of an appropriately defined coupling time, employing a set of drift conditions, adapted from an earlier work by [32]. Applications of the model result were then presented to study the convergence of random walk Hastings Metropolis algorithm for generating super-exponential target functions and general state-space models. Like the MCMC, the Metropolis-Hastings algorithm is used to generate serially correlated draws from a sequence of probability distributions [26]. The sequence converges to a given target distribution. Explicit bounds for f-ergodicity were given for the zip code leukemia model for an appropriately defined control function f. As a corollary, we provided results on confidence estimation.

The expected variance  $\hat{v}$  was the expectation for the ensemble of the sequences as robustly parsimoniously rendered from the MCMC estimators, which in this experiment we expressed as:

$$\begin{aligned}\sigma_{\hat{v}}^2 &= \text{var}(\hat{v}) = E\{(\hat{v} - E\{\hat{v}\})^2\} \\ &= E\left\{\frac{1}{N_k} \sum_j (v_j - \bar{v}) \frac{1}{N_k} \sum_k (v_k - \bar{v})\right\} \\ &= \frac{1}{N_k^2} \sum_{jk} \{(v_j - \bar{v})(v_k - \bar{v})\}\end{aligned}$$

where  $\bar{v} = E\{v\} = E(\hat{v})$ . The autocovariance of the sequence was definable as:  $E\{(v_k - \bar{v})(v_{k+l} - \bar{v})\}$ . The asymptotical normalized, non-zero autocovariance was  $\rho(l) = (\sigma^2)^{-1} E\{(v_k - \bar{v})(v_{k+l} - \bar{v})\}$ , where  $\sigma^2$  was the variance of v and  $\rho(l)$  did not depend on k. The length of the non-zero, derived, autocovariance explanatory values was then optimally determinable by  $\sigma_{\hat{v}}^2 = \frac{1}{N_k^2} \sum_l \sum_{k=1}^{N_k-1} E\{(v_k - \bar{v})(v_{k+l} - \bar{v})\}$ .

Here the normalized autocovariance was a symmetric function, i.e.,  $\rho(-l) = \rho(l)$ . The sequence sufficiently converged to the target PDF. The variance of the distribution of the non-skew, homoscedastic, non-multi-collinear, asymptotically normalized, aggregation / non-aggregation-oriented, leukemia-related stratified, diagnostic estimators was generated employing  $\sigma_v^2 \approx S^2 = \frac{1}{N_k-1} \sum_{j=1}^{N_k} (v_k - \bar{v})^2$ . The normalized autocovariance was derivable from the sequence employing:  $\rho(l) \approx \frac{1}{S^2} \frac{1}{N_k-l-1} \sum_{k=1}^{N_k-l} (v_k - \hat{v})(v_{k+l} - \hat{v})$  for lag  $l \geq 0$ .

The Poisson probability model revealed the most important leukemia-related covariates were being White and 65+. The model did not violate the assumption that the variance was equal to the mean. Our model revealed that 33647 and the urban residential landscape were the most important stratifiable covariate for a georeferenced hotspot in 33647.

The second-order eigenfunction eigendecomposition revealed multiple non-zero autocorrelated clusters throughout various zip codes in Hillsborough County. The hot spots were in 33647, 33578, and 33511, and the cold spots were in 33621, 33503, and 33530.

As observed below in Table 1, the interaction terms within the Poisson model suggest that being African American and uninsured does not have a statistically significant relationship with developing leukemia. The interaction term has a z-value of 1.136 and a p-value of 0.256123, implying that this relationship is not statistically significant.

**Table 1: Statistical Analysis of Covariates: Beta Values, Errors, and Significance Levels.**

| Co-variate                        | Beta-value | Estimated Std. Error | z-value | p-value    |
|-----------------------------------|------------|----------------------|---------|------------|
| White                             | 7.184e-05  | 4.394e-08            | 1634.9  | <2e-16 *** |
| American Indian and Alaska Native | -2.277e-03 | 2.563e-06            | -888.5  | <2e-16 *** |
| Asian                             | 2.686e-05  | 6.153e-08            | 436.5   | <2e-16 *** |
| African-American                  | 5.161e-05  | 3.832e-08            | 1346.8  | <2e-16 *** |
| Hispanic/Latino                   | 2.921e-05  | 3.830e-08            | 762.7   | <2e-16 *** |
| Pacific Islander                  | -2.351e-03 | 1.425e-05            | -165.0  | <2e-16 *** |
| Above 65                          | -3.168e-05 | 9.388e-08            | -337.5  | <2e-16 *** |
| Uninsured                         | 1.626e-04  | 1.640e-07            | 991.3   | <2e-16 *** |

---Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Table 2: Poisson Model Covariates: Without Negative Binomial.**

|                                   |          |
|-----------------------------------|----------|
| White                             | 5.756875 |
| American Indian and Alaska Native | 2.898379 |
| Asian                             | 2.308341 |
| African-American                  | 2.504663 |
| Hispanic/Latino                   | 3.664907 |
| Pacific Islander                  | 3.587691 |
| Above 65                          | 2.655138 |
| Uninsured                         | 8.531651 |

Poisson model table: VIF AIC: 4596172

**Table 3: Poisson Model Covariates: Applying Negative Binomial.**

|                                   |          |
|-----------------------------------|----------|
| White                             | 5.752834 |
| American Indian and Alaska Native | 2.897992 |
| Asian                             | 2.306731 |
| African-American                  | 2.504733 |
| Hispanic/Latino                   | 3.665841 |
| Pacific Islander                  | 3.587018 |
| Above 65                          | 2.654414 |
| Uninsured                         | 8.532876 |

Negative binomial table: VIF: AIC: 4606902

As observed in Table 2 and Table 3, negative binomial regression could not improve the VIF results due to the absence of overdispersion in the model. Therefore, Poisson regression with a logarithm link would be appropriate for the full model.

## DISCUSSION

Currently, the treatment of leukemia in Hillsborough County at the zip code level is not homogenous. Early treatment protocols can be very expensive, especially for those who do not have medical insurance [33]. If there is no insurance coverage, the average estimated cost for a complete blood count test typically ranges from \$140 to \$622 but may vary [34]. Based on the analysis and this experiment, zip code 33647 was the most prominent in terms of predictive rates for leukemia among Whites and Asians above 65. Based on the landscape classification, urban residential was the most prominent in zip code 33647. The average household price in 33647 in urban residential locations is \$487,749 in mean value [35, 36]. Hence, it may be correlated that the price of \$487,749 could be the house value of a patient suffering from leukemia. Furthermore, the prolonged search for an unrelated donor, with a median time of 2 to 3 months, along with a median time from diagnosis to HSCT of 12.9 months in the US, underscores the challenges in accessing appropriate matches for transplantation [37].

There are approximately 19 hospitals that offer ionizing radiation in Hillsborough County, Florida. The total cost of leukemia may include initial care, continuing care, and end-of-life care, depending on the course of the patient's leukemia. It can incur a cost of \$47,264 for initial care, \$12,701 for continuing care, and \$169,588 for end-of-life care [38]. As such, this would prohibit people who are homeless, low-income, or uninsured from obtaining optimal treatment for leukemia. However, there exists a paradox, as there is a level of exposure risk from oncological treatment of leukemia. At certain times, there is a threat of dysplasia due to exposure to rad ions during oncological therapy [39], which may suggest that a non-insured homeless leukemia victim above 65 may have a better outcome than one who is being treated for leukemia.

Unfortunately, there is some disagreement in the literature regarding leukemia treatment. In 2010, [37] claimed that even though there was not a link between symptom management and the radiation therapy dosage, individuals who received more than 25 Gy (or BED10 30 Gy) exhibited a higher overall response rate. However, according to [3], there was an elevated risk of leukemia-related death even with minimal accumulated radiation doses (<5 mGy). Since there are controversial literature contributions concerning treatment for leukemia, we recommend early diagnostics for males and females above 65. We recommend using social media platforms to alert potentially vulnerable populations (Urban residential homesteads above \$479,749 in Hillsborough County). Creating online messaging for pre-leukemia treatment would enable patients to forego exposure to rad ions, hence lowering the probability of acquiring leukemia.

The sociodemographic Poissonian prognosticative model output revealed that the most significant covariate was being White. The non-dispersed model did not violate the assumption that the variance was equal to the mean. There was no need for a negative binomial regression to employ a non-homogenous, gamma-distributed mean to quantitate the outliers in the regression model, which is common in epidemiological, empirical, count variable models. A Poissonian model can be employed to determine parameter estimator hierarchy amongst leukemia and non-time series-dependent sociodemographic covariates.

In this experiment, we utilized an Eigen-spatial filtering approach, to generate a specific subset of eigenvectors derived from a non-linear modified matrix in Python to capture the aggregation/non-aggregation-oriented, spatial uncertainties amongst a leukemia specified, sampled estimator dataset. However, we assumed that the optimal subset in the proposed eigen-spatial filter, eigenvector, diagnostic paradigm would be identifiable more intuitively by an objective function that minimizes residual, erroneous non-zero autocorrelation rather than maximize a diagnostic model fit. The proposed objective function we assumed would have the advantage of optimally quantitating deviant estimator variable inclinations [e.g., non-Gaussian, residual zero autocorrelation] for optimizing space-time series [i.e., inducing non-multicollinearity, and/or non-skew, non-heteroscedasticity] in vulnerability-oriented, epidemiologic georeferenceable zip code stratified forecasts [e.g., clustering, sub-county, geolocation of a non-insured epicenter i.e., homeless shelter, 'hot/cold spot' ground centroid coordinates]. We assumed employing smaller subsets of frugally selected, eigenfunction, eigen-spatial filters derived from a geographically weighted regression would reveal multicollinearity, skew heteroscedasticity non-asymptoticalness, and other non-Gaussianism due to violations of regression assumptions in the leukemia-related, diagnostic, sociodemographic, stratified, estimator determinants in Bayesian Poissonian eigenvector eigen-geospace.

To verify the accuracy of spatial eigenvector mapping of the sociodemographic zip code stratified, georeferenced, leukemia-related covariates, model "checking" procedures are advocated. For example, in this research, an eigenfunction eigen-decomposition eigen-algorithm yielded eigen-orthogonal eigenvectors describing positive spatial autocorrelation, but only after statistical restrictions were applied (e.g.,  $MC > 0.25$ ), reducing the candidate set to a more manageable number for describing a county-level leukemia estimator determinant model. Restricting the set of eigenvectors in a leukemia model over which a stepwise selection search is performed is sensible for several reasons. First, if a homogeneous process underlies a set of georeferenced leukemia-related epidemiological data capture points, then the nature of spatial autocorrelation components should not vary at the zip code level. A Moran scatterplot may help assess the feasibility of this characterization. Second, if a parsimonious set of eigenvectors is to be selected for leukemia mapping, then eigenvectors depicting near-zero spatial autocorrelation should be avoided in a zip code stratified, georeferenced prognosticative model since they fail to capture any geographic information. Third, as  $n$  increases, the number of eigenfunctions increases, and hence, the numerical intensity involved in executing a regression also increases in a leukemia zip code model. This issue is of paramount importance when dealing with remotely sensed data where the number of eigenfunctions is in the hundreds, thousands, millions, or billions. Finally, selection experience with positive spatial autocorrelation analyses in a leukemia model suggests that virtually all prominent eigenvectors are contained in the set of eigenfunctions whose eigenvalues constitute the top quartile.

The outcome of our inferential Bayesian Poissonian risk analyses was a systematic segment partitioned into two sub-segments, namely, the Spatially Structured Random Effect (SSRE), which in the epidemiological, vulnerability-oriented, leukemia-related, zip code stratified model was related to the autocorrelation models that represented contagion and hierarchical diffusion mechanisms, and the spatially unstructured RE (i.e. SURE). Our SURE output was geographically random. The SURE furnished clues about aspatial, omitted, potentially

significant, epidemiological, diagnostic, indicator variables for controlling for optimizing, delineating higher rates (e.g., zip code stratified regions with higher urban residential show a higher probability of leukemia-related sociodemographic covariates). The SURE term in our model constituted an explanatory response variable in a linear regression, with substantive attribute, homoscedastic, normalizeable, Gaussian, aggregation-oriented, diagnostic determinants (i.e., independent variables), which precisely revealed the geolocation of the spilled over, hierarchical, diffusion-related, determinant, attribute features (e.g., stratifiable, leukemia-related, epidemiological hot spots) in Hillsborough County.

A conventional SURE term in the epidemiological, zip code stratified, empirical vulnerability, forecast-oriented model represented not only aspatial, hierarchical, diffusion-related co-factors that were in play in Hillsborough County but also connected geographic influences, such as sociodemographic linkages based on Euclidean distance measurements in an urban residual capture point scaled up to a potential zip code level hot spot. The inferential hierarchical Bayesian paradigm achieved the SSRE-SURE partitioning by including repeated, epidemiological, georeferenced measures, which for the geosampled, diagnostic, leukemia data were characterizable in the second order, eigen-autocorrelation, eigenvector, eigen-Bayesian model in R. The daily counts of cases for a georeferenced, stratified, potential, hyper/hypoendemic, aggregation-oriented geolocation, coupled with additional geosampled, residual, asymptotically normalized, non-heteroscedastic, non-multicollinear, unbiased, hierarchical, diffusion-related, information identified multiple hot spots in varying geolocations, (e.g. extensive, seasonal, climatic flooding conditions causing unnecessary crowding in internally displaced migrant farmer refugee camps).

A key ingredient of our Bayesian approach for determining non-normality in the epidemiological estimator feature attributes in the geosampled estimator determinants was the choice of the prior distribution. We derived two versions of Jeffreys prior, the Jeffreys rule prior and the Independence Jeffreys prior, which has not yet been developed in the literature for non-normal, diagnostic, uncertainty-oriented, semi-parametric, non-asymptotical models. These priors may be employed for generating an inferential Bayesian analysis of an aggregation-oriented, hierarchical, diffusion-related, epidemiological, county-level, diagnostic oncological model when prior information is completely unavailable. Moreover, we propose an informative and weakly informative prior for the autocorrelation parameter based on an extensive literature review of empirical applications of the autocorrelation model and our experimental findings. Finally, we provide efficient MCMC algorithms to sample from the resulting posterior distributions from iteratively simulated, aggregation-oriented, hierarchical, diffusion-related, geo-spatiotemporally, spilled over, subcounty, county-level, semi-parameterized, hyper/hypo-endemic, zip code stratified, leukemia-related, diagnostic determinants in Bayesian Poissonian eigenvector eigen-geospace. Simulation results here suggest that the model may outperform the ML model concerning bias and frequentist coverage of credible and confidence intervals. Frequentist methods do not demand the construction of a prior and depend on the uncertainty-oriented potential probabilities of observed and unobserved data [29].

## CONCLUSION

The outlined linear and non-linear prognosticative modeling strategies offer a strong basis for decreasing leukemia rates in Hillsborough County. Detecting leukemia early through regular

blood tests is crucial, as it frequently results in better treatment outcomes. Tailored outreach and educational initiatives, such as leveraging social media platforms to distribute information about risk factors and the significance of early detection, can effectively reach vulnerable populations like seniors in retirement homes and uninsured individuals utilizing homeless shelters or food banks. Strategic resource allocation, guided by risk assessment models pinpointing high-risk regions, improves the effectiveness of intervention programs. The interplay between these strategies is clear: heightened awareness from the social media campaign prompts more people to undergo blood tests, potentially leading to swifter diagnoses, while risk assessment models inform targeted communication and outreach efforts. Collaboration with community groups and ensuring affordable healthcare access are vital factors for success. Overall, this multifaceted modeling approach holds considerable promise for reducing leukemia prevalence in Hillsborough County. In the future, leukemia data for the state of Florida might be incorporated into a model created with Python signatures. Through the use of improved sampling strategies and pertinent healthcare data analysis, this model would seek to identify relationships between leukemia incidence, demographic characteristics, and healthcare access. We were able to determine that being White and above 65 were the most important covariates in the hotspot in zip code 33647. An over-dispersed Poissonian leukemia regression model can be treated due to outliers using a negative binomial with a non-homogenous distributed mean. An eigenfunction, eigendecomposition spatial filter algorithm can discern potential hyper/hypoendemic, leukemia-stratified clusters. A Bayesian hierarchical model analysis can determine causation covariates in a non-frequentistic leukemia model.

## References

- [1] Centers for Disease Control and Prevention. (2022, September 13). *Hematologic cancer incidence, survival, and prevalence*. Centers for Disease Control and Prevention. <https://www.cdc.gov/cancer/uscs/about/data-briefs/no30-hematologic-incidence-surv-prev.htm>
- [2] Leukemia - *Cancer Stat Facts*. SEER. (n.d.). <https://seer.cancer.gov/statfacts/html/leuks.html>
- [3] *Leukemia*. Hematology.org. (n.d.). <https://www.hematology.org/education/patients/blood-cancers/leukemia#:~:text=Leukemia%20is%20a%20type%20of,red%20blood%20cells%20and%20platelets>
- [4] The Biology of Leukemia: The Cancer of the Blood Author: Emily Jenkins
- [5] Moen MD, McKeage K, Plosker GL, Siddiqui MA. Imatinib: a review of its use in chronic myeloid leukaemia. *Drugs*. 2007;67(2):299-320
- [6] Davis AS, Viera AJ, Mead MD. Leukemia: an overview for primary care. *Am Fam Physician*. 2014 May 1;89(9):731-8. PMID: 24784336.
- [7] *Exposure to ionizing radiation and leukemia risk: Is there a safe threshold?* Clinical News | American Society of Hematology. (2021, December 30). <https://ashpublications.org/ashclinicalnews/news/2301/Exposure-to-Ionizing-Radiation-and-Leukemia-Risk>
- [8] Leuraud, K., Richardson, D. B., Cardis, E., Daniels, R. D., Gillies, M., O'Hagan, J. A., Hamra, G. B., Haylock, R., Laurier, D., Moissonnier, M., Schubauer-Berigan, M. K., Thierry-Chef, I., & Kesminiene, A. (2015). Ionising radiation and risk of death from leukaemia and lymphoma in radiation-monitored workers (INWORKS): An international cohort study. *The Lancet Haematology*, 2(7). [https://doi.org/10.1016/s2352-3026\(15\)00094-0](https://doi.org/10.1016/s2352-3026(15)00094-0)

- [9] Little, M. P., Wakeford, R., Borrego, D., French, B., Zablotska, L. B., Adams, M. J., Allodji, R., de Vathaire, F., Lee, C., Brenner, A. V., Miller, J. S., Campbell, D., Pearce, M. S., Doody, M. M., Holmberg, E., Lundell, M., Sadetzki, S., Linet, M. S., & Berrington de González, A. (2018). Leukaemia and myeloid malignancy among people exposed to low doses (<100 msv)
- [10] Jacob, B. G., Izureta, R., Bell, J., Parikh, J., Loum, D., Casanova, J., Gates, T., Murray, K., White, L., & Aceng, J. R. (2023). Approximating Non-Asymptoticalness, Skew Heteroscedascity and Geo-spatiotemporal Multicollinearity in Posterior Probabilities in Bayesian Eigenvector Eigen-Geospace for Optimizing Hierarchical Diffusion-Oriented COVID-19 Random Effect Specifications Geosampled in Uganda. *American Journal of Mathematics and Statistics* 2023, 13(1), 1–43. <https://doi.org/10.5923/j.ajms.20231301.01>
- [11] Griffith, D. A., Chun, Y., & Li, B. (2019). Spatial regression analysis using eigenvector spatial filtering. Academic Press.
- [12] Griffith DA (2003). Spatial autocorrelation and spatial filtering gaining understanding through theory and scientific visualization first edition 2003. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-24806-4>
- [13] Wang, Q., Dong, W., Yang, K., Ren, Z., Huang, D., Zhang, P., & Wang, J. (2021). Temporal and spatial analysis of covid-19 transmission in China and its influencing factors. *International Journal of Infectious Diseases*, 105, 675–685. <https://doi.org/10.1016/j.ijid.2021.03.014>
- [14] Kaiser, M. S., & Cressie, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statistics & Probability Letters*, 35(4), 423–432. [https://doi.org/10.1016/s0167-7152\(97\)00041-2](https://doi.org/10.1016/s0167-7152(97)00041-2)
- [15] Hosmer, D. W., & Lemeshow, S. (2004). *Applied logistic regression*. Wiley.
- [16] Griffith, D. A. (2005). Effective Geographic Sample Size in the Presence of Spatial Autocorrelation. *Annals of the Association of American Geographers*, 95(4), 740–760. <https://doi.org/10.1111/j.1467-8306.2005.00484.x>
- [17] Bureau, U. C. (2023, September 21). *2020 census results*. Census.gov. <https://www.census.gov/programs-surveys/decennial-census/decade/2020/2020-census-results.html>
- [18] USF Water Institute, S. of G. (n.d.). *Welcome to the Hillsborough County Water Atlas*. Welcome - Hillsborough.WaterAtlas.org. <https://hillsborough.wateratlas.usf.edu/>
- [19] *Hillsborough County Florida ZIP codes - map and full list*. Zipdatamaps.com. (n.d.). <https://www.zipdatamaps.com/en/us/zip-maps/fl/county/borders/hillsborough-county-zip-code-map>
- [20] Kasim, M. F., Galligan, T. P., Topp-Mugglestone, J., Gregori, G., & Vinko, S. M. (2019). Inverse problem instabilities in large-scale modeling of matter in extreme conditions. *Physics of Plasmas*, 26(11), 112706. <https://doi.org/10.1063/1.5125979>
- [21] Robert, C. P., & Casella, G. (2004). Monte Carlo Statistical Methods. *Springer Texts in Statistics*. <https://doi.org/10.1007/978-1-4757-4145-2>
- [22] Geyer, C. J., & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3), 657–683. <https://doi.org/10.1111/j.2517-6161.1992.tb01443.x>
- [23] Jacob, B.G. and hastik, R.J. (2014) Integrating a Trimble Recon X 400 MHz Intel PXA255 Xscale CPU® Mobile Field Data Collection System Using Differentially Corrected Global Positioning System Technology and a Real-Time Bidirectional Actionable Platform within an ArcGIS Cyberenvironment for Implementing Mosquito Control. *Advances in Remote Sensing*, 3, 141-196. <http://dx.doi.org/10.4236/ars.2014.33012>

- 
- [24] Hastie, T., & Tibshirani, R. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics*, 46(4), 1005. <https://doi.org/10.2307/2532444>
- [25] Freedman, D. A. (2008). On regression adjustments to experimental data. *Advances in Applied Mathematics*, 40(2), 180–193. <https://doi.org/10.1016/j.aam.2006.12.003>
- [26] Gelman, A. (2005). Analysis of variance—why it is more important than ever. *The Annals of Statistics*, 33(1). <https://doi.org/10.1214/009053604000001048>
- [27] Jacob Benjamin G., Fiorella Krapp, Mario Ponce, Nanhua Zhang, Semiha Caliskan, Daniel A. Griffith, Eduardo Gotuzzo and Robert J. Novak (2013), A Bayesian Poisson specification with a conditionally autoregressive prior and a residual Moran's coefficient minimization criterion for quantitating leptokurtic distributions in regression-based multi-drug resistant tuberculosis treatment protocols, *Journal of Public Health and Epidemiology*. 5(3): 122-143.
- [28] Torres-Flores, J., Espinoza-Zamora, R., Garcia-Mendez, J., Cervera-Ceballos, E., Sosa-Espinoza, A., & Zapata-Canto, N. (2020). Treatment-related mortality from infectious complications in an acute leukemia clinic. *Journal of Hematology*, 9(4), 123–131. <https://doi.org/10.14740/jh751>
- [29] Cressie, N. A. (1993). *Statistics for Spatial Data*. *Wiley Series in Probability and Statistics*. <https://doi.org/10.1002/9781119115151>
- [30] Gelman, A. (2014). The connection between varying treatment effects and the crisis of Unreplicable Research. *Journal of Management*, 41(2), 632–643. <https://doi.org/10.1177/0149206314525208>
- [31] Roberts, G. O., & Tweedie, R. L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Processes and Their Applications*, 80(2), 211–229. [https://doi.org/10.1016/s0304-4149\(98\)00085-4](https://doi.org/10.1016/s0304-4149(98)00085-4)
- [32] Tuominen, P., & Tweedie, R. L. (1994). Subgeometric rates of convergence of  $f$ -ergodic markov chains. *Advances in Applied Probability*, 26(3), 775–798. <https://doi.org/10.2307/1427820>
- [33] Health, W. by Enhance. (nanak, January 3). *How much does blood work cost without insurance?* Enhance Health. <https://enhancehealth.com/how-much-does-blood-work-cost-without-insurance/>
- [34] Silamlak Birhanu Abegaz, "Human ABO Blood Groups and Their Associations with Different Diseases", *BioMed Research International*, vol. 2021, Article ID 6629060, 9 pages, 2021. <https://doi.org/10.1155/2021/6629060>
- [35] 33647 home values. Zillow. (n.d.). <https://www.zillow.com/home-values/72737/tampa-fl-33647/>
- [36] Redfin. (n.d.). 33647 housing market: House Prices & Trends. <https://www.redfin.com/zipcode/33647/housing-market>
- [37] Bartenstein M, Deeg HJ. Hematopoietic stem cell transplantation for MDS. *Hematol Oncol Clin North Am*. 2010 Apr;24(2):407-22. doi: 10.1016/j.hoc.2010.02.003
- [38] City of Hope. (2023, July 17). *Cancer treatment cost: What to expect & how to manage*. City of Hope. <https://www.cancercenter.com/community/blog/2023/07/managing-cancer-treatment-cost>
- [39] Braunstein, S., & Nakamura, J. L. (2013). Radiotherapy-induced malignancies: Review of clinical features, pathobiology, and evolving approaches for mitigating risk. *Frontiers in Oncology*, 3. <https://doi.org/10.3389/fonc.2013.00073>
-