International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

# Spatial Temporal Modelling of Tuberculosis in Kenya Using Small Area Estimation

**Kipruto H[1], Mung'atu J[2], Ogila K[3], Adem A[3], Mwalili S[5], Kibuchi E[6], Ong'ang'o JR[7], Sang G[8]**

[1, 2, 3, 5]Jomo Kenyatta University of Agriculture and Technology, Nairobi Kenya, Department of Statistics and Actuarial Sciences. P.O BOX 62000-002000, Nairobi

[4]Mathematics and Physics Department, Technical University of Mombasa, Kenya, P.O Box 90420-80100 G.P.O ,Mombasa, Kenya

[6, 8] KEMRI-Wellcome Trust Research Programme, P.O Box 43640 - 00100, 197 Lenana Place, Off Lenana Road, Nairobi, Kenya.

[7]Kenya Medical Research Institute, Centre for Respiratory Disease Research, Mbagathi Rd. Nairobi, Kenya, P.O. BOX 54840 – 00200 Nairobi

**Abstract:** ***Tuberculosis (TB) is a disease which continues to be of public health importance, and is characterized with varying distribution across regions depending on socio-economic status, HIV burden and efficiency of the health system. Kenya is one of the 22 high burden countries that contribute 80% of global TB burden. This study aims to assess spatial and temporal distribution of TB using small area estimation methodology. The spatial reference regions considered were the 47 Kenyan counties. The small area estimates were mapped to produce smear positive TB and favorable treatment outcomes maps. The covariates considered were gender, HIV positive proportion, directly observed Treatment (DOTs), average weight, average Body Mass Index (BMI) and average age. The significant covariates were used in the model to generate the relative risks, posterior probability means and the associated standard deviations which were then used to generate the spatial temporal maps. The spatial temporal maps generated showed distribution clustering of TB cases in a number of counties over the years (2012-2014). From the results of all notified TB cases, only average BMI was excluded from the spatial temporal model since it was not statistically significant (P-value > 0.05). The variables gender, HIV+, DOTs by, Weight, BMI and age were included in the spatial temporal model. The estimated risk of case notification rates per 100,000 found to be highest in the following counties Marsabit, Isiolo, Nairobi, Lamu, Mombasa, Machakos, Kajiado, Makueni, Kisumu, Siaya and Homabay The risk maps from the small area estimates can be used by policy makers to target and develop interventions which address the real challenges which occur in the public health arena.***

**Keywords:** Tuberculosis, HIV, Small area estimation, spatial, temporal, Poisson, Generalized linear models, Kenya

## 1. Introduction

Tuberculosis has had a long history with the human race. Mostly Tuberculosis (TB) is likely to have affected humans for most of their history, [1]. It remains a major cause of death worldwide despite the discovery of effective and affordable treatment regimens more than 60 years ago, [1].

Tuberculosis (TB) continues to be one of the major public health threats globally [2]. It is caused by *bacillus* bacteria and the most common causative organism is the *Mycobacterium tuberculosis*. Other causative agents that have been noted are *Mycobacterium bovis* which is transmitted through contaminated milk and *Mycobacterium africanum*, [3]. Bacteria causing TB is transmitted through infectious aerosolized droplet nuclei generated by coughing, laughing, talking, sneezing and singing, [4]. The ability to generate infectious aerosolized droplet nuclei is dependent on the infectivity of the patient where a sputum smear positive patient is considered most infectious, [5]. Infection with the *Mycobacterium* does not always lead to development of disease as the immune system is able to contain the infection and the bacilli remain dormant. The risk of infection is dependent on the extent to which exposure happens; longer durations of exposure to infected persons who are not on treatment increase the chance of infection, [5]. The most at-risk to develop TB include children under five, the old and those who are immunosuppressed, [4].

Globally there were 1.3 million TB deaths (including TB deaths in HIV positive individuals) in 2012 [2]. TB and HIV continue to be the top causes of death from a single infectious agent globally, [6]. Unfortunately tuberculosis continue to take its toll amongst the most productive members of the community in both morbidity while still remaining to be a leading killer among adults in the most economically productive age groups and people living with HIV, [4] and [7]. Even though curable, the disease significantly reduces the quality of life and at times it becomes a lifetime sequelae that substantially reduce their quality of life, [8]. Recognition of these facts has kept TB control high on the global infectious disease public health agenda.

In the African region which constitutes 54 countries it contributes 26% of the Global burden of Tuberculosis making the African region to be ranked second after Asia which contributes 59% of the Global case load, [9].

Kenya is among the twenty two (22) TB high burden countries in the world which contribute 80% of the global TB burden, [9]. The absolute number of TB cases notified increased exponentially in the nineties recording an almost tenfold increase since 1990 by the end of the year 2007. TB incidence has increased from below 50 per 100,000 in 1990 to 329 per 100,000 populations in 2008, [10]. The HIV epidemic in Kenya has been the single most contributing factor for this unprecedented increase in the burden of TB in Kenya. From the Kenya AIDS Indicator Survey 2013, the prevalence of HIV in Kenya currently stands at 5.6%. [11],

while the TB HIV co-infection rate was at 39% in 2012, [10]. To combat the challenges of TB epidemic in Kenya there has been massive scale up of treatment and diagnostic facilities, [10].

In Kenya, TB is not distributed uniformly with certain regions recording higher notification rates than others, the Kenyan TB program does not provide services based on those areas with the greatest notifications but rather on a uniform strategy, [10]. A uniform approach to TB control is not the most efficient or effective approach hence the need to have a more in-depth look and design strategies that fits different epidemiological patterns and risk factors customised to each the region. Recent studies have suggested problems with treating an entire region with one control strategy, rather than targeting high-risk areas with more effective control measures, [12]-[15].

Research to evaluate the spatial distribution of TB and to identify high-risk areas still remains quite limited especially in developing countries, [16], [15] and [17]. It's widely known that TB control efforts are best formulated and implemented when areas of high notifications are known and well documented. It is equally important to identify areas where notification rates are abnormally high given underlying risk factors. In particular, county-based surveillance systems have demonstrated that the distribution of endemic diseases is also determined by social processes that are linked and closely related to the space where they are notified to occur in, [14] and [17]. Tools in spatial statistics have advanced our understanding of the geographic distribution of disease and improved the focus of public health actions, [12]-[15] and [17]. In this study, we investigated the county spatial patterns of Tuberculosis notifications in Kenya and the relationship between TB and HIV status, and used these data to identify those areas in Kenya that are at risk of highest TB incidence.

## 2. Materials and Methods

This study was carried out through detailed and extensive review of existing TB/HIV data from the National TB control program databases. The Kenyan National TB control program is well known for maintaining a robust surveillance system which maintains data on notification from the whole country. Since the year 2012, Kenya has maintained a case based surveillance system. A 3–year retrospective County based surveillance study of all reported TB cases was conducted in Kenya between 2012 and 2014. Kenya is divided into 47 administrative units referred to as counties and covers an area of 582,650 $km^2$ with an estimated population of 44 Million people in 2013.

Figure 1 shows the distribution of administrative counties in Kenya. Analysis of the data and generation of the maps were based on the administrative counties as shown in figure.


**Figure 1:** Distribution of Administrative Counties in Kenya

After retrospectively extracting anonymized data the National TB control Program database for the periods 2012 to 2014, the data was subjected to exploratory data analysis (EDA) to look at the emerging trends that was inherent in the data. The datasets that did not have all the requisite variables were omitted from the analysis. The data was managed using MS access and excel was used to create graphs, and tables to depict the changes in TB epidemiology over the years in the presence of HIV.

## 3. Methodology

The models used for describing the distribution of an observation in space and time are usually formulated within a hierarchical Bayesian framework with two main approaches namely empirical and full Bayes. Empirical Bayes approach provides parameter estimates by maximizing their posterior distribution using penalized quasi-likelihood (PQL) techniques while the full Bayes provides the posterior distribution of the target parameters by assigning hyperpriors which take care of model uncertainties [18]. In this study the Bayesian Hierarchical Generalized Linear Mixed Models (BHGLMMs) which are used in small area estimation because of their ability to incorporate multiple levels of model dependencies [19]-[20] was considered.

The BHGLMMs falls in subclass of structured additive regression (STAR) models known as latent Gaussian models and their response variable is usually non Gaussian and belongs to an exponential family [21]-[22]. The posterior marginals of the latent Gaussian models are not available in closed form for non-Gaussian observation models [23]. In such models the common approach to inference is the Markov Chain Monte Carlo (MCMC) which exhibits poor performance in terms of convergence and computation time. Integrated nested Laplace approximation (INLA) developed by [23] and based on nested Laplace approximations is a new approach for Bayesian inference on latent Gaussian models. It has an excellent performance in terms of good accuracy and reduced computational time, [24]- [26] and [23].

## 4. Generalized linear Mixed Model

The main assumption of GLMMs is that the distribution of the response variable $Y_i$ belongs to an exponential family of the form $Y_i/\theta_i,\phi_1 \sim p(\cdot)$ where $p(\cdot)$ is a member of the exponential family defined as,

$$p(y_i/\theta_i,\phi_1)=\exp\left(\frac{y_i\theta_i-b(\theta_i)}{a(\phi_1)}+c(y_i,\phi_1)\right) \quad (1)$$

for $i=1,...,n$ observations and $\theta_i$ is the scalar canonical parameter. The mean $\mu_i=\mathrm{E}(y_i/\boldsymbol{\beta}, f^i(\cdot),\phi_1)$ can be linked to the structured additive predictor which accounts for various covariates $\eta_i$ by a monotonic link function $g(\cdot)$ such that

$$g(\mu_i)=\eta_i=\beta_0+\sum_{i=1}^{n} f_i(\mu_i)+\sum_{k=1}^{m}\beta_k x_{ki}+\varepsilon_i \quad (2)$$

where $f_i(\cdot)$ are unknown functions of the covariates used to model temporal and spatial dependencies and also used to relax the linear relationships of the covariates. The $\beta_k$'s represents the linear effects of the covariates $x$'s and $\varepsilon_i$'s are unstructured terms.

Latent Gaussian model a flexible and large class of statistical models is obtained by assigning a Gaussian prior to $\beta_0$ , $f_i(\cdot)$, $\beta_k$ and $\varepsilon_i$. This can be represented as $\Theta=(\beta_k\text{'s}, f_i\text{'s},...)$ where $\Theta$ is unobserved multivariate Gaussian random variable, whose density $\pi(\Theta/\phi)$ is controlled by a vector of hyperparameters $\Phi$ (Rue & Martino, 2007). The latent Gaussian field $\Theta$ is assumed to have a Gaussian distribution with zero mean and variance covariance matrix $Q(\phi_2)$; with vector of hyperparameters defined as $\Phi=(\phi_1,\phi_2)$ which are not necessarily Gaussian [27], [23] and [20].

Latent Gaussian model is composed of three elements namely; the likelihood of the data $\pi(\mathbf{y}/\Theta)$, the Gaussian density of the random vector $\Theta$, $\pi(\Theta/\Phi)$ and the prior distribution of the parameter vector $\pi(\Phi)$. The posterior is therefore defined as

$$\pi(\Theta,\Phi/\mathbf{y})\propto\pi(\Phi)\pi(\Theta/\Phi)\prod_{i=1}\pi(y_i/x_i,\Phi) \quad (3)$$

The main inferential interest involves computing the posterior marginals for $x_i$ and posterior marginals for $\Phi$ or some $\Phi_j$. The approaches for Bayesian inference on latent Gaussian models are MCMC sampling and INLA. The high dimensionality of the latent field $\Theta$ and the strong correlation within $\Theta$ and between $\Theta$ and $\Phi$ especially when the numbers of observations are many leads to problems in convergence and computation time. INLA developed by [23] bypasses MCMC entirely by basing inferences on closed form approximations making it computationally efficient compared to MCMC. In the next section we describe INLA methodology.

### INLA Methodology

This is an approximate inference based method for approximating the posterior marginals of the latent Gaussian field $\pi(x_i/\mathbf{y}), i=1....n$ in three steps.

The posterior marginals of the latent effects $\Theta$ and hyperparameters $\Phi$ are written as

$$\pi(x_i/\mathbf{y})=\int\pi(x_i/\Phi,\mathbf{y})\pi(\Phi/\mathbf{y})d\Phi \quad (4)$$

$$\pi(\Phi_i/\mathbf{y})=\int\pi(\Phi/\mathbf{y})d\Phi_{-j} \quad (5)$$

The posterior marginals $\tilde{\pi}(x_i/\mathbf{y})$ and $\tilde{\pi}(\Phi_j/\mathbf{y})$ can be approximated using the Laplace approximation. The first approximation to $\pi(\Phi/\mathbf{y})$ using Gaussian distributions is constructed as follows

$$\tilde{\pi}(\Phi/\mathbf{y})\propto\left.\frac{\pi(\Theta,\Phi,\mathbf{y})}{\tilde{\pi}_G(\Theta/\Phi,\mathbf{y})}\right|_{\Theta=\Theta^*(\Phi)} \quad (6)$$

where $\tilde{\pi}_G(\Theta/\Phi,\mathbf{y})$ is a Gaussian approximation to the full conditional of $\Theta$ and $\Theta^*(\Phi)$ is the mode of the full conditional for $\Theta$, for a given value of $\Phi$ [23]. It involves locating the mode of $\tilde{\pi}(\Phi/\mathbf{y})$ which is used to integrate out the uncertainty with respect to $\Phi$ when approximating the posterior marginal of $x_i$.

The posterior marginals of the latent field are supposed to start from $\tilde{\pi}_G(x_i/\Phi,\mathbf{y})$ and approximate the density of $x_i/\Phi,\mathbf{y}$ with the Gaussian marginal derived from $\tilde{\pi}_G(\Theta/\Phi,\mathbf{y})$, i.e.

$$\pi(x_i/\Phi,\mathbf{y})=N(x_i;\mu_i(\Phi),\sigma_{ii}^2(\Phi)) \quad (7)$$

The marginals of the interest can be computed using numerical integration over a multidimensional grid of values of $\Phi$

$$\tilde{\pi}(x_i/\mathbf{y})=\sum_k\tilde{\pi}(x_i/\Phi_k,\mathbf{y})\tilde{\pi}(\Phi_k/\mathbf{y})\Delta_k \quad (8)$$

where the sum is over the values of $\Phi$ with area weights $\Delta_k$ [28].

The first step in INLA computation involves approximating the posterior marginal of $\Phi$ by using Laplace approximation in equation (6)

The second step involves computing the Laplace approximation of $\tilde{\pi}(x_i/\Phi,\mathbf{y})$ for selected values of $\Phi$ which improves the Gaussian approximation in equation (7).

$$\tilde{\pi}_{LA}(x_i/\Phi,\mathbf{y})\propto\left.\frac{\pi(\Theta,\Phi,\mathbf{y})}{\tilde{\pi}_{GG}(\Theta_{-i}/x_i,\Phi,\mathbf{y})}\right|_{\Theta_{-i}=\Theta^*_{-i}(x_i,\Phi)} \quad (9)$$

where $\tilde{\pi}_{GG}(\Theta_{-i}/x_i,\Phi,\mathbf{y})$ is a Gaussian approximation to $\Theta_{-i}/x_i,\Phi,\mathbf{y}$ around its mode $\Theta_{-i}(x_i,\Phi)$. An improved version of $\tilde{\pi}_{LA}(x_i/\Phi,\mathbf{y})$ known as Simplified Laplace

approximation was developed by [23]. It involves a series of expansion of $\tilde{\pi}_{LA}(x_i/\Phi,\mathbf{y})$ around $x_i=\mu_i(\Phi)$ which corrects for skewness and location and it is also less computationally expensive [23].

The third step involves combining steps 1 and 2 using numerical integration in equation 8 (For more details, [23].

**Spatial Effects**

Suppose that the index $s\in(1,2,...,S)$ represents the geographically connected regions. The spatially correlated effects in INLA are introduced by assuming that neighboring regions are more alike than two arbitrary regions. Two regions $s$ and $s'$ are neighbors if they share a common boundary. The spatial smoothness prior for the function evaluation $f(s)$ is given by

$$f(s)/f(s'),s\neq s',\tau^2\sim N\left(\frac{1}{N_s}\sum_{s\sim s'}f(s'),\frac{\tau^2}{N_s}\right) \quad (10)$$

where $N_s$ the number of neighbours of region where $s\sim s'$ indicates that two regions $s$ and $s'$ are neighbours and $\tau^2$ is the variance parameter [29]-[30].

**TB Distribution in Kenya**

In this study a generalized linear mixed model assuming a Poisson distribution with spatial and temporal random effects was used to characterize the relationships between TB cases and covariates. In this model the response variable is generated by a Poisson process:
The Poisson regression model is

$$y_i\sim Poisson(\mu_i)$$
$$\mu_i=\exp(\mathrm{X}_i\beta+offset_i)$$

where $y_i$ is the observed TB cases , $\mathrm{X}_i$ are the covariates for the $i$th observation and offset term represented the $i$th population (StataCorp, 2013).
The generalized linear mixed model used to describe the TB cases $y_j$ is of the form:

$$g(\mu_i)=\beta_0+\sum_j\beta_jX_{ij}+f_{trend}(time)+f_{str}(S_i)+f_{unstr}(S_i) \quad (11)$$

Where $g(.)$is a monotonic link function, $\beta_j$ represents the parameter vector of the covariates $X_{ij}$, $f_{trend}$ is trend component, $f_{\text{str}}$ and $f_{\text{unstr}}$ are structured and unstructured spatial effects of the county. The $f_{\text{str}}$ and $f_{\text{unstr}}$ were assigned a Markov random field prior and Gaussian i.i.d respectively. The spatial effects were estimated at county level in which a household was located and Kenya counties' boundaries were used to compute the neighborhood information. The prior for $f_{trend}$ was first order random walk. The covariates were assigned default INLA Gaussian priors while the default inverse gamma hyperpriors were assigned for the random effects. The structured spatial effects considered were of the Besag Intrinsic conditional autoregressive (ICAR) model while the unstructured spatial effects were of Unstructured IID model and temporal effect. The parameters of interest which then were found to be significant were used to calculate the relative risks which were mapped into the different geographical areas for the different years.

## 5. Results

To have a clear understanding of regional localized variations in the burden of Tuberculosis, the study looked at the variations from 3 different perspectives namely; distribution of TB cases, smear positive cases and the treatment outcomes in the presence of risk factors. To obtain the covariates that were included in the spatial temporal model used to obtain relative risks, a generalized linear model was used in each of the three models.

**All TB cases**

Table 1 shows that only average BMI was excluded from the spatial temporal model since it was not statistically significant (P-value > 0.05). The variables gender, HIV+, DOTs by, Weight, BMI and age were included in the spatial temporal model. Table 2 shows the predicted mean risk, 2.5% quadrant, 97.5% quadrant and its associated standard deviation. Table 3 shows the model with, temporal effect and both structured and unstructured spatial effects. Table 4 shows exponentiated model results of fixed effects.

**Table 1:** GLM model All TB CASES Results

| | Estimate | Std Error | z-value | P value |
|---|---|---|---|---|
| Intercept | -13.595 | 3.986 | -3.411 | 0.001 |
| Gender (props) | 4.321 | 0.232 | 18.593 | 0.001 |
| HIV+(props) | 1.245 | 0.102 | 12.256 | 0.001 |
| Dots by props | 8.4 | 4.003 | 2.098 | 0.036 |
| Average Weight | -0.024 | 0.006 | -4.135 | 0.001 |
| Average BMI | 0.005 | 0.004 | 1.211 | 0.226 |
| Average Age | -0.105 | 0.005 | -20.013 | 0.001 |

**Spatial Temporal Model**

**Table 2:** Fixed Model Results

| | Mean | Sd | 0.025quant | 0.975quant |
|---|---|---|---|---|
| Intercept | -6.428 | 0.848 | -6.420 | -4.787 |
| Gender (props) | 0.515 | 0.554 | 0.515 | 1.603 |
| HIV+(props) | 0.346 | 0.353 | 0.352 | 1.025 |
| Average Weight | 0.004 | 0.013 | 0.004 | 0.030 |
| Average Age | -0.041 | 0.015 | -0.042 | -0.042 |
| DOTS BY | 0.001 | 0.000 | 0.002 | 0.001 |

**Table 3:** Time: Random Walk 1 model

| Model Hyper parameters | Mean | Sd | 0.025quant | 0.975quant |
|---|---|---|---|---|
| structured spatial effect | 4.077 | 1.002 | 3.971 | 6.338 |
| Time | 5616.823 | 8358.246 | 27073.973 | 555.377 |
| unstructured spatial effect | 18406.828 | 18286.296 | 66605.713 | 3358.016 |

Effective number of parameters =47.92: Deviance Information Criterion=407.49

**Table 4:** Exponentiated Model Results

| | Mean | sd | 0.025quant | 0.975quant |
|---|---|---|---|---|
| Intercept | 0.002 | 2.334 | 0.000 | 0.008 |
| Gender (props) | 1.674 | 1.741 | 0.564 | 4.969 |
| HIV+(props) | 1.414 | 1.424 | 0.695 | 2.788 |
| Average weight | 1.004 | 1.013 | 0.979 | 1.031 |
| Average age | 0.959 | 1.015 | 0.932 | 0.988 |
| DOTS BY | 1.001 | 1.002 | 1.000 | 1.002 |

**Smear Positive Tuberculosis cases**

Smear positive cases are the infectious forms of TB and it is important to understand the variations in the relative risks among the different counties. From table 5 it shows that gender, HIV+ and average age covariates were statistically significant to be included spatial models (P-value <0.05). While average BMI, average weight and DOTS were not statistically significant (P-value > 0.05).

Table 6 shows the predicted mean 2.5% quadrant and 97.5% quadrant, and its associated standard deviation. Table 7 shows the model with both structured and unstructured random effects, table 8 shows the model results when the fixed effects models results in table 6 were exponentiated.

**Table 5:** Smear Positive TB GLM model Results

| | Estimate | Std Error | z-value | P-value |
|---|---|---|---|---|
| Intercept | -0.388 | 4.233 | -0.092 | 0.927 |
| Gender (props) | 4.238 | 0.216 | 19.605 | 0.001 |
| HIV+(props) | 1.320 | 0.105 | 12.580 | 0.001 |
| DOTs by props | -6.360 | 4.233 | -1.502 | 0.133 |
| Average Weight | -0.006 | 0.005 | -1.157 | 0.247 |
| Average BMI | 0.005 | 0.004 | 1.103 | 0.270 |
| Average Age | -0.092 | 0.006 | -15.725 | 0.001 |

**Table 6:** Spatial and temporal model Fixed Model Results

| | Mean | sd | 0.025quant | 0.975quant |
|---|---|---|---|---|
| Intercept | -5.976 | 0.585 | -7.129 | -4.835 |
| Gender (props) | 0.021 | 0.481 | -0.924 | 0.962 |
| HIV+(props) | 0.621 | 0.396 | -0.171 | 1.387 |
| Average Age | -0.038 | 0.016 | -0.038 | -0.007 |

**Table 7:** Time: Random Walk 1 model

| Model Hyper parameters | Mean | s.d | 0.025quant | 0.975quant |
|---|---|---|---|---|
| structured spatial effect | 3.064 | 0.725 | 1.87 | 4.7 |
| Precision for Time | 9403.734 | 41500 | 182.19 | 60883.48 |
| unstructured spatial effect | 18854.859 | 18500 | 1258.555 | 67555.95 |

Effective number of parameters =48.52 Deviance Information Criterion=417.88

**Table 8:** Exponentiated Model Results

| | Mean | s.d | 0.025quant | 0.975quant |
|---|---|---|---|---|
| Intercept | 0.003 | 1.794 | 0.003 | 0.008 |
| Gender (props) | 1.0207 | 1.617 | 1.021 | 2.618 |
| HIV+(props) | 1.870 | 1.486 | 1.882 | 4.001 |
| Average age | 0.963 | 1.016 | 0.962 | 0.993 |

**Favorable treatment Outcomes**

In Tuberculosis control favorable outcomes play an important part in providing an understanding on the performance of the TB control programs. It is important to understand the distribution of favorable outcomes throughout the county. From table 9 it shows that all the covariates were statistically non-significant (P-value > 0.05). Thus all the covariates were omitted from the spatial temporal model.

Table 10 presents the results of the 2.5% quadrant and 97.5% quadrant, its associated mean and standard deviation. Table 11 shows the model with both structured and unstructured random effects. Table 12 shows the model results when the fixed effects models results in table 10 were exponentiated.

**Table 9:** GLM model Results for favorable treatment outcomes

| | Estimate | Std Error | z-value | P-value |
|---|---|---|---|---|
| Intercept | 0.034 | 0.418 | 0.081 | 0.935 |
| Gender (props) | -0.079 | 0.290 | -0.273 | 0.785 |
| HIV+(props) | -0.148 | 0.123 | -1.205 | 0.228 |
| Average Weight | -0.003 | 0.007 | -0.037 | 0.971 |
| Average BMI | 0.0005 | 0.004 | 0.118 | 0.906 |
| Average Age | -0.003 | 0.007 | -0.456 | 0.648 |
| Nutrition props | 0.029 | 0.059 | 0.482 | 0.630 |

**Fixed Model Results**

**Table 10:** Spatial and temporal fixed effects model

| | Mean | sd | 0.025quant | 0.975quant |
|---|---|---|---|---|
| Intercept | -0.1502 | 0.0112 | -0.1722 | -0.1282 |

**Table 11:** Time: Random Walk 1 model

| Model Hyper parameters | Mean | sd | 0.025quant | 0.975quant |
|---|---|---|---|---|
| spatial effect | 19777.28 | 18826.60 | 69638.15 | 2.913 |
| Time | 20138.93 | 20142.33 | 2649.11 | 75694.55 |
| unstructured spatial effect | 23138.93 | 19033.72 | 1665.00 | 70555.62 |

Effective number of parameters =2.108 Deviance Information Criterion=611.46

**Table 12:** Exponentiated Model Results

| | Mean | sd | 0.025quant | 0.975quant |
|---|---|---|---|---|
| Intercept | 0.861 | 1.011 | 0.842 | 0.880 |

**Spatial Temporal Maps**

To display the distribution of notified cases, case notification rates, relative risks and TB/HIV distribution spatial temporal maps were produced. The calculations of the relative risks were premised on the significant covariates found to be significant and used in spatial temporal models used to generate the relative risks, posterior probability means and the associated standard deviations which facilitated the generation of the spatial temporal maps.

Figure 2 shows the spatial temporal distribution over the three years (2012-2014) for all forms of TB and smear positive TB show similar pattern for the notified TB cases with close to 10 counties showing a huge TB burden which remained constant over the three years. These counties could benefit from more focused TB control strategies.

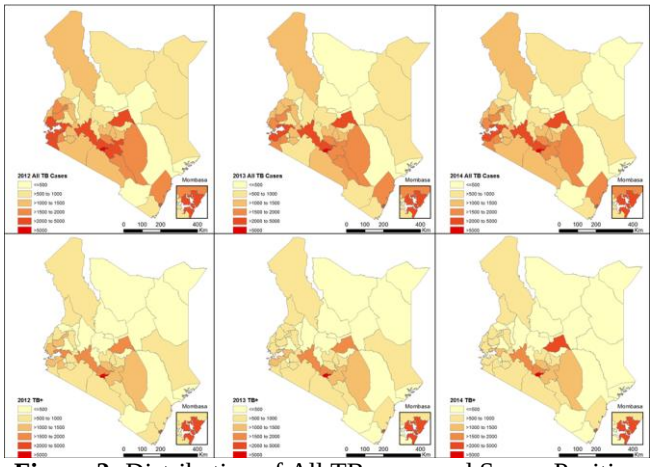

**Figure 2:** Distribution of All TB cases and Smear Positive TB Cases: 2012-2014

Figure 3, shows maps of the case notification rates for all TB forms and smear positive tuberculosis were considered the picture changes significantly for additional number of counties, clearly showing where more effort should be put by the national TB control program. It shows that the highest case notification rates were in the following counties; Nairobi, Mombasa, Lamu, Machakos, Kisumu, Isiolo, West Pokot increasing in the last two years.

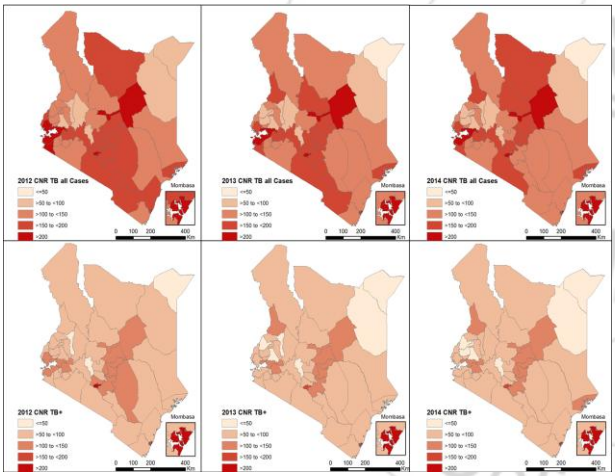

**Figure 3:** Case Notification Rate for All TB cases and Smear Positive TB Cases: 2012-2014

From Figure 4, it shows that the estimated risk of case notification rates per 100,000 was highest in the following counties Marsabit, Isiolo, Nairobi, Lamu, Mombasa, Machakos, Kajiado, Makueni, Kisumu, Siaya and Homabay and it shows that over the three years the estimated risk has remained unchanged and it seems the dynamics of TB disease has not been addressed in these counties.

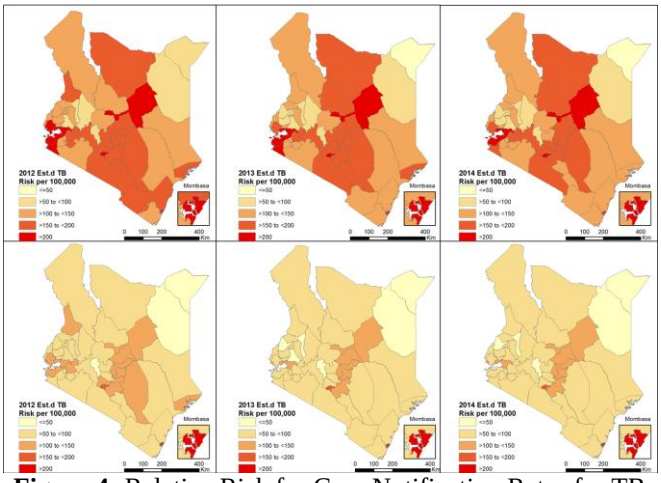

**Figure 4:** Relative Risk for Case Notification Rates for TB cases and Smear Positive TB Cases: 2012-2014

Figure 5, shows that the estimated risk all forms of TB and smear positive TB were consistent with the results of relative risk for the case notification rates (figure 4) shows that the highest relative risk being reported in the following counties Marsabit, Isiolo, Nairobi, Lamu, Mombasa, Machakos, Kajiado, Makueni, Kisumu, Siaya and Homabay. It further shows that over the three years it seems the dynamics of TB disease has not been addressed in these counties.

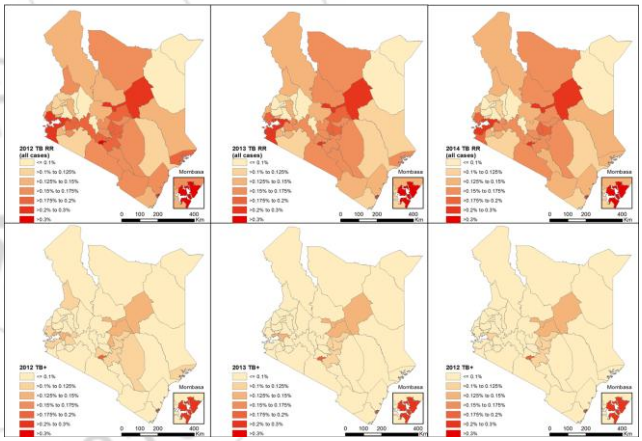

**Figure 5:** Relative Risk for All TB cases and Smear Positive TB Cases: 2012-2014

In figure 6 it shows that HIV continues to be a major factor in spatial temporal distribution of tuberculosis given its major contribution to the burden of TB on HIV patients. The distribution shows that most of the counties with high burden of TB have also high burden of HIV. It is however notable that although not all high burden counties are high HIV it is important to then understand the dynamics of what is driving the TB epidemic in those counties.

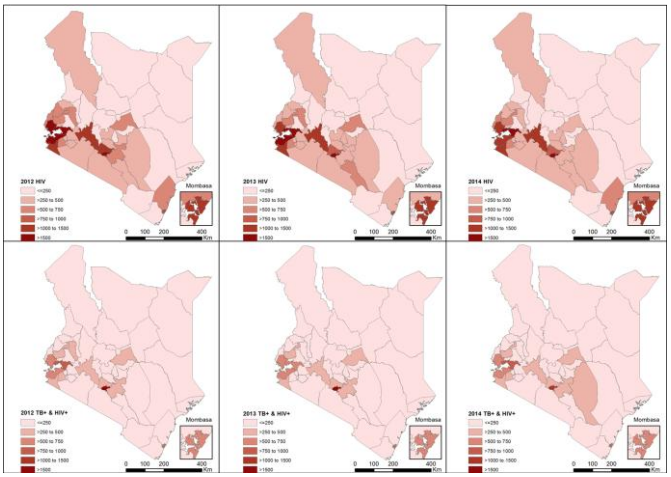

**Figure 6:** TB HIV distribution

In this figure 7, shows that all the counties continue to post very good treatment outcomes herein referred to as recovery rates. In the two years where the treatment outcomes were available all counties reported recovery rate greater than 85%. This result is very encouraging given the fact that for successful TB control you require to cure or successful treat almost all you TB patients if you are to significantly cut down the rates of transmission in the communities.

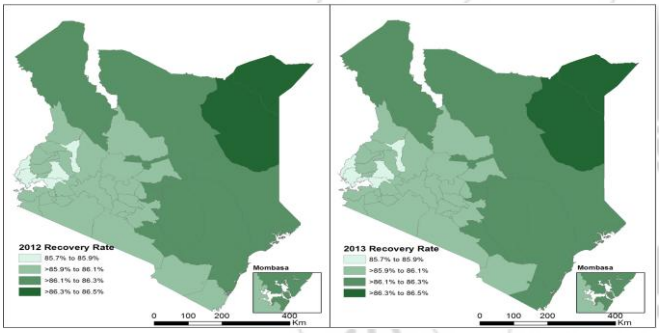


**Figure 7:** Distribution of Recovery Rates: 2012-2013

## 6. Discussion

The results show the inequality of TB disease distribution; both all TB forms and smear positive in Kenya. The results clearly show the counties (Nairobi, Mombasa, Isiolo, Homa bay, Kisumu and Siaya at risk of severe TB disease in the future. In this study it has been shown that some of the key covariates that were found to be significant were HIV+, gender and age. These covariates are some of the known risk factors for developing TB disease. Although this study did not examine the role of socio economic status it showed that TB tends to be more pronounced in those areas where population is highest and also where there is high incidence of HIV. Other studies have shown that there is a strong correlation between the measures of income, education and social vulnerability [31]-[33].

While several studies have attempted to utilise spatial analytical tools using the GIS framework to describe the pattern of various infectious diseases in the African region [34] -[38], only a handful of studies have attempted to obtain spatial temporal review of tuberculosis.

It has been noted that with improved surveillance and improvement in the access to and utilization of TB control services could lead to decrease in the transmission ongoing in the communities since with access and utilization of services, individuals who are sick are able to access care quickly hence cutting down the rates of transmission. One possible hypothesis is that those areas showing high burden to TB could be having challenges of access and utilization of TB services, [39]. [40]. In this study it has been demonstrated that national programs can utilize spatial temporal tools to identify hot spots for TB for improved interventions. With this illustration of cluster analysis and mapping of the disease distribution it adds value beyond that which can be obtained by presenting the disease notification rates or notification numbers reported in a table, as a cluster analysis helps identify areas with unusually high disease rates, which have less likely occurred by chance, [41].

It has been documented that improving TB control efforts could help reduce the transmission and change the geographic distribution of TB, [42]. Further, despite different intervention programs aimed at reducing disease transmission and improving case detection over many years, the unusually high rates of the disease could persisted in the same places, with the most likely spatial clusters showing a stable pattern if the actual dynamics of disease transmission are not addressed, [43]. Most of the studies reviewed did not include the mapping of recovery rates which provides an idea of cure rates which provides insights into the removal rates of the infectious cases in the population; figure 7 shows also that those counties with high relative risks also have lower recovery rates although higher than the target of 85%.

Despite the interesting results, this study was faced with a number of challenges which included understanding the relationship between the economic status and tuberculosis distribution. Further there was need for understanding the health seeking behavior of patients, access to and utilization of health services in particularly TB control services. These are the areas that require further research to further understand the dynamics with govern the occurrence of Tuberculosis transmission.

## 7. Conclusion

It is very clear from this study that the growing field of small area estimation epidemiology helps to account for changing TB risk in space and time.. The results show that small area estimation enables mapping of TB risk, which can be used by policy makers to target and develop interventions which address the real challenges which occur in the public health arena.

The results indicate that TB risk distribution in Kenya is different among the counties which demand a differentiated set of interventions. Special focus should thus be given to the identified “hotspots” and more focused interventions geared towards these areas. The study joins a small but growing number of studies with utilising spatial and temporal effects in mapping the underlying factors that influence Tuberculosis dynamics.

## References

[1] Holloway KL, Henneberg RJ, deBarros Lopes M, Henneberg M. "Evolution of human tuberculosis: A systematic review and meta-analysis of paleopathological evidence". Homo 62: 402–458, 2011
[2] WHO. Global tuberculosis report. Geneva: World Health Organization, 2013
[3] Cadmus S, Palmer S, Okker M, Dale J, Gover K, . "Molecular analysis of human and bovine tubercle bacilli from a local setting in Nigeria." *Journal of clinical microbiology* **44**: 29–34, 2006
[4] Sitienei J, Nyambat V, and Borus P, "The Epidemiology of Smear Positive Tuberculosis in Three TB/HIV High Burden Provinces of Kenya (2003–2009)," Epidemiology Research International, vol. 2013, Article ID 417038, 7 pages, 2013.doi:10.1155/2013/417038
[5] Dooley S.W.; Villarino M.E, Lawrence M, Salinas L, Amil S, Rullan J.V, Jarvis W.R, Bloch A.B, Cauthen G.M. Nosocomial transmissions of tuberculosis in a hospital unit for HIV-infected patients. *Journal of the American Medical Association.;* **267**:2632-2634.1992
[6] Ortblad, K. F., Lozano, R., & Murray, C. J. L. "The burden of HIV: insights from the Global burden of Disease Study 2010". *AIDS (London, England)*, **27**(13), 2013
[7] Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. "Global and regional burden of disease and risk factors, 2001: Systematic analysis of population health data." Lancet 367: 1747–1757, 2006.
[8] MillerTL,McNabb SJ, Hilsenrath P, Pasipanodya J, Weis SE. "Personal and societal health quality lost to tuberculosis". PLoS ONE4, 2009: e5080.
[9] WHO. Global tuberculosis report. Geneva: World Health Organization, 2012
[10] DLTLD, Division of Leprosy, TB and Lung Disease Annual Report. Nairobi: Kenya Ministry of Health, 2013
[11] GoK, Kenya Aids Indicator Survey, 2013
[12] Assunção RM, Barreto SM, Guerra HL, Sakurai E. "Mapas de taxas epidemiológicas: uma abordagem Bayesiana." Cad Saúde Pública;14:713–723, 1998
[13] Anselin L. "Local indicators of spatial association—LISA." *Georg Anal*. 1995;**27**: 91–115q, 1995
[14] Bailey TC, Gatrell AC. Interactive spatial data analysis. Harlow, UK: Longman Scientific & Technica, 1995;
[15] Bishai WR, Graham NM, Harrington S, et al. "Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy." *JAMA*. ;**280**:1679–1684, 1998
[16] Vieira RC, Prado TN, Siqueira MG, Dietze R, Maciel EL. "Spatial distribution of new tuberculosis cases in Vitória, state of Espírito Santo, between 2000 and 2005." *Rev Soc Bras Med Trop*. ;41:82–86, 2008
[17] Verver S, Warren RM, Munch Z, et al. "Transmission of tuberculosis in a high incidence urban community in South Africa". Int J Epidemiol. 33:351–357, 2004.
[18] Gómez-Rubio, V., & López-Quılez, A. Empirical and Full Bayes estimators for disease mapping. In International workshop on spatio-temporal modelling (METMA3), 2006
[19] Cnaan, A., Laird, N.M., & Slasor, P. "Using the General Linear Mixed Model to analyze unbalanced repeated measures and longitudinal data." *Statistics in Medicine*. V.**16**,2349-2380, 1997
[20] Fong, Y., Rue, H. & Wakefield, J. "Bayesian inference for generalized linear mixed models." *Biostatistics,V.0*, n.0, 1-16 , 2009
[21] Umlauf, N., Adler, D., Kneib, T., Lang, S., & Zeileis, A. Structured additive regression models: An R interface to BayesX (No. 2012-10). Working Papers in Economics and Statistics, 2012
[22] Klinker, F. Generalized linear mixed models for ratemaking: A means of introducing credibility into a generalized linear model setting. In Casualty Actuarial Society E-Forum, Winter 2011 Volume 2, 2010
[23] Rue, H., Martino, S., & Chopin, N. "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." Journal of the royal statistical society: Series b (statistical methodology), 71(2), 319-392, 2009
[24] Grilli, L., Metelli, S., & Rampichini, C. "Bayesian estimation with integrated nested Laplace approximation for binary logit mixed models." Journal of Statistical Computation and Simulation, (ahead-of-print), 1-9, 2014
[25] Taylor, B. M., & Diggle, P. J. "INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes." Journal of Statistical Computation and Simulation, (ahead-of-print), 1-19, 2013.
[26] Cameletti, M., Lindgren, F., Simpson, D., & Rue, H. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. AStA Advances in Statistical Analysis, 97(2), 109-131, 2013
[27] Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. Bayesian computing with INLA: new features. Computational Statistics & Data Analysis,67, 68-83, 2013
[28] Rue, H., & Martino, S. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of statistical planning and inference*, **137**(10), 3177-3192, 2007
[29] Martino, S., & Rue, H. Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the Inla program.Department of Mathematical Sciences, NTNU, Norway. 2009
[30] Brezger, A., Kneib, T., & Lang, S. "BayesX: Analyzing Bayesian structured additive regression models." *Journal of statistical software*, **14** (11), 1-22, 2005
[31] Roza, DL, Caccia B, Martinez EZ, "Spatio-temporal patterns of tuberculosis incidence in Ribeirão Preto, State of São Paulo, southeast Brazil, and their relationship with social vulnerability: a Bayesian analysis*." Rev. Soc. Bras. Med. Trop*., vol.**45**, n.5, pp. 607-615. ISSN 0037-8682, 2012
[32] Corbett EL, WattCJ, Walker N,MaherD,Williams BG, Raviglione MC,et al. "The growing burden of tuberculosis: global trends and interactions with the HIV epidemic." *Archives of internal medicine*.; **163**(9):1009–21, 2003. PMID:12742798
[33] Houben RM, Crampin AC, MallardK, Mwaungulu JN, Yates MD, Mwaungulu FD, et al. "HIV and the risk of tuberculosis due to recent transmission over 12 years in Karonga District, Malawi. "*Transactions of the Royal Society of Tropical Medicine and Hygiene*.

**103**(12):1187–9, 2009.doi: 0.1016/j.trstmh.2009.03.013PMID:19362727

[34] Rogers DJ & Williams BG. Monitoring trypanosomiasis in space and time. *Parasitology* **106** (Suppl), S77–S92, 1993

[35] Beyers N, Gie RP, Zietsman HL et al. "The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community." South African Medical Journal 86, 41–44, 1996

[36] Munch Z, Van Lills W, Booysen CN, Zietsman HL, Enarson DA &Beyers N. "Tuberculosis transmission patterns in a high incidence area: a spatial analysis." *The International Journal of Tuberculosis and Lung Disease* **7**, 271–277,2003

[37] Gaudart J, Poudiougou B, Ranque S, Doumbo O. Oblique decision trees for spatial pattern detection: optimal algorithm and application to malaria risk. *BMC Medical Research Methodology* **5**, 22, 2005

[38] Gaudart J, Poudiougou B, Dicko A et al. (2006). Space-time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village. *BMC Public Health* **6**, 286, 2006

[39] Tanrikulu' AC, Acemoglu H., Palanci' Y, Danli. CE. "Tuberculosis in Turkey: high altitude and other socio-economic risk factors." *Public Health Journal of The Royal Institute of Public Health* 2008;122,:613–9, 2007. doi: 10.1016/j.puhe.2007.09.005

[40] Vargas MH, Furuya ME, Perez-Guzman C. "Effect of altitude on the frequency of pulmonary tuberculosis." The international journal of tuberculosis and lung disease: the official *Journal of the International Union against Tuberculosis and Lung Disease*; **8**(11):1321–4, 2004. pmid:15581199

[41] Cromley EK, McLafferty SL .GIS and Public Health. Second edition ed. New Y ork: The Guilford press; 2012

[42] Jacobson LM, deLourdes Garcia-Garcia M,Hernandez-Avila JE, Cano-Arellano B, Small PM, Sifuentes-Osornio J, et al. "Changes in the geographical distribution of tuberculosis patients in Veracruz, Mexico, after reinforcement of a tuberculosis control programme." Tropical medicine &international health: TM &IH.10(4):305–11, 2005

[43] Dangisso MH, Datiko DG, Lindtjørn B. "Spatio-Temporal Analysis of Smear-Positive Tuberculosis in the Sidama Zone, Southern Ethiopia." PLoS ONE 10 (6), 2015: e0126369. doi:10.1371/journal.pone.0126369