Predictive Mapping of Mycobacterium Tuberculosis at the County Level in the State of Florida

Ali Moradi

University of South Florida, amoradi@health.usf.edu

Follow this and additional works at: http://scholarcommons.usf.edu/etd

Part of the Public Health Education and Promotion Commons

Scholar Commons Citation
http://scholarcommons.usf.edu/etd/6544

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Predictive Mapping of Mycobacterium Tuberculosis at the County Level in the State of Florida

by

Ali Moradi

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Public Health Department of Global Health College of Public Health University of South Florida

Co-Major Professor: Ricardo Izurieta, M.D., Ph.D., DTMH
Co-Major Professor: Benjamin Jacob Ph.D.
Boo H. Kwa Ph.D.
Ismael Hoare Ph.D.
Teresa Hunter-Pettersen M.D., MPA, MS

Date of Approval:
November 4, 2016

Keywords: public health, mixed binomial, spatial autocorrelation, random effects

Copyright © 2016, Ali Moradi
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>iii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>iv</td>
</tr>
<tr>
<td>Abstract</td>
<td>v</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Aim of Study</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>1</td>
</tr>
<tr>
<td>Diagnosis, Model of Treatment, and Current Control</td>
<td>2</td>
</tr>
<tr>
<td>Predicting Encounters</td>
<td>4</td>
</tr>
<tr>
<td>Income as a Risk Factor</td>
<td>4</td>
</tr>
<tr>
<td>Co-infections as a Risk Factor</td>
<td>6</td>
</tr>
<tr>
<td>HIV</td>
<td>6</td>
</tr>
<tr>
<td>Diabetes</td>
<td>7</td>
</tr>
<tr>
<td>Social Behaviors as Risk Factors</td>
<td>9</td>
</tr>
<tr>
<td>Tobacco smoke</td>
<td>9</td>
</tr>
<tr>
<td>Alcohol consumption</td>
<td>10</td>
</tr>
<tr>
<td>The PRECEDE-PROCEED Planning Model</td>
<td>13</td>
</tr>
<tr>
<td>Phases of the PRECEDE-PROCEED Model</td>
<td>13</td>
</tr>
<tr>
<td>Validation and Future Applications of the PPM</td>
<td>21</td>
</tr>
<tr>
<td>Analysis of Mycobacterium Tuberculosis, Using GIS</td>
<td>22</td>
</tr>
<tr>
<td>The Effect of Elevation: Respiration Rate and UV Irradiance</td>
<td>23</td>
</tr>
<tr>
<td>Reassessing the Directions of This Study</td>
<td>26</td>
</tr>
<tr>
<td>An Examination of UV Light Irradiance</td>
<td>26</td>
</tr>
<tr>
<td>The Use of Geographic Information Systems</td>
<td>31</td>
</tr>
<tr>
<td>Chapter 2: Literature Review</td>
<td>41</td>
</tr>
<tr>
<td>Chapter 3: Results</td>
<td>60</td>
</tr>
<tr>
<td>Chapter 4: Selection of Strategies for Intervention Based on a GIS Prioritization</td>
<td>122</td>
</tr>
<tr>
<td>Understanding Tuberculosis Transmission and its Endemicity</td>
<td>122</td>
</tr>
<tr>
<td>Global Vision</td>
<td>123</td>
</tr>
<tr>
<td>Types of Tuberculosis</td>
<td>124</td>
</tr>
<tr>
<td>Pulmonary Tuberculosis</td>
<td>124</td>
</tr>
<tr>
<td>Extra-pulmonary Tuberculosis</td>
<td>125</td>
</tr>
<tr>
<td>Transmission Parameters</td>
<td>127</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Florida Tuberculosis Case Report</td>
<td>128</td>
</tr>
<tr>
<td>State of Florida Tuberculosis Program Act</td>
<td>130</td>
</tr>
<tr>
<td>Adopted Tuberculosis Treatment Guidelines</td>
<td>132</td>
</tr>
<tr>
<td>Common Obstacles</td>
<td>133</td>
</tr>
<tr>
<td>Prevention Policy and Screening Procedures</td>
<td>134</td>
</tr>
<tr>
<td>Screening</td>
<td>137</td>
</tr>
<tr>
<td>Screening Intervention</td>
<td>138</td>
</tr>
<tr>
<td>Intervention Strategies to Secure Healthy Outcomes for Screened Subjects</td>
<td>143</td>
</tr>
<tr>
<td>Objective of screening</td>
<td>144</td>
</tr>
<tr>
<td>GIS Technology Intervention</td>
<td>144</td>
</tr>
<tr>
<td>Intervention Structure</td>
<td>146</td>
</tr>
<tr>
<td>Level of clinical services intervention based on GIS</td>
<td>146</td>
</tr>
<tr>
<td>Community-field services program based on GIS</td>
<td>147</td>
</tr>
<tr>
<td>Community surveillance services based on GIS</td>
<td>147</td>
</tr>
<tr>
<td>Chapter 5: Conclusion</td>
<td>149</td>
</tr>
<tr>
<td>Implications for Public Health Research</td>
<td>153</td>
</tr>
<tr>
<td>Limitations</td>
<td>154</td>
</tr>
<tr>
<td>References</td>
<td>156</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: The time series regressed endemic TB regression-based model parameter estimate ..........................................................63

Table 2: Summary of backward elimination ..................................................73

Table 3: Improvement of fit of the PROC MCMC, hierarchical Bayesian model (HBM) endemic, county-level TB model .................................................118

Table 4: Counties with most predictive value .............................................121

Table 5: Category of pulmonary and extra-pulmonary tuberculosis ..................125

Table 6: Homeless. Information in this table is adapted from: Florida Department of Health Tuberculosis Report (9) ........................................131

Table 7: Alcohol-related. Information in this table is adapted from: Florida Department of Health Tuberculosis Report (9) ........................................131

Table 8: Non-injected drug use. Information in this table is adapted from: Florida Department of Health Tuberculosis Report (9) ..........................132

Table 9: Injected drug use. Information in this table is adapted from: Florida Department of Health Tuberculosis Report (9) ..........................132

Table 10: Counties with high potential for hyper-endemicity without intervention ....134

Table 11: Florida tuberculosis health outcomes and targeted goals for 2020. Florida System of Tuberculosis Care - Florida Department of Health, March 1, 2013 ....145
LIST OF FIGURES

Figure 1: TB incidence rate per 100,000 GNI per Capita ........................................5
Figure 2: Risk of TB recurrence by smoking status ..............................................10
Figure 3: Alcohol-related DALYS by disease category .......................................12
Figure 4: Map of Florida with Hillsborough County ........................................34
Figure 5: Digital elevation of central and southern Florida ..............................35
Figure 6: Digital elevation of Hillsborough County .........................................35
Figure 7: Three-Dimensional zed models of Florida, depicting elevation in meters ....36
Figure 8: An example view of Hillsborough County as seen in Google Earth .......37
Figure 9: An image of Hillsborough County obtained from the LANDSAT 8 satellite ....39
Figure 10: Top left (a): random effects term ................................................99
Figure 11: Fit diagnostic for cases ......................................................................100
Figure 12: Top left (a): normal quantile plot for the random effects term ...........119
Figure 13: Predicted probabilities of tuberculosis in Florida by county ...............120
Figure 14: Forecasts of rank-ordered largest expected numbers of tuberculosis cases based upon projected county population .........................................................133
Figure 15: A TB screening test ...........................................................................139
ABSTRACT

Introduction: One of the major barriers to developing an accurate tuberculosis (TB) surveillance program for Florida is the design and implementation of a sampling system that will adequately monitor and predict varying sizes and characteristics of county-level vulnerable endemic sub-populations and their explanatory covariates (e.g., living or working in a residential care facility). The aim of this research study is to envision an endemic, tuberculosis-related web-based interface for use by public health officials in the State of Florida which includes generating essential information such as a real-time syndrome-based reporting to regulate automated and immediate 'Alerts' to public health officials, doctors, hospitals and local community in ArcGIS. This study demonstrates the capability of an autocorrelation, time series, epidemiological, interpolative, and vulnerable predictive ArcGIS model to target tuberculosis at the county-level in the state of Florida.

Methodology: The data for constructing an autocorrelation, probabilistic paradigm was acquired from the Centers for Disease Control and Prevention [CDC] in Atlanta, Georgia. The full dataset contained two points in time, allowing estimation of a mixed binomial model that aided in predicting the probability of tuberculosis by county. The random effects term in the ArcGIS model was comprised of spatially structured and stochastic effects (i.e., spatially unstructured) terms. These terms substituted for covariates in the model. The assumption was that random effects term in the endemic, TB-related, explanatory, county-level, risk model had a frequency distribution that was bell-shaped (i.e., normally/Gaussian distributed) with a mean of zero.
Results: The results indicated the empirical estimate had a mean of 0.0197 with a Shapiro-Wilk normality probability of 0.0027. The mean in the model was not exactly zero, although the forecasts indicated 0.06, which was not significantly different from zero. It was noted that the frequency distribution deviated from a bell-shaped curve. This random effects term accounted for roughly 41% of the variability in the observed probability of TB by county and yielded an under dispersed binomial model. An eigenvector spatial filter description of the random effects term involved 5 of 18 total eigenvectors, which portrayed noticeable positive spatial autocorrelation. The decomposition algorithm also revealed 4 of 25 eigenvectors portraying noticeable negative spatial autocorrelation. These two spatial filter components accounted for, respectively, roughly 16% and 10% of the variability in the probability of TB by county. The spatially unstructured random effects component accounted for roughly 15% of this variability. The final model revealed that from 2015 to 2020, Duval, Orange, and Broward counties would require immediate intervention in order to prevent TB transmission. The model also revealed that from 2025 to 2040 Hillsborough and Palm Beach counties could become hyper-endemic without implementation of control strategies.

Conclusion: An endemic, TB-related, ArcGIS, autocorrelation eigenanalyses forecast, paradigm may be employed by public health officials in Florida to target, vulnerable, county-level populations. A Precede-Proceed model-based reporting mechanism may help disseminate the ArcGIS model results and help regulate automated and immediate 'Alerts' to public health officials, doctors, hospitals and local community at the county-level. An ArcGIS, web-based, epidemiological tool for data entry and communication can also allow real-time, predictive, real-time mapping of any TB county outbreaks Precede–Procede model may be employed by county-level public health officials in Florida to disseminate and prioritize county-
level, TB model, epidemiological, information to their constituents. In so doing, factors regulating outbreaks of county-level TB may be accurately identified.
CHAPTER 1
INTRODUCTION

Aim of Study

The intended goal for this project is to implement a robust demonstration by employing a technology project for forecasting hyper-endemicity of tuberculosis at county level in the state of Florida, which can be replicable in other states with higher incidence of tuberculosis than the national average. This project provides capacity to demonstrate means and accuracy in utilization of mapping technology in combination with geographical information system to produce a competent model for forecasting infectious disease endemicity. Therefore, these projects provide greater accuracy than currently existing models and have a great potential to be simulated globally.

Statement of the Problem

- Forecast Risk modeling endemic TB using Geographic Information Systems (GIS)
- Analyzing Risk Factors of Mycobacterium tuberculosis through Social Determinants

Tuberculosis (TB) is an infectious disease caused by a germ, or bacterium, called Mycobacterium tuberculosis (M. tuberculosis). It is more commonly known to affect the lungs; however, it can affect any organ system. There is no specific age range for this disease due to that fact that it can affect anyone regardless of age. The biggest issues concerning this disease are dissemination of the disease and treatment. TB is spread through aerial dispersal. The disease mode of transmission is through the act of coughing, talking and spraying the bacterium. However, according to the National Jewish Health, TB is not easily spread and typically involves weeks of
indoor contact with a person who is infected with TB. Other issues at hand, concern the medication. There are many medications that treat and cure this disease; however, throughout the years TB has become resistant to antibiotics, due to the inconsistency of patients taking the medications.

The cases of TB at most risk would be those that are left untreated, these cases can cause high levels of mortality and morbidity. Nevertheless, inadequate treatment can lead to the development of a more serious kind of TB. Drug-resistant strains of M. tuberculosis have caused serious cases of mortality within the population it affected, due to its inability to take in effective treatment. Co-infections have also worsened the treatment of tuberculosis. Once the immune system has been compromised with a previous disease, it allows for TB to worsen the affected. In many cases, not everyone who inhales the germ develops the disease. In most tuberculosis infections, the body's natural defenses are able to control the infection. Only about 10 percent of those infected will develop active disease in their lifetimes. Instead, the bacteria persist as a latent TB infection, which cannot be spread to other people. Active disease can occur in an infected person when the body's resistance is low or if a large or prolonged exposure to the germs overcomes the body's natural defenses.

**Diagnosis, Model of Treatment, and Current Control**

All Tuberculosis Sanitarium Facilities have been closed in the state of Florida since July 2012. Since then all cases of tuberculosis are controlled by Florida System of Tuberculosis Care (FSTBC) for rate reduction and treatment. Florida System of Tuberculosis Care (FSTBC) goal is set to reduce the number of tuberculosis cases in Florida to reach no more than two cases per 100,000 populations by year 2020. Their framework or model of care is based on two elements of patient-centered and community-based. The Florida System of Tuberculosis Care (FSTBC) model,
does not require hospitalization (reduces admission) of tuberculosis infected individual instead it will develop community-based care option.

The Florida System of Tuberculosis Care (FSTBC) program utilizes the following key components:

- **Treat to cure** all active cases of tuberculosis infected individual must be treated completely and kept in safe environment to protect the public
- **Protect** all individuals with close contact with active tuberculosis disease
- **Prevent** advancement of all tuberculosis cases by identification and proper treatment of all cases of latent tuberculosis infection
- **Ensure** that all tuberculosis infection prevention policy, guideline and practices are installed and practiced in all faculties to protect other with high risk of transmission
- **Monitor and evaluate** program performance continuously to ensure that intervention program activities are working properly

Furthermore, the Florida System of Tuberculosis Care (FSTBC) program for controlling tuberculosis activities provides three levels of care;

**Level 1**: Local public health system provides management for identification; accurate diagnosis, proper treatment as well as medical and social matters and patient follow up in home - county. This level provides almost 90% of care through 67 local county health departments.

**Level 2**: Area tuberculosis network provide access to advance level of care such as infectious disease specialty for specialized medical care beyond level 1 as well as mental care and cohort review. This level provides management for approximately 5% of all cases of active tuberculosis through 8 multicounty health departments.
Level 3: Level care are provided through contracted hospital services to remaining 5% cases of active tuberculosis from moderate cases to severely complicated through in-patient hospitalization services with highly trained or specialized pulmonary and infectious disease physicians.

Treatment: The main commitment is to provide each patient with confirmed cases of tuberculosis infection with direct observed therapy (DOT). All county health departments in Florida adhere to the policy for teaming up with local private physicians to provide direct observation therapy (DOT) services.

Most common barriers to treatment are defiance, socioeconomic status, illiteracy, homelessness, lifestyle, non-adherence to complete therapy protocols, limited access to health care and lost to follow up.

Predicting Encounters

Income as a Risk Factor

Levels of poverty have always had correlation to the burden of any disease. In the case of TB this remains the same. According to the World Bank, there is a correlation that relates to the lower the national income per capita the higher the TB incidence. Therefore, one can analyze that income plays a major role in the population it affects.

In the household individual level, prevalence of TB is mostly seen in low and middle-income countries. According to the World Health Organization, South Africa, Zambia and Bangladesh are known to have prevalence of at least 6 times higher than any other country. Socioeconomic status increases the risk for TB.
In the U.S. there have been 6 indicators that have been identified in the Morbidity and Mortality Weekly Report, CDC, IDSA and ATS.

Six Identified Indicators in the United States

1. Overpopulation
2. Education
3. Income
4. Poverty
5. Public Assistance
6. Unemployment

In addition to the identified indicators, there was a noticeable increased risk for TB associated with race and ethnicity. African Americans, Hispanics and American Indians, are found to have a higher prevalence of TB in the U.S. There have been many assumptions as to why this relation exists. When it comes to socioeconomic status, these populations have been prone to have a lower income level and therefore the indicators would apply to many within this specific demographic. As seen in the picture below, the World Health Organization produced a pathway showing the socioeconomic status of a community and the connection it has with the influential risks of possible TB infections.

Socioeconomic factors are not the only ones that affect TB and the population at most risk. Other factors, such as co-infections make this disease such a serious illness that burdens the most disadvantaged population. Social behaviors and lifestyle choices also make a great impact on the population at most risk for TB infections. These factors create a massive effect on the prevention and control of TB. Not only must health professionals treat the disease but they must also take into consideration the social aspects that contribute to the emergence and reemergence of the disease in the population they are attending. Policy makers must also take into account these social behaviors in order to create policies that can help ease the burden of tuberculosis among their populations.

**Co-infections as a Risk Factor**

**HIV.** According to the Centers for Disease Control and Prevention (CDC) there has been
a decrease in TB cases among the U.S. population. However, TB is still seen as a threat due to vulnerability that immune compromised people have with this disease. Worldwide, TB is the leading cause of death among people living with HIV.

In the U.S. people living with HIV/AIDS are extremely vulnerable to co-infections with TB. It is estimated that about 4.2% of Americans (13 million individuals) are infected with TB bacteria. As of 2011, CDC estimated 6% of all TB cases and 10% of TB cases among people aged 25–44 occurred among people who were HIV-positive. Due to the gravity and the vulnerability co-infected people have, CDC recommends that all HIV positive people get tested for TB.

HIV co-infection is not the only risk factor among TB infection; there are many other risk factors that also play a role in the transmission of the disease. These two diseases can be correlated through their social aspect of transmission, both can be said to have social factors that influence and highly increases prevalence among the population it affects.

**Diabetes.** In recent studies there has been a finding in the correlation between diabetes and TB infection. People with diabetes are at higher risk of developing tuberculosis (TB) than those without diabetes. TB affects mainly low- and middle-income countries, however now a day's these countries are not only vulnerable to infectious diseases. Presently, chronic disease has been on the rise in developing countries. Regions, such as in Africa and Asia that are affected by tuberculosis and are also those that have some of the highest numbers of people with diabetes and will experience the biggest increases by 2030.

The growing prevalence of diabetes in low-income countries has become a challenge for TB control as uncontrolled diabetes leads to a greater risk of developing TB. A recent study indicated that countries that saw an increase in diabetes prevalence had a significant increase in the number of people with TB. These trends and correlation in prevalence among the diseases
mentioned show an important link between diabetes and TB infection. Several studies have looked at the association between diabetes and tuberculosis in developed countries and found that people with diabetes are around 2.5 times more likely to develop tuberculosis. These findings were found to be accurate when analyzed in developing regions such as Africa and Asia. In Africa one study showed that the prevalence (%) of diabetes was twice as high in people with tuberculosis as in people without tuberculosis.

The methods used to analyze the correlation between diabetes and TB infections were through individual-level analyses using the World Health Survey (n = 124,607; 46 countries). In which an estimation of the relationship between TB and diabetes was examined, adjusting for gender, age, body mass index, education, housing quality, crowding and health insurance. A longitudinal country-level analysis using data from 1990 to 1995 and 2003 to 2004 on per-capita GDP, TB prevalence, incidence and diabetes prevalence in 163 countries was used to estimate the relationship between increases in diabetes prevalence and TB, identifying countries at risk for disease interactions.

The findings led to the conclusion that in countries where diabetes cases are relatively high, for example Mexico, Egypt, Saudi Arabia and the U.S., diabetes is a significant contributor to the number of cases of tuberculosis. However, where rates of tuberculosis are high and diabetes is relatively low, diabetes contributes to a smaller proportion of the TB burden. These findings also established that in lower income countries, individuals with diabetes are more likely than non-diabetics to have TB [univariable odds ratio (OR): 2.39; 95% confidence interval (CI): 1.84-3.10; multivariable OR: 1.81; 95% CI: 1.37-2.39]. Increases in TB prevalence and incidence over time were more likely to occur when diabetes prevalence also increased (OR: 4.7; 95% CI: 1.0-22.5; OR: 8.6; 95% CI: 1.9-40.4). One can assume that, given this information, countries which are
prevalent with TB cases and have projected increases in diabetes are more likely to increase in tuberculosis infection cases. This study mentioned countries like India, Peru and the Russia Federation as areas of particular concern.

Social Behaviors as Risk Factors

In many aspects social behaviors have always been the least likely to be analyzed when it comes to diseases. However, the burden of any disease cannot be ascertained without first examining how it affects the person, their daily habits and their life style. Social scientists have been able to contribute to the control of disease dispersion because of the links that have been found between diseases and the social behaviors of the population.

Tobacco smoke. There have been recent studies that have found association between TB and tobacco smoking. According to the article published in the International Journal of Tuberculosis and Lung Disease by the University of Taipei in Taiwan, tobacco smoking doubles the risk that people who have been successfully treated for TB will develop TB again. This is a condition known as "recurrent" TB. The study is the most robust ever conducted which illustrates how smoking tobacco increases the risk of recurrent TB. As seen in the graph below, the probability of TB reoccurrence grows as the amount of smoking increases. The studies showed the difference between smoking more than 10 cigarettes per day, smoking 1 to 10 cigarettes per day as well as never smoked or have quitted smoking overall. The differences can be seen when compared to each other against the probability of TB infection reoccurrence.
The results obtained from this study showed that from the 5567 adults who were followed for recurrence after successful TB treatment. The mean age was 58.5 years; 62.9% were male. Overall, 84 (1.5%) had a recurrence of TB during follow-up. The incidence of TB recurrence was 4.9 episodes/1000 person-years of follow-up. Cox proportional hazards regression showed that after controlling for other variables, the risk of TB recurrence among subjects who smoked more than 10 cigarettes a day was double that of never/former smokers. Other independent risk factors significantly associated with TB recurrence were homelessness (aHR 3.75, 95%CI 1.17–12.07), presence of comorbidities (aHR 2.66, 95%CI 1.22–5.79) and a positive acid-fast bacilli smear (aHR 2.27, 95%CI 1.47–3.49). Therefore, one can assume that smoking more than 10 cigarettes a day had a significant correlation with TB infection recurrence.

**Alcohol consumption.** Alcohol consumption has been described as a risk factor for infection with Mycobacterium tuberculosis, but its contribution to tuberculosis has been difficult to separate from other socioeconomic factors. However, many studies have shown that there are significant associations between one and the other. There are studies showing pathogenic effect of
alcohol on the immune system causing susceptibility to TB among heavy drinkers. There are many potential social pathways linking alcohol use disorders and TB. Heavy alcohol use strongly influences both the incidence and the outcome of the disease and was found to be linked to altered pharmacokinetics of medicines used in treatment of TB.

The graph below is a testament for the prevalence of alcohol usage and/ or alcohol use disorder in association with TB infection in the general population. Prevalence of TB infections among persons with alcohol use disorders is drastically shown to differ.

Numerous studies show the pathogenic impact of alcohol on the immune system causing susceptibility to TB among heavy users. In addition, there are potential social pathways linking alcohol use disorder and TB. Heavy alcohol use strongly influences both the incidence and the outcome of the disease and was found to be linked to altered pharmacokinetics of medicines used in treatment of TB, social marginalization and drift, higher rate of re-infection, higher rate of treatment defaults and development of drug-resistant forms of TB.
Studies have also shown that the risk of active TB is substantially elevated in people who use more alcohol than 40g per day or have been diagnosed with an alcohol use disorder. Among these vulnerable populations, alcohol usage has also been found to aggravate the progression of TB infection.

Social marginalization has also been a trait found in both TB infected people and alcohol consumers. The same studies have also found that along social marginalization, there has been higher rates of re-infection, higher rates of treatment cessation and development of drug-resistant forms of TB. These drug resistant forms of TB are the hardest to treat especially because of vulnerability that the immune system is already encountering due to the prevalence of other malignancies affecting the body.

There are many factors that contribute to TB infection, however as explained and seen in
the sections previously mentioned, most of these factors can be modified with the proper usage of public health methodology as well as governmental policies. In order to eradicate or eliminate burden of TB infection among the population we must take into account all forms of possible associations to the disease.

The PRECEDE-PROCEED Planning Model

The PRECEDE-PROCEED Planning Model (PPM), is currently used as a planning model that serves as an ecological approach to health promotion. The model recognizes that every aspect of a person’s environment is a potential target for intervention, along with the individual’s behaviors, skills and thoughts. PRECEDE (Predisposing, Reinforcing, and Enabling Constructs in Educational/Environmental Diagnosis and Evaluation) refers to the first four stages of the model, whereas PROCEED (Policy, Regulatory and Organizational Constructs in Educational and Environmental Development) refers to the portion of the model that begins with step 5. Per Crosby and Noar (2011), the PPM is a method of guiding the target community to developing appropriate, fundamentally sound, and ecologically based approaches to determining solutions to the problems they face. Additionally, the PPM framework offers the support of theory (although the use of theory is not mandatory in a PPM-based intervention plan) and also incorporates evaluation mechanisms. The PPM is divided into a number of phases, which will be detailed individually (Crosby & Noar, 2011).

Phases of the PRECEDE-PROCEED Planning Model

The PRECEDE-PROCEED Model begins with Phase 1, which dictates that the interventionists should work with community groups with the intention of identifying common
problems and goals of a community, while mobilizing resources and developing and implementing intervention strategies. Of note is the importance of ensuring that community mobilization is not focused on the “process,” rather, it should be community-driven with a focus on problem identification, needs assessment and program design. The phase begins with the collection of demographic data which are then presented to the members of the community with the hope of aiding them in deciding on their priorities. This demographic data is usually acquired in one of the following methods, including community surveys; focus groups, phone or face-to-face interviews, and questionnaires. Applied to the issue of tuberculosis, it is recommended that both individuals affected directly by TB and their communities alike be involved in the planning, implementation, and evaluation of health interventions.

For example, it is possible for a TB intervention team to use a GIS-based, remotely-sensed model to predict the prevalence of TB in an area and thereby proceed to enter the communities predicted to be most affected by tuberculosis and engage the members of the community in developing a TB management and control plan. Using our model, this intervention team could identify a specific neighborhood or collection of streets which need to be targeted for intervention and contact community leaders to cooperatively establish a program of information collection, including surveys taken door-to-door, by telephone, or via email, concerning the neighborhoods concerns regarding the presence of TB in their community. By first going to community leaders, the interventionists can determine which method of contact is most efficient and most strongly preferred in this area, saving valuable time and resources. These surveys should focus on ascertaining what the community feels are the most glaring weaknesses of their neighborhood when dealing with members of the community who are suffering from TB, or when attempting to deal with an ongoing spread of TB. Additionally, demographic information concerning the
members of the community should also include whether or not they have many (or any) contacts with TB, along with the requisite information regarding their socio-economic status, health, and personal characteristics.

Phase 2 of the PPM addresses epidemiological, environmental, and behavioral assessments. In this phase, the interventionists must identify the health priorities of the community and their behavioral and environmental determinants. Specifically, the epidemiological assessment must focus on identifying health problems, aspirations, and issues upon which the intervention program will focus. In addition, the epidemiological assessment must uncover the environmental and behavioral factors that are most likely to influence the prioritized health concerns identified in Phase 1 by the community.

Following this, behavioral determinants must be identified, on the basis of three levels of proximity: most proximal, more distal, and most distal. The most proximal behavioral determinants are behaviors and lifestyles that contribute to the severity of a health problem, while the more distal determinants are the behaviors of others that can impact the behaviors of those individuals at risk. Finally, the most distal factor includes the actions undertaken by individuals of authority that may affect the physical or social environment that influences the individual at risk.

Subsequently, environmental determinants must be assessed, and these include those social and physical factors that are external to the individual and are often beyond their control. However, these factors can be modified to support the behavior or influence the health outcome, albeit requiring a strategy other than education, such as community mobilization. At this level, it is very likely that involving stakeholders and community members will help the planner in acquiring the best information possible concerning which issues to emphasize and, additionally, helps to prevent
mistakes due to ignorance of the community’s history, culture and the interpersonal relationships found within.

Applied to a TB intervention scenario, the epidemiological, environmental and behavioral assessments should focus on what the community deems most important in terms of health priorities, such as the known prevalence of a number of TB cases in a community. The interventionists can then attempt to identify, with the community’s assistance, the behavioral and environmental determinants that are contributing to the presence of TB in the community, focusing on the proximal, the more distal, and the most distal.

At the proximal level, the interventionist is assessing the behavioral tendencies of an individual that contributes to the severity of the spread of TB, such as the tendency of infected individuals to delay seeking treatment or venture to public areas where many others often congregate, such as a park or a concert. These actions and others like them help spread TB to individuals at risk, an outcome that is easily avoidable given appropriate modifications to a TB-positive individual’s behavior, and the subsequent durability of these behavioral modifications.

At the more distal level, the behaviors of others who are in a position to influence those with TB are assessed. A family member or a friend may inadvertently make an individual with TB feel stigmatized for seeking treatment or admitting any manner of health issue, leading this person to either delay or avoid seeking medical care and participate in group activities without taking precautions to prevent spreading the disease to others.

At the most distal level, the actions of an authority can influence the social and physical environment surrounding an individual at risk, which may in turn lead to a greater likelihood of contracting TB. As an example, the closing of a hospital in the area due to budgetary concerns may lead the community’s residents to have to commute a much farther distance to seek treatment,
which may in turn discourage them from following up with their physicians and/or continuing their treatment regimen, thereby increasing the likelihood that they will stay contagious and will only help to worsen the situation in the community.

Several useful theories can be applied at this level, including interpersonal theories of behavioral change, which emphasize the interaction between the individual and the environment such as the Social Cognitive Theory, developed by renowned psychologist, Albert Bandura, and his team at Stanford in 1986 (Bandura, 1991). The Social Cognitive Theory emphasizes the cognitive processes that mediate learning. The theory posits that human behavior is both motivated and regulated by self-influencea perpetually ongoing process. Bandura (1991) claimed that self-regulation occurs through three avenues: self-monitoring of one’s own behavior, its determinants, and its effects, judgment of one’s behavior through the prism of personal standards and environmental circumstances, and affective self-reaction.

Per the interactionist perspective of the Social Cognitive Theory, it is posited that social factors affect the operation of an individual’s self-regulative systems, thereby allowing for a TB intervention planner to examine the interactions between a TB patient and his or her environment by way of evaluation of the social structure within which he or she is involved. Theories can be explored concerning organizational change, which are useful when a noted concern of the community is a disconnect between the current policies and practices of formal organizations and the needs of the community.

Another important theory that may be applied at this level is the Diffusion of Innovations Theory, which aims to describe and predict the process by which new ideas are adopted by a community.
Moving to the third phase of the PPM, the community and intervention team must consider an educational and ecological assessment following the selection of behavioral and environmental factors for intervention. This is followed by identification of antecedent and reinforcing factors that must be in place in order to initiate and sustain the change process, referred to as predisposing factors, reinforcing factors, and enabling factors. Predisposing factors are antecedents to a behavior that provide rationale or motivation for that behavior, and can include a person’s knowledge, beliefs, attitudes, skills, and self-efficacy beliefs. Reinforcing factors are factors following a behavior that provide continued reward and incentive for repetition of that behavior, and include social support, peer influence, and family influence. In addition, enabling factors are another category of antecedent and reinforcing factors; specifically, as antecedents to behavioral or environmental change that allow a motivation or environmental policy to be realized. Examples of enabling factors include the availability of resources such as mental health treatment services, accessibility of services leading to quick treatment, community and government laws, policies and commitment to the issue at hand, and issue-related skills.

In the application of the third phase of the PPM to a tuberculosis intervention, predisposing factors include an individual’s knowledge of TB, its symptoms and epidemiology, and of treatment regimens, and prevention techniques. Additionally, an individual believing in his or her ability to stave off infection and thereby leading to the belief that prevention techniques are unnecessary could lead to this individual not observing the proper precautions that will help prevent the spread of TB in the community. There are reinforcing factors when considering the context of TB interventions where social support can be included for individuals who are seeking treatment and who may feel stigmatized as a result. Counseling for these individuals and their families will greatly aid in ensuring that these individuals seek treatment and that their families will support
them throughout the process, rather than simply ostracizing them. Enabling factors are also crucial in a TB intervention scheme. If a TB treatment facility is readily available, providing free screening and treatment options, affected individuals will be far more likely to take advantage of these resources and, most importantly, seek help rather than simply not take any action.

The fourth phase of the PPM is administrative and policy assessment and intervention alignment, a phase in which the planner will select and align the program’s components, with the highest priority being given to the determinants of change that were previously identified. This additionally includes identifying resources, organizational barriers and facilitators, and policies that are needed for the program’s implementation. This also involves two levels of consideration: the macro and the micro. The macro phase involves organizational and environmental systems, while the micro level focuses on the individual, his or her peers, friends, family, and others who are capable of influencing desired change. Interventions that occur at this level are aimed directly at the predisposing, reinforcing, and enabling factors. A TB interventionist can identify the community’s infrastructure and determine the number of clinics and hospitals within accessible distance of the affected neighborhood, and their resources and capabilities with regard to treating TB. Additionally, the interventionist can examine the organizational and policy barriers to the implementation of the TB intervention, such as local laws that may restrict their activities. On a micro level, the interventionist should also consider the power of the individual, his or her peers, and family to bring about change through word-of-mouth or the starting of a grassroots campaign to address organizational and social barriers that are impeding progress on a TB intervention’s ability to help a community that needs it.

Thus, given the information presented with the four previous phases, the following recommendations can be given to an intervention planner concerning matching, mapping, pooling,
and patching: Ecological levels should be matched to the broad program, specific interventions based on theory and prior research should be mapped to identify predisposing, enabling, and reinforcing factors, previous work and interventions done in the area should be pooled together, and finally interventions should be patched so as to fill in the gaps and reflect evidence-based best practices. The main theory at work at this level is the organizational change theory at the community level, which addresses processes and strategies for creating change.

At phase 5, the PPM shifts over towards the PROCEED Model, with this phase specifically concerning implementation of the intervention. The final three phases are carried out as the intervention progresses, allowing for the planner to monitor the intervention and perform adjustments as necessary. Phase 6 concerns process evaluation, emphasizing the effectiveness of the procedure of the intervention with the goal of determining whether or not the intervention is succeeding in proceeding as it was originally planned. Phase 7 addresses impact evaluation, wherein the planner evaluates the initial successes of the intervention efforts. Effectively, this phase seeks to determine whether or not the intervention is doing what is expected.

The last phase is phase 8. Phase 8 addresses outcome evaluation, focusing on whether or not the intervention is successfully leading to the outcome identified by the community in phase 1. A potential source of frustration at this step is the knowledge that the intervention proceeded exactly as it was designed, only to not lead to any noticeable or significant effect on the issue at hand. Further, some outcomes may not make themselves apparent for years, even decades, such as chronic disease outcomes including stroke and heart disease. With respect to TB, the interventionist can re-visit the community and determine to have suggestions made and implemented.
The following questions could be used to obtain the determinants in question. Are TB patients and their families regularly seeking treatment, and then maintaining this regimen? Are local governments appropriating sufficient capital and resources to help reduce the spread of TB in the affected neighborhoods? The answers to all of these questions will, together, help the interventionist determine whether the TB intervention truly succeeded in controlling and reducing the prevalence of TB in an affected community.

**Validation and Future Applications of the PPM**

The PRECEDE-PROCEED Model, at its core, serves to utilize the power of community engagement in an effort to combat the unique challenges faced by the community in question at a number of levels, ranging from the individual to the organizational to the legislative. The importance of the fundamental belief underlying the model, the idea that the community being affected by a specific issue should be not only involved with, but in some ways spearheading the efforts to obtain a solution to this problem, cannot be understated. From the perspective of the members of the community, a visiting team of clinicians and public health experts, as skilled and knowledgeable as they may be, may simply be a group of outsiders who are unfamiliar with the history and dynamics of the community’s culture, attempting to tell them what’s best for them. In essence, the public health workers who enter the community with the best of intentions simply work to push the community further away from accepting their help if their suggestions and advice is given from a position of implied superiority over those individuals who are seeking assistance. If the members of the community are actively involved in the planning of the intervention at all stages, to include the crafting of its cultural characteristics, dynamics and history; the outcome of
such effort would be mutual acceptance for treatment as well as placing an emphasis on the prevention so that others may benefit.

The Geographic Information Systems (GIS) identifies locations that are most heavily affected by a specific medical condition of the investigator’s interest, such as tuberculosis. Here, counties in Florida were assessed with a projection of future cases of tuberculosis assigned for each county. Counties can then be ranked, giving a focused list which prioritizes counties for intervention and assistance based on the number of cases of tuberculosis. With this information, for example, investigators can go to the counties most heavily affected or projected to be affected by tuberculosis and identify the specific neighborhoods and communities that are of highest importance for urgent interventional care. With these communities identified, investigators can proceed to use the PRECEDE-PROCEED Model to ascertain the community’s history and culture, along with its social dynamics, and work with community leaders and members to develop an intervention that has the support and the interest of the community itself. This helps to promote greater adoption and acceptance of the intervention by the community, as the community itself was partly responsible for designing the program that is aimed at alleviating the burden they face due to the presence of tuberculosis.

**Analysis of Mycobacterium Tuberculosis, Using Geographic Information Systems**

Mycobacterium tuberculosis/tuberculosis disease (TB) is spread from person to person through the air-infected individuals who can transmit droplet nuclei of TB bacteria via coughing, sneezing or by spitting. Only a few of these organisms, need be inhaled by another in order to cause transmission (Anderson, 1958) It has been postulated that changes in breathing can in turn change the rate at which TB is spread by infected individuals. Specifically, a significant increase
in altitude is known to cause hyperventilation in humans; indeed, up to 3500m the tidal volume is increased, with breathing rate significantly increasing above this altitude (Baker, Harries, & Jeon, 2011). In addition to the reduction of barometric and inspiratory oxygen pressure, the decreasing air density, temperatures, humidity, and pollution all contribute to ever faster breathing: the reduced air density permit less-resistant air flow, the decreased humidity and temperature contribute to the hyperventilation of dry, cold air and the minimization of pollution reduces the amount of airway inflammation (Baker et al. 2011).

Consequently, the question must be asked: is an individual infected with TB capable of spreading the bacteria more effectively at these higher altitudes due to faster and more frequent respiration? Is there a method of using geographic information systems to identify areas that, if this association between altitude and TB was documented, are people vulnerable to the spread of TB due to high elevation and the proximity of many cases of TB to these areas?

The Effect of Elevation: Respiration Rate and UV Irradiance

A literature review was conducted to determine whether or not higher elevations truly impact tuberculosis infection rates, especially whether or not higher elevations lead to greater respiration and subsequently higher transmission of TB bacteria. A study by the Department of International Health of Johns Hopkins Bloomberg School of Public Health showed the Comparison of Altitude Effect on Mycobacterium Tuberculosis Infection Between Rural and Urban Communities in Peru, examines the distribution of TB through an endemic country. Peru is noted to have an estimated prevalence of 233 cases per 100,000 inhabitants as of 2004 (Balasubramanian, Wiegeshaus, Taylor, & Smith, 1994). This study examined 9 communities across Peru: two rural highlands, one urban lowland, two rural lowlands, two urban non-highland
shantytowns (Sachaca and Cerro Colorado), and finally two highland peri-urban shantytowns (San Jeronimo and Huascahura). In the context of this study, “highland” was defined as an altitude equal to or above 3,000 meters (Balasubramanian et al., 1994). The authors noted that residents of the shantytowns are typically impoverished poverty having been identified as a noted predictor of TB infection primarily earning money by establishing temporary small businesses, selling seasonal fruits and vegetables, or performing labor for local construction companies (Balasubramanian et al., 1994). It is also important to note that these individuals live in adobe or stone houses with few windows, giving rise to the increased probability of transmission between cohabitants of these dwellings.

This study utilized trained field workers who visited households in the target villages and administered structured questionnaires concerning their recent (< 5 years) exposure to active cases of TB and their overall health, and also administered the tuberculin skin test (TST); an induration size greater than or equal to 10 mm being considered TST positive. These individuals who tested positive were then examined for further evidence of active TB infection (Balasubramanian et al., 1994). In addition to the data collected regarding TB infection and the results of the TST, demographic information was collected including history of living with individuals with active TB, the presence/absence of BCG scar, the number of individuals in a household, education level, and whether or not the individuals were living in urban or rural areas (Balasubramanian et al., 1994). These data were used in multiple regression analyses, which revealed that the protective factor of high altitude on prevalence of positive TST in rural highland populations was utterly neutralized in urban highland populations due to crowding and the possibility of increased contact between active TB cases and vulnerable individuals (Balasubramanian et al., 1994). Further, it was noted that rural highland villages experience very little migration inwards from lowland areas
where TB disease is common, and that high altitude was strongly associated with decreased TST positivity after the adjustment for household active TB contacts, BCG vaccination, age, household positive TST contacts, and community clustering (Balasubramanian et al., 1994).

Subsequently, it was concluded that greater risk of TB infection in rural areas at low altitudes (and the opposite at high altitude, with greater risk in urban areas) could be due to the high humidity in the jungle and lack of adequate healthcare facilities (Balasubramanian et al., 1994). Indeed, while the authors acknowledge that higher altitudes do make it possible for TB infection to be prevented due to lower oxygen tensions or stronger UV light exposure, they conclude that the most likely location for TB transmission is within the household itself, which in these areas are typically small and lack windows, diminishing UV exposure and increasing the probability of transmission (Balasubramanian et al., 1994). Finally, it was concluded that the protective effects of UV light destroying TB bacteria at high altitudes (marked by dry climates) do not apply to urban areas at these altitudes due to the overwhelming effects of population density.

This study establishes two things: that there is evidence that higher altitudes are capable of inhibiting TB distribution due to environmental covariates, and that this effect is only valid in rural areas because higher population density in urban areas overpowered this protective effect and led to increased transmission of TB (on the same level as lowland areas).

After consulting several changes must be considered to the scope of this research: thus far, it has been posited that elevation has a direct effect on the respiration rate of individuals who travel to areas of high altitude, thereby increasing the amount of TB bacilli released into the air and subsequently increasing the exposure of TB bacteria to uninfected individuals. However, while simply breathing in the air in which an infected person has spoken or respired is sufficient for transmission of TB, there are only a few situations in which an infected individual would be placed
into a hypoxic state necessitating hyperventilation, such as a first-time visit to an area of high altitude. While this could mean that visiting tourists or businesspeople who are infected with TB are potential risks for the spread of TB in a highlands area, this does nothing to explain the cases of individuals who are either frequent travelers to these areas or who live locally, and are therefore acclimated to the altitude and do not respire abnormally. Therefore, due to the rarity of the situations in which TB infected individuals may reach highland areas and spread the disease due to hyperventilation, focus should instead switch to the effect of elevation on TB through a different proxy variable: Ultraviolet (UV) radiation.

The Department of International Health of Johns Hopkins Bloomberg School of Public Health noted several mechanisms inhibit TB infections at higher altitudes, including lower oxygen tensions or stronger exposure to UV light, with the drier climate at these altitudes increasing susceptibility of *M. tuberculosis* to UV light, thereby reducing TB transmission (Balasubramanian et al., 1994).

**Reassessing the Direction of This Study**

The questions driving this research now become: what specific intensity of UV light adversely impacts TB bacilli growth, and how can GIS software be utilized to identify regions which receive this protective benefit?

**An Examination of UV Light Irradiance**

The concept of ultraviolet germicidal irradiation (UVGI) is a documented methodology for disinfection and the prevention of transmission of infectious diseases such as influenza and tuberculosis, most commonly utilizing low pressure mercury discharge lamps which emit
ultraviolet-C radiation (UV-C, 100-280 nanometer (nm)) at a wavelength of 254nm (Bandura, 1991). UVGI has been attempted in a variety of conditions and avenues, including water, air, and surface disinfection, with water disinfection widely considered the most advanced and commonplace application of UVGI treatment (Bandura, 1991). On the other hand, surface disinfection is plagued with difficulties due to the limitations created by the presence of micro-shadows and absorptive layers atop the surface (Bandura, 1991). Finally, air-based disinfection is accomplished in a number of ways, including upper-room air irradiation, full room irradiation, and irradiating air that passes through enclosed heating, ventilation, and air-conditioning (HVAC) systems (Bandura, 1991). For the scope of this research, the air-based irradiation systems will be examined in the hopes of identifying the requisite dose of radiation required to neutralize airborne M. tuberculosis particulates.

Upper-room UVGI is primarily designed to be implemented in enclosed spaces (such as within residences) and employs louvered UVGI lamps that are either suspended from the ceiling or mounted high upon the walls to focus the germicidal radiation to the upper portions of the entire room, keeping the UV radiation above the heads of the occupants and attempting to minimize the risk of exposure (Bandura, 1991). This method also relies upon effective vertical movement of air between the upper and lower portions of the room, generated by an HVAC system, fans, or in the case of areas with limited resources, natural convection (Bandura, 1991).

The alternative UVGI method based on airflow is the in-duct method, which attempts to disinfect air within the building’s HVAC system and before it enters the room or is recirculated using very high intensities of UV radiation that is not exposed to the occupants of the room itself. This method requires efficient circulation of the air through the duct (with priority placed on a system that maximizes the amount of air, by volume, that passes through this duct) and the velocity
at which this air circulates through the system (Bandura, 1991). While both of these methods hold
promise as methods of eliminating TB bacteria, their limitations are straightforward: they can only
be used in buildings with HVAC systems or with sufficient technological capacity and availability
to use hanging UV lamps, fans to distribute irradiated air, and adequate electricity to power all of
these systems. The reality of the situation, however, is that many areas endemic for TB are also
the areas with some of the highest rates of poverty in the world, wherein such simple luxuries as
electrical fans are rarities. Nonetheless, the development of a cheap, easy-to-use method of
irradiating the interiors of the homes of individuals diagnosed with TB has the potential to greatly
diminish transmission of TB between cohabitants. Several potential solutions will be explored in
brief in subsequent sections of this paper, but would have to be researched separately to be given
their due attention.

Nonetheless, the importance of UVGI on TB transmission has been recognized, subsequently explored and ultimately summarized by Reed (2010). The first study examined, the Tuberculosis Ultraviolet Shelter Study (TUSS) was the first study to be conducted out “in the field” and examine the use of upper-room UVGI at 14 homeless shelters spread across 6 cities in the United States (Bandura, 1991). While the results of this study were found to be inconclusive, Reed (2010) notes the gain of much practical knowledge as a result of carrying out this study. In addition, Reed (2010) notes a study conducted which was inspired by the classic study of UVGI’s effect on TB transmission in the 1950s, wherein air was ventilated from an occupied HIV-TB ward in Peru to colonies of guinea pigs housed within rooftop chambers for 535 days (Bandura, 1991). By alternating UV exposure by the day (that is, a UV exposure day followed by one without UV exposure), the treatment group of guinea pigs breathed air from the TB ward treated with upper-room UVGI, while the control group breathed air from the TB ward with this upper-room UVGI
turned off. Results of this study demonstrated a 34.9% infection rate in the control group (without UVGI), but only 9.5% infection rate in the group with UVGI-treated air (Bandura, 1991). TB disease, however, was then confirmed in 8.6% of the control group subjects, but only 3.6% of the UVGI-treated group (Bandura, 1991).

Another study, conducted by Xu et al. (2003), sought to determine the effectiveness of UVGI treatment in inactivating bacterial spores in a controlled laboratory environment made to simulate real-world situations, using *Mycobacterium parafortuitum* and *Mycobacterium bovis* BCG, a surrogate for *M. tuberculosis*. The test room developed for this study, at the Joint Center for Energy Management’s Larson Building Systems Laboratory at the University of Colorado at Boulder, was connected to a computer-managed ventilation system to deliver 2-8 ACH of HEPA-filtered air from the outside (Bishai et al., 1998; Xu et al., 2003). Of particular note is the UVGI system employed in the study, which consisted of 4 UV luminaries, one in each corner of the room, and a fifth hung from the ceiling in the center of the room (consisting of 4 18W lamps, yielding a total output of 72W for this center luminary) (Bishai et al., 1998; Xu et al., 2003). The authors of this study then piped in airborne bacteria and cultured samples acquired from the indoor atmosphere both before and after activation of the UVGI luminaries to determine the reduction in concentration of the bacteria following exposure to UVGI. Xu et al. (2003) determined a 46%-80% reduction of *B. subtilis* spore concentration, an 83%-98% reduction for *M. parafortuitum*, and most crucially, a 96%-97% reduction for *M. bovis* BCG, the *M. tuberculosis* substitute.

This study demonstrated the capability of the upper-room UVGI system to inactivate airborne mycobacteria and significantly decrease the concentrations of culturable airborne cells, while also identifying factors that can influence the efficacy of the UVGI system, including UV irradiance levels, bacterial physiology, room ventilation rates, photo-reactivation, and relative
humidity (Bishai et al., 1998; Xu et al., 2003). Specifically, it was noted that increasing ventilation rate from 0 to 6 ACH decreased the effectiveness of UVGI treatment on both B. subtilis and M. parafortuitum, albeit not to a statistically significant level; the authors do note, however, that while these results were not significant they were also not conclusive enough to determine that ventilation rates do or do not affect UVGI effectiveness (Bishai et al., 1998; Xu et al., 2003). It stands to reason, however, that increasing the ventilation rate in a room will result in a higher amount of airborne contaminants being removed from the room, but this is not necessarily a desirable outcome: removing airborne TB bacteria from a room will only spread it to adjacent rooms or the air outside the room, giving rise to the potential for further transmission. Additionally, when considering the effectiveness of upper-room UVGI treatment, flushing the bacteria from the room being treated more frequently will only result in less time of exposure of the bacteria to the UV irradiation, thereby reducing the amount of time that the bacteria could be inactivated by UV exposure. Examining the linear relationship between UVGI inactivation rates and irradiance, Xu et al. (2003) suggested that increasing irradiance from UVGI systems provide superior inactivation of airborne bacteria, with the stipulation that there must exist some threshold of limited returns wherein a higher level of irradiance does not appreciably lead to significant inactivation of the bacteria. Additionally, Xu et al. (2003) derived Z values to parameterize the effect of UVGI on microbes as it depends solely on the physiology of the microbe itself (Bishai et al., 1998; Xu et al., 2003). Of particular note is the Z value derived for M. bovis BCG, the stand-in for M. tuberculosis, which was determined to be $1.9 \times 10^{-3} \text{ cm}^2 \text{ } \mu \text{W}^{-1} \text{s}^{-1}$ (Bishai et al., 1998; Xu et al., 2003).
The Use of Geographic Information Systems

With the efficacy of UV-based irradiation as a viable mechanism of the inactivation of TB bacteria established, the focus must now shift to the interplay between a study focusing on this phenomenon and the incredibly powerful tool of geographic information systems (GIS). When one thinks of GIS, what usually comes to mind is a map identifying the distribution of cases across an area, reported either in terms of absolute number of cases or as an incidence or prevalence rate per 100,000. The problems with this type of map are manifold: first, it relies entirely on data of pre-existing or newly confirmed cases in order to paint a picture of areas that are currently suffering from the disease in question. What if, however, this information is difficult, if not impossible, to accurately obtain? With the outbreak of Ebola in West Africa in 2014, it became evident that due to the stigmatization of Ebola patients and the general panic of the citizens of Sierra Leone, Guinea, and Liberia that many families chose to hide the information that someone within their household was infected for fear that they would be removed from the premises and, if and when they were to succumb to the virus, cremated or given a burial that did not align with the cultural beliefs of the people in that area. Although an extreme example, this scenario highlights that there are many complications that may arise when a person attempts to obtain accurate case data.

Furthermore, a case or incidence map suffers from another fundamental problem: it only tells you what has happened up to a certain point in time; it gives no predictions about the future. Looking at such a map, it may be possible to identify a certain region that requires attention (for example, if the majority of cases of TB in Hillsborough County were centered around the University of South Florida area), but this does not tell you if an area is currently undergoing the gradual changes necessary for it to become a hotspot for TB in the future.
A study by Moonan, et al. (2004) exemplifies the issue. Here, the authors sought to identify locations that require targeted testing and treatment for persons “most likely to develop tuberculosis,” a note that is of particular importance and which will be addressed below. To accomplish this, the authors performed a cross-sectional analysis using molecularly characterized clinical isolates to identify patients infected with the same strain of TB, with this information cross-referenced with residential addresses of these patients at the time of diagnosis, thereby identifying which strain of TB was located within each zip code in the Dallas-Fort Worth area. The conclusion reached indicates that the use of molecular strain characterization along with GIS leads to the identification of risk factors associated with the transmission of TB, such as homelessness, alcoholism, drug use, along with highlighting the importance of transmission that occurs outside of the household (Moonan et al., 2004). This study, however, only tells us that certain counties in this area of Texas are at risk of increased TB transmission due to the fact that there are TB positive individuals in the area, a conclusion that could have been reached without geocoding the addresses of patients based on their clinical isolates. In fact, the conclusion effectively states that as there are TB-positive individuals in an area, and TB is an airborne disease, others in the area are at risk. However, this information could be far more useful and informative if the authors did not use GIS to simply map out where TB has shown itself to accumulate, but rather gone a step further and mapped out the distribution of the factors that are most commonly associated with TB transmission: the distribution of homeless shelters, drug dens, crowded urban-residential housing developments, an analysis of the population based on their federal income level and other measures of poverty.

In so doing, it is possible to identify locations - down to a street level - that show the highest potential for the development and spread of tuberculosis throughout its entity. Too often, public
health and medicine are reactionary; they identify areas that are currently suffering and seek to treat, rather than identifying areas that show signs of potential suffering in the future, which could be prevented. That, in essence, is the direction of this project. By identifying the social and environmental factors most strongly associated with TB transmission, it will be possible to use GIS technology to determine the areas that suffer from the highest rates of drug use, alcoholism, crowding, and the link between all of these factors: poverty, to determine the areas that are at greatest risk for an outbreak of TB over 10, 20, or 30 years into the future. With this information in hand, and with knowledge of the areas local culture and customs, it will be possible to craft social and medical interventions that aim to prevent the spread of TB into these areas, thereby sparing the population from a great deal of emotional, physical, and financial suffering and in turn allows healthcare workers to then be able to treat the pre-existing cases of TB without having to worry about that number continuing to rise despite all of their efforts.

There are multiple facets of GIS that make it an appealing approach to combating TB by focusing on the risk factors that contribute to its’ spread. Firstly, there is the capacity of GIS to model an area in a plethora of ways. Using Hillsborough County, once again, as our example it is possible to depict the county in almost any way imaginable. Should a map be required that demonstrates the basic administrative boundaries of the region, focusing on where the county is demarcated, it’s relative location within the state of Florida, and potentially more detailed aspects including roads, rivers, and the locations of known risk factors such as homeless shelters and known drug dens, it can be created.
If, however, a map is needed depicting the vegetation profile of the county, known as a Normalized Difference Vegetation Index (NDVI), the county can be depicted in this manner as well. Further, the county can also be depicted as a Digital Elevation Model (DEM) which shows the elevation contours of the county. In Hillsborough County, which is not particularly hilly, this type of map does not contribute the same amount of information as it would in another country - such as Peru, Ecuador, or Uganda – that is marked by dramatic rises and falls in elevation due to the presence of mountain ranges. This will prove especially useful when exploring the importance of ambient UV radiation on TB bacteria death in the outside air. Specifically, it is theorized that cities and villages built in the mountains will be exposed to greater amounts of solar radiation that will inactivate TB bacteria in the air, leading to decreased transmission of TB in these areas and suggesting that UV exposure, through elevation, is a protective factor against TB.
Figure 5: Digital elevation of central and southern Florida

Figure 6: Digital elevation of Hillsborough County
In addition to the digital elevation models that can be generated using standard GIS software, such as ArcMAP\textsuperscript{TM} 10.1, these models can be three-dimensional as above to depict the individual contours of the location’s elevation profile, allowing for easier visualization of the peaks and valleys of the region. These maps can be used as a base map for the information of interest: for example, the maps on the previous page can be used as a base upon which the boundaries of Hillsborough county can be depicted, along with the city of Tampa and, if desired, the locations of areas strongly associated with the transmission of TB. The elevation of Hillsborough county, as stated previously, is minimal enough such that it is unlikely that there will be any strong associations found between the clustering of TB-risk factors as mediated by their elevation, however it is possible that this could be an important factor in countries where there is a significant difference in altitude due to the proximity of large mountain ranges (such as Ecuador, for example). Therefore, elevation will be considered as a potential covariate affecting TB transmission in these
countries, most notably when considering the effect of UV radiation, which is hypothesized to have a higher average intensity at high altitudes than at sea level in TB-endemic countries that are marked by widely varying elevations.

Additionally, satellite imagery can be used to visualize a study area. There are a number of options that can be utilized: the most basic option would be to use Google Earth™ to identify a study area and save a JPEG image of this view. While this is a quick way of showing a location using satellite images, the limitation in so doing is that Google Earth™’s images do not come with band data, the layers of data required to perform more complex calculations or manipulations to truly study the image. The only satellite sources which come with this data are LANDSAT™, IKONOS™, and QUICKBIRD™.

Figure 8: An example view of Hillsborough County as seen in Google Earth
LANDSATTM is a free data source maintained by the United States Geological Survey (USGS) that provides imagery from the LANDSAT satellite, albeit with mediocre resolutions and with no input from the user regarding off-nadir (points located at an angle from a scanner’s sensors), time of image acquisition, or viewing angle. Of note is the fact that this imagery contains band data, allowing for the researcher to calculate image properties such as pixel values and perform other calculations that would not be possible without such data. For example, if a hypothetical predictor of a disease is a certain type of vegetation, it is possible to obtain the numerical values of the pixels associated with this foliage in the satellite image, then use this data in ArcMAPTM software to conduct a Kriging analysis, which takes the pixel information and scans the rest of the satellite image for areas that have a high probability of showing the same pixel value combination. Thus, it is possible to identify other areas in the image that contain this type of foliage and, therefore, may be sites associated with the disease in question. The limitation of conducting a Krig analysis on a LANDSATTM image, however, is the fact that the resolution of a LANDSATTM image is quite poor: images are either captured at 30m resolution (each pixel of the image representing a 30 meter by 30 meter square) or 15m resolution. While this latter option is certainly an upgrade, it pales in comparison to IKONOS™ and QUICKBIRD™ image resolutions, both of which are sub-meter (each pixel comprising <1m by <1m squares).

Most importantly, however, is the ability to depict any location with extreme fidelity and sharpness, achieved by utilizing imagery of resolutions and clarity such that each pixel of the satellite image represents .61m by .61m, also known as sub-meter resolution, allowing for a skilled GIS technician to identify details as small as rocks or specific types of leaves. This can be done using QUICKBIRD™ imagery, which provides the most detailed resolution possible albeit at a steep price as the consumer is given full control of the off-nadir angle, the time of acquisition, and
the flyby angle for the satellite itself. As such, the consumer can request a fly-by of a study site that requires the satellite to be repositioned based on his or her preferences, a major factor associated with the high cost of the imagery produced by this satellite.

![Image of Hillsborough County](image.png)

Figure 9: An image of Hillsborough County obtained from the LANDSAT 8 satellite

For this study, a combination of these sources will be employed to determine the distribution of risk factors throughout Hillsborough county, known cases of TB, and the areas that stand to be affected by TB throughout the next few years and decades. The goal of this study is to identify and visualize Hillsborough County based on the amount of solar radiation that the county is exposed to on a monthly basis, thereby identifying specific areas, if any, which are conferred
protection from airborne TB due to the amount of ambient UV radiation. Given that Hillsborough County does not have any particularly high hills or mountains, it is expected that there will be nearly uniform dispersion of UV radiation, although the effect of elevation on the intensity of UV exposure will be explored in greater detail in the other countries that are targeted as part of the scope of this study, which all boast more widely varying elevation profiles due to the presence of numerous mountain ranges. In these countries, the amount of UV exposure in highland areas will be compared to that in the lowland regions to determine if areas that are built at higher altitudes, should they receive higher intensities of UV radiation, should be given a lower prioritization for TB treatment schema than lower-lying regions which may not benefit from this potentially protective factor.

In addition, in both Hillsborough County and these other countries, the predictive factors associated with TB transmission will be spatially identified and mapped using a variety of tools available through GIS, including choropleths maps, digital elevation models, land use/land cover models, and point prevalence maps that will allow the researcher to identify “hot spots” marked by a high concentration of the factors that are thought to lead to TB transmission, allowing the researcher to know where TB is currently a problem and, potentially more significantly, make predictions about where TB will be an issue in the near and distant future.
CHAPTER 2
LITERATURE REVIEW

Perez-Guzman et al. (2014) conducted a case-control study, generally descriptive in nature, including 86 patients diagnosed with Tuberculosis (TB) in medical facilities of the state of Aguascalientes, Mexico during 2008. The primary purpose of the study was to identify demographic and other health-related characteristics associated with extra-pulmonary TB (EXPTB) with respect to pulmonary TB (PTB). Subsequent ecological analyses examined the relationship between elevation, immigration, time and the incidence of both forms of TB among 32 Mexican states. The clinical records of all patients recorded as an active TB case by the national TB control program (NTP) were retrieved and interviews were performed in with patients or family members in order to obtain additional information as necessary. Data regarding the nationwide epidemiology of TB, the prevalence of bovine TB, and the health and socioeconomic status of Aguascalientes and adjacent regions was obtained from government websites and publications. Clinical diagnoses of TB were made in conjunction with international guidelines set by the World Health Organization and implemented by the NTP, that is, each TB case exhibited clinical features and/or imaging data congruent with TB and met one of the following criteria: a positive sputum smear for acid-fast bacilli (AFB); a positive culture for *Mycobacterium tuberculosis*; a biopsy suggestive of TB infection and/or a full response to anti-tubercular treatment. Initially, 93 participants were included in the study, but 7 were excluded due to incomplete or untraceable records, bringing the study population to 86. Of these 52 were classified as EXPTB cases and 34 were classified as PTB cases, however, one should note that at least 4
cases (and perhaps as many as 7 cases) among those classified as EXPTB also appeared to be co-infected with PTB as they exhibited sputum smears positive for AFB. Among all EXPTB cases, only 2 diagnoses were made on clinical grounds while the remaining were determined by histological evidence (47), positive culture (7), and positive sputum smears (4), lending credence to the diagnoses. EXPTB and PTB cases were compared according to demographic characteristics, such as age, sex, and occupation, as well as other health-related variables including smoking status, drug addiction, and alcohol dependence. Subsequent analyses compared the incidence of PTB and EXPTB in 32 Mexican states with the percentage of immigrants who moved that state within the last five years and the mean elevation of each state in terms of meters above sea level. Finally, temporal trends of PTB and EXPTB incidence were evaluated for Aguascalientes and Guerro, a state with hyper-endemic PTB, and compared to the national average between the years of 1997 to 2011. The analyses were performed within Epi-Info 6.0 and were largely descriptive as they assessed the relationship between the outcome and nominal variables utilizing the Fisher exact test and the Chi square test with Yates correction. Interval variables were evaluated using the Student’s T-test or Mann-Whitney U-test for pairwise differences, or by means of ANOVA and Tukey tests or Kruskal-Wallis and Dunn tests for multiple comparisons. Subsequent analyses of the outcome against immigration rates, elevation, and time were performed using Pearson’s correlation coefficients.

The most commonly affected sites of patients with EXPTB were the lymph nodes (42.3%); abdominal cavity, including the peritoneal and intestinal TB (15.4%); skin (11.5%); pleura (7.7%); kidney and genitourinary system (7.7%); meninges (5.8%); disseminated (5.8%); and joints (3.8%). There were no statistically significant associations between EXPTB and sex, hypertension, diabetes, undernourishment, HIV sero-status, chronic renal insufficiency, immigration, and time.
Between the years of 1997 and 2011 there was a relatively stable incidence of EXPTB in Aguascalientes as well as Mexico’s national average. There also appeared to be no relationship between either form of TB and immigrations rates. There was a statistically significant positive association between PTB and obesity. Specifically, while 26.9% of participants possessed a BMI greater than or equal to 30 kg/m², there were no subjects with PTB who fit those criteria. There were statistically significant negative associations between EXPTB and the following variables: age, smoking, and alcohol dependence. With respect to age, those with EXPTB tended to be younger than those with PTB (mean ages were 31.6 years and 53.5 years, respectively). However, the ages of those with certain forms of EXPTB, namely pleural and abdominal forms, tended to be older and did not differ significantly from the ages of those with PTB. Thus, the results suggest that only certain forms of EXPTB, such as cutaneous, lymphatic, and renal/genitourinary are inversely associated with age. No case among those with EXPTB had a history of smoking or alcoholism while 21% and 12% of those with PTB did, respectively. Additionally, those with EXPTB experienced a shorter lag time between the onset of symptoms and clinical diagnosis in contrast to those with PTB (median lag time was 4.5 months vs. 3.0 months, respectively.). Although the difference in lag time was not statistically significant, it does suggest that the scope of PTB may be misrepresented at any given point of time. However, this is not believed to affect case control studies, which generally oversample cases from the onset (Baker et al., 2011). While no relationship between EXPTB and elevation was evident ($r = -0.76, p<0.41$), there appeared to an inverse association between the incidence of PTB and elevation ($r = -0.76, p<0.001$) (see Figure below).

With regard to limitations, there are two notable sources of information bias affecting this study. For one, smoking and drug addiction are measured dichotomously, which ignores the range
of variation that exists within the variable and results in residual confounding as the effects of each level of the variable are mixed (Balasubramanian et al., 1994). For instance, all participants were labeled as either smokers/ex-smokers or those who have never smoked; however, this measurement method does not distinguish between individuals who smoked for varying amounts of time (e.g., 1 year, 2 years, 3 years, etc.) and instead collapses them into a single group. Consequently, the varying health effects that accompany the varying durations of the exposure are blended together, being forced into two groups, and the resultant effect estimate may be overestimated, underestimated, nullified, or falsely significant depending on the magnitude of the true effect and the unique composition of the sample’s exposure status (Balasubramanian et al., 1994). The second source of error relates to the misclassification of the subjects’ outcome status. Although four of the study participants displayed signs and symptoms congruent with EXPTB, they also tested positive for AFB in their sputum, which is indicative of PTB. In other words, despite co-infection with both forms of TB, these subjects were singularly categorized as being affected by EXPTB without explanation. In turn, the effect estimates of the various independent variables derived from the analyses related to the demographic and health-related characteristics may have been biased towards significance (Balasubramanian et al., 1994). One should note, however, that the analyses related to immigration, elevation and time would be minimally affected if at all by such misclassification.

Care must be taken whilst interpreting the results of this study, particularly the results regarding demographic and other health-related characteristics. First, all cases of TB, whether EXPTB or PTB, are active cases, thus the results do not necessarily apply to the large fraction of the population in which latent infections exist. Albeit a fraction of active infections begin as latent infections, systematic differences in the demographic and/or other health-related characteristics of
those who progress to active infection could prevent generalization to those with latent TB. Secondly, generalization of the findings likely applies to residents of Aguascalientes, Mexico that have access to healthcare and not extremely poor or rural populations who are less likely to receive medical attention. Third, this investigation is characteristically descriptive, being comprised entirely of bivariate analyses and lacking a multivariate model, which describes the effect of a particular independent variable while controlling for the influence of others (Bandura, 1991; Bishai, 1998). Consequently, the reported associations may be distorted by other confounding or concomitant variables (Bandura, 1991). Finally, this study compares those with EXPTB to those diagnosed with PTB – not the general population. Thus, one should interpret the various risk and protective factors as variables that increase or decrease the risk of EXPTB among all those experiencing active TB infections; the study does not necessarily reflect factors that affect the risk of the general population acquiring EXPTB.

With respect to the analyses of EXPTB and PTB incidence against immigration and elevation, other possible regional confounders that may explain variation in the incidence of the disease are not included in the bivariate models. In addition, the rates are statewide estimates of TB incidence over a broad geographic region, and the average elevation of each state is a summarization of a range of elevations at which cases of EXPTB and PTB occur. Therefore, generalization of this set of results must not be applied narrower than the state level (Baker, 2011), and one should bear in mind mean elevation measures can be misleading. For example, an excess of TB cases may be concentrated in the lowlands of a state that is exceptionally mountainous, which would falsely associate a high TB incidence with higher elevation. Despite such limitations, a clear, inverse, linear association exists between PTB and elevation among 32 states, which suggests that higher elevations may decrease the risk of broad populations, either directly or
indirectly, and further research is necessary to elucidate this relationship by specifying the scale of this study to a municipal or individual level and/or including possible confounders in a multivariate analysis.

Effect of altitude on the frequency of pulmonary TB by Vargas, Furuya, and Perez-Guzman (2004), this state-level, ecological study investigated the relationship between altitude and pulmonary Tuberculosis (PTB) incidence among 32 Mexican states between the years of 1998 and 2002. Annual notification rates of TB incidence were taken from the Ministry of Health through the General Directory of Epidemiology, which is a database containing national statistics for all health services provided by all health institutions in Mexico. In order to control for the influence of other possible confounders, summary data regarding living conditions as well as the economic, the medical, the demographic, the geographic, and the educational status of each state was obtained through a combination of sources including the aforementioned source; the 2000 Population and Housing Census; the Instituto Nacional de Estadisticas, Geografia e Informatica; and the Consejo Nacional de la Poblacion.

Specifically, the following variables were included: the percentage of the working population with less than 2 minimum salaries; the diabetes mellitus incidence rate; the percentage of the population with social security; population density; the percentage of the rural population; the percentage of immigrants; the percentage of the population 65 years of age or older; the percentage of single-room housing; and the state-level marginalization index, which is a measure of poverty derived by weighting the sum of nine variables including the percentage of the illiterate population; the percentage of the population with less than a full, primary education; the percentage of occupants of houses without drainage or an exclusive toilet; the percentage of the population lacking access to electricity; the percentage of the population enduring crowded
conditions; the percentage of the population with earth floors; the percentage of the population living in communities with less than 5,000 members; and the percentage of the working population with less than 2 minimum salaries.

The incidence rate of PTB in each state was derived by averaging all of the PTB rates stratified by age and sex between the years of 1998 and 2002. Similarly, diabetes was taken as the average diabetes mellitus incidence rate of each state within the same time frame. The altitude of each state, which was measured as the meters above sea level, was calculated by averaging the altitudes of the counties therein, each one weighted by the county’s population size. The relationship between altitude and PTB incidence was examined in SPSS 10.0 and Prophet 5.0 utilizing two statistical analyses: Pearson’s correlation coefficient and multiple linear regression analysis. Forward stepwise procedures were implemented in order to specify the regression model.

According to the results of the correlation analysis, only altitude (r = -0.74, p < 0.0001) was statistically significant, and the percentage of single-room housing (r = 0.31, p < 0.0001) was marginally significant. Comparably, altitude was the only statistically significant variable in the final regression model, exhibiting an inverse relationship with PTB incidence.

This investigation possesses several limitations which keep the results suggestive and in need of further investigation. First, several independent variables and most notably the dependent variable are computed as averages over a 4-year time span, and valuable information regarding the existing variation within a variable is generally lost when averages are calculated and used as inputs in multivariate models (Cogo, 2011). As opposed to averaging rates across an entire time interval and inputting them into a single multivariate model, single measures for each year could be taken for each variable and inserted into multiple multivariate models, one for each year. As a result, multiple effect estimates would be generated for each variable and utilized to test a set of
null hypotheses $H_0: \mu_i = 0$ in a series of one mean $t$-tests where $\mu_i$ is the true effect of each independent variable and the point estimate is the average of the effect estimates from each multivariate model (Crosby & Noar, 2011). As a result, the resolution of the data would be increased, finer effect estimates would be obtained and more accurate conclusions could be drawn about the population parameters. Second, although averages related to PTB incidence, altitude and diabetes are computed, there is no mention as to whether or not other variables are averaged in the same consistent fashion.

Third, a multiple linear regression model is arguably inappropriate in light of other more robust models, such as Poisson regression. Linear regression is theoretically most suitable when the dependent variable is truly continuous, that is, its values can assume any real number within a given range (Bishai et al., 1998); Poisson regression, on the other hand, is best suited for dependent variables whose values are discrete integers (Cuzick & Edwards, 1990). Since incidence rates are proportions of count data comprised of discrete integers, their discrete probability distributions are best described by the Poisson probability distribution underlying the Poisson regression model. Given the utilization of a linear model, no review of the model’s fundamental assumptions is made. Linear models yield valid effect estimates under a narrow set of conditions among which the following are chief: the distribution of residuals must be independent of one another and normally distributed possessing a mean of zero and a constant variance throughout the domain of the data (Bishai et al., 1998). Additionally, the authors did not appear to assess the model for multicollinearity or possible interactions between independent variables. The questionability of the results is further heightened by the lack of any table or record of point estimates generated by the final regression model generated to evaluate the hypothesis.
The ecological nature of this investigation reduces the generalization of the results to the scale at which the study was performed (Fleiss, Levin, & Paik, 2003), which in this case is the state level. In other words, individuals do not necessarily incur an increased risk from residing at higher altitudes; rather, states at higher altitudes are more likely to experience higher incidence rates of PTB. Also, one should note an important limitation of ecological studies: they generally utilize indicators or summary measures for large areas (Baker et al., 2011). In doing so, the correlation between the dependent and independent variables can become obscured if there exists relatively small regions within each area that contribute significantly to the dependent variable, and this is perhaps one explanation for lack of any correlation between PTB incidence and variables that are known to be related, such as socioeconomic status, in this study. For example, most of a state’s PTB cases may have come from a small, impoverished area while the rest of the state may have been relatively wealthy. As such the coarseness of the study’s scale would fail to adequately link PTB cases to impoverished conditions, as the larger or more populated wealthy areas would elevate the coincident socioeconomic measure.

Despite the limitations, there appears to be a clear, linear trend between PTB incidence and altitude (see Figure below). The authors of this study posit that the association may be due to the predisposition of *M. tuberculosis* to higher oxygen concentrations based on the results of in vitro and animal studies (Goldhaber-Fiebert, Jeon, & Cohen, 2011; Goodwin & Des Prez, 1983; Goswami et al., 2012; Greenland & Morgenstern, 1989). As elevation increases the relative aveolar oxygen pressure in the lungs decreases, and this is believed to lead to a less favorable environment for bacterial multiplication (Blair & Taylor, 2007). Future ecological studies of this kind may be able to obtain more accurate effect estimates by decreasing the scale of the study from state-level
to county-level, utilizing well suited models for count data such as the Poisson, and compiling the
effect estimates of multiple multivariate models over multiple time intervals.

The ecological study cited by Tanrikulu, Acemoglu, Palanci, & Dagli (2008) investigated
the association between TB incidence and altitude as well as a number of socioeconomic variables
among 56 cities in Turkey during a 7-year time period between 1999 and 2005. A total of 81 cities
were randomly selected from seven regions in Turkey, including the Black Sea region, East
Anatolia, Middle Anatolia, Southeast Anatolia, the Aegean, and the Mediterranean. Data regarding
TB incidence for each city was obtained from the Ministry of Health of Turkey to whom mandatory
reports of TB cases are due. All cases possessed at least two of the following clinical features:
weakness, fever, sputum, night sweats, weight loss, and at least one of the following: positive
culture, positive sputum for acid-fast bacilli (AFB), histological findings or radiological evidence.
Due to incomplete data records 25 studies were removed from the analysis. TB incidence was
measured as the average incidence occurring over the 7-year time frame and then roughly
dichotomized about the mean at 23.8 cases per 100,000 populations. Similarly, data regarding a
number of socioeconomic variables, including the urbanization rate, the number of people living
in the same household, the literacy rate, the number of doctors per 10,000 populations, the infant
mortality rate, the fertility rate and possession of a social insurance card, which were obtained
from the state planning organization’s website, were dichotomized about their respective means.
Univariate, bivariate, and multivariate analyses were conducted in SPSS 10.0 utilizing Chi-square
tests, Pearson’s correlation coefficient, and logistic regression.

There was a statistically significant positive association between the TB incidence and
population density and possession of a social insurance card. Specifically, the multivariate model
demonstrated that the odds of a city developing a TB incidence greater than 24 cases per 100,000 population with a population density greater than 80 people/km$^2$ was 4.18 times greater than the odds of the outcome occurring with population density less than that (95%CI 2.30, 7.60). These results were also confirmed by the correlation analysis ($r = 0.51, p < 0.0001$).

Also, the odds of a city developing a TB incidence greater than 24 cases per 100,000 population with a percentage of the population owning a social insurance card greater than 19.8% was 2.28 times greater than the odds of the outcome occurring with percentage of the population owning a social insurance card less than that (95%CI 1.26, 4.15). The correlation analysis yielded a statistically significant inverse association between TB incidence and altitude ($r = -0.57, P < 0.0001$) (see Figure below). This relationship was duplicated in the multivariate analysis in which the odds of a city developing a TB incidence greater than 24 cases per 100,000 population at an altitude below 750 meters was 3.28 times greater than the odds of the city experiencing the outcome above an altitude of 750 meters (95%CI 1.83, 5.88).

Similarly, the multivariate model demonstrated a significant inverse relationship between the outcome and annual income as the odds of a city developing a TB incidence greater than 24 cases per 100,000 population with an annual income per person less than 1,400 NTL was 3.16 times greater than the odds of the outcome occurring with an annual income less than that (95%CI 1.41, 7.07). No associations were detected between the outcome and the following variables: urbanization rate, household size, literacy rate, infant mortality rate, number of doctors per 10,000 population, and fertility rate.
With regard to limitations, the results of the study may have been biased if the 26 studies that were removed were systematically different from those that remained. Given that studies were removed due to insufficient data records, one may reasonably assume that the quality of healthcare in such areas is relatively poor, thus the effect estimates may only be reflective of those populations, which have access to at least moderate quality medical services. One should note that although the dependent variable is a summary measure of TB incidence occurring over the study time period, many of the dependent variables are singular rates taken from a population census in the year 2000. Therefore, there is an implicit assumption that all rates represented by the independent variables remained constant between 1999 and 2005. Since this is unlikely, the data collected on each city is not entirely coincident and in turn the results may be biased.

Logistic regression is but one of a handful of generalized linear models that must follow a unique set of assumptions in order to yield valid effect estimates. Among the most important are the following: the log of the odds of the outcome must be a linear combination of each independent variable and the true conditional probabilities are a logistic function of each independent variable (Horsburgh et al., 2010). Also, each observation must be independent of all others.

Arguably, a logistic regression model may be an inappropriate choice for this study in contrast to a Poisson model. Logistic regression is theoretically most suitable when the dependent variable is truly dichotomous, that is, its values can assume only assume one of two nominal values (Horsburgh et al., 2010); Poisson regression, on the other hand, is best suited for dependent variables whose values are discrete integers (Cuzick & Edwards, 1990). Since incidence rates are proportions of count data comprised of discrete integers, their discrete probability distributions are best described by the Poisson probability distribution underlying the Poisson regression model (Cuzick & Edwards, 1990). Furthermore, when ratio or interval variables are collapsed into
categorical variables, existing variation within the variable is lost and mixing of effects at each level of the variable takes place, resulting in residual confounding (Jeon, Harries, & Baker, 2010). However, albeit information is lost in the process of categorizing continuous variables, it can be helpful in terms of clearly depicting a general change in risk when moving from one level of an exposure to another, especially in circumstances where the effect of the independent variable is hardly perceptible, given the cut-off point is aptly chosen. For this reason, although the exact effect estimate of altitude and other covariates is likely not precise, one may reasonably assume they are indicative of general positive or negative relationships that exist with the dependent variable. However, the authors did not appear to utilize a model specification strategy in order specify another vital assumption of logistic regression, that is, no extraneous or unnecessary variables are included in the model (Horsburgh et al., 2010). There are many model specification strategies that exist, but they function similarly to reduce the amount of statistical noise and increase the precision of effect estimates by eliminating variables that are not required to explanation variation in the dependent variable. Given that many of the covariates in this study are various proxies of the same concept, which is socioeconomic status, the insignificance of some variables, such as literacy rate and infant mortality, is expected as other variables, such as the percentage of the population with a social insurance card, adequately explain enough variation. Retaining extraneous variables in a multivariate model has the consequence of reducing precision, inflating the confidence intervals, and yielding unstable effect estimates. Furthermore, retaining extraneous variables that are proxies of the same concept risks will violate the fundamental assumption of the absence of multicollinearity. Understanding the whole of the limitations, the effect estimates provided by these results should be interpreted cautiously and seen as suggestive of possible trends that require finer examination.
Maciel et al. (2010) cites that the purpose of this 4-year, retrospective study was to investigate spatial patterns of pulmonary TB incidence, to examine the relationship between TB and socio-economic status, and to identify geographic areas with elevated risk of the disease within the city of Vitoria, Espirito Santo, Brazil during 2002-2006. Residential addresses of TB cases were geocoded in and surveillance records were reviewed in order to minimize misclassification of disease status. The socio-economic status for 78 neighborhoods was determined by computing its index of quality of urban municipality (IQU), a single score which encompasses the quality of education, environment, housing and finances in a given area. Census data from the year 2000 was utilized in combination with case data in order to determine incidence rates. Smoothed empirical Bayes estimates, model-predicted incidence rates, and spatial clustering statistics, such as Anselin’s local indicators of spatial association (LISA) and Getis-Ord Gi* statistics were used to compare spatial patterns of TB incidence. The association between socioeconomic status and TB incidence was investigated using fitted spatial Poisson models. Strong spatial autocorrelation among incidence rates (0.399, P < 0.0001) was indicated by Moran’s I, and LISA and Gi* statistics identified four areas with elevated incidence rates, all of which were low-income, overcrowded slum communities. Smoothed spatial empirical Bayes estimates demonstrated that two of these areas ranged from 70 to 90 cases per 100,000 populations, while the other two ranged from 40 to 70 cases per 100,000 populations. There was a statistically significant curvilinear relationship between TB incidence and socioeconomic (p = 0.02). This study did not include molecular genotyping, as many studies often do, in order to distinguish between recent transmission and long-standing, latent infection.

Investigation of geospatial hotspots for the occurrence of TB in India by Tiwari, Adhikari, Tewari, and Kandpal (2006) cites the purpose of this study was to utilize GIS with the spatial scan
statistic developed by Martin Kulldorff in order to detect statistically significant hotspots, or elevated clusters, of TB in Uttaranchal, India during 2003-2005. With a maximum spatial cluster size less than or equal to 50% of the total population, two clusters of high TB prevalence were located at DTC, Alomora and Chaukhutiya with overall relative risks (RR) of 4.042 (p = 0.001) and 1.648 (p=0.001) within the clusters, respectively. Subsequent space-time analyses revealed that these clusters existed during 2004 and 2003, respectively. In order to detect smaller clusters, the maximum spatial cluster size was adjusted to less than or equal to 25% of the total population, and one additional cluster was detected at Dhauladevi with a RR of 1.366 (p=0.007). Although this study quantifies and clarifies the burden of TB in the Uttaranchal region, which is advantageous for surveillance purposes, it neglects to investigate factors associated with these geographic clusters.

Touray et al. (2010) cites the purpose of this prospective study was to examine the spatial patterns of TB occurrence in Greater Banjul Are (GBA) of the Gambia; where over 80% of Gambian TB cases occur, during March 2, 2007 to February 29, 2008. Demographic information, clinical information and medical history were collected by way of questionnaire and medical records from patients registered at local chest clinics. Residential addresses were geo-referenced, and a spatial scan statistic was utilized to identify spatial and space-time clusters. First, the maximum spatial cluster size set to 50% of the population at risk, and then it was reduced to 25% for subsequent analyses. Statistically significant high rate spatial clusters were detected in two communities with high population density: one spanned Brufut, Sukuta, Sakkuta Sanchaba and the Bijilo areas, and another was detected among the Dippa Kunda and Serekenda settlements. Space-time spatial analysis indicated that September of 2007 was the significant time period for the former cluster. Furthermore, a cluster of low TB incidence rates was identified among the
settlements of Wellingara, Sinchu Baliya, Sinchu Sorry and Kunkujang. Similar as before, although this study quantifies and clarifies the burden of TB in the Uttaranchal region, which is advantageous for surveillance purposes, it neglects to investigate factors associated with these geographic clusters.

Using GIS technology to identify areas of tuberculosis transmission and incidence by Moonan et al. (2004) in an effort to identify high-risk populations for targeted screening and treatment of tuberculosis, this cross-sectional study examined socio-demographic factors as well as spatial characteristics associated with geographic clusters of incident cases infected with identical strains of *Mycobacterium tuberculosis*. Data regarding culture positive residents of Tarrant County, Texas was collected from the county health department during the period between January 1, 1993 and December 31, 2000. DNA fingerprinting was performed on tuberculosis isolates at the Texas Department of Health Mycobacteriology Laboratory by way of clinical isolate IS6110-based RFLP and spoligotyping analyses. Subsequent interviews collected data on current and past employment, alcohol and illicit drug use, housing, incarceration history, psychiatric history and sexual orientation, and residential addresses, including zip code, were geocoded in ArcView 4.0. Logistic regression analyses determined that the mean age among genotypically clustered cases ($\mu=44.1; 95\%$CI 27.5, 60.7) was significantly lower than those of unique strains ($\mu=48.5; 95\%$CI 30.9, 66.1). Regarding race, the odds of a clustered strain were significantly greater among African Americans (OR=2.7; 95\%CI 1.8, 4.0) whereas the odds of a unique strain significantly greater among Asians (OR=3.9; 95\%CI 2.3, 6.0) and Hispanics (OR=1.9; 95\%CI 1.2, 2.9) with respect to Caucasians. Additionally, males (OR=1.9; 95\%CI 2.1, 2.8), those born in the U.S. (OR=5.3 95\%CI 3.5, 7.9) and those with a history of homelessness (OR=12.4; 95\%CI 2.9, 52.1) were positively correlated with clustering. Spatial analyses highlighted a distinct geographic
distribution of disease and indicated that zip codes with elevated incidence rates were characterized by high unemployment, low socioeconomic status, drug use, homelessness and poor quality housing conditions. Furthermore, Geno-typically clustered cases exhibited a heterogeneous geographic distribution, often occurring in zip codes with higher incidence rates. Incidence rates were determined, in part, by taking the average population number per zip code between the 1990 and 2000 US Census. Geographic clusters were defined as at least two cases with the same strain.

Outcome of Targeted TB Screening Based on Genotyping & Location by Moonan et al. (2006) cites the purpose of this study was to increase the case detection rate of those with recent yet latent transmission of tuberculosis through the utilization of a GIS-based screening approach targeting geographic areas where genotypic clusters of Mycobacterium tuberculosis infections existed. In Tarrant County, Texas, community-based organizations established screening locations in three zip codes that exhibited the highest incidence of TB and genotypic clusters of tuberculosis isolates. These zip codes were identified in a previous study utilizing methods outlined by Moonan et al. (2004). Demographic information, housing and employment history were collected on individuals at each screening. The case detection rate among strategically located screening sites (1 active case per 83 screened and 1 latent case per 5 screened) far exceeded expected rates from non-targeted screening locations in the county, which has a TB incidence of 5.7 per 100,000 populations. There is a lack of an appropriate control group, that is, the results of screening locations in zip codes with high incidence and low molecular clustering were not compared to the results of screening locations in zip codes with low incidence and low molecular clustering.

GIS-based screening for TB, HIV and Syphilis - Cross Sectional Study by Goswami et al. (2012) cites the purpose of this study was to test the feasibility of a GIS-based screening strategy aimed at TB, HIV and Syphilis and compare its case detection rates with. The residential addresses
of incident cases between January 1, 2005 and December 31, 2007 in Wake County, North Carolina were geocoded in ArcMap 9.3, and ten screening sites were established in those areas with a higher density of cases, defined as 10 cases or more per square mile. For comparison, STD and TB clinics located at the Wake County Health Department, which is located outside of any designated “hot spots,” were designated as controls. From February 6, 2009 to March 11, 2011, residents were offered incentives, such as snacks, beverages and $5 grocery gift cards, in order to encourage participation in screenings for TB, HIV and syphilis. Each test was performed with a single blood draw, blood was stored in an incubator and samples were processed by the Duke Clinical Immunology Laboratory. Questionnaires gathered information regarding demographics, comorbidities, and perceived risk of infection and utilization of health care services. Chi-square analyses were used to compare case detection rates between screening facilities. HIV prevalence was significantly greater among those screened at a community-based sites (3%; 95%CI 4%, 6.5%) versus the Wake County Health Department (0.4%; 95%CI 0.3%, 0.5%), p<0.001. Similarly, case detection rates of TB were significantly greater among community-based sites (15%; 95%CI 11%, 21.7%) compared with the health department (6%; 95%CI 5.6%, 6.6%), p<0.001. Molecular & Geographic Patterns of TB after 15 years in Baltimore MD by Bishai et al.(1998) cites the purpose of this study was to determine the patterns of tuberculosis transmission, with particular attention to spatial clustering of tuberculosis strains, within Baltimore, Maryland remaining after 15 years of successful implementation of community-based directly observed therapy (DOT), which contributed to an approximate 58% decrease in the tuberculosis rate over the span of 14 years. During a citywide, prospective study conducted between January 1994 and June 1996, 182 cases of culture-positive tuberculosis were identified through the Baltimore City Health Department and interviewed regarding illicit drug use, homelessness and employment history. Traditional contact
investigations were also conducted in order to identify anyone sharing a minimum of four hours of airspace with a primary case. DNA fingerprinting of *Mycobacterium tuberculosis* isolates was performed by means of IS6110-based RFLP and 2 DNA probes. Statistical methods included Fischer’s Exact Test, multivariate regression utilizing the least squares method and spatial clusters were tested using the approach of Cuzik and Edwards as well as Simes modified Bonferroni method (Maciel et al., 2010). 84 (64%) of cases exhibited molecular clustering. 58 (32%) cases were recently transmitted, and epidemiologic evidence of recent contact was present in 20 (24%) cases with clustered DNA fingerprints. Spatial analyses demonstrated significant geographic clustering of 20 cases occurring in areas of low socioeconomic status and elevated drug use.

Previous Studies Applying GIS to TB Research (Inside of Florida) cite the previous investigations conducted in Florida have examined factors associated with TB occurrence, including factors associated with genotypic clusters. Such factors have included homelessness, incarceration, substance abuse with 1 year of TB diagnosis, HIV infection (Mason, Dobard, Zhang, & Nelson, 2004). However, to the author’s knowledge, there has been little or no scientific investigation of the spatial patterns of TB incidence that exist within the state of Florida.
CHAPTER 3
RESULT AND ASSESSMENT

The results presented initially with the CLUSTER procedure in SAS hierarchically aggregated the geosampled, georeferenceable, endemic TB field and remote-specified, explanatory, endemic, transmission-oriented, predictive, field-operationizable, time-series, risk-based, county-level observations. PROC CLUSTER computed all Euclidean distances in the dataset based on the flexible-beta method. PROC CLUSTER then generated the number of clusters in the geosampled, county-level TB positive population. PROC CLUSTER also created an output dataset which used the TREE procedure in SAS to draw a diagram of the cluster hierarchy. To obtain the five-cluster solution, we first used PROC CLUSTER with the OUTTREE= option, and then employed the output, explanatory, endemic, transmission-oriented, field and remote-specified dataset as the input dataset to the TREE procedure. Within PROC TREE, NCLUSTER specified the number of clusters based on the geosampled, georeferenceable, time-series, elucidative county-level parameterizable categorical and continuous covariates and the OUT= options to obtain the final solution and draw a tree diagram. Since we considered all the geosampled, georeferenceable, iterative interpolative, field-operationizable, data points to be equally important, we used the STD option in PROC CLUSTER to standardize the endemic, transmission-oriented, expository geosampled variables to mean 0 and standard deviation 1. We removed the outliers before using PROC CLUSTER with the STD option. The STDIZE procedure provides additional methods for standardizing variables and imputing missing values (www.sas.edu).
The relationship between county-level prevalence and each explanatory, individual, potential, time-series, endemic, transmission-oriented, TB, transmission-oriented, geosampled regressors was investigated by employing single variable regression analysis in PROC NL MIXED. The first line of the code began the PROC NL MIXED command. The second line specified the fixed portion of each clinical, field and remote-specified, endemic, transmission oriented, epidemiological, risk model, [i.e., the model without the random intercept, value (i.e., xb)]. The third line of code created a value (i.e., rand) that was equal to the fixed part of the model (xb) plus a random intercept term $u$. The model statement specified that the parameterizable covariate, geo-spatiotemporal estimators were distributed (\sim) normally with a mean of $xb$ and variance $s^2$. The random statement defined the random effect $u$ whilst quantitating the normally, distributed, operationizable, time series data with mean zero and a variance term. In so doing, $s^2u$ was optimally solved. The level 2 units were identified by subject = $id$ in PROC NL MIXED.

Importantly, the last two lines of the command in PROC NL MIXED were predictive statements. While traditionally only a single set of predicted values are created when constructing time-series, clinical, field and/or remote-specified, vector, iterative interpolative, endemic TB, elucidative, geosampled, georeferenceable, forecasting epidemiological, risk models, we generated two. The predict statement in SAS rendered the clinical, field and remote-specified, explanatory, time-series, predicted values for the fixed portion of the model. The model identified $xb$, and output a dataset called output-fixed. The second predict statement generated the time-series, regressed, endemic, county-level TB, transmission-oriented, predicted values that included the estimate of the random intercept in addition to the quantitated, fixed portion, randomized estimates.
The regression line employed \((y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)\) to generate a pseudo \(R^2\) value where the first term was the total variation in the response \(y\) (county-level TB prevalence) and the second term was the variation in mean response based on the geosampled, asymptotical, normalized, parameterizable covariate estimators. The third term was the rendered residually forecasted, elucidative regressed clinical, field and/or remote-specified, geosampled, endemic, transmission-oriented, derivative values in the operationizable, time-series, iteratively interpolative, endemic transmission-oriented risk model derivatives. Squaring each of these terms and adding over all of the geosampled, georeferenceable, county-level TB, observations generated the equation \(\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2\). This equation was written as SST = SSM + SSE, where SS was notation for sum of squares and \(T, M,\) and \(E\) were the notation for total quantized model error estimates. The square of the sample correlation was then equal to the ratio of the estimates while the sum of squares was related to the total sum of squares: \(r^2 = \frac{SSM}{SST}\). This formalized the interpretation of \(R^2\) for explaining the fraction of variability in the geosampled, county-level epidemiological data explained by the regression model. The sample variance \(s_y^2\) was equal to \(\sum \frac{(y_i - \bar{y})^2}{n-1}\), which in turn was equal to the SST/df, the total sum of squares divided by the total df. A regression equation was constructed by employing the mean square model (i.e., MSM) = \(\sum \frac{(\hat{y}_i - \bar{y})^2}{l}\), which was equal to the SSM/df. The corresponding MSE was \(\sum \frac{(y_i - \hat{y}_i)^2}{n-2}\) which was determined to be equal to SSE/df and the quantitated, time-series, operationizable, clinical, field and remote-specified, county-level, TB, endemic, transmission-oriented, explicatory, georeferenceable estimate of the variance about the regression line (i.e., \(\sigma^2\)). The MSE is an estimate of \(\sigma^2\) for determining whether or not the null hypothesis is true [10].
Robustly, parsimoniously, quantizing, the geo-spatiotemporal, geosampled, operationizable, clinical, field and remote-specified, transmission oriented, explanatory, stochastically/deterministically, iteratively interpolative, endemic, county-level TB, predictor variables, \((p)\) a DFM, was generated which we noted was equal to \(p\) and the error degrees of freedom (DFE). This product was also equal to \((n - p - 1)\), and the total degrees of freedom (DFT) which was subsequently equal to \((n - 1)\) The sum of DFM and DFE was determined. The relationship between the mean of the response variable (i.e., county-level prevalence count) and the level of the geosampled, explantorial, parameterizable, covariate coefficients in the regression equation were assumed to be approximately linear (i.e., straight line). The corresponding table generated classified each time-series, clinical, field and remote geosampled, asymptotical, unbiased, covariate estimator in SAS (see Table 1).

Table 1: The time series regressed endemic TB regression-based model parameter estimates

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>(\sum (y_j - \hat{y}_j)^2)</td>
<td>SSM/DFM</td>
</tr>
<tr>
<td>Error</td>
<td>(\sum (y_j - \hat{y}_j)^2)</td>
<td>SSE/DFE</td>
</tr>
<tr>
<td>Total</td>
<td>(\sum (y_j - \bar{y})^2)</td>
<td>SST/DFT</td>
</tr>
</tbody>
</table>

In the multiple regression analyses, the test statistic \(MSM/MSE\) had an \(F(p, n - p - 1)\) distribution. The null hypothesis was \(\beta_1 = \beta_2 = ... = \beta_p = 0\), and the alternative hypothesis was evaluated by encompassing the county-level, endemic, transmission-oriented, predictive, epidemiological, risk parameters \(\beta_j \neq 0, j = 1, 2, ..., p\). The \(F\) test did not indicate which of the
parameters $\beta_j \neq 0$ nor, which was not equal to zero only that at least one of them was linearly related to the response variable. The ratio $\text{SSM}/\text{SST} = R^2$ (i.e., squared multiple correlation coefficient) was thereafter the proportion of the variation in the response variable that was explained by the county-level TB data. The square root of $R^2$ (i.e., the multiple correlation coefficient) was the correlation between the explanatorial, time-series, empirical observations (i.e., $y_i$) and the fitted values (i.e., $\hat{y}_i$). Additionally, from the sampling distribution generated from the $t$ parameters, the probability of obtaining an $F$ was calculated. There were only two means to compare, the $t$-test and the $F$-test, which coincidentally were equivalent. The relation between ANOVA and $t$ was then given by $F = t^2$. Thereafter, significant differences by ANOVA were noted for the quantitated mean numbers of explicative, clinical, field and remote specified, endemic, transmission-oriented, operationizable, time series, iteratively interpolative, asymptotical, normalized, georeferenceable, county-level, data, feature attributes captured throughout the sampling frame ($F = 44.7, DF = 1$).

A Poisson regression analyses was constructed in PROC NL MIXED to determine the relationship between the endemic TB, count data and county-level, prevalence. The Poisson models were built by employing the, time-series, explanatory, clinical, field and remote, multivariate, endemic, transmission-oriented, county-level TB, parameterizable, covariate coefficients. A negative binomial regression with a non-homogeneous, gamma distributed mean had to be employed, since examination of the data indicated that over dispersion was a significant problem in the Poisson model. The Poisson distribution is a special case of the negative binomial distribution where the mean approximates the standard deviation (Crosby & Noar, 2011). We assumed that the log of the mean $\mu$ was a linear function of independent variables, [i.e., $\log(\mu) = \text{intercept} + b1^*X1 + b2^*X2 + \ldots + b3^*Xm$] in the county-level TB model which implied that $\mu$
was the exponential function of the independent variables when 
\[ \mu = \exp(\text{intercept} + b_1^* X_1 + b_2^* X_2 + \ldots + b_3^* X_m). \]
Therefore, instead of assuming that the distribution of the geosampled, georeferenceable, parameter estimates (i.e., \( Y \)) was Poisson, we were able to assume that \( Y \) had a negative binomial distribution. We relaxed the assumption about equality of mean and variance (i.e., Poisson distribution property), since the variance of negative binomial was equal to \( \mu + k\mu^2 \), where \( k > 0 \) was a dispersion parameter. The ML method was then used to estimate \( k \), as well as the county-level parameters for quantitating \( \log(\mu) \). For the negative binomial distribution; the variance was equal to the mean \( + k \cdot \text{mean}^2 \) (i.e., \( k > 0 \)) as the negative binomial distribution reduced to Poisson when \( k = 0 \).

In the regression analyses, of the county-level, georeferenceable, explanatory, endemic, TB, transmission-oriented, iteratively interpolative, asymptotical, normalized, operationizable, time-series, data the null hypothesis was: \( H_0 : k = 0 \) and the alternative hypothesis was: \( H_a : k > 0 \). We recorded the log-likelihood (i.e., LL) for the models. We employed the likelihood ratio (LR) test to compute the LR statistic using \(-2(\text{LL})\) (Poisson) and the LL (i.e., negative binomial). The asymptotic distribution of the LR statistic had a probability mass of one half at zero and one half - Chi-square distribution with 1 df. To test the null hypothesis at the significance level \( \alpha \), the critical value of Chi-square distribution corresponding to significance level \( 2\alpha \), whereby there was a rejection of \( H_0 \) if LR statistic \( > \chi^2(1-2\alpha,1\ \text{df}) \). The log of the mean, \( \mu \) was generated using a linear function of independent variables whereby, \( \log(\mu) = \text{intercept} + b_1^* X_1 + b_2^* X_2 + \ldots + b_3^* X_m \), in the explanatively regressed, clinical, field and remote- specified, time-series, endemic, transmission-oriented, county-level, predictive, operationizable, iteratively interpolative epidemiological, risk model which implied that \( \mu \) was the exponential function of the independent variables when
The SAS model data was then log-transformed and loaded as:

\[ \mu = \exp (\text{intercept} + b_1 \cdot X_1 + b_2 \cdot X_2 + \ldots + b_3 \cdot X_m) \]

libname data "F:\TB Project\Datasets";
run;

PROCIMPORT OUT= DATA.TB2008
DATAFILE= "F:\TB Project\Original Datasets\2008 State TB Inc. idence 4-17.xlsx"
DBMS=EXCEL REPLACE;
   RANGE="Monthly Morbidity$";
   GETNAMES=YES;
   MIXED=NO;
   SCANTEXT=YES;
   USEDATE=YES;
   SCANTIME=YES;
RUN;

PROCIMPORT OUT= DATA.TB2009
DATAFILE= "F:\TB Project\Original Datasets\2009 State TB Inc. idence 4-17.xlsx"
DBMS=EXCEL REPLACE;
   RANGE="Monthly Morbidity$";
   GETNAMES=YES;
PROCIMPORT OUT= DATA.TB2010
DATAFILE= "F:\TB Project\Original Datasets\2010 State TB Inc. idence 4-17.xlsx"
DBMS=EXCEL REPLACE;
   RANGE="Monthly Morbidity$";
   GETNAMES=YES;
   MIXED=NO;
   SCANTTEXT=YES;
   USEDATE=YES;
   SCANTIME=YES;
RUN;

PROCIMPORT OUT= DATA.TB2011
DATAFILE= "F:\TB Project\Original Datasets\2011 State TB Inc. idence 4-17.xlsx"
DBMS=EXCEL REPLACE;
   RANGE="Monthly Morbidity$";
   GETNAMES=YES;
   MIXED=NO;
SCANTEXT=YES;
USEDATE=YES;
SCANTIME=YES;

RUN;

PROCIMPORT
OUT= DATA.TB2012
DATAFILE= "F:\TB Project\Original Datasets\2012 State TB Incidence 4-17.xlsx"

DBMS=EXCEL REPLACE;
RANGE="'Monthly Morbidity$'";
GETNAMES=YES;
MIXED=NO;
SCANTEXT=YES;
USEDATE=YES;
SCANTIME=YES;

RUN;

*Adding variable "YEAR" to each dataset;

data data.tb2008a;
   set data.tb2010;
   YEAR = 2010;
run;

data data.tb2011;
   set data.tb2011;
YEAR = 2009;
run;

data data.tb2012a;
    set data.tb2012;
    YEAR = 2010;
run;

data data.tb2013a;
    set data.tb2013;
    YEAR = 2013;
run;

*Merge each of the five SAS datasets into a single SAS dataset, dropped unnecessary variable, added a label to variable "YEAR," and renamed remaining variables;

data data.data1 (drop = Case_Rank
    rename = (Age_greater_than_or_equal_to_65 = Age_65_More
        American_Indian_ = Indian
        Cases_Reported_as_Residents_of_C = Jail
        Excessive_alcohol_useage__age___ = Alcohol
        Homeless__Only_Age_greater_than_ = Homeless
        Incidence_Rate__per_100_000_ = Incidence
        Injecting_Drug_Use__Age_15____ = Inject_Drug
        Native_Hawaiian_or_Other_Pacific = Islander
        Long_Care = Long_Care
        Older_Age_greater_than_or_equal_to_65 = Age_65_More
       primirial_Hawaiian = Islander
        jail = Jail
        American_Indian_ = Indian
        Cases_Reported_as_Residents_of_C = Jail
        Excessive_alcohol_useage__age___ = Alcohol
        Homeless__Only_Age_greater_than_ = Homeless
        Incidence_Rate__per_100_000_ = Incidence
        Injecting_Drug_Use__Age_15____ = Inject_Drug
        Native_Hawaiian_or_Other_Pacific = Islander
        Long_Care = Long_Care
        Older_Age_greater_than_or_equal_to_65 = Age_65_More
        primar
Non_Hispanic_Black_or_African_Am = Black
Non_Hispanic_White = White
Non_Injecting_Drug_Use__Age_15__ = NonInject_Drug
Hispanic_or_Latino = Hispanic

county = state));


label Year = "Year Cases Occurred In A Given State";

run;

proccontentsdatal=datad.data1;
run;

procprintdata=datad.data1 (firstobs=1obs=10);
run;

*Preliminary check of the data to check for completeness;

*One observation with missing data;

procmeanndata=datad.data1 nnmiss;
run;

procfreqdata=datad.data1;

table state / missing;

Run;

*Divided all relevant numerical variables by the variable "CASES"
in order to obtain the proportion of cases that belongs to each variable;

data data.data2;

    set data.data1;

    Age_15_24 = Age_15_24 / Cases;
    Age_25_44 = Age_25_44 / Cases;
    Age_45_64 = Age_45_64 / Cases;
    Age_5_14  = Age_5_14  / Cases;
    Age_65_More= Age_65_More/ Cases;
    Age_under_5= Age_under_5/ Cases;
    Alcohol   = Alcohol   / Cases;
    Asian     = Asian     / Cases ;
    Black     = Black     / Cases ;
    Foreign_Born= Foreign_Born/ Cases ;
    HIV_Positive = HIV_Positive/ Cases ;
    Hispanic     = Hispanic  / Cases ;
    Homeless     = Homeless  / Cases ;
    Indian       = Indian    / Cases ;
    Inject_Drug  = Inject_Drug/ Cases ;
    Jail         = Jail       / Cases ;
    Long_Care    = Long_Care / Cases ;
    Multiple_Races= Multiple_Races/ Cases ;
    Native_Born  = Native_Born/ Cases ;
    NonInject_Drug= NonInject_Drug/ Cases ;
White_ = White / Cases ;
run;
procmeans data=data.data2 nnmiss;
run;
procprint data=data.data2 (firstobs=1 obs=10);
run;
proccontents data=data.data2;
run;

*Transformed the varaible "CASES" by taking its natural log in order to linearize model;
data data.data3;
    set data.data2;
    Cases_T=log(Cases);
run;

*ANALYSIS;
procreg data=data.data3;
    model Cases_T = Latine Age_25_44 Age_45_64 Age_5_14 Age_65_More Age_under_5 Alcohol Asian Black Foreign_Born HIV_Positive Hispanic

Homeless Indian Inject_Drug Islander Jail Long_Care Multiple_Races

Native_Born NonInject_Drug White_ / selection=backward pcorr2;

run;

Table 2: Summary of backward elimination

<table>
<thead>
<tr>
<th>Step</th>
<th>Variable</th>
<th>Number</th>
<th>Partial</th>
<th>R-Square</th>
<th>Model R-Square</th>
<th>C(p)</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Homeless</td>
<td>21</td>
<td>0</td>
<td>0.5154</td>
<td>21.002</td>
<td>0</td>
<td>0.968</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Foreign born</td>
<td>20</td>
<td>0</td>
<td>0.5154</td>
<td>19.017</td>
<td>0.02</td>
<td>0.902</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>HIV positive</td>
<td>19</td>
<td>0</td>
<td>0.5153</td>
<td>17.036</td>
<td>0.02</td>
<td>0.889</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Black</td>
<td>18</td>
<td>0.0025</td>
<td>0.5128</td>
<td>16.239</td>
<td>1.22</td>
<td>0.271</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Long_Care</td>
<td>17</td>
<td>0.0027</td>
<td>0.5101</td>
<td>15.517</td>
<td>1.29</td>
<td>0.257</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Age_25_44</td>
<td>16</td>
<td>0.0034</td>
<td>0.5067</td>
<td>15.146</td>
<td>1.65</td>
<td>0.201</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Age_45_64</td>
<td>15</td>
<td>0.0043</td>
<td>0.5024</td>
<td>15.185</td>
<td>2.06</td>
<td>0.153</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Age_65_More</td>
<td>14</td>
<td>0.0041</td>
<td>0.4983</td>
<td>15.147</td>
<td>1.97</td>
<td>0.162</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Latin</td>
<td>13</td>
<td>0.0041</td>
<td>0.4942</td>
<td>15.12</td>
<td>1.97</td>
<td>0.162</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>NonInject_Drug</td>
<td>12</td>
<td>0.0046</td>
<td>0.4896</td>
<td>15.31</td>
<td>2.18</td>
<td>0.141</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Alcohol</td>
<td>11</td>
<td>0.0049</td>
<td>0.4847</td>
<td>15.648</td>
<td>2.32</td>
<td>0.129</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Jail</td>
<td>10</td>
<td>0.0049</td>
<td>0.4798</td>
<td>15.992</td>
<td>2.31</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

All residual estimates from the model were then evaluated in a spatial error (SE) model. Whereas the spatially lagged dependent variable model sees spatial dependence as substance, in
the sense that the $y_i$ is influenced by the value $y_j$ for ($j \neq i$), the spatial error time-series, predictive, risk, model treats uncertainty correlations primarily as a nuisance, much like how statistical approaches often treat temporal serial correlation as something to be eliminated and solely as an estimation problem. An autoregressive iteratively interpolative, georeferenceable, asymptotical, normalized, explanatory, epidemiological autoregressive TB, endemic transmission-oriented, risk model in AUTOREG quantitated the parameter estimator significance levels. This process used a variable $Y$ as a function of nearby sampled $Y$ values [i.e., an autoregressive response (AR) or spatial linear (SL) specification] and the error residuals of $Y$ as a function of nearby $Y$ residuals [i.e., an AR or SE specification]. Distance between geosampled predictors was then defined in terms of an $n$-by-$n$ geographic weights matrix, $C$, whose $c_{ij}$ values were 1 if the geosampled, endemic TB geolocations $i$ and $j$ were deemed nearby, and 0 otherwise. Adjusting this matrix by dividing each row entry by its row sum then rendered the row sums given by $C1$ which then converted the $C$ matrix to matrix $W$. The $n$-by-$1$ vector $x = [x_1 \ldots x_n]^T$ contained measurements of a quantitative variable for $n$ spatial units and $n$-by-$n$ spatial weighting matrix $W$. The formulation for the Moran’s index of spatial autocorrelation employed was: 

$$I(x) = \frac{n \sum_{(2)} w_{ij} (x_i - \bar{x}) (x_j - \bar{x})}{\sum_{(2)} w_{ij} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where $\sum_{(2)} \sum_{i=1}^{n} \sum_{j=1}^{n}$ with $i \neq j$. The values $w_{ij}$ were spatial weights stored in the symmetrical matrix $W$ [i.e., $(w_{ij} = w_{ji})$] that had a null diagonal $(w_{ii} = 0)$. The matrix was generalized to an asymmetrical matrix $W$. Matrix $W$ can be generalized by a non-symmetric matrix $W^*$ by using $W = (W^* + W^{*T})/2$ [11]. Moran’s $I$ was then parsimoniously quantitated by employing the rewritten matrix notation:

$$I(x) = \frac{n}{1^T W_1 x^T HHWHx}{1^T W_1 x^T HHx} = \frac{n}{1^T W_1 x^T HHx}{1^T Hx}$$

where $H = (I - 11^T/n)$ was an orthogonal projector verifying that $H = H^2$ (i.e., $H$ was independent). Features of matrix $W$ for analyzing the elucidative,
endemic, transmission-oriented, expositorial, geosampled, time-series, stochastically/deterministically, iteratively interpolative, parameterizable, covariate coefficients of the georeferenceable, county-level, endemic, TB data $S$ included that it was a stochastic matrix which expressed each observed value $y_i$ as a function of the average of endemic TB geolocation $i$’s nearby county-level linearly determined, high, endemic transmission, foci count variables whilst simultaneously allowing for a single geospatialized, autoregressive, explanatory, parameter $\rho$ to have a maximum value of 1.

A SAR model specification was employed to describe the explanatory, time-series, predictive, asymptotical, endemic, transmission-oriented, clinical, field and remote specified, TB endemic, model, autoregressive, variance, uncertainty estimates. A spatial filter (SF) model specification was also used to describe both Gaussian and Poisson random variables. The resulting SAR model specification took on the following form: $\mathbf{Y} = \mu(1 - \rho)\mathbf{1} + \rho \mathbf{WY} + \mathbf{\varepsilon}$, [eqn.3.1]where $\mu$ was the scalar conditional mean of $Y$, and $\mathbf{\varepsilon}$ was an $n$-by-1 error vector, whose elements were statistically independent normally randomized variates. Jacob et al. [2] employed simple scalar mixing models coupled to a Lagrangian stochastic trajectory model to calculate a range of conditional scalar statistics for three different source configurations, a uniform gradient source, a line plume and a scalar mixing layer for parsimoniously qualitatively regressing an empirical, geosampled, dataset of time-series, dependent, explanatory, clinical, field and remote-specified, county-level, endemic TB, operationizable, georeferenceable, asymptotical, unbiased, parameterizable, county-level, covariate coefficients.

The spatial covariance matrix for Equation (3.1) incorporated the georeferenceable, endemic, transmission-oriented, explanatory, stochastically/deterministically, iteratively interpolative, operationizable, time-series covariates in $E[(\mathbf{Y} - \mu\mathbf{I})(\mathbf{Y} - \mu\mathbf{I})] = \Sigma = \sum_{i,j=1}^{n} \Sigma[i,j]$.
\[ \{(I - \rho W)(I - \rho W)^{-1}\sigma^2 \], where \( E(\cdot) \) denoted the calculus of expectations, \( I \), which was parsimoniously optimally quantitated employing the \( n \)-by-\( n \) identity matrix denoting the matrix transpose operation where \( \sigma^2 \) was the error variance. A matrix for which all the column vectors are probability vectors is called transition or stochastic matrix (Anderson, 1958). A stochastic matrix for constructing a robust, elucidative, endemic, transmission-oriented, county-level TB-related, interpolatable, geosampled, georeferenceable, epidemiological, operationizable, risk model is the transition matrix for a finite Markov chain (i.e., a Markov matrix) which must have elements of the matrix that are real numbers (e.g., clinical, field and remote-sampled, time series, explanators) within a closed interval \([0, 1]\) (Anderson, 1958). Markov chain is collection of random variables (where the index runs through \(0, 1, \ldots\)) having the property that, given the present, the future may be predicted employing conditionally independent quantitated data variables of the past (Anderson, 1958). However, since a mixture of positive spatial autocorrelation PSA and negative spatial autocorrelation (NSA) was present in the county-level dataset of endemic, TB explanatory, predictive, endemic, transmission-oriented, clinical, field and remote-specified, epidemiological, risk model in AUTOREG, a more explicit representation of both effects lead to a more accurate interpretation of the empirical results. Alternately, the excluded values were set to zero, although this was done assuming the mean and variance in the explanatory, county-level, endemic TB, autoregressive, transmission-oriented, stochastically/deterministically, iteratively interpolative, operationizable, time-series, ecological, georeferenceable, risk model would have to be adjusted.

Different georeferenceable autospatial autoregressive parameters appeared in the covariance matrix, expositional, forecasting, endemic TB epidemiological, operationizable, risk model specifications which were parsimoniously quantitated using: \( \Sigma = [ (I - \rho >\text{diag} \ W) \]
where the diagonal matrix of autoregressive parameters, \(< \rho >_{\text{diag}}\), contained a minimum of two geosampled TB, transmission-oriented, clinical field and remote-specified parameters. \(\rho_+\) was employed to represent specific county-level endemic, geolocations. The pairs displayed positive spatial dependency (i.e., \(\rho_+\) for those county-level pairs displaying negative spatial dependency). For instance, by letting \(\sigma^2 = 1\) and employing a 2-by-2 regular square tessellation rendered

\[
\Sigma = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
\end{pmatrix} - \begin{pmatrix}
\rho_+ & 0 & 0 & 0 \\
0 & \rho_+ & 0 & 0 \\
0 & 0 & \rho_- & 0 \\
0 & 0 & 0 & \rho_- \\
\end{pmatrix} \begin{pmatrix}
0 & \frac{1}{2} & \frac{1}{2} & 0 \\
\frac{1}{2} & 0 & 0 & \frac{1}{2} \\
\frac{1}{2} & 0 & 0 & \frac{1}{2} \\
0 & \frac{1}{2} & \frac{1}{2} & 0 \\
\end{pmatrix}. \\
\]

For the vector \(\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}\), this matrix enabled positing a positive relationship between the geosampled, explanatory, clinical, field and remote specified, county-level TB by geospatially adjusting the endemic, transmission-oriented, stochastically/deterministically, iteratively interpolatable covariates, \(y_1\) and \(y_2\). A negative relationship between the covariates, \(y_3\) and \(y_4\), and, no relationship between covariates, \(y_1\) and \(y_3\) and between \(y_2\) and \(y_4\) were noted. This covariance specification yielded:

\[
Y = \mu(I - \rho_+ < I_+ >_{\text{diag}} - \rho_- < I_- >_{\text{diag}})I + (\rho_+ < I_+ >_{\text{diag}} + \rho_- < I_- >_{\text{diag}})WY + \xi, \\
\]

Equation (3.2) where \(I_+\) was a binary 0-1 indicator variable. This equation also denoted those elucidative, time-series dependent, clinical, field and remote geosampled, explanatorial, endemic, transmission-oriented, county-level stochastically/deterministically, operationizable, parameterizable covariates displaying positive spatial dependency where \(I_- \) was a binary 0-1 indicator variable denoting those county-level geolocations displaying negative spatial dependency using \(I_+ + I_- = 1\). Expressing the preceding 2-by-2 example in terms of Equation (3.2) yielded:

\[
(I_- < \rho >_{\text{diag}} W)^{-1}\sigma^2 \\
\]
The Jacobian generalized the gradient of a scalar valued function of the quantitated variables by generalizing the derivative of a scalar. A more complex endemic TB county-level model specification was posited by generalizing binary indicator variables. 

\[
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4
\end{pmatrix}
= \mu
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4
\end{pmatrix}
- \rho_+ \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4
\end{pmatrix}
- \rho_- \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4
\end{pmatrix}
\]

\[
\begin{pmatrix}
1 \\
1 \\
1 \\
1
\end{pmatrix}
+ \rho_+ \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4
\end{pmatrix}
+ \rho_- \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4
\end{pmatrix}
\]

If either \( \rho_+ = 0 \) (and hence \( I_+ = 0 \) and \( I_- = I \)) or \( \rho_- = 0 \) (and hence \( I_- = 0 \) and \( I_+ = I \)) appeared in the model then Equation (3.2) reduced to Equation (3.1). This indicator variable classification was made in accordance with the quadrants of the corresponding Moran scatterplot generated by employing the regressed, time-series, predictive, explanatory, endemic, transmission-oriented, operationizable, time-series, stochastically/deterministically, interpolative TB, covariate coefficients geosampled in the counties.

In the model forecast, PSA and NSA processes counterbalanced each other in a mixture such that the sum of the two spatial autocorrelation parameters - \( (\rho_+ + \rho_-) \) were close to 0. The function from Euclidean \( n \)-space to Euclidean \( m \)-space which was simulated in SAS/GIS employing the county-level distance measurements
between the georeferenced, endemic transmission foci. The partial derivatives of all these functions were organized in an \( m \)-by-\( n \) matrix whereby, the Jacobian matrix \( J \) of \( F \) was as follows:

\[
J = \begin{bmatrix}
\frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n}
\end{bmatrix}.
\]

This matrix was denoted by \( J_F(x_1, a_n) \) and \( \frac{\partial (y_1, \ldots, y_m)}{\partial (x_1, \ldots, x_n)} \). The \( i \)-th row \( (i = 1, m) \) of this matrix was the gradient of the \( i \)-th component function \( y_i : (V, y_i) \). In these analyses, \( p \) was a georeferenced, endemic, transmission-oriented, forecast able, operationizable, time-series, clinical, field and/or remote specified, stochastically/deterministically, iteratively interpolatable, parameterizable covariates in \( \mathbb{R}^n \) when \( F \) was differentiable at \( p \). A derivative was given by \( J_F(p) \). The model described by \( J_F(p) \) was the best linear approximation of \( F \) near the point \( p \) in the sense that:

\[
F(x) = F(p) + J_F(p)(x - p) + o(\|x - p\|).
\]

The spatial structuring was achieved by constructing a linear combination of a subset of the eigenvectors of a modified geographic weights matrix using \((\mathbf{I} - 11'/n)\mathbf{E}(\mathbf{I} - 11'/n)\) that appeared in the numerator of the MC spatial autocorrelation which was then indexed with a product moment correlation coefficient.

A subset of eigenvectors was selected with a stepwise regression procedure. Because \((\mathbf{I} - 11'/n)\mathbf{E}(\mathbf{I} - 11'/n) = \mathbf{E}\mathbf{E}'\), where \( \mathbf{E} \) is an \( n \)-by-\( n \) matrix of eigenvectors and \( \mathbf{A} \) was an \( n \)-by-\( n \) diagonal matrix of the corresponding eigenvalues, the resulting endemic, county-level, time series, explanatory, time series, TB model specification was given by: \( Y = \mu \mathbf{1} + \mathbf{E}_k \beta + \epsilon \). In the iteratively interpolative, georeferenceable, asymptotical, normalized, autoregressive, clinical, field and remote specified, endemic, transmission-oriented, county-level Tb operationizable, time-series, predictive, epidemiological, risk model, \( \mu \) was the scalar mean of \( Y \), \( \mathbf{E}_k \) was an \( n \)-by-\( k \) matrix
containing the subset of \( k << n \) eigenvectors selected with a stepwise regression technique where \( \beta \) was a \( k \)-by-1 vector of regression coefficients.

A number of the eigenvectors were extracted from \((I - 11'/n)C(I - 11'/n)\) which we noted were affiliated with the county-level geographic patterns of the endemic, transmission-oriented, stochastically/deterministically, iteratively interpolative, operationizable, georeferenceable, TB, clinical field and remote-specified, time-series, parameterizable covariate. These covariates portrayed a negligible degree of spatial autocorrelation. Consequently, only \( k \) of the \( n \) eigenvectors was of interest for generating a candidate set for conducting a stepwise regression procedure. Candidate eigenvector represents a level of spatial autocorrelation which can account for the redundant information in an orthogonal risk map for cartographically delineating stochastically/deterministically, iteratively interpolatable, time-series, risk-related, endemic, transmission-oriented, temporally dependent, residual patterns (Baker et al., 2011).

Of note, the 2-by-2 square tessellation rendered a repeated eigenvalue in the epidemiological risk model. To identify spatialized clusters of county-level high density endemic transmission-oriented, georeferenceable clusters, Thiessen polygon surface portioning was performed in ArcGIS. By doing so, we generated a robust geographic weighted neighborhood-matrix. Each grid cell had a unique identifier. This identifier was the weighted geosampled, county-level, prevalence count, These matrices were also employed in the spatial autocorrelation analysis. Entries in matrix were 1, if two geosampled, endemic transmission, parameterizable 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 county-level, parameterizable covariate estimators sharing a common Thiessen polygon boundary. Attention was restricted to those stochastically/deterministically,
iteratively interpolative georeferenceable, risk map patterns associated with at least a minimum level of spatial autocorrelation which, for implementation purposes, was defined by \(|MC_j / MC_{\text{max}}| > 0.25\), where \(MC_j\) denoted the \(j\)-th value and \(MC_{\text{max}}\) was the maximum value of MC. This threshold value allowed two candidate sets of eigenvectors to be considered for spatial quantitation of substantial PSA and substantial NSA, respectively. These statistics indicated that the detected NSA may be considered to be statistically significant in a time-series, regressed dataset of predictive, county-level, endemic, transmission-oriented, operationizable, clinical, field and remote-specified, explanatory, stochastically/deterministically, iteratively interpolative covariate coefficients based upon a randomization perspective.

The ratio of the PRESS (i.e., predicted error sum of squares) statistic to the sum of squared errors from the MC scatterplot trend line was 1.27. Fortunately, this value fell within two standard deviations of the average standard prediction error value (roughly 1.18) for a georeferenceable, county-level endemic, transmission-oriented, TB, forecaster. Because prevalence counts were being analyzed, a Poissonian spatial filter model specification was then employed in AUTOREG.

The iterative interpolative, autoregressive, time-series, expositorial endemic, transmission-oriented, epidemiological, predictive, risk model specification was written as follows:

\[
\ln(\hat{\mu}_i) = \alpha 1 + E_k\beta, \quad \sigma_i^2 = \mu_i(1 - \eta\mu_i), \text{in AUTOREG}
\]

where \(\hat{\mu}_i\) was the expected prevalence count for a county-level geosampled geolocation \(I\) where \(\mu\) was an \(n\)-by-1 vector of the expected count variables. LN denoted the natural logarithm (i.e., the generalized linear model link function), \(\alpha\) was an intercept term and \(\eta\) was the negative binomial dispersion parameter. This log-linear equation had no error term. Estimation was executed assuming a negative binomial random variable.
The eigenfunctions of a spatial, weighted, explanatory, time-series, endemic, county-level TB, autoregressive, weight matrix was generated. The upper and lower bounds for a spatial matrix was derived by using Moran’s indices \( I \) given by \( \lambda_{\text{max}}(n/1^TW_1) \) and \( \lambda_{\text{min}}(n/1^TW_1) \), where \( \lambda_{\text{max}} \) and \( \lambda_{\text{min}} \) were the extreme eigenvalues of \( \Omega = HWH \). Hence, 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) autocorrelation (Goldhaber et al., 2011).

PSA and NSA spatial filter component pseudo-\( R^2 \) values were reported. These values did not exactly sum for the complete spatial filter. However, they were very close to their corresponding totals suggesting that any induced multicollinearity during the regression exercises was quite small. We noted that our spatial filter model specification could describe selected Gaussian and Poissonized time-series, explanatory, endemic, county-level TB-related, transmission-oriented, stochastically/deterministically, operationizable, time-series, iterative, interpolatable, random variables rendered from the autoregressive model.

A GLM was extended to account for latent, non-spatial, correlation effects. This epidemiological, risk-related, georeferencable, geosampled, clinical, field and remote specified, endemic transmission-oriented predictive, risk model allowed inferences to be drawn for a much wider range of geographic sampling configurations generated from the georeferenced, operationizable, explanatory, covariate coefficients than those utilized by employing a GLMM alone. The time-series, endemic, transmission-oriented, epidemiological, predictive, county-level TB, risk-related model included a random effect, which was specified as a random intercept that was assumed to be normally distributed with a mean of zero, a constant variance and zero spatial
autocorrelation. This varying intercept term compensated for the non-constant mean associated with the negative binomial regression which had a gamma distributed, non-homogenous mean, generalized specifications. The spatial structuring of random effects was then implemented with a conditional autoregressive model, and was achieved with a spatial filter. All of this work was done in the univariate format and employed an improper specification.

The assumption was that by moving the stochastic/deterministic, iteratively interpolative, elucidative, time-series, geosampled, endemic, transmission-oriented, parameterizable, TB, covariate estimators to a multivariate, asymptotical, normalized, time series, clinical, field and remote specified, conditional autoregressive model framework, the residually forecasted derivatives would then yield proper normalized distributions.

The approach was to introduce spatial autoregression parameters to the regression-related risk analyses. First clarified was what classes could be developed and contrasted. Novel parametric linear transformation was evaluated which provided an extension with attractive interpretation. These models were employed as specifications for parsimoniously quantitating second-stage spatialized effects in the hierarchical models. An application was performed for further quantizing the time-series, epidemiological, risk-related model, and the clinical, field, and remote-specified, explanatory, iteratively interpolative, georeferenceable, endemic, transmission-oriented explanatory time series, patterns. The spatial autocorrelation components revealed 11% redundant information in the estimation results.

The improvements of these fit in the adjusted and unadjusted models for all model specifications and random error in the spatial analyses. The unadjusted model compared the univariate model to a model containing only the intercept term. Interactions were examined, and significant interactions were included in the residually, forecasted, explanatively iteratively
interpolative, asymptotical, unbiased derivatives. Improvement of fit was also calculated for the first-order interaction models to determine whether including significant interactions improved fit compared to the full main effects model. Convergence problems prevented obtaining results of a saturated model to determine whether the presented model fit, as well as the saturated model.

The inverse Wishart distribution revealed a probability distribution defined by the positive-definite matrix generated from the explanatory, time-series, georeferenced, endemic, transmission-oriented, predictive, risk-related, clinical, field and remote-specified, stochastically/deterministically, iteratively interpolative, epidemiological covariates.

In the Bayesian model we employed the inverse Wishart distribution to generate the conjugate prior for constructing a viable covariance matrix for a multivariate normalized distribution. In our hierarchical generalized Bayesian probabilistic estimation model, we employed the inverse Wishart distribution to generate the conjugate prior for the covariance matrix of a multivariate normal distribution. In Bayesian probability theory, if the posterior distributions $p(\theta|x)$ are in the same 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 (Cogo, 2011).

In our research, the pdf of the inverse Wishart was:

$$\frac{|\Psi|^{m/2}}{2^{np/2} \Gamma_p(m/2)} \exp\left(\frac{-\text{trace}(\Psi B^{-1})}{2}\right) e^{-\frac{1}{2} \text{trace}(\Psi)}$$

where $B$ and $\Psi$ were quantified employing $p \times p$ positive definite matrices, where $\Gamma_p(\cdot)$ was the multivariate Gamma function. In mathematics, the multivariate gamma function, $\Gamma_p(\cdot)$, is a generalization of the gamma function. It is useful in multivariate statistics, appearing in the probability density function of the Wishart and inverse Wishart distributions (Cogo, 2011).

In the georeferencable, geosampled, endemic, transmission-oriented, explanatory, endemic, transmission-oriented, TB, county-level, Bayesian model, we had two explanatory, time series,
outcomes. One model outcome was \( \Gamma_p(a) = \int_{S>0} \exp \left( -\text{trace}(S) \right) |S|^{(p-1)/2} dS \), where \( S \geq 0 \) was defined when \( S \) was positive-definite. The other model outcome was \( \Gamma_p(a) = \pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma(a + (1/2)/p \sum_{i=1}^{p} \Gamma(a + (1 - i)/2) \). from this, model output we developed the recursive relationships between the geosampled, county-level, endemic TB-related, operationizable, parameterizable, covariate estimators employing: \( \Gamma_p(a) = \frac{\pi^{p-1/2}}{\Gamma(a)} \Gamma_p(a-1) \), and \( \Gamma_1(a) = \frac{\pi^{1/2}}{\Gamma(a)} \Gamma_1(a-1/2) \). Thus, the residualized clinical, field and remote regressed stochastically/deterministically, iteratively interpolative covariates were parsimoniously quantized by \( \Gamma_1(a) = \Gamma(a) \), \( \Gamma_2(a) = \pi^{1/2} \Gamma(a) \Gamma(a-1/2) \),and \( \Gamma_3(a) = \pi^{3/2} \Gamma(a) \Gamma(a-1/2) \Gamma(a-1) \).

The multivariate digamma illustrated below shows the function in the epidemiological, operationizable, risk model \( \psi_p(a) = \frac{\partial \log \Gamma_p(a)}{\partial a} = \sum_{i=1}^{p} \psi(a + (1 - i)/2) \) and the general polygamma function as \( \psi_p^{(n)}(a) = \frac{\partial^n \log \Gamma_p(a)}{\partial a^n} = \sum_{i=1}^{p} \psi_p^{(n)}(a + (1 - i)/2) \). Thereafter, it followed that \( \frac{\partial \Gamma_p(a)}{\partial a} = -\psi(a + (1 - i)/2) \Gamma(a + (1 - i)/2) \). By definition of the digamma function, \( \psi \),

\[ \frac{\partial \Gamma(a + (1 - i)/2)}{\partial a} = \psi(a + (i - 1)/2) \Gamma(a + (i - 1)/2). \]

Thus it followed that

\[ \frac{\partial \Gamma_p(a)}{\partial a} = \frac{\pi^{p-1/2}}{\Gamma(a)} \prod_{j=1}^{p} \Gamma(a + (1 - i)/2) \sum_{i=1}^{p} \psi(a + (1 - i)/2) = \Gamma_p(a) \sum_{i=1}^{p} \psi(a + (1 - i)/2) \]

Irt WAS noted that in the explanatory, georeferenced, field and remote-specified, time-series, autoregressive, endemic,transmission-oriented,stochastically/deterministically iteratively, interpolative, operationizable, risk model that, if \( B \) followed an inverse Wishart distribution, denoted as \( B \sim W^{-1}(\psi, m) \), its inverse \( B^{-1} \) had a Wishart distribution \( W^{1}(\psi^{-1}, m) \).

The multivariate Gamma function, \( \Gamma_p() \), was a generalization of the Gamma function in the explanatory, time-series, endemic, transmission-oriented, predictive, iteratively interpolative,
asymptotical, normalized, epidemiological, county-level TB, risk model. The Gamma function is an extension of the factorial function with its argument shifted down by 1 to real and complex numbers and, as such, \( n \) is a positive integer: \( \Gamma(n) = (n-1)! \) (Anderson, 1958). The gamma function can be seen as a solution to the following interpolation problem: find a smooth curve that connects the points \((x, y)\) given by \( y = (x - 1)! \) using positive integer values for \( x \) (Anderson, 1958).

A plot of the first few factorials derived from an empirical, georeferenceable, dataset of time-series, expository, clinical, field and remote specified, operationizable, endemic, transmission-oriented, parameterizable, covariate coefficients makes it clear that a such a smooth curve can be drawn, but it would be preferable to have a formula that precisely describes the curve, in which the number of operations does not depend on the size of \( x \) (Baker et al., 2011). The simple formula for the factorial, \( n! = 1 \times 2 \times \ldots \times n \), however, cannot be used directly for parsimoniously, quantitating, regressed, time-series, dependent, stochastically/deterministically, iteratively interpolative, fractional values of \( x \) since it is only valid when \( x \) is a positive integer (Anderson, 1958). Negative integer values are common in vector, arthropod-related, epidemiological, data analyses (Baker et al., 2011). There are, relatively speaking, no such simple solutions for factorials. Any combination of sums, products, powers, exponential functions or logarithms for constructing a robust, explanatory, time-series, operationizable, epidemiological, predictive, TB risk model, with a fixed number of terms may express \( x! \) (Anderson, 1958). However, the Gamma function appears commonly in the pdf of the Wishart and inverse Wishart distributions (Cogo, 2011).

The calculation of moments of complex Wishart and complex inverse Wishart distributed random matrices in the time-series, clinical, field and remote-specified, county-level, TB, endemic, transmission-oriented, operationizable, explanatory, predictive, risk, geosampled, georeferenceable, risk model, iteratively interpolative, covariate estimators was addressed in SAS/GIS.
applications such as infectious disease epidemiological modeling, complex Wishart and complex inverse Wishart distributed random matrices are used to model the statistical properties of complex sample covariance matrices and complex inverse sample covariance matrices, respectively (Anderson, 1958). Moments of random matrices are needed for spatiotemporally robustly, parsimoniously quantitating the asymptotic properties of the empirically geosampled, endemic, TB-related, transmission-oriented, elucidative, time series, explicative, clinical, field and remote-specified, endemic, transmission-oriented, iteratively interpolative, asymptotical, normalized, , parameterizable covariate, coefficient estimates (Baker et al., 2011). A derivation of the pdf of complex inverse Wishart distributed random matrices was attained. Furthermore, strategies were outlined for the calculation of the moments of both complex Wishart and complex inverse Wishart distributed matrices. The distribution generated from the geosampled, explanatory, time-series, endemic, transmission-oriented, clinical, field and remote-specified, iteratively interpolatable, georeferenceable, data had an inverse Wishart distribution $\mathbf{A} \sim W^{-1}(\Psi, m)$.

Successfully partitioned was the matrices $\mathbf{A}$ and $\Psi$ using $\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, $\Psi = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}$, where $A_{ij}$ and $\Psi_{ij}$ were $p_i \times p_j$ matrices in AUTOREG. The next step was to generate the conjugate distribution of the geosampled, georeferenceable, county-level, endemic, transmission-oriented, explanatory, iterative interpolative, parameterizable covariates by employing a covariance matrix $\sum$ whose prior $p(\sum)$ had a $W^{-1}(\Psi, m)$ distribution. Using PROC AUTOREG, can aid in testing for the stability of the regression coefficients and test for stationarity or unit roots in the time series. (https://support.sas.com/). We specified initial parameter values for GARCH and heteroscedasticity models. Exact gradients were used for GARCH-type model estimation.
Consider the series \( y_t \), which follows the GARCH process. The conditional distribution of the series Y for time \( t \) is written \( y_t | \Psi_{t-1} \sim N(0, h_t) \), where \( \Psi_{t-1} \) denotes all available information at time \( t - 1 \). The conditional variance \( h_t \) is
\[
h_t = \omega + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j} \quad \text{where} \quad \rho \geq 0, q > 0 \]
\( \omega > 0, \alpha_i \geq 0, \gamma_j \geq 0 \). The GARCH\((p,q)\) model reduces to the ARCH\((q)\) process when \( p = 0 \). At least all the endemic TB, county-level, ARCH parameters was nonzero \( (q > 0) \). The GARCH regression model was written \( y_t = x_t^\prime \beta + \epsilon_t \), where \( \epsilon_t \sim \text{IN}(0, 1) \). In addition, you can consider the model with disturbances following an autoregressive process and with the GARCH errors. The \( \text{AR}^m \cdot \text{GARCH}(p, q) \) regression model was denoted \( y_t = x_t^\prime \beta + v_t \), and the series \( y_t \), which follows the GARCH process in AUTOREG to quantitate spatial heteroskedascity in the count-level endemic TB-related epidemiological, forecast-oriented, county-level, epidemiological, risk model. The conditional distribution of the series Y for time \( t \) was written in AUTOREG as \( y_t | \Phi_{t-1} \sim N(0, h_t) \), where \( \Phi_{t-1} \) denoted all available county-level information at time \( t-1 \). The conditional variance \( h_t \) was
\[
h_t = \omega + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j} \quad \text{where} \quad p \geq 0, q > 0, \alpha_i \geq 0, \gamma_j \geq 0. \]
The GARCH \((p, q)\) model reduces to the ARCH \((q)\) process when \( p = 0 \). At least one of the ARCH parameters must be nonzero \( (q > 0) \). The endemic TB-related, GARCH regression model then was written
\[
y_t = x_t^\prime \beta + \epsilon_t \quad \text{where} \quad \epsilon_t \sim \text{IN}(0, 1). \]
GARCH errors. The AR (m)-GARCH (p, q) regression model was denoted \( y_t = \chi_t \beta + \nu_t \).

\[ \nu_t = \epsilon_t - \varphi_1 \nu_{t-1} - \ldots - \varphi_m \nu_{t-m}, = \sqrt{h_t} \epsilon_t, \]

The GARCH (p, q) model was written in ARCH \((\infty)\) form as

\[ h_t = \left( 1 - \sum_{j=1}^{p} \gamma_j \beta^j \right)^{-1} \left[ \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \right] \]

model was written in ARCH \((\infty)\) form as

\[ = \omega^* + \sum_{i=1}^{\infty} \phi_i \epsilon_{t-i}^2 \]

Where \( \beta \) was a backshift operator. Therefore, \( h_t \geq 0 \) if \( \omega^* \geq 0, \alpha_i \geq 0, \forall i \). The coefficient \( \phi_i \)

\[ \phi_0 = \alpha_1 \]
\[ \phi_1 = \gamma_1 \phi_0 + \alpha_2 \]
\[ \ldots \]
\[ \phi_{n-1} = \gamma_1 \phi_{n-2} + \gamma_2 \phi_{n-3} + \cdots + \gamma_{n-1} \phi_0 + \alpha_n \]

was written as \( \phi_k = \gamma_1 \phi_{k-1} + \gamma_2 \phi_{k-2} + \cdots + \gamma_n \phi_{k-n} \) for \( k \geq n \)

where \( \alpha_i = 0 \) for \( i > q \) and \( \gamma_j = 0 \) for \( j > p \).

Nelson and Cao (1992) proposed the finite inequality constraints for GARCH \((1, q)\) and GARCH \((2, q)\) cases. However, we found that it was not straightforward to derive the finite inequality constraints for the general GARCH \((p, q)\), endemic, transmission-oriented, epidemiological, TB-related, county risk model. For the GARCH \((1, q)\) model, the nonlinear inequality constraints were \( \phi_k \geq 0 \) for \( k = 0, 1, \ldots, q - 1 \). For the GARCH \((2, q)\)
where $\Delta_1$ and $\Delta_2$ are the roots of $\left( \zeta^2 - \gamma_1 \zeta - \gamma_2 \right)$. For the GARCH(p,q) model with $p > 2$, only $\max(q-1,p)+1$ nonlinear inequality constraints ($\phi_k \geq 0$ for $k=0$ to $\max(q-1,p)$) were imposed, together with the in-sample positivity constraints of the conditional variance $h_t$.

In the endemic, TB model $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \gamma_j < 1$ implied that the GARCH process was weakly stationary since the mean, variance, and autocovariance are finite and constant over time. However, this condition was not sufficient for weak stationarity in the presence of autocorrelation in the model. For example, the stationarity condition for an AR(1)-GARCH(p,q) process was

$$\frac{1}{1-\varphi^2} \sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \gamma_j < 1$$

Since the GARCH process was stationary, the unconditional variance of $\epsilon_t$ is computed as

$$V(\epsilon_t) = \frac{\omega}{\left(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \gamma_j \right)}$$

where $\epsilon_t = \sqrt{h_t} \epsilon_t$ and $h_t$ is the GARCH($p,q$) conditional variance. Sometimes, the multistep forecasts of the variance do not approach the unconditional variance when the model is integrated in variance; that is,

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \gamma_j = 1$$

(Baker et al., 2011).

The EGARCH model argues that the nonnegativity constraints in the linear GARCH model are too restrictive. The GARCH model imposes the nonnegative constraints on the parameters, $\alpha_i$ and $\gamma_j$, while there are no restrictions on these parameters in the EGARCH model.
model. In the EGARCH model, the conditional variance, $h_t$, was an asymmetric function of lagged disturbances $\varepsilon_{t-i}$:

$$\ln(h_t) = \omega + \sum_{i=1}^{q} \alpha_i g(|z_{t-i}|) + \sum_{j=1}^{p} \gamma_j \ln(h_{t-j})$$

where

$$g(z_t) = \theta z_t + \gamma \left[ |z_t| - \mathbb{E}|z_t| \right] z_t = \varepsilon_t / \sqrt{h_t}.$$ The coefficient of the second term in $g(z_t)$ was set to be 1 ($\gamma = 1$) in our formulation. Note that $\mathbb{E}|z_t| = (2/\pi)^{1/2}$ if $z_t \sim \mathcal{N}(0, 1)$. The properties of the EGARCH, county-level, endemic, TB, forecasting risk model were summarized as follows:

- The function $g(z_t)$ was linear in $z_t$ with slope coefficient $\theta + 1$ if $z_t$ was positive while $g(z_t)$ was linear in $z_t$ with slope coefficient $\theta - 1$ if $z_t$ was negative.

- If $\theta = 0$.

- Large innovations increased the conditional variance if $|z_t| - \mathbb{E}|z_t| > 0$ and decreased the conditional variance if $|z_t| - \mathbb{E}|z_t| < 0$.

- The innovation in variance, $g(z_t)$, was positive if the innovations $z_t$ were less than $(2/\pi)^{1/2}/(\theta - 1)$. Therefore, the negative innovations in returns, $\varepsilon_t$, causes the innovation to the conditional variance to be positive if $\theta$ is much less than 1.

The likelihood function was maximized via either the dual quasi-Newton or trust region algorithm. The default is the dual quasi-Newton algorithm. The starting county-level TB value for the regression parameters $\beta$ were obtained from the OLS estimate. The variance-covariance matrix is computed using the Hessian matrix. The dual quasi-Newton method approximates the Hessian matrix while the quasi-Newton method gets an approximation of the inverse of Hessian. The trust region method uses the Hessian matrix obtained using
numerical differentiation. When there are active constraints, that is, $q(\theta) = 0$, the variance-covariance matrix is given by $\mathbf{V}(\hat{\theta}) = \mathbf{H}^{-1}[\mathbf{I} - \mathbf{Q}'\mathbf{Q}\mathbf{H}^{-1}\mathbf{Q}'\mathbf{Q}\mathbf{H}^{-1}]$, where $\mathbf{H} = -\partial^2 i / \partial \theta \partial \theta'$ and $\mathbf{Q} = \partial q(\theta) / \partial \theta'$. Therefore, the variance-covariance matrix without active constraints reduces to $\mathbf{V}(\hat{\theta}) = \mathbf{H}^{-1}$.

The operationizable, endemic, transmission-oriented, time series, explanatory, clinical, field and remote-specified, eigenvectors associated with eigenvalues with extremely small absolute values corresponded to low spatial autocorrelation which were not suitable for defining spatial structures in the explanatory, time-series, stochastically/deterministically, iteratively interpolative, operationizable, county-level, endemic, transmission-oriented, risk model. The diagonalization of the spatial weighting matrix rendered from the field and remote geosampled, endemic, transmission oriented, georefernceable, regressed, asymptotical, normalized, explanatory, covariate coefficients instead consisted of finding the normalized vectors $u_i$ stored as columns in the matrix $U = [u_1 \cdots u_n]$, for satisfying: $\Omega = \mathbf{HWH} = U\Lambda U^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T$, where

$$\Lambda = \text{diag}(\lambda_1 \cdots \lambda_n), \quad u_i^T u_i = \|u_i\|^2 = 1 \quad \text{and} \quad u_i^T u_j = 0 \quad \text{when} \quad i \neq j.$$ Note that double centering of $\Omega$ implied that the endemic, transmission-oriented, time –series, explanatory, field and remote-specified eigenvectors $u_i$ generated from the ecological, geosampled, regressed, stochastically/deterministically, iteratively interpolative, endemic TB, county-level, covariate coefficients were centered whereby at least one eigenvalue was equal to zero. Introducing these eigenvectors in the original formulation of Moran’s index lead to

$$I(x) = \frac{n}{1^T W_1} \frac{x^T \mathbf{HWH} x}{x^T H x} = \frac{n}{1^T W_1} \frac{x^T U \Lambda U^T x}{x^T H x} = \frac{n}{1^T W_1} \frac{\sum_{i=1}^{n} \lambda_i x^T u_i u_i^T x}{x^T H x}.$$
Considering the centered vector \( z = Hx \) and using the properties of idempotence of \( H \), Equation (3.1) was then tabulated as being equivalent to the formulation

\[
I(x) = \frac{n}{1^T W_1} \sum_{i=1}^{n} \lambda_i z^T u_i u_i^T z = \frac{n}{1^T W_1} \sum_{i=1}^{n} \lambda_i \|u_i^T z\|^2 \quad (3.4)
\]

Next that was transformed was the autocorrelation indicators to correlation, time series, explanatory coefficients. The remote specified eigenvectors \( u_i \) and vector \( z \) were then centered, and the time-series, explanatory, iteratively interpolatable, county-level TB-related, predictive equation was rewritten as

\[
I(x) = \frac{n}{1^T W_1} \sum_{i=1}^{n} \lambda_i \text{cor}^2(u_i, z) \var(z) n = \frac{n}{1^T W_1} \sum_{i=1}^{n} \lambda_i \text{cor}^2(u_i, z) \quad (3.5)
\]

where \( r \) was the number of null eigenvalues of \( \Omega(r \geq 1) \).

In linear algebra for a matrix \( A \), there may not always exist a full set of linearly independent eigenvectors that form a complete basis – a matrix may not be diagonalizable. This happens when the algebraic multiplicity of at least one eigenvalue \( \lambda \) is greater than its geometric multiplicity when quantitating the nullity of the matrix, or the dimension of its null space. In such cases, a generalized eigenvector of \( A \) is a nonzero vector \( \mathbf{v} \) which is associated with \( \lambda \) having algebraic multiplicity \( k > 1 \). Thus, a linear operator of dimension \( n \) in a robust, time-series, explanatory, county-level TB-specified, clinical, field and remote geosampled, stochastically/deterministically, iteratively interpolative, operationizable, epidemiological, risk model has a non-trivial null space (i.e., \( A \cdot x = 0 \)). Further, suppose the dimension of the null space in the residually forecasted risk model derivatives is \( k < n \), then \( 0 < k \) would reveal linearly
independent vectors each of which would yield the 0 vector when the linear operator $A$ is applied to it. It would be more standard to quantitate 0 as an eigenvalue of geometric multiplicity $k$ in the endemic, transmission-oriented, county-level, TB model output. This would then imply that that the model algebraic multiplicity was greater than or equal to $k$. The characteristic equation would then have a factor $x_n$ for $n \geq k$ rather than for “$k$ eigenvalues” for all of the geosampled, explanatory, predictive values of $k$ in the empirical dataset of eigenvalues that happened to have the same value. In this research, these eigenvalues and corresponding eigenvectors were removed from $\Lambda$ and $U$, respectively. The explicative, county-level TB-related, regressed residual forecasts were then strictly equivalent to $I(x) = \frac{n}{1^T W 1} \sum_{i=1}^{n-r} \lambda_i \text{cor}^2(u_i, z)$. 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 explained by $u_i$ in the autoregressive, time-series, predictive, endemic, transmission-oriented, operationizable, endemic, transmission-oriented, model $[\text{i.e., } z = \beta_i u_i + e_i]$. This quantity was equal to $\beta_i^2 / n \text{ var}(z)$. By definition, the eigenvectors $u_i$ were orthogonal and therefore the regression coefficients of the linear models $z = \beta_i u_i + e_i$ were those of the multiple regression model $z = U \beta + \varepsilon = \beta_i u_i + \ldots + \beta_{n-r} u_{n-r} + \varepsilon$.

Next, the distribution of the time-series, expositorial, TB, clinical field and remote-specified, georeferenceable, erroneous residuals in the explanatory, time-series, county-level, operationizable, iteratively interpolative, endemic transmission-oriented georeferenceable, autocovariance matrix revealed that the maximum value of $I$ was obtained by all of the variations of $z$, as explained by the eigenvector $u_1$, which coincidently corresponded to the highest
eigenvalue \( \lambda_1 \) in the spatial autocorrelation, diagnostic, error matrix.

Thereafter several Lagrange Multiplier tests were generated for constructing a robust, panel data, regression model including a joint test for serial correlation, spatial autocorrelation and random effects. We drew upon two strands of earlier work. The first was the LM tests for the autoregressive error correlation model, the second was the LM tests for the error component, panel data model with serial correlation. Hence, the joint LM test optimally derived for quantizing the county-level, TB, clinical, field and remote-geosampled, operationizable, georeferenceedable, empirical, explanatory, data encompassed those derived in both strands of earlier works. In fact, in the context of the generalized, county-level, endemic transmission- oriented, explanatory, county-level, clinical, field and/or remote-specified, geosampled, georeferenceable, TB model, the earlier LM tests became marginal LM tests that ignored either serial correlation over time or uncertainty correlation effects.

Conditional LM and LR tests were then derived that did not ignore any uncertainty correlations and contrasted them with their marginal LM and LR counterparts. \( \text{corr}^2(u_i, z) = 1 \) and \( \text{corr}^2(u_i, z) = 0 \) for \( i = 1 \) was noted. The maximum value of \( I \) was deduced for Equation (3.3) which was equal to \( I_{\text{MAX}} = \lambda_1(n/1^TW1) \). This minimum value in the error matrix was parsimoniously regressively quantitated based on all the variation of \( z \). This value was explained by the eigenvector \( u_{n-r} \) corresponding to the lowest eigenvalue \( \lambda_{n-r} \) which was optimally generated in the epidemiological, predictive TB, risk model. This minimum value was equal to \( I_{\text{MIN}} = \lambda_{n-r}(n/1^TW1) \). If the geosampled, georeferenceable, asymptotical, unbiased, normalized, predictor variable was not spatialized, the part of the variance explained by each eigenvector was then equal to \( \text{corr}^2(u_i, z) = 1/n - 1 \). Because the explanatory, time-series, clinical, field and remote-
specified, stochastically/deterministically, iteratively interpolative, predictor variables in \( z \) were randomly permuted, it was assumed that we would obtain this result. The set of \( n! \) random permutations revealed that

\[
E_R(I) = \frac{1}{1^T W_1(n-1)} n \sum_{i=1}^{n} \lambda_i = \frac{1}{1^T W_1(n-1)} \text{trace}(\Omega).
\]

It was easily demonstrated that \( \text{trace}(\Omega) = -\frac{1}{n} \sum_{i=1}^{n} \lambda_i \); thus, it followed that \( E_R(I) = -\frac{1}{n-1} \).

The georeferenced, geosampled, time-series, endemic, TB county-level, transmission-oriented, epidemiological, risk-related parameterizable covariates were then input into an eigenfunction decomposition algorithm in AUTOREG to quantitate any latent autocorrelation error coefficients in the forecasted stochastically/deterministically, iteratively interpolative, operationizable, explanatory, \( r \) time-series, variance estimates. Results indicated negligible PSA was detected for the county-level, georeferenceable, explanatory, time-series, epidemiological, geosampled data. Eigenvectors were then extracted from the matrix \((I - \mathbf{1}{1'}/n)^C(I - \mathbf{1}{1'}/n)\) employing the ecological, geosampled, variables. It was noted that the autoregressive, georeferenceable, explanatorial, parameterizable, covariate estimators captured the latent spatiotemporal autocorrelation in the seasonal, multivariate, endemic, TB transmission-oriented, epidemiological, predictive, risk model, residually, forecasted, operationizable, clinical, field and remote-specified, stochastically/deterministically, iteratively interpolative derivatives. This quantification involved \( \rho \), a conditional autoregressive covariance specification and the matrix \((\mathbf{1} \mathbf{1}^T)\), where \( \mathbf{I} \) was an \( n \)-by-\( n \) identity matrix. In linearized matrix algebra, the identity matrix of size \( n \) is the \( n \times n \) square matrix with ones on the main diagonals and zeros may be denoted by \( \mathbf{I}_n \), or simply by \( \mathbf{I} \), if the size is immaterial (Anderson, 1958). The matrix may be trivially determined by the context. However, when \( \mathbf{A} \) is \( m \times n \), it is a property of matrix multiplication (Cogo, 2011). In particular, the identity matrix serves as the unit of the ring of all \( n \times n \) matrices, and as the
identity element of a generalized group $GL(n)$ consisting of all invertible $n \times n$ matrices where the identity matrix itself is invertible, being its own inverse (Cogo, 2011). In our research $n \times n$ matrices were employed to represent linearized transformations from an $n$-dimensional vector space of the robust, georeferenceable, geosampled, time-series, explanatory, field and remote-specified, endemic, transmission-oriented, stochastically/deterministically, iteratively interpolatable, operationizable, endemic, epidemiological, risk model, covariate, parameterizable estimators from an $n$-dimensional vector space to itself. $I_n$ represents the identity function regardless of the basis identity function regardless of the basis (Baker et al., 2011). Thereafter, the residualized, county-level, endemic, transmission-oriented, autocorrelation, error, components were calculated as the matrix $C$ raised to the power 1.

Since adjacent geosampled, time-series, explanatory, endemic, transmission-oriented, stochastically/deterministically, iteratively interpolatable, predictive, operationizable, clinical, field and remote-specified, time-series data were involved in the autoregressive function, we used a first-order specification, with the autoregressive term being $CY$. An important matrix was optimally derived from $C1$ which was the vector of the number of geosampled, georeferenceable, county-level, high density foci neighbors... The inverse of the elements of $C1$ were inserted into the diagonal of a diagonal matrix (i.e., $D^{-1}$) rendering matrix $W = D^{-1}C$, which became a stochastic matrix (i.e., each of its row sums equaled 1). One appealing feature of this matrix was the autoregressive term became $WY$, which generated averages, rather than sums of the neighboring geosampled, TB, endemic, parameterizable, covariate estimate values. Because a covariance matrix for a robust, county-level, TB distribution model must be symmetrical [2], we employed a matrix $W$ specification with a conditional autoregressive model by making the individual, geosampled, estimator variance $w$, non-constant using $(I - \rho D^{-1}C)D^{-1} = (D^{-1} - \rho D^{-1}CD^{-1})$. An appealing feature
of this model version was that it restricted, explanatory, time-series, county-level, hyper/hypo endemic, transmission-oriented, stochastically/deterministically iteratively interpolative, predictive, operationizable, epidemiological. risk-related, distribution, model values of the autoregressive parameter to the more intuitively interpretable range of \(0 \leq \hat{\rho} \leq 1\). The epidemiological, risk-related, epidemiological, risk-related, predictive model furnished an alternative specification which was also written in terms of matrix \(W\). The spatial covariance was then a function of the matrix 
\[
(I - \rho CD^{-1})(I - \rho D^{-1}C) = (I - \rho W^T)(I - \rho W),
\]
where \(T\) denoted the matrix transpose. The resulting matrix was symmetric and was considered a second-order specification as it included the product of two spatial structure matrices (i.e., \(W^T W\)). This matrix restricted values of the autoregressive parameter to the more intuitively

One assumption about the random effects term is that its frequency distribution was bell-shaped (i.e., normally/Gaussian distributed) with a mean of zero. The empirical estimate has a mean of 0.0197, with a Shapiro-Wilk normality probability of 0.0027. In other words, the mean was not exactly zero, although its t statistic of 0.06 indicates it was not significantly different from zero, and the frequency distribution deviated from a bell-shaped one, but not dramatically. This random effects term accounts for roughly 41\% of the variability in the observed probability of tuberculosis by county in Florida, and yielded an under dispersed binomial model.

An eigenvector spatial filter description of the random effects term involves five (of 18) eigenvectors portraying noticeable PSA (Fig. 1b), and four (of 25) eigenvectors portraying noticeable NSA (Fig. 1c). These two spatial filter components accounted for, respectively, roughly 16\% and 10\% of the variability in the probability of tuberculosis by county. The spatially unstructured random effects component (Fig. 10d) accounts for roughly 15\% of this variability.
Figure 10: Top left (a): random effects term. Top right (b): positive spatial autocorrelation eigenvector spatial filter. Bottom left (c): negative positive spatial autocorrelation eigenvector spatial filter. Bottom right (d): spatially unstructured random effect. The counties were ranked by predicted probability of tuberculosis, these figures allowed the counties to be ranked by number of future cases.
The georeferenced, explanatory, clinical, field and remote specified, endemic, transmission-oriented, county-level regressors were independent $p$-variate Gaussian variables which were drawn from a $\mathcal{N}(0, \sum)$ distribution. If the observations are independent $p$-variate Gaussian variables
drawn, then the conditional distribution has a distribution which may be optimally quantitated employing the sample covariance matrix (Cogo, 2011). Because the prior and posterior distributions are the same family, we say the inverse Wishart distribution is conjugate to the multivariate Gaussian (Anderson, 1958). Due to its conjugacy to the multivariate Gaussian, it is possible to marginalize out the Gaussian’s parameter (Anderson, 1958). The endemic, TB, georeferenceable, time-series, asymptotical, unbiased, explanatory, epidemiological, operationizable, predictive, county-level, risk model had a conditional distribution [i.e., \( p(\sum |X|) \)] which had a \( W^{-1}(A + \Psi, n + m) \) distribution. Thereafter, \( A = XX^T \) was used to generate the sample covariance matrix in AUTOREG.

In this research, the inverse Wishart distribution was conjugate to the multivariate Gaussian. Due to its conjugacy to the multivariate Gaussian it was possible to “integrate out” the Gaussian-based time-series, endemic TB, county-level, stochastically/deterministically, iteratively interpolative, explanatory parameters [i.e., \( \sum \)] from the other transmission-oriented, operationizable, georeferenceable, predictor variables. We used

\[
P(X|\Psi, m) = \int P(X|\Sigma)P(\Sigma|\Psi, m) d\Sigma = \frac{|\Psi|^{\frac{m}{2}} \Gamma_p\left(\frac{m + n}{2}\right)}{\pi^\frac{m}{2} |\Psi + A|^{\frac{m + n}{2}} \Gamma_p\left(\frac{m}{2}\right)}.
\]

The variance matrix \( \sum \) for the geosampled, georeferenceable, interpolative, time-series, endemic, transmission-oriented, county-level, TB explanatory covariates was \( \Psi \) (i.e., the priori) and, as such, \( A \) was directly obtained from the coefficient indicator values. The meant the, time-series, autoregressive, predictive, endemic, transmission-oriented, interpolatable, epidemiological, risk-related, operationizable, TB model was then \( \mathbf{E}(B) = \frac{\Psi}{m - p - 1} \). The variance of each element of \( B \) was then
\[ \text{var}(b_{ij}) = \frac{(m - p + 1)\bar{y}_{ij}^2 + (m - p - 1)y_{ij}y_{jj}}{(m - p)(m - p - 1)^2(m - p - 3)}. \]

The variance of the diagonal employed in the distribution model was also rendered by using the same formula as above with \( i = j \), which further simplified the model to:

\[ \text{var}(b_{ii}) = \frac{2\bar{y}_{ii}^2}{(m - p - 1)^2(m - p - 3)}. \]

Defined was a right eigenvector as a column vector \( \mathbf{X}_R \) satisfying \( A\mathbf{X}_R = \lambda_R \mathbf{X}_R \), where \( A \) was a matrix so \( (A - \lambda_R I)\mathbf{X}_R = 0 \), which meant the right eigenvalues had zero determinant in our geospatially, autoregressive, clinical, field and remote-specified, explanatively stochastically/deterministically, iteratively interpolative, operationizable, georeferenceable, asymptotical, normalized, time-series, epidemiological, risk model. Similarly, we defined a left eigenvector as a row vector, \( \mathbf{X}_L \), satisfying \( \mathbf{X}_L A = \lambda_L \mathbf{X}_L \). Taking the transpose of each side rendered \((\mathbf{X}_L A)^T = \lambda_L \mathbf{X}_L^T\) which was rewritten as \( A^T \mathbf{X}_L^T = \lambda_L \mathbf{X}_L^T \). We rearranged this equation once again to obtain \( (A^T - \lambda_L I)\mathbf{X}_L^T = 0 \) which generated \( \det(A^T - \lambda_L I) = 0 \). The epidemiological forecasting equation, in turn, generated \( 0 = \det(A^T - \lambda_L I) = \det(A^T - \lambda_L I)^T \), \( \det(A - \lambda_L I)^T \), and \( \det(A - \lambda_L I) \), where the last step was derived from the identity of \( \det(A) = \det(A^T) \). We equated these equations to 0 for \( A \) and \( X \), which required that \( \lambda_R = \lambda_L = \lambda \) (see 7).

\( \mathbf{X}_R \) be a matrix formed by the columns of the right eigenvectors \( \mathbf{X}_L \), which was actually a matrix formed by the rows of the left eigenvectors. We let \( D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \), and, as such,

\[ A\mathbf{X}_R = \mathbf{X}_R D \quad \text{and} \quad \mathbf{X}_L A = DX_L X_L A X_R \quad \mathbf{X}_L X_R = X_L X_R D \quad \text{while}, \quad \mathbf{X}_L A X_R = DX_L X_R, \quad \text{so} \quad \mathbf{X}_L X_R D = DX_L X_R. \]

But this equation was of the form \( \mathbf{C} = \mathbf{D} \mathbf{C} \), where \( \mathbf{C} \) was a diagonal matrix, so, \( \mathbf{C} = \mathbf{X}_L X_R \) was
also diagonal. If $A$ is a symmetric matrix, then the left and right eigenvectors are simply each other’s transpose, and if $A$ is a self-adjoint matrix (i.e., Hermitian), then the left and right eigenvectors are adjoint matrices.

In predictive, autoregressive, endemic, transmission-oriented, explanatory, iterative interpolative, time-series, time series, clinical, field and remote-specified, epidemiological, risk modeling, a Hermitian matrix is a square matrix with complex entries that is equal to its own conjugate transpose – that is, the element in the $i$-th row and $j$-th column is equal to the complex conjugate of the element in the $j$-th row and $i$-th column, for all indices $i$ and $j$: $a_{i,j} = \overline{a_{j,i}}$ (Crosby & Noar, 2011). Using the matrix with eigenvectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ and corresponding eigenvalues $\lambda_1$, $\lambda_2$, and $\lambda_3$, an arbitrary vector $\mathbf{y}$ was then written as $\mathbf{y} = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + b_3\mathbf{x}_3$. In this research matrix $A$ was generated $A\mathbf{y} = b_1A\mathbf{x}_1 + b_2A\mathbf{x}_2 + b_3A\mathbf{x}_3 = \lambda_1\left(b_1\mathbf{x}_1 + \frac{\lambda_2}{\lambda_1} b_2\mathbf{x}_2 + \frac{\lambda_3}{\lambda_1} b_3\mathbf{x}_3 \right)$: so $A^n\mathbf{y} = \lambda_1^n \left(b_1\mathbf{x}_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^n b_2\mathbf{x}_2 + \left(\frac{\lambda_3}{\lambda_1}\right)^n b_3\mathbf{x}_3 \right)$. Further, since $\lambda_1 > \lambda_2$, $\lambda_3$, ..., and $b_1 \neq 0$, it followed that

$$\lim_{n \to \infty} A^n\mathbf{y} = \lambda_1^n b_1\mathbf{x}_1,$$

so repeated application of the matrix to our vector in the geosampled, clinical, field and remote-specified, endemic transmission-oriented, county-level, georeferenced, asymptotical, TB unbiased, elucidative, stochastically/deterministically, interpolative parameter estimator epidemiological, dataset resulted in a vector proportional to the eigenvector with largest eigenvalue.

Probabilities for the autoregressive model. We used $\frac{dS_t}{S_t} = \mu dt + \sigma dW_t$ where $W_t$ was a $P$-standard Brownian motion (i.e., a standard Brownian motion under the probability measure $P$). We generated: $\frac{dS_t}{S_t} = (r - g)dt + \sigma(dW_t + \lambda dt)$. Then we used the Black and Scholes model whereby, the
We then generated the inverse underlying motion rendered:

\[ C(t) = e^{-rt} e^{r t} S_t \text{ and } e^{-rt} e^{r t} M_t \]

Then defined the \( S \) set by: \( E = \{ \omega \in \Omega | S(T)(\omega) \geq k | F_t \} \), which rendered

\[ C(t) = e^{-r(T-t)} E_Q[S_T I_E|F_t] - k e^{-r(T-t)} E_Q[I_E|F_t]. \]

This quantity was computed by splitting each of its terms. The second term in the model generated \( E_Q[I_E|F_t] = P_Q(E) = P_Q(S_T \geq k | F_t) \) employing

\[ S(T) = S(t) \exp \left( (r - \frac{\sigma^2}{2}) (T-t) + \sigma \hat{W}(T-t) \right). \]

We employed the condition \( S_T \geq k \) to spatiotemporally quantitate the TB, endemic, transmission-oriented, stochastically/deterministically, iteratively interpolative operationizable, time-series, ecological, parameter estimators which defined

\[ Y = -\frac{\hat{W}(T-t)}{\sqrt{T-t}} \leq d_2. \]

The properties of the Brownian motions allowed us to write the expression

\[ Y \sim N(0,1) \]

for the distribution, epidemiological, risk model which rendered

\[ P_Q(S_T \geq k) = P_Q(Y \leq d_2) = N(d_2). \]

Using the first term \( E_Q[S_T I_E] = \int_k^\infty x f_{S_T}(x) dx \), we were able to subsequently generate

\[ S(T) = S(t) \exp \left( (r - \frac{\sigma^2}{2}) (T-t) + \sigma \hat{W}(T-t) \right). \]

The log-normal property of the underlying motion rendered:

\[ L_Q((r - \frac{\sigma^2}{2}) (T-t) + \sigma \hat{W}(T-t)) = N((r - \frac{\sigma^2}{2}) (T-t), \sigma \sqrt{T-t}) \]

and

\[ e^{-r(T-t)} E_Q[S_T I_E|F_t] = S(t) N(d_1). \]

We then generated the inverse-Gamma distribution which was a univariate specialization of the inverse-Wishart distribution summarized by using the regressed, georeferenced, TB-related,
endemic, transmission-oriented, stochastically/deterministically, iteratively interpolatable, predictive, risk-related, seasonal, geosampled, and explanatory covariates. The pdf was

\[ \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(\frac{-\beta}{x}\right) \]

while the mean of the model was \( \frac{\beta}{\alpha - 1} \) for \( \alpha > 1 \). The variance was

\[ \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} \]

for \( \alpha > 2 \). The skewness was \( \frac{4\sqrt{\alpha - 2}}{\alpha - 3} \) for \( \alpha > 3 \), while the kurtosis was

\[ \frac{30\alpha - 66}{(\alpha - 3)(\alpha - 4)} \]

for \( \alpha > 4 \) and the entropy was \( \alpha + \ln(\beta \Gamma(\alpha)) - (1 + \alpha)\Psi(\alpha) \). The moment generating function was

\[ 2\left(\frac{-\beta t}{\Gamma(\alpha)}\right)^{\alpha/2} K_\alpha\left(\sqrt{-4\beta t}\right), \]

while the characteristic function was

\[ \frac{2(-i\beta t)^{\alpha/2}}{\Gamma(\alpha)} K_\alpha\left(\sqrt{-4i\beta t}\right). \]

The stochastically/deterministically, interpolatable, county-level TB model revealed that with \( p = 1 \) (i.e., univariate) and \( \alpha = m/2, \beta = s/2 \), and \( x = y \) the pdf of the inverse-Wishart distribution was

\[ p(x|\alpha, \beta) = \frac{\beta^\alpha x^{-\alpha-1} \exp(-\beta/x)}{\Gamma_1(\alpha)} \]

The pdf of the Gamma distribution was \( f(x) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)} \). We found:

\[ f_Y(y) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) = \frac{1}{\theta^k \Gamma(k)} \left(\frac{1}{y}\right)^{k-1} \exp\left(-\frac{1}{\theta y}\right) \]

Replacing \( k \) with \( \alpha \); \( \theta^{-1} \) with \( \beta \); and \( y \) with \( x \) resulted in the inverse-Gamma pdf: \( f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right) \). The inverse Gamma distribution’s pdf was then defined over the support \( x > 0 \) using the equation \( f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right) \) with shape parameter \( \alpha \) and scale parameter \( \beta \). The cumulative distribution function was quantified using the regularized Gamma function \( F(x; \alpha, \beta) = \frac{\Gamma(\alpha, \frac{\beta}{x})}{\Gamma(\alpha)} = Q(\alpha, \frac{\beta}{x}) \) where the numerator in the operationizable,
time-series, epidemiological, risk-related model was the upper incomplete Gamma function and
the denominator was the Gamma function. The regularized Gamma functions were defined in the
explanatory, time-series, TB, endemic, transmission-oriented, stochastically/ deterministically,
iteratively interpolative, georeferenceable, geosampled, predictive, epidemiological, risk model by
\[
\frac{\gamma(a, z)}{\Gamma(a)} \quad \text{and} \quad \frac{\Gamma(a, z)}{\Gamma(a)}
\]
where \(\gamma(a, z)\) and \(\Gamma(a, z)\) were incomplete Gamma functions, and \(\Gamma(a)\) was a
complete Gamma function.

Method employed in PROC MCMC was to calculate themultivariate Gamma function for
quantitating the explanatory, field and remote-specified, endemic, transmission-oriented, predictive,
epidemiological, risk model stochastically/ deterministically interpolative, time-series, georeferenceable, asymptotical, unbiased, covariate coefficients. The function was constructed using
\[
\Gamma_p(a) = \int_{S > 0} \exp(-\text{trace}(S))(S^{a-(p+1)/2})dS \quad \text{where} \quad S > 0 \quad \text{thus,} \quad S \quad \text{was positive-definite. We employed the}
\]
Gamma function to determine the recursive relationships in the geosampled, endemic, transmission-
oriented, georeferenced, operationizable, time-series, explanatory covariates using
\[
\Gamma_p(a) = \pi^{(p-1)/2}\Gamma(a)\Gamma_{p-1}(a-\frac{1}{2}) = \pi^{(p-1)/2} \Gamma_{p-1}(a)\Gamma[a + (1 - p)/2].
\]
Next, we quantified \(\Gamma_1(a) = \Gamma(a), \Gamma_2(a) = \pi^{1/2}\Gamma(a)\Gamma(a-1/2), \quad \text{and} \quad \Gamma_3(a) = \pi^{3/2}\Gamma(a)\Gamma(a-1/2)\Gamma(a-1). \quad \text{We defined the field and remote-specified multivariate digamma function in the explanatory, model as}
\[
\psi_p(a) = \frac{\partial \log \Gamma_p(a)}{\partial a} = \sum_{i=1}^{P} \psi(a + (1 - i)/2) \quad \text{and the general polygamma function as}
\]
\[
\psi_p^{(n)}(a) = \frac{\partial^n \log \Gamma_p(a)}{\partial a^n} = \sum_{i=1}^{P} \psi^{(n)}(a + (1 - i)/2). \quad \text{The digamma function in the explanatory, time-series,}
\]
county-level TB regression-based, georeferenceable, asymptotical, normalized, epidemiological,
risk model was defined as the logarithmic derivative of the Gamma function: \(\psi(x) = \frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}.\)
This equation then calculated the endemic, transmission-oriented, digamma function which was expressed as \( \frac{\partial \Gamma(a + (i - 1)/2)}{\partial a} = 
abla(a + (i - 1)/2)\Gamma(a + (i - 1)/2) \). In mathematics, the digamma function is defined as the logarithmic derivative of the gamma function (Anderson, 1958). The roots of the digamma function are the saddle points of the complex-valued gamma function (Cogo, 2011); thus, all the residually, forecasted, georeferenceable, explanatory, time-series dependent, clinical, field and remote-specified, stochastic/deterministic, iteratively interpolatable, optimizable derivatives generated, we assumed would lie on the real axis. The endemic, transmission oriented digamma function is often denoted as \( \psi_0(x), \psi^0(x) \) (Cogo, 2011).

We then generated the following expression:

\[
\frac{\partial \Gamma_p(a)}{\partial a} = \pi^{p(2p-1)/4} \prod_{j=1}^{p} \Gamma(a + 1 - j/2),
\]

it followed that

\[
\frac{\partial \psi_p(a)}{\partial a} = \pi^{p(2p-1)/4} \sum_{i=1}^{p} \frac{\partial \Gamma(a + 1/2)}{\partial a} \prod_{j=1, j \neq i}^{p} \Gamma(a + 1 - j/2).
\]

The pdf of the inverse Wishart for the interpolative, explanatory, epidemiological, county-level TB, predictive, risk model was found to be \( \frac{|y|^{\pi/2}}{2^{p \pi/2} \Gamma_p(\frac{p}{2})} |x - \frac{y+1}{2} e^{-\frac{1}{2}tr(X^{-1})} \) when \( X \) and \( \psi \) were \( p \times p \) positive definite matrices, and \( \Gamma_p() \) was the multivariate Gamma function.

In this research, when \( f \) was convex in the explanatory, time-series, endemic, transmission-oriented, interpolatable, county-level TB, epidemiological, risk model and twice continuously differentiable and the georeferenceable, asymptotical, unbiased, sublevel set \( \{x : f(x) \leq f(x^0)\} \) was bounded, then the sequence of function values generated by the BFGS method with an inexact Armijo-Wolfe line search converged at the minimal value (i.e., county-level prevalence count) of \( f \). Interestingly, the performance of a modified Armijo line search rule related to BFGS gradient
type method. Although the modified Armijo rule does require as much computational cost as the other rules, it shows more efficient local minima for resolving unconstrained optimization problems (Cogo, 2011). The sensitivity of the time-series, explanatory, stochastic/deterministic, operationizable, elucidative, parameterizable covariate estimators employed in the line search rules was also analyzed.

Newton’s method was very fast (e.g., quadratic convergence), but had no global convergence property. This result did not follow directly from the standard Zoutendijk theorem as commonly quasi-Newton methods do (Cogo, 2011), but instead the eigenvalues of the inverse Hessian approximation \( H_k \) did not grow too large or too small. In our line search algorithm, the sequence \( \{x_n\} \) was constructed iteratively at each step by choosing a search direction \( p_k \) for minimizing the objective function along the line (ray) for quantitating direction in the endemic, transmission-oriented, time-series explanatory, epidemiological, operational, field and remote-geosampled, georeferenceable, risk, model. A sequence of one dimensional stochastically/deterministically, interpolatable, explicative, county-level TB-related estimates was rendered where \( X_k \) was given by the basic recurrence: \( X_{k+1} = X_k + \alpha_k p_k \) and where the step-length \( \alpha_k > 0 \) was parsimoniously estimated. The key to having a sequence \( X_k \) which converges rapidly is to construct an algorithm which makes effective use of the information about \( x \) from the previous iterates to choose a good \( p_k \) and step-length \( \alpha_k \) (Anderson, 1958). To ensure that progress was made for each iteration, we imposed a natural condition on \( p_k \) by decreasing the function, at least locally. We then quantitated \( p_k \) as a descent direction when \( \nabla f \cdot p_k < 0 \). The descent direction in the epidemiological, predictive, risk model had the form \( p_k = -B^{-1} k \nabla f_k \). Since in our model, \( B_k \) was position definite this ensured that \( p_k \) was a descent direction quantitated geo-spatiotemporally by \( \nabla f_k \cdot p_k = -\nabla f_k^T B^{-1} k \nabla f_k < 0 \).
The objective was to minimize the multi-variable nonlinear (convex or nonconvex) function $\min f(x)$, where $f$ was twice differentiable and $x \in R^n$ by efficiently, robustly, quantizing, the explanatory, interpolatable, asymptotical, normalized, time-series, field and remote-specified, georeferenced, TB-related, endemic, transmission-oriented covariate coefficients. We defined our risk model by $g(x)$ employing $g$ the gradient of $f(x)$ and by $H(x)$ for quantitating the Hessian of $f(x)$. We denoted $bH > 0$, if a matrix $H$ was positive definite and/or by $H^0$, if $H$ was positive semidefinite. We employed a subscript $k$ for the $k$-th iteration, hence, $x_0$ represented the initial, geosampled, georeferenceable, operationizable, county-level capture points. According to Booth and Hobert [10], if $x_{k+1} = x_k + _k p_k^k$ is an iteration, where $k$ satisfies the Wolfe condition, then the Zoutendijk theorem states that $f(x)$ will be bounded below while still being continuously differentiable. One way to prove global convergence of a method is to use Zoutendijk theorem (Anderson, 1958). In our iteratively interpolatable, georeferenceable, explanatory, time-series, endemic, transmission-oriented, epidemiological, county-level TB model, the differentiable variable was expressed as $\{x \mid f(x) \leq f(x_0)\}$, where $\nabla f$ was Lipschitz continuous, (i.e., $\forall x, y \in Nk\nabla f(x) - \nabla f(y) \leq Lkx - yk$). In our model this procedure rendered $x_k$. The search direction $p_k$ at stage $k$ was given by the solution of the analogue of the Newton equation $B_k p_k = -\nabla f(x_k)$, where $B_k$ was an approximation to the Hessian matrix, which was updated iteratively at each stage in the seasonal, predictive, endemic, county-level, operationizable, TB time-series, explanatory, iteratively Interpolatable, epidemiological, risk-related, predictive risk-related model construction process when $\nabla f(x_k)$ was the gradient of the function evaluated at $x_k$. A line search in the direction $p_k$ was employed to find the next geosampled, clinical, field and/or remote-specified, geosampled, explanatory, endemic,
transmission-oriented, georeferenceable county-level sampled point $x_{k+1}$. Instead of requiring the full Hessian matrix at the point $x_{k+1}$ to be computed as $B_{k+1}$, the approximate Hessian at stage $k$ was updated by the addition of two matrices: $B_{k+1} = B_k + U_k + V_k$. Both $U_k$ and $V_k$ were then symmetric rank-one matrices but with different matrix bases. Fortunately, the symmetric rank assumption meant that we could write $C = ab^T, U_k$ and $V_k$ which in this research we did by employing a rank-two update matrix.

Usage of the Sherman-Morrison-Woodbury formula updated the linearized systems after low rank modifications of the system matrix were performed in SAS/GIS. However, it is well known that this formula can lead to serious instabilities in the presence of round off error (Anderson, 1958). If the system matrix is symmetric positive definite, it is almost always possible to use a representation based on the Cholesky decomposition which renders the same results (in exact arithmetic) at the same or less operational cost for numerically stabilizing regressed, time-series, stochastically/deterministically, geosampled, iteratively interpolatable, covariate coefficients (Anderson, 1958).

The Cholesky decomposition updated to incorporate low rank additions for robustly conducting low rank subtractions. We quantitated a special case of an indefinite update of rank two. The methods discussed here are well-known in the numerical mathematics literature, and the code for most of them can be found in the LINPACK suite. Matlab MEX implementations for most of the techniques we employed for qualitatively regressing the geosampled, explanatory, county-level TB, time-series interpolatable, georeferenceable, asymptotical, unbiased, explicative, covariate parameterizable estimators were downloaded at [http://www.kyb.tuebingen.mpg.de/bs/people/seeger/](http://www.kyb.tuebingen.mpg.de/bs/people/seeger/). We found our spatiotemporal, endemic, transmission-oriented, predictive, explanatory, clinical, field and remote-specified, endemic TB.
model was robust against the scale problem often suffered in the gradient descent searching (e.g., in Broyden’s method). The quasi-Newton condition imposed on this update was

$$B_{k+1}(x_{k+1} - x_k) = \nabla f(x_{k+1}) - \nabla f(x_k).$$

From an initial guess $X_0$ and approximate Hessian matrix $B_0$, estimators were repeated until $x$ converged to solution in the spatiotemporal, autoregressive, endemic, transmission-oriented, operationizable, TB, time-series, explanatory, iteratively interpolatable, georeferenceable, epidemiological, predictive stochastically/deterministically risk model which helped us render a direction $p_k$ by solving: $B_k p_k = -\nabla f(x_k)$. We performed a line search to find an acceptable step size $\alpha_k$ in the direction found in the first step, which then updated $X_{k+1} = X_k + \alpha_k p_k$. We set $s_k = \alpha_k p_k$, and solved for $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, $B_{k+1} = B_k + \frac{y_k y_k^T}{s_k}$, $B_k s_k y_k^T B_k - \frac{y_k y_k^T}{s_k y_k^T B_k s_k}$. The $f(x)$ denoted the objective function to be minimized in the explanatory, time-series, autoregressive, endemic, transmission-oriented, epidemiological, risk-related, endemic TB, predictor, model residually clinical, field and remote-specified stochastically/deterministically, iteratively interpolatable, georeferencable derivatives. Convergence was checked by observing the norm of the gradient, $|\nabla f(x_k)|$. Practically, $B_0$ can be initialized with $B_0 = I * x$, so that the first step will be equivalent to a gradient descent, but further steps are more and more refined by $B_k$ using the approximation to the Hessian (Anderson, 1958). The first step of the algorithm was thereafter carried out using the inverse of the matrix $B_k$ which in this research was obtained efficiently by applying the Sherman-Morrison formula to the fifth line of the algorithm. The formula rendered

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k^T y_k + y_k^T B_k^{-1} y_k)(s_k s_k^T)}{(s_k^T y_k)^2} - \frac{B_k^{-1} y_k s_k^T + s_k y_k^T B_k^{-1}}{s_k^T y_k}.$$

The Sherman-Morrison formula is a formula that allows a perturbed matrix to be computed for a change to a given matrix $A$. If the change
can be written in the form \( u \otimes v \) for two vectors \( u \) and \( v \), then the Sherman-Morrison formula for the two vectors \( u \) and \( v \) using the formula is \( (A + u \otimes v)^{-1} = A^{-1} - \frac{(A^{-1}u) \otimes (v \cdot A^{-1})}{1 + \lambda} \) (Cogo, 2011). For the two vectors \( u \) and \( v \), the Sherman-Morrison formula was \( (A + u \otimes v)^{-1} = A^{-1} - \frac{(A^{-1}u) \otimes (v \cdot A^{-1})}{1 + \lambda} \), where \( \lambda = v \cdot A^{-1}u \). In statistical estimation problems such as ML or Bayesian inference, credible intervals or confidence intervals for the solution can be estimated from the inverse of the final Hessian matrix (Baker et al., 2011). However, in our interpolatable, asymptotical, normalized, explanatory, time-series, predictive, autoregressive, endemic, transmission-oriented, county-level epidemiological, predictive, TB, risk model these intervals were technically defined by the true Hessian matrix, as such, BFGS approximation did converge to the true Hessian matrix.

The Poisson regression for the interactional, spatiotemporal, interpolative, explanatory, predictive, operationizable, time-series, clinical, field and remote-specified, endemic, transmission-oriented, risk-related, TB model, explanatory stochastically/deterministically, interpolatable, covariate coefficients. These results provided information for the time-series, explanatory, field and remote-specified estimates of the prior distribution based on the regressed, georeferenceable main effect coefficients for the Bayesian analysis. The values for parameter estimates and standard errors were used as mean values and standard errors to parameterize prior expected values for the geosampled endemic, transmission, oriented, stochastically/deterministically interpolatable county-level covariates. The prior expected mean value for the error term was assumed to be zero (0), with a standard deviation of 0.01. Initial values for the MCMC chains were generated.

This research for spatiotemporally, autoregressively, quantizing, the time-series, endemic, transmission-oriented, explanatory interpolatable, TB-related, georeferenceable, epidemiological,
risk model, three MCMC chains were then estimated for the intercept, which appeared to converge within the first 1,000 samples. The first 1,000 samples were discarded to allow the model to stabilize (i.e., known as “burn in”), and the next 10,000 samples were used to derive optimal stochastically/deterministically, parameter estimates based on statistical significance. The MCMC was able to numerically calculate multi-dimensional integrals. The multiple integral is a generalization of the definite integral to functions of more than one real variable, for instance, $f(x, y)$ or $f(x, y, z)$ (Anderson, 1958; Cogo, 2011). Just as the definite integral of a positive function of one variable represents the area of the region between the graph of the function and the $x$-axis, a positive function of two variables may also represent the volume of the region or between the surface defined by the function on the three-dimensional Cartesian plane where $z = f(x, y)$ and the plane contains a domain (Cogo, 2011). As such, we assumed that the same volume could be obtained via the integral of a function in three variables of the constant function $f(x, y, z) = 1$ over a region between the surface and the plane in the georeferenceable dataset of county-level TB, epidemiological, risk model interpolatable residually forecasted derivatives. If there were more explanatory, time-series, endemic, transmission-oriented, clinical, field and remote specified, operationizable TB variables, a multiple integral we assumed would yield hyper-volumes of stochastically/deterministically, multi-dimensional, functions. Multiple integration of a function in $n$ variables: $f(x_1, x_2, ..., x_n)$ over a domain $D$ is most commonly represented by nested integral signs in the reverse order of execution (the leftmost integral sign is computed last), followed by the function and integrand arguments in proper order (Anderson, 1958). Interestingly, the integral with respect to the right most argument in our MCMC risk model was computed last.

In the PROC MCMC methods, an ensemble of “walkers” moved around randomly. At each geosampled, time-series, operationizable, interpolatable, clinical, field and remote specified,
georeferenced, explanatory, county-level, TB, endemic, transmission-oriented, point where the walker stepped, the integrand value at that point was counted towards the integral. The walker then made a number of tentative steps around the area looking for a place with reasonably high contribution to the integral. Random walk methods are a kind of random simulation or Monte Carlo method (Anderson, 1958). However, whereas standardized random samples of the integrand use a conventional Monte Carlo integration which are generally statistically independent, those used in our PROC MCMC specified explanatory, time-series, dependent, clinical, field and remote-specified, autoregressive, stochastically/deterministically, operationizable, endemic, TB, transmission-oriented, georeferenceable, asymptotical, normalized, epidemiological, risk model were correlated. A Markov chain was constructed in such a way as to have the integrand as its equilibrium distribution. A improvement in model fit, was noted as variables were added to the Bayesian paradigm.

This specification moved the investigation towards a Bayesian map analysis given that the entire georeferenceable, empirical, dataset of time-series, georeferenced, clinical, field and remote-specified, county-level TB, epidemiological, endemic, transmission-oriented, predictive, risk-based, explanatory, stochastically/deterministically, iteratively interpolative, operationizable, covariates, with the exception of the intercept, were treated as single-valued. Intercept was treated as a distribution of values and was estimated by using empirical Bayes techniques.

Next, the difference in the deviances between a simple, explanatory, georeferenceable, predictive, endemic, transmission-oriented, epidemiological, explanatory risk model and the more complex model provided the improvement $\chi^2$ values. We examined all interaction between the geosampled, TB, time-series, clinical, field and remote, ecological covariates and found that an interaction model did not improve the fit therefore, no interaction terms were included in the
stochastically/deterministically, interpolative, operationizable, finalized model. We could not examine the improvement of fit between a saturated model and the full effects model as the number of the geosampled parameters that needed to be estimated exceeded the maximum number that could be parsimoniously regressed.

To derive the improvement of fit values listed in Table 5, the posterior mean deviance values were obtained with deviance information criterion (DIC) spatial analytical tools. We focused on a spatial consideration of the localized DIC measure for model selection and goodness-of-fit evaluation. We employed a partitioning of the DIC into the local DIC, leverage and deviance residuals, to assess the local model fit and influence of the geosampled, endemic, transmission-related, explanatory, stochastically/ deterministically, interpolatable time-series, county-level, endemic, TB observations in the Bayesian framework. The NL MIXED procedure computed three kinds of residuals. Residuals are available for all generalized linear models except multinomial models for ordinal response data, for which residuals are not available in SAS (http://support.sas.com). Raw residuals and Pearson residuals are available, however, in SAS for parsimonious model fit with generalized estimating equations. The raw residualized forecasts rendered from our time-series, explanatory, clinical, field and remote-specified, endemic, transmission-oriented, interpolatable, georeferenceable, asymptotical, normalized, risk-related, predictive, epidemiological, risk model was defined where the \( n \)th response was the corresponding predicted mean. We requested raw residuals in an output dataset using an OUTPUT statement. The Pearson residual then was the square root of the \( n \)th contribution to the Pearson’s chi-square. We then requested the operationizable time-series, clinical, field and remote-specified, explanatory, endemic, transmission-oriented Pearson residuals in an output dataset.

Finally, the deviance residual was defined as the square root of the contribution of the \( n \)th
and geosampled, observation to the deviance, with the sign of the raw, residual quantitated in the
OUTPUT statement. The adjusted Pearson, deviance and likelihood residuals were then defined.
These residuals were useful for outlier detection and for assessing the influence of single
observations on the fitted model. For our generalized, linearized, epidemiological, county-level
TB operationizable, time-series, explanatory, iteratively, interpolatable, risk model, the variance
of the \( n \)th individual observation was then given by the risk model where the dispersion parameter
was a user-specified prior weight and where the mean, and the variance function was
parsimoniously quantized. We also constructed a robust, diagonal, georeferenceable, matrix for
denoting the \( n \)th diagonal element in the empirical dataset. The weight matrix was then employed
in computing the expected information matrix. The Pearson residual were standardized to have
unit asymptotic variance. We requested standardized Pearson residuals in our output dataset
employing STDRESCHI in the OUTPUT statement. The deviance endemic, transmission-
oriented, residualized forecasts were standardized to have a unit asymptotic variance as tabulated
by the total deviance from our geosampled, regressed, explanatory, clinical, field and remote-
specified, TB, stochastically/deterministically, operationizable, time-series, explanatory, iteratively
interpolatable observations. We requested standardized deviance residuals in our output dataset
employing “STDRESDEV” in the OUTPUT statement. The likelihood residuals were thereafter
defined. We requested likelihood georeferenceable residuals in an output dataset “RESLIK” in the
OUTPUT statement. We then quantitated the local DIC to assist in model selection to visualize
the global and local impacts of adding covariates and other model parameters to the Bayesian
estimation matrix. DIC statistics were generated to identify the best fitting model.

The deviance was defined as \(-2 \log (\text{likelihood})\), where the ‘likelihood’ was defined as
\( p(y \mid \theta) \). This included all the georeferenceable normalizing constants where \( y \)

\[ p(y \mid \theta) \]
comprised all stochastic node values and theta stochastic parents of $y$ in the explanatory, autoregressive, time-series, endemic, transmission-oriented, clinical, field and remote-specified, county-level, stochastically/ deterministically operationizable, iteratively interpolatable, epidemiological, risk model. The definition of deviance is $-2^{*}\log$ (likelihood): ‘likelihood’ is defined as $p(y \mid \theta)$, whereby comprises all stochastic nodes given values (i.e., data), and comprises the of $y$ – ‘stochastic parents’ were the stochastic nodes are dependent upon the distribution of $y$ (Anderson, 1958). ‘Stochastic parents’ are the stochastic nodes upon which the distribution of $y$ depends upon when collapsing over all logical relationships (Crosby & Noar, 2011). In this research when $y$~$\text{Dnorm}(\mu, \tau)$, then $\tau$ was a function of a parameter $\phi$ which then defined the prior distribution in the endemic, transmission-oriented, georeferenceable, operationizable, interpolatable, epidemiological, predictive, endemic TB, risk model. Thereafter, the likelihood was defined as a function of $\phi$ in the risk model. The expectation $\mathcal{D} = E^{\theta}[D(\theta)]$ was used as a measure of model fitness based on the values of the endemic, transmission-oriented, interpolative asymptotical, normalized, parameterizable, covariate coefficient values. The effective number of parameters included in the model was computed as $p_{D} = \mathcal{D} - D(\bar{\theta})$, where $\bar{\theta}$ was the expectation of $\Theta$. The DIC generated the following conclusions: (1) the $D_{bar}$, was the posterior mean of the deviance, (2) the $D_{hat}$, was the point estimate of the deviance (i.e., $-2^{*}\log$ (likelihood)) obtained by substituting the posterior means theta bar of theta which then rendered $D_{hat} = -2^{*}\log (p(y \mid \text{theta. bar})$; and, 3) $p_{D}$ was the effective number of endemic, transmission-oriented, explanatory, time-series, stochastically/ deterministically operationizable, iteratively interpolative parameters provided by $p_{D} = D_{bar} - D_{ha}$ and $p_{D}$ employing the posterior mean of the deviance minus the deviance of the posterior means. In normal hierarchical models, $P_{D} = \text{TR}(H)$, where $H$ is the ‘hat’ matrix maps the observed data to their fitted values (Crosby & Noar,
2011). The DIC was then calculated as: $DIC = p_D + D$. The DIC value for the finalized, interpolative, stochastically/deterministically, operationizable, explanatory, time-series, county-level, endemic transmission-oriented, georeferenced, clinical, field and remote specified, epidemiological, risk model was 931.6.

Table 3: Improvement of fit of the PROC MCMC, hierarchical Bayesian model (HBM) endemic, county-level TB model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unadjusted effects</th>
<th>Adjusted effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Improvement</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>HOMELESS</td>
<td>1</td>
<td>1.133</td>
</tr>
</tbody>
</table>

Median parameter values, as well as the 95% credibility intervals (2.5 percentile and 97.5 percentile values), were generated in the residually, forecasted, autoregressive, endemic, transmission-oriented, clinical, field and remote-specified, stochastically/deterministically, operationizable, interpolatable, asymptotical, normalized, covariate, parameterizable, estimator dataset. As the sampling sites increased based on the georeferenced, explanatory, covariate, personal income of <20,000, the median log-count increased. The adjusted model robustly quantized the independence among the time-series, field and remote-geosampled, endemic, transmission-oriented, explanatory, operationizable, covariates representing the count data. We noted this model fit better that the model that adjusted for correlation within the study site based on the RMSE. All L components closely conform to a bell-shaped curve.
Figure 12: Top left (a): normal quantile plot for the random effects term. Top right (b): normal quantile plot for the spatially structured random effects term. Bottom left (c): normal quantile plot for the spatially unstructured random effects term.
Figure 13: Predicted probabilities of tuberculosis in Florida by county

The number of homeless was reviewed in each county during the study period. A random effects term was then specified using the sampled coefficients. This random effects term displayed no latent uncertainty autocovariate effects. The model’s forecasted residual error variance however, implied a substantial variability in the forecasted county level regressed seasonal prevalence rates.
The counties were then ranked by probability estimates (Fig. 10a). This allowed the counties to be ranked by number of future cases (Fig. 10b). The final model revealed that from 2015 to 2020 Duval, Orange and Broward counties would require immediate intervention in order to prevent TB-related transmission. The model also revealed that from 2025 to 2040 Hillsborough and Palm Beach counties could become hyper-endemic without implementation of control strategies.

Table 4: Counties with most predictive value

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Miami-Dade</td>
<td>Duval</td>
</tr>
<tr>
<td>Orange</td>
<td>Broward</td>
</tr>
<tr>
<td>Hillsborough</td>
<td>Palm Beach</td>
</tr>
</tbody>
</table>
CHAPTER 4

SELECTION OF STRATEGIES FOR INTERVENTION BASED ON A GIS PRIORITIZATION

Understanding Tuberculosis Transmission and its Endemicity

Intervention and strategies to intervene are generally aimed at reduction of incidence with possibility of elimination by specific and consistent intervention. In general, various strategies for intervention are collectively required to establish the following criteria:

- understanding the mode of transmission
- population at risk
- environment for transmissibility
- susceptible host
- identification of chain of infection
- strategies for interruption of chain of infection
- intervention with utilization of specific tools and enhanced/modern technology

Tuberculosis endemicity may be defined as a situation in which there is a greater number of infected individuals than would normally be expected, within a specific geographic area or specific population during a particular time period, while there is a great evidence of recent transmission of \textit{M. tuberculosis} among those cases or individuals.

In such situations, understanding the current incidence rate or knowing the exact number of incidence in comparison to the current prevalence is the key information for implementation of
intervention. Incidence is defined as number or cluster of new cases of TB infection whereas prevalence corresponds to the number of individuals already identified and diagnosed with tuberculosis infection.

The techniques used to detect and provide strategies to control, halt or eliminate are guided by knowledge of the transmission and pathogenesis of *M. tuberculosis* (in this case), and by the principles of effective intervention policy. Thus, in order to understand how to detect and intervene effectively, it is important to have a good understanding of the transmission and pathogenesis of tuberculosis and population at risk as well as available technology.

In the state of Florida, the Department of Public Health had developed a plan for tuberculosis elimination policy in conjunction with the Centers for Disease Control and Prevention (CDC) to bring the Florida tuberculosis incidence rate to 2 cases per 100,000 populations by the year 2020 through enhanced screening and a sturdy treatment plan (Baker et al., 2011).

**Global Vision**

The overall vision for control and elimination of global tuberculosis infection is not only considered one of the priorities for most developing countries, but it is also a great challenge in developed countries with significant incidence. It also presents as a major task expected to present its presence in industrialized nations. Many international organizations, private non-governmental organizations, as well as scholar and academic associations, provided various dates and deadlines to celebrate the elimination or control the spread of this infectious agent/disease. Unfortunately, these goals and the overall vision are still far from reaching reality.
Types of Tuberculosis

Tuberculosis is mainly divided into two major categories:

- Pulmonary tuberculosis
  
  Primary tuberculosis: Mainly diagnosed in a person who has never been diagnosed with a tuberculosis infection. Infection may spread to other organs hematogenously.
  
  Post-primary tuberculosis: Usually present in a person with a previous infection due to lack of proper treatment or reactivation of tuberculosis within five years after clinical diagnosis of primary infection.

- Extra-pulmonary tuberculosis
  
  This type mainly occurs in other organs or systems outside of the lungs. Infection may spread to other organs hematogenously. The most common system is the lymphatic system, but also includes other systems such as skeletal, gastrointestinal, genitourinary, and central nervous system.

Pulmonary Tuberculosis

Pulmonary tuberculosis infection in humans is mainly categorized in two types:

- Active Tuberculosis
- Latent Tuberculosis

Normally, tuberculosis spreads by droplet nuclei when an infected person has an unprotected cough, sneeze, and close respiratory contact. The chance of getting infected depends on the level of infectiousness of an infected individual, open or closeness of the environment, strength of the mycobacterium organism and duration of first hand exposure.
However, there are other forms of tuberculosis infections which spread to other organs and are grouped in the following table (Anderson, 1958).

Table 5: Category of pulmonary and extra-pulmonary tuberculosis

<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary TB Pneumonia</td>
<td>More common in immunocompromised individuals and people older than 60 years of age</td>
</tr>
<tr>
<td>Tuberculosis pleurisy</td>
<td>Plural space surrounding the lung is destroyed and causes chest pain</td>
</tr>
<tr>
<td>Tuberculosis pericarditis</td>
<td>Accumulation of excessive body fluid around the heart affects the general function</td>
</tr>
<tr>
<td>Tuberculosis meningitis</td>
<td>Very dangerous type of tuberculosis, usually affects the functioning of the brain</td>
</tr>
<tr>
<td>Osteal tuberculosis</td>
<td>Usually affects the bone: Mycobacterium organisms make bone weak and brittle with very unhealthy consequences</td>
</tr>
<tr>
<td>Peritonitis, gastrointestinal and laryngeal tuberculosis</td>
<td>Affect the intestinal lining</td>
</tr>
<tr>
<td>Lymph node infection</td>
<td>The lymph nodes become inflamed and painful</td>
</tr>
<tr>
<td>Adrenal tuberculosis</td>
<td>Causes major endocrine disturbances by altering normal functioning of the adrenal glands</td>
</tr>
<tr>
<td>Miliary tuberculosis</td>
<td>Invade extra-pulmonary organs such as the spleen, kidneys, liver and eventually the lymphatic and circulatory systems</td>
</tr>
<tr>
<td>Cavitary tuberculosis</td>
<td>Usually affect the upper part of the lungs but rapidly spread to the other part of the lungs</td>
</tr>
</tbody>
</table>

Extra-pulmonary Tuberculosis

While extra-pulmonary TB can occur in various areas of the body, a common risk factor for all of these various types is the HIV infection and status of the patient (immunocompromised), along with the previously diagnosed person with LTBI (Latent TB Infection) that had not presented with symptoms or a typical clinical presentation.
Various manifestations of extra-pulmonary tuberculosis symptoms and clinical presentations are as follows, but are not limited to outline below.

A. Lymphatic presentation: Lymph node tuberculosis affects the lymph nodes (small glands which are a part of the immune system) causing persistent swelling of the nodes (usually in the neck), which can lead to fluid release through the skin over time.

B. Skeletal TB: Particularly affects the spine and end of most long bones and joints (especially children), and is marked by bone pain, a curving of the affected joint or bone, a loss of movement or sensation in the affected joint or bone, and a weakening of the bone which may lead to easier fracturing (spinal tuberculosis).

C. Gastrointestinal TB: Symptoms of gastrointestinal TB include abdominal pain, diarrhea, and rectal bleeding. GI tuberculosis is considered a major problem in tuberculosis control in underdeveloped countries. Ulcerative types are the most common presentation.

D. Genitourinary TB: Include the entire genitourinary and reproductive systems. Common presenting symptoms include a burning sensation during urination, blood in the urine, frequent need to urinate during the night (often leading to misdiagnosis with Crohn’s Disease), and groin pain.

E. Central Nervous System (CNS) TB: Is usually an uncommon form of infection with a very disturbing result. Symptoms include headaches, stiffness, confusion, blurred vision, and possible seizures (8).

Common general factors such as poor living conditions, poverty, overcrowding, malnutrition, HIV co-infection, and living in proximity with infected animals are important
vehicles contributing to the spread of tuberculosis in general and specifically, increasing the rate of extra-pulmonary type.

**Transmission Parameters**

Transmissibility: Tuberculosis is an upper respiratory infection with very high transmissibility within any given population. The estimated risk of transmissibility by close contact with a patient with positive tuberculosis sputum is greater than 50%. On the other hand, the risk of transmissibility with a patient with negative sputum but low infectivity is greater than 5%. There is also approximately a 10% chance for an immunocompromised patient to get infected with tuberculosis within the first two years of infection and this rate increases as the immunosuppressive status carries on with more susceptibility each year. Factors such as time or duration of contact, closeness to the source of infection, and closed environment verses open environment are very influential parameters.

In the state of Florida, transmissibility mainly takes place among the following:

- Foreign born
- Substance abuse
- Homelessness
- Hazardous occupation (point of contact)
- Long term facility residences
- Correctional facility residences
- Immunocompromised
Florida Tuberculosis Case Report

Further analyzing the state of Florida’s tuberculosis information and available statistics, a marked dichotomy is reported or observed in cases over the past between reported cases from 2008 to 2012.

All case data obtained from the entire state of Florida Department of Public Health for the period encompassing 2008-2012 are all showing a decreasing trend in the number of TB cases reported in the state of Florida. All cases of tuberculosis reported are composed of the following categories as pulmonary tuberculosis, extra-pulmonary tuberculosis and a combination of both conditions (Cogo, 2011).

1. In 2008, a total number of 957 cases of tuberculosis were reported
   a. Pulmonary tuberculosis consists of 777 cases, which is equal to 81%.
   b. Extra-pulmonary tuberculosis consists of 152 cases, which is equal to 16%.
   c. The rest of cases are reported as a combination of both types of tuberculosis.

2. In 2009, a total number of 822 cases of tuberculosis were reported.
   a. Pulmonary tuberculosis consists of 667 cases, which is equal to 81%.
   b. Extra-pulmonary tuberculosis consists of 127 cases, which is equal to 15%.
   c. The combination of both types of tuberculosis was 28 cases, which is equal to 3%.

3. In 2010, a total number of 835 cases of tuberculosis with slight increases in comparison to last year were reported.
   a. Pulmonary tuberculosis consists of 652 cases, which is equal to 78%.
   b. Extra-pulmonary tuberculosis consists of 160 cases, which is equal to 19%.
   c. The combination of both types of tuberculosis was 23, which are equal to 3%.
4. In 2011, a total number of 749 cases of tuberculosis were reported. In this year, there was a significant trend of decreasing reported.
   a. Pulmonary tuberculosis consists of 562 cases, which is equal to 75%.
   b. Extra-pulmonary tuberculosis consists of 174 cases, which is equal to 23%.
   c. The combination of both types of tuberculosis was 2, which is equal to <1%.

5. In 2012, a total number of 679 cases of tuberculosis were reported with a slight increase in comparison to last year reported cases.
   a. Pulmonary tuberculosis consists of 543 cases, which is equal to 80%.
   b. Extra-pulmonary tuberculosis consists of 111 cases, which is equal to 16%.
   c. The combination of both types of tuberculosis was 25, which is equal to 4%.

Different levels of care accountability for tuberculosis are provided by the State of Florida Health Department (Goldhaber-Fiebert et al., 2011).

1. Function at level 1: **Local public health system**
   A. consists of local a public health system to identify, treat and manage 90% of all active cases of tuberculosis
   B. provides diagnosis, treatment, contact identification and follow up in home-county
   C. manages medical and social or cultural issues for moderately multifaceted patients

2. Function at level 2: **Local area tuberculosis network system**
   A. provides the care and management of 5% of active cases of tuberculosis
   B. supports special needs such as social and mental health further than activity of level 1
   C. makes access available to tuberculosis expert for consultation
   D. organizes individual cases and provides cohort study review

3. Function at level 3: **provide hospital services**
A. makes hospital services available to 5% of active cases of tuberculosis
B. provides inpatient hospitalization for non-compliant and moderately and severely complicated cases
C. makes access available to a highly specialized tuberculosis specialist

State of Florida Tuberculosis Program Act

The State of Florida Department of Public Health targets the objective tuberculosis elimination mission to reduce the state rate to 2.0 cases per 100,000 populations within next 5 years (2020). The following measures are taken in order to achieve the stated goal by 2020:

A. target for case rate by subset of population: Reducing the tuberculosis rate among the foreign-born despite being below the national rate.
B. reducing the rate of tuberculosis among the US born individual: This is currently higher than the national average; therefore, intervention is set by employing an enhanced quality improvement process.

Risk factors

1. Foreign born: Number of morbidity cases has increased, but the number of foreign born incidence in general, has decreased (Cuzick & Edwards, 1990).
2. Homeless population: Department of Health reported a 2% increase of tuberculosis case rate reported for the homeless population in general with comparison to previous years in table (7) (Cuzick & Edwards, 1990).
Table 6: Homeless. Information in this table is adapted from: Florida Department of Health Tuberculosis Report

<table>
<thead>
<tr>
<th>Data sources 2008 - 2012</th>
<th>Homeless cases of Tuberculosis within past year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2008</td>
</tr>
<tr>
<td>Year</td>
<td></td>
</tr>
<tr>
<td>Cases</td>
<td>64</td>
</tr>
<tr>
<td>Percent</td>
<td>7</td>
</tr>
</tbody>
</table>

3. **Substance abuse**: Alcohol related tuberculosis infection has increased in general with comparison with injection and non-injection drug users. Such comparisons are demonstrated in table 7, table 8 and table 9 respectively.

Table 7: Alcohol-related. Information in this table is adapted from: Florida Department of Health Tuberculosis Report.

<table>
<thead>
<tr>
<th>Data sources 2008 - 2012</th>
<th>Tuberculosis cases associated with excess alcohol use within past year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2008</td>
</tr>
<tr>
<td>Year</td>
<td></td>
</tr>
<tr>
<td>Cases</td>
<td>167</td>
</tr>
<tr>
<td>Percent</td>
<td>17</td>
</tr>
</tbody>
</table>
Table 8: Non-injected drug use. Information in this table is adapted from: Florida Department of Health Tuberculosis Report (9).

<table>
<thead>
<tr>
<th>Year</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>103</td>
<td>89</td>
<td>69</td>
<td>73</td>
<td>66</td>
</tr>
<tr>
<td>Percent</td>
<td>11</td>
<td>11</td>
<td>8</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 9: Injected drug use. Information in this table is adapted from: Florida Department of Health Tuberculosis Report (9).

<table>
<thead>
<tr>
<th>Year</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cases</td>
<td>16</td>
<td>5</td>
<td>16</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Percent</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

**Adopted Tuberculosis Treatment Guidelines**

The Florida Department of Public Health had established the tuberculosis control program policy with the main concern of aiming to cure all individuals diagnosed with active tuberculosis condition. The therapeutic completion rate for the State of Florida is currently above the national average rate with 89.2% in comparison to the national average rates of 87.7% (Fleiss et al., 2003). The future goal is aiming to reach the department of health tuberculosis treatment goal of 93% by 2015 (Fleiss et al., 2003).
Common Obstacles

There are various forms of obstacles presented in most interventions that make this endeavor almost unfeasible. Many factors such as deficient funding for community education and screening techniques are challenging factors for developing countries. The combination of low screening measures, utilization of traditional methods of disease surveillances and lack of predicting future incidence at the county level may be greatly contributing to the currently high prevalence of tuberculosis in the State of Florida. Therefore, the 2015 goal has not been successfully completed.

Figure 14: Forecasts of rank-ordered largest expected numbers of tuberculosis cases based upon projected county population

These counties were ranked by probability estimates (Fig. 2a). This allowed the counties to be ranked by the number of future cases (Fig. 2b). This project clearly demonstrates a better understanding of the current tuberculosis condition and clearly illustrates the predictive
information in the final model that from 2015 to 2020 Duval, Orange, and Broward counties would require immediate intervention in order to prevent TB-related transmission.

Table 8 Forecasts of rank ordered largest expected numbers of tuberculosis cases based upon projected county population.

This project clearly demonstrates the endemicity of tuberculosis infection at the county level in the entire state of Florida, with especial illustration of the hyper-endemicity of the following counties: Miami-Dade, Duval, Broward, Orange, Hillsborough and Palm Beach counties. The model also revealed that from 2025 to 2040, Hillsborough and Palm Beach counties could become hyper-endemic without implementation of control strategies.

Table 10: Counties with high potential for hyper-endemicity without intervention

<table>
<thead>
<tr>
<th></th>
<th>2015</th>
<th>2020</th>
<th>2025</th>
<th>2030</th>
<th>2035</th>
<th>2040</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miami-Dade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duval</td>
<td></td>
<td>Orange</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broward</td>
<td></td>
<td></td>
<td>Duval</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orange</td>
<td></td>
<td></td>
<td></td>
<td>Broward</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hillsborough</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palm Beach</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Prevention Policy and Screening Procedures**

To better understand prevention policy, it is essential to understand the general concept of prevention in public health practice. All prevention strategies are implemented based on three methods of prevention with a specific purpose or goal to achieve in each level.
- **Primary prevention**: Educate the public from getting infected with tuberculosis.

  This concept is the foundation of intervention, because the intervention strategy will target public education about deterrence of tuberculosis infection through schools, churches, faith groups, working places, and through general and social media to inform the entire population.

- Another intervention strategy that must be implemented in hyper-endemic counties is the use of Bacillus Calmette Guerin (BCG) vaccination for high risk groups. The BCG vaccine is not used and it is not recommended in childhood or adulthood required vaccination schedules recommended by Center for Disease Control and Prevention (CDC); therefore, it is not used in the United States. BCG is a live-attenuated vaccine and initially structured from mycobacterium bovis.

- Intervention strategies in hyper-endemic counties must include providing or implementing BCG vaccinations for all identified high risk groups with an increased chance of acquiring tuberculosis. Implementing BCG vaccination intervention not only reduces prevalence but also prevents complicated tuberculosis in at risk and high risk populations.

- BCG is used worldwide as a part of primary prevention. There is some concern about its efficacy, but one point is certain that BCG will help in prevention of severe complicated or disseminated tuberculosis specifically cerebral tuberculosis in children. Children under 2 years of age without use of the BCG vaccination are at an increased risk for developing Miliary and Meningitis tuberculosis, which is a life threatening situation.

- **Secondary prevention**: Intervention strategy is mainly focused on implementing screening procedures for early detection and proper referral to a specialized center for an accurate treatment model. Intervention strategy in hyper-endemic counties must target
tuberculosis diagnosed individuals and making sure that each patient goes through a 
completed treatment plan as scheduled or directed, with the highest degree of compliance 
either by hospitalization or by drug observation therapy (DOT) until a full disease-free 
state is documented. The use of medication for treatment of tuberculosis is mainly based 
on following drug therapies:
1. Isoniazid 
2. Rifampin 
3. Pyrazinamide 
4. Ethambutol 
5. Streptomycin 

- There is also a common therapeutic model of four drugs therapies composed of Isoniazid, 
  Rifampin, Pyrazinamide and Streptomycin. To complete the final cycle of intervention in 
  this step requires appropriately and effectively employing pharmacological treatment 
  either to stop the progression of this infective agent or completely cure the infection.

- **Tertiary prevention**: Intervention strategy is mainly concerned with rehabilitation of the 
  affected organ to evade incapability or loss of function. Intervention in hyper-endemic 
  counties must include compliance with a therapeutic regimen for prevention of drug 
  resistance tuberculosis along with post treatment evaluation to assure complete eradication 
  with no chance of latent reactivation. Another intervention strategy in hyper-endemic 
  counties must include post evaluation of treated individuals with co-morbidities such as 
  substance abuse, diabetes, or immunocompromised status. These groups must be evaluated 
  for each condition and tuberculosis infection separately because of an increased risk of 
  tuberculosis re-infection (approximately 10% each year).
Intervention strategy will include implementation of all three levels of preventions in all identified hyper-endemic counties to decrease prevalence rate and incidence rate. Utilization of such measures will result in the Florida Department of Health achieving the stated goal in combating tuberculosis for 2020.

Screening

To combat transmission of tuberculosis, the very first and essential step is identification of infected individuals through acceptable and effective screening procedures. Screening methods must have high sensitivity and high specificity properties to include the true positive individuals and exclude the true negative subjects among large populations. Therefore, implementing this project recommendation in all hyper-endemic counties will result in increased accuracy in screening procedures which is a comprehensive and accurate tool or method for identification of infected subjects and prediction of future epidemics.

There are different types of screening procedures commonly employed for identification of all subjects suspected of tuberculosis infection. Identification of true positive subjects through screening procedures is essential, because such procedures assist in providing proper treatment to prevent progression to irreversible conditions and it should be the mainstay of any program aiming for elimination or reducing incidence of tuberculosis in the community.

Resources are limited in most situations; therefore, it necessitates all preventive tuberculosis programs to employ the new technology that yields the most benefits such as Geographical Information System (GIS) for forecasting the future incidence.

Screening methods in tuberculosis programs will target populations at higher risk among the main population of any local community. The goal of ultimate elimination encompasses the
objective aiming to reduce the incidence, which requires proper screening techniques and rapid case identification.

Screening Intervention

1. Screening for identification of tuberculosis disease: Suggested screening is recommended for populations with a high prevalence of active pulmonary tuberculosis. Such populations include:

   - individuals born outside of the United States
   - seasonal workers and common immigrants especially from endemic countries
   - homeless populations, as well as individuals living in shelters and prisons
   - immunocompromised individuals
   - long-term facilities residents

Screening for tuberculosis disease must include existence of clinical presentation and common symptoms suggestive of tuberculosis infection. In this type of screening, the suspected patient must go under further clinical assessment and evaluation including chest radiograph and sputum culture examination for definite confirmation.

In some situations, screening and confirmation of tuberculosis are made based on chest radiography only. This is mainly due to sternness of patient's clinical presentation because sputum culture will take a longer time for confirmation of organism than chest radiograph.

In such instances, patient, family member, co-worker and public are at risk of acquiring infection, therefore a prompt treatment plan of anti-tuberculosis drug therapy is initiated according to the most recent recommendation guideline provided by the Center for Disease Control and Prevention (CDC) and American Lung and Thoracic Society, unless proved otherwise.
2. Screening intervention for Latent Tuberculosis Infection (LTBI): This intervention method targets screening for individuals in hyper-endemic counties suspected of previously diagnosed tuberculosis infection, but who did not complete the prescribed therapeutic course.

- Screening for LTBI is by use of the Mantoux Tuberculosis skin test (PPD test)
- Blood test by use of an Interferon Gamma Release Assay (IGRA)

After careful evaluation and assessment of clinical presentation, all suspected population groups such as immigrants, homeless and shelter groups, foreign born individuals from endemic countries, individuals with substance abuse issues, and immunocompromised individuals, must be evaluated for probability of compliance and completing the full treatment as recommended.

Mantoux Purified Protein Derivative or PPD test: Tuberculosis screening test, such as the Mantoux Purified Protein Derivative (PPD) skin test, is usually administered subcutaneously in the anterior surface area of the forearm in the amount of 5TU which is equal to 0.1cc of purified protein derivatives. General reaction of sensitivity of notable redness with well distinguished border on skin will appear on the injected site within two to three days. Tuberculosis disease must be excluded by chest radiography as well.

Figure 15: A TB screening test. Adopted from Center for Disease Control and Prevention: www.cdc.gov
Current guidelines for screening with PPD skin exam:

1. Indurations of $\geq 5$mm are mostly considered positive in the following patients:
   a. Subjects with Human Immunodeficiency Viral infection (HIV/AIDS)
   b. Patients with any recent organ transplant
   c. Patients with evidence of pulmonary fibrosis on chest radiograph

2. Indurations of $\geq 10$mm are mostly considered positive in the following patients:
   a. Patients with a history of recreation drug users especially injection drug users
   b. Patients who recently migrated (less than 6 years) to the United States from an endemic region
   c. Persons from the following locations are considered high risk populations:
      i. employees and residents of nursing home and long-term care facilities
      ii. employees and residents of local and federal penitentiaries and jails
      iii. employees and residents of healthcare facilities with patients diagnosed with HIV/AIDS
      iv. employees of infectious disease/ micro-bacteriology laboratory
   d. Persons diagnosed with following pathology:
      i. person with unexplained weight loss (10% of body mass or higher)
      ii. persons diagnosed with diabetes mellitus
      iii. persons diagnosed with chronic renal failure or hematological disorder
      iv. patients diagnosed with silicosis or other malignancies
      v. persons diagnosed with gastrointestinal tract surgery
      vi. persons living in crowded places such as shelters or homeless
vii. children less than 5 years of age or infants being exposed to parents or a family member with diagnosed tuberculosis

3. Indurations of $\geq 15$ mm are strongly considered positive in patients without any known risk factor for developing tuberculosis.

Other confirmatory screening laboratory tests and procedures for screening of LTBI:

- **Interferon Gamma Release Assay (IGRA)** is utilized by blood sample:
  
  This test can help screen and identify individuals infected with TB bacteria by measuring the immune system response to organism mycobacterium. The United States Department of Food and Drug Administration (FDA) approved and supported the utilization of two different IGRA$s$, QuantiFERON-TB Gold test and T-SPOT TB test.

  All positive patients identified with utilization of mentioned screening test strongly suggest the presence of TB infection; therefore, additional testing and screening is required to clearly distinguish infected individuals with active cases of pulmonary tuberculosis from individuals with latent tuberculosis infection (LTBI). Another way to utilize the IGRA test is to use it in patients with previous administration of Bacillus Calmette-Guerin (BCG) Vaccine. IGRA is also very essential and useful in rapid identification of population groups or individuals with difficulty for future follow-up to measure the PPD skin exam.

**3. Screening in patient with HIV infection:** Screening individuals diagnosed with HIV/AIDS infection is strongly recommended because tuberculosis co-infection with immunodeficiency status is extremely high. The screening procedure must initiate immediately after diagnosis of HIV infection with the tuberculosis skin test and IGRA test. Any suspected subjects with positive laboratory results must be followed up with a chest radiographic
examination and a sputum culture for exclusion. Any patient with a positive skin test or IGRA test but no abnormal radiographic finding must undergo preventive tuberculosis treatment.

4. Screening for mobile/moving groups such as seasonal workers and homeless: Screening for tuberculosis among such populations is extremely difficult due to the nature of their activity and mobility of residence status. Seasonal workers are usually living in crowded congregations and are continuously on the move to various rural areas and homeless are also on the move between various shelters and street places. An intervention strategy in all hyper-endemic counties for asymptomatic either homeless or seasonal workers without any immune deficiency diseases is to implement the Interferon Gamma Release Assay (IGRA) test due to lack of compliance and difficulty in follow up.

5. Screening for students of all ages from preschool age students to higher education students: Intervention strategy in all hyper-endemic counties screenings procedure for preschool students should be implemented. This intervention helps with identification of household infection and prevention of classroom transmission. Intervention strategy also applies to screening by PPD skin test for all university students and health care workers.

6. Screening for health care workers and residents of highly populated facilities and occupational exposure: Intervention strategy for long-term care facility, health care provider facilities and nursing homes in all hyper-endemic counties must screen all of their employees and residents for possible tuberculosis infection on unannounced basis sooner than regular timeline currently recommended by Center for Disease Control and Prevention (CDC).

7. Screening for patients with previous history of either infection or disease with cured treatment: Intervention strategy for all subjects in hyper-endemic counties that have been correctly diagnosed and treated successfully in the past, screening must be implemented on annual basis or
shorter if they were identified with other communicable diseases. To prevent re-infection, routine screening such as skin test or chest radiograph or use of Interferon Gamma Release Assay (IGRA) must be done even if they are asymptomatic.

8. Screening pregnant patients and screening during prenatal period: Intervention strategy for pregnant individuals in hyper-endemic counties is not recommended because pregnancy by itself dose does not categorize individuals in high risk groups for tuberculosis infection, but pregnancy with communicable disease or positive HIV mother screening test is recommended. Screening tests such as a skin test and IGRA blood tests are usually safe to administer during pregnancy but starting the patient with prevention therapy is delayed until after delivery. Therefore, intervention must include avoidance of close contact with family members and immunocompromised individuals, occupational transmission, public transportation, and public congregation until proper treatment is initiated and completed. However, in pregnant women such as those with HIV co-infection or close contact with active pulmonary tuberculosis must be screened with a skin test or IGRA blood test. If the result is positive, the preventive therapy must be initiated immediately without any delay despite pregnancy status.

**Intervention Strategies to Secure Healthy Outcomes for Screened Subjects**

Intervention strategies in hyper-endemic counties are tailored to suit each patient based on socioeconomic status, occupation, living condition, transmissibility and other co-infection or chronic disease such as diabetes. Infected low income individuals must be provided with all means of treatment, social support, transportation, and monitoring of therapy with regular and constant disease status evaluation until complete recovery.
**Objective of screening.** The main and clear stated objective of screening is to identify or detect the tuberculosis infection as early as possible. Early detection of infection will help the healthcare provider and diagnosed patient with the following benefits:

- Enhancement of provided treatment
- Prevention of poor treatment and long term health consequences
- Prevention of social and economic/loss of income consequences
- Reduction of patient’s morbidity, community prevalence and overall mortality rate
- Reduction in transmission in community due to short duration of infective tuberculosis of infected subjects
- Help to rule out active tuberculosis infection from latent tuberculosis infection
- Help for identification of all patients/subject at risk or at higher risk than general population
- Tuberculosis case review committee will monitor treatment plan data quarterly to ensure the efficacy of treatment policy.

**GIS Technology Intervention**

This research proposes the use of GIS technology intervention to enhance the screening procedure not only to target high-risk populations but also employ GIS technology to predict the future county level incidence throughout the entire state of Florida.

This study in turn will help to assist in better preparation and efficient use of available resources such as budgetary protocol and policy, for advance training, enhanced screening, and improved case identification and treatment plan.

The preliminary data analyzes revealed that counties in Florida can be ranked by predicted probability of tuberculosis related explanatory covariates. These probabilities, which had the
random effects term common through time, allowed determining optimal epidemiological forecasts.

These counties were ranked by probability estimates. Counties requires immediate attentions are Duval County, Orange County, and Broward County. Duval County, Orange County, and Broward County: The counties were the ranked by probability estimates (Fig. 2a). This allowed the counties to be ranked by number of future cases (Fig. 2b). The final model revealed that from 2015 to 2020 Duval, Orange and Broward counties would require immediate intervention in order to prevent TB-related transmission. The project model also revealed that from 2025 to 2040 Hillsborough and Palm Beach counties could become hyper-endemic without implementation of appropriate control strategies.

Table 11: Florida tuberculosis health outcomes and targeted goals for 2020. Florida System of Tuberculosis Care - Florida Department of Health, March 1, 2013

<table>
<thead>
<tr>
<th>Tuberculosis program performance Target</th>
<th>2012 Target</th>
<th>2013 Target</th>
<th>2014 Target</th>
<th>2015 Target</th>
<th>2020 Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>TB case rate per 100,000</td>
<td>3.8</td>
<td>3.7</td>
<td>3.6</td>
<td>3.5</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Florida Tuberculosis Health Outcome and Targeted Goal for 2020 in above table is clearly illustrate the performance target for past four years. It is clear that by simple calculation which consists of (0.1) percent per year improvement, the health department will have a long way to reach their stated goal of 2.0 cases per 100.00 populations by year 2020.
If we subtract 0.1 for the next 4 years from current 2015 case rate of 3.5 the remaining will be 3.0 tuberculosis case rate per 100,000 population; therefore, the Florida Department of Health is short another 10 years to reach the announced target goal for 2020. This analysis, that employs these project findings, will help health department tremendously in reaching stated goal and set more objective goals.

**Intervention Structure**

This project will ultimately help the Department of Public Health to achieve the Florida State goal with certainty. This project with its strong and proper tool (use of GIS technology) seamlessly intervenes and prioritizes strategies to assist all hyper-endemic counties by implementation of the following strategies at different levels. Intervention structures are done by levels:

1. Clinical services based on GIS prioritization
2. Community field services based on GIS prioritization
3. Community surveillance services based on GIS prioritization

**Level of clinical services intervention based on GIS prioritization.** At this level all identified hyper-endemic neighborhoods by this project via use of geographical information system (GIS) must be supplemented with health services clinical personnel such as public health practice nurses, physicians with specialty, microbiology laboratory personnel and field epidemiologists for an appropriate identification and diagnosis of infected individual with active tuberculosis in order to immediately initiate treatment at the point of identification. Services should include social and emotional support as well drug observational therapy (DOT) and follow up patient to achieve free disease state. Provide prompt tuberculosis therapy to all high risk groups
including patients with latent tuberculosis infection. Intervention strategies include monitoring the progression of implemented therapy by re-examination of patients with radiological services for latent tuberculosis infection and patients with primary infection under current therapy. All testing and diagnostic laboratory services such as culture of body fluids, sputum culture as well as other laboratory diagnostic measurement used for monitoring of therapeutic adverse reaction must be done within same community for accuracy and fast response.

**Community-field services program based on GIS prioritization.** At this level of intervention, the strategy is to have social services available and responsible for all community outreach activity designed for targeted population or neighborhoods at county level. The stated objective at this level of intervention is to correctly identify persons or families recently exposed to tuberculosis or at higher risk of exposure. Also, monitor compliance of patients currently under treatment by attending Direct Observational Therapy (DOT) method for assurance of therapeutic delivery. Other intervention strategy is to bridge the gap with community organizations such as institutional and correctional facilities, long term care facilities, homeless shelters, refugee shelters, migrant/seasonal workers, congregations, religious facilities. In addition, there is a need to provide age and educational level, specific tuberculosis health education materials in the forms of live presentation, social media, via public and private sectors for the entire county, among specifically targeted neighborhoods.

**Community surveillance services based on GIS prioritization.** At this level of intervention, the strategy focused on gathering and buildup the most important component of this project the Surveillance Report. The aim of intervention at this level is to collect and analyses all reported tuberculosis cases at the county level with special emphasis in targeted population or neighborhoods. All reported communicable diseases from private to public physician, diagnostic
laboratories, residential facilities, and hospitals are collected and analyzed for specific trend and chain of infection. Intervention must include cooperation link with all providers to ensure compliance with stated standard of pharmacotherapy for all cases of tuberculosis. Build up complete data base for each collected tuberculosis cases from all sort of social life including immigrants, seasonal workers, human trafficking, refugees, and follow up to ensure a disposition of given care for each diagnosed case. Other strategy is working collaboratively with department of health epidemiologist to further evaluate demographic trends.

Ultimately the model also revealed that from 2025 to 2040 Hillsborough and Palm Beach counties could become hyper-endemic without implementation of control strategies. Future population covariates associated to TB prevalence in Florida can be utilized by using Florida projected population by each county 2015 – 2040. Projections of Florida Population by County, 2015-2040, with estimates for 2014 which may furnish figures that support more invasive elucidative, explanatory forecasts (e.g. county-level hyper-endemic clusters). These figures may then allow the counties to be ranked by number of future cases. With the use of ArcGIS analysis combined with epidemiological surveillance data may aid in identifying geographic locations of localized transmission at the county level. These methods may then be employed to enhance targeted screening and control efforts with the goal of interruption of disease transmission and ultimately incidence reduction.
CHAPTER 5
CONCLUSION

Community intervention in general, requires knowledge competency, accuracy of data and utilization of available highly developed technology. This project illustrates the presence of proper and systemic method along with high accuracy tool (geographic information system model) and trained personnel to accurately identify and predict the community at risk but it is not widely employed currently in the state of Florida as well as other industrialized nations despite the existence of such technology.

Understanding the current endemic situation of tuberculosis within a community is extremely important factor because, such information illustrates the sickness or healthiness of our living community. One of the enhanced methods with most accuracy and accountability tools utilized to identify such information in a given community is to employ epidemiological surveillances by process of Geographical Information System (GIS).

GIS is an automated system to capture, store, recovery and examineas well as exhibit longitudinal or spatial data, to note in addition specific information relating to variables that accept different standards at diverse places (Anderson, 1958). GIS-based utilized intensely provides spatial analyses by which a space-time scan can statistically characterize geographic distribution patterns of TB endemics suspected on a county-level clustered geographic location within the entire state of Florida. Spatial, temporal, and space-time scan statistics are now commonly used to detect and evaluate disease clusters for a variety of
diseases, including different types of cancers, Creutzfeldt-Jakob disease, granulocytic ehrlichiosis, sclerosis and common chronic condition such as diabetes.

In using the spatial regression modeling, it assists in clear demonstration of the fact that there is an association that exists in a spatial dependency at the county level. TB incidences in Florida are results from confounding socioeconomic variables. These variables are including the percentages of recent immigrants, homelessness, average household income and other contributory factors.

This project also utilize the temporality of the quantified clustering TB infectious processes at the county-level employing a spatial-time lag model for calculating TB incidences from specified time frames based on cumulative effects of local TB transmission with contagious diffusion among the county-level communities.

GIS analysis and spatial statistics combined with epidemiological surveillance are considered an effective method for identifying tuberculosis transmission not identified during regular standard contact tracing methods currently practiced in most counties in the state of Florida.

The application of these methods can be utilized in areas where contact tracing is routinely performed within the counties. These methods may enhance targeted screening and control efforts, with the goal of interruption of disease transmission and ultimately incidence reduction.

The GIS modelling approach of this research may thus demonstrate that using GIS and spatial statistics can identify previously undetected TB transmission at the county-level in Florida. These results may also be used to design new targeted screening efforts. Studies of these efforts may be then employed to demonstrate if identifying focal areas for targeted screening has utility in reducing TB transmission at the county-level in Florida. For example, the knowledge about the
geographic location and the presence of county-level forecasted high prevalence clusters of TB in Florida can help the Department of Public Health to develop and implement more effective Tuberculosis Control program. This strategy will help to intensify their remedial measures in identifying geolocation areas of neighbouring county-level with high tuberculosis prevalence clusters and construct future strategies for more effective TB control at the county level.

This project eventually will be a valuable research tool for problems involving distance, location and population health dimensions in the field of epidemiological study, medical geography, public and environmental health, community medicine and particularly health service development in targeted community.

Utilization of geographical information system (GIS) ultimately can serve as a superior tool to add to current TB-related molecular techniques and conventional epidemiology to exemplify pathways and illustrate geographical locations with high prevalent clusters. Geographical information system (GIS) provide valuable resources for helping with rational development of community-based care by providing epidemiological predictive risk maps to locate potential hyper-endemic transmission-focal points.

In providing health care services and community planning, geographical information system (GIS) may thus greatly assist in the development of strategy, implementation of theory and model of delivery systems. GIS may also be utilized for determining predictor variables that are associated with TB for further statistical analysis. In using GIS analysis, combined with epidemiological surveillance, data may help to identify geographic location of localized transmission.

This method can then be used to enhance statistical targeted screening and control efforts, with the goal of interruption of disease transmission and ultimately incidence reduction in our study.
population. Multiple linear regression analysis techniques, coupled with normal probability models, have been historically the standard epidemiological tool to examine sampled clinical and environmental predictor variables associated with TB, specifically for identifying important covariates associated with high-risk populations (e.g., homelessness, alcoholism, immigrants, seasonal workers etc.)

In summary this project provides

- Better understanding of the spatial epidemiology of endemic tuberculosis will help health department in Florida to adapt and to construct enhanced policy in formulating county-level prevention and control strategies.
- Recognizing geographic clustering of tuberculosis in early stage will provide implications for investigating underlying etiology, and real-time intervention.
- Produces new methodology for rapid identification of new and under-detected cases.
- The power of mapping and GIS sensing will result in discovery of preciseness of locality available high and low clusters at each county level.

The final aim of this research is to envision an endemic, TB-related web-based interface for use by public health officials in Florida. This will include generating essential information such as real-time syndrome-based reporting tool. Such web-based information will regulate automated and immediate 'Alerts' to public health officials, doctors, hospitals and local community.

It serves as specific web based tool for tuberculosis data entry and professional data sharing among health care providers. It provides accurate reflection of existing information to assist policy makers in development of public health policy.

The use of geographical information system (GIS) enables researchers to entirely appreciate the power of mapping with greater efficiency. Ultimately enhances clear decision-making and
certainly improve record keeping and communications, therefore; this method provide great potential to be utilized as an effective global model for future research and intervention.

**Implications for Public Health Research**

In future, epidemiological, county-level, TB, ArcGIS research should focus on a likelihood approximation which may be parsimoniously obtainable for the matrix logarithm in a SAS/GIS (i.e., PROC MCMC) autoregressive quantification of the covariance matrix, via Bellman's iterative solution. The Bellman–Ford algorithm is an algorithm that computes shortest paths from a single source vertex to all of the other vertices in a weighted digraph. In endemic, probabilistic, epidemiological, county-level TB, forecast modelling and more specifically in graph theory, a directed graph (or digraph) is a graph (that is a set of vertices connected by edges), where the edges have a direction associated with them. In formal terms, a directed, county-level, diagnostic, autoregressive, county-level TB graph is an ordered pair \( G = (V, A) \) where: 1) \( V \) is a set whose elements are called vertices, nodes, or points, 2) \( A \) is a set of ordered data pairs of vertices, called arrows, directed edges (sometimes simply edges with the corresponding set), directed arcs, or directed lines (see www.sas.edu). Such an endemic, time series, epidemiological, TB, forecast-oriented, vulnerability map would differ from an ordinary or undirected TB graph, in that the latter may be optimally definable in terms of unordered pairs of vertices, which may be cartographically illustratable as edges, arcs, or lines in ArcGIS. The aforementioned definition would not allow an endemic, county-level, georeferenceable directed graph to have multiple arrows with same source and target nodes, but some data analyst or epidemiologists may consider a broader definition that allows directed TB graphs in ArcGIS to have such multiple arrows (namely, they allow the arrows set to be a multiset). More specifically, these entities may be addressed as directed, time series,
county-level, predictive, endemic, epidemiological, TB forecast graphs (e.g., multidigraphs). On the other hand, the aforementioned definition may allow a directed, endemic, county-level, TB graph to have loops (that is, arrows that connect nodes with themselves). More specifically, directed endemic, county-level, vulnerability, forecast, TB graphs without loops may be addressed as simple directed graphs, while directed graphs were loops are addressed as loop-digraphs.

**Limitations**

While conjugate Bayesian inference in our diagnostic, explanatory, geo-spatiotemporal, Gaussian-oriented, graphical forecast, vulnerability, county-level, TB models is largely solved, employing non-decomposable, epidemiologically geosampled clinical, field or remote-specified variables, these explanators may pose difficulties with the specification of suitable priors and the evaluation of normalizing constants for targeting, vulnerable, county-level, classified populations (e.g., migrant worker camp sites) inconspicuously. A data analyst or epidemiologist may derive the DY-conjugate prior for county-level, non-decomposable, forecast-oriented, endemic, TB-related, probabilistic paradigms which may reveal geosampled, vulnerability, geo-spatiotemporal, parameterizable covariates as a generalization to an arbitrary graph G of the hyper inverse Wishart distribution. In particular, if G is an incomplete, prime, endemic, TB, diagnostic, time series mutigraph it would constitute a non-trivial generalization of the inverse Wishart distribution. Inference based on marginal likelihood would require that the county-level, forecasted, TB data evaluation of a normalizing constant require an importance sampling algorithm for its computation. Indeed, this would be a limitation to the usage of an autoregressive, forecasting, probabilistic, geo-spatiotemporal paradigm. Hence, propagational inconspicuous uncertainties (e.g., latent negative autocorrelation coefficients) would be embedded in the Bayesian regression
values. In such circumstances it may be advisable to treat the model output with an eigenfunction decomposition algorithmic, weighted paradigm. In so doing, the algorithmic would validate all non-zero autocorrelation forecasts which would then geostatistically confirm georefernced, geolocations of vulnerable TB county-level population. Negative autocorrelation refers to aggregation is geospace of dissimilar georeferenceable attributes. Examples of structural learning involving non-decomposable, endemic, time series, TB vulnerability, epidemiological, forecast model residual’s may deal efficiently with the set of all positive definite matrices with non-decomposable, zero-autocorelation pattern whilst introducing the operation of triangular completion of an incomplete triangular matrix. Such a device may turn out to be extremely useful both in validating results for the implementation of the Monte Carlo procedure without negative autocorrelation traits constricting a county-level, endemic, TB, predictive, county-level map.
REFERENCES


