Statistical Analysis of Spatial Distribution of Tuberculosis in Ethiopia

Getahun Dejene[1*], Geremew Muleta[2]

[1]Department of Statistics, Mizan-Tepi-University, Tepi Ethiopia
[2]Department of Statistics, Jimma University, Jimma Ethiopia

ABSTRACT

Tuberculosis is the world's main source of death from an irresistible sickness. In the year 2018 Ethiopia enrolled 146,172 instances of TB.The principle objective of this examination is to survey the spatial dispersion of TB in Ethiopia by utilizing spatial factual techniques. A Sum of 157,548 Tuberculosis cases were accounted for in Ethiopia from February 2018 to January 2019 the investigation period. The most elevated dissemination of Tuberculosis cases was seen in Oromia (38.98%) while the least was in Harar (0.62%). The cross country yearly frequency pace of informed Tuberculosis was 167 for every 100,000 populaces. From the worldwide Moran's I (- 0.575909091) and Geary's C (2.6437818) test insights show critical different grouping among adjoining Districts in Ethiopia. The analyst Presume that TB was spatially grouped (Scattered) in Ethiopia. In this way, the concerned body should focus on the areas of interest to control Tuberculosis.

Watchword: TB, Spatial Reliance, Poisson Relapse Model**Author Corresponding:getahundejene9@gmail.com*

**Author:yehagor@yaho.com*

INTRODUCTION

Tuberculosis is an irresistible sickness, and is most usually brought about by microorganisms called Mycobacterium tuberculosis (MTB). TB is engendering from one individual to another through the air. At the point when individuals with lung TB hack, sniffle or spit, they drive the TB germs into the air. Individual requirements to in breathe a couple of these germs to get tainted. These qualities are what make analyzing the spatiality of TB so significant. Not exclusively does air transmission happen through nearness collaboration, however likewise, only a couple of undetected patients can represent a potential wellbeing hazard to the local area (WHO, 2011).

As indicated by information from the World Wellbeing Association Worldwide Tuberculosis Report (WHO, 2018).In 2017, an expected 10 million individuals (5.8 million men, 3.2 million ladies, and 1 million youngsters) created tuberculosis and 4 million individuals with tuberculosis stayed undiscovered and untreated. There were 558,000 new instances of medication safe TB, 82% of which were multi-drug safe tuberculosis (MDR-TB-protection from isoniazid and rifampicin), and 8.5% has broadly drug safe TB (XDR-TB - MDR-TB in addition to protection from both a fluoroquinolone and an injectable) (WHO, 2018).

In 2017, 2.5 million People got wiped out with TB in the African region, account ring for a fourth of new TB cases all throughout the planet. A normal 417,000 people kicked the pail from the contamination in the African area (1.7 million universally). Over 25% of TB passing occur in the African Region. Seven countries address 64% of the new TB cases.

An update to measures for TB cases and passing in the African, the most recent WHO checks for Ethiopia are: yearly TB recurrence (tallying HIV positive) of 261 for each 100,000; inescapability (checking HIV positive) of 394 for each 100,000 and mortality (excepting HIV) of 35 for each 100,000 people (WHO; 2011.). In the year 2018 Ethiopia enrolled 146,172 cases of TB. Among these, 139,261 were new cases; 46,132 new smear-positive (33.1%); 49,037 new smear-negative (35.2%); 44,092 new extra-pneumonic TB (31.6%) (FMoH: 2018).

Mycobacteriumtuberculosis (MTB) transmission regularly happens inside a family or little local area in light of the fact that suffering term of contact is ordinarily needed for disease to happen, making the potential for confined groups to create (Verma, Schwartzman;2014). Be that as it may, geospatial TB bunches are not generally because of continuous individual to-individual transmission yet may likewise result from reactivation of inert disease in a gathering of individuals with shared danger factors (Verma, Schwartzman: 2014). Spatial investigation and distinguishing proof of regions with high TB rates (bunches), trailed by portrayal of the drivers of the elements in these groups, have been advanced for designated TB control and escalated utilization of existing TB control instruments (Theron,et al:2015).

Spatial autocorrelation might be characterized as the relationship among upsides of a solitary variable that comes from the geographic course of action of the spaces where these qualities happen. It's anything but a spatial weight lattice that mirrors the power of the geographic connection between perceptions in an area. Spatial autocorrelation measurements, for example, worldwide Moran's I and Geary's C are gauges the general level of spatial autocorrelation for an informational collection. The chance of spatial heterogeneity recommends that the assessed level of autocorrelation may differ essentially across geographic space. Nearby markers spatial autocorrelation (LISAs) give disaggregated to the degree of spatial examination units. Worldwide spatial autocorrelation measurements, for example, the worldwide Moran's I and Geary's C portray the in general spatial reliance of TB by and large the whole district, nearby spatial autocorrelation insights mostly the neighbourhood Moran's I recognized from the Moran dissipate plot (Anselin, 1995) is valuable in distinguishing nearby examples or problem areas.

MATERIALS AND METHODS

Study Area and Data Source: The study was conducted in Ethiopia. It is located in the North Eastern part of the African continent or what is known as the "Horn of Africa." Ethiopia is bounded by Sudan on the west, Eritrea and Djibouti on the northeast, Somalia on the east and southeast, and Kenya on the south. Ethiopia lies between the Equator and Tropic of Cancer, between the 30 N and 150N Latitude or 330 E and 480 E Longitude. The country is administratively divided into 9 regions and 2 administrative cities (Addis Ababa and Dire Dewa), and has a tiered administrative system consisting of regional states (first-level), zones (second-level), districts (woreda), and neighbourhoods (kebeles). A spatial analysis was conducted using National Tuberculosis data from February 2018 to January 2019 obtained from Ethiopian public health institute.

Spatial Analysis: Spatial analysis is analysis of data in which the location or distance between objects is taken into consideration. Spatial analysis includes techniques for visualizing phenomenon, determining if data exhibit spatial autocorrelation and modelling spatial relationships (Anselin, 1995; Fotheringham, Brunsdon and Charlton, 2000). The researcher begins with the question of whether or not spatial autocorrelation exists and then focus on techniques for modelling spatial relationships.

Spatial Dependence: Spatial dependency is a key concept on understanding and analysing a spatial phenomenon. Such notion stems from what Waldo Tobler calls the first law of geography: “everything is related to everything else, but near things are more related than distant things. Observations from different locations or geographical areas are usually related. i.e., observations located nearby are more related than the observations located farther apart. This type of dependency of observation based on the location position is termed as spatial autocorrelation, which measures the correlation of a variable with itself through space. ‘Spatial autocorrelation’ is the correlation among values of a single variable strictly attributable to their relatively close positions (location) on a two-dimensional (2D) surface, introducing a deviation from the independent observations assumption of classical statistics. Spatial autocorrelation coefficients indicate whether and to what extent the observations influence each other via the structure of the network. Spatial autocorrelation tools test whether the observed value of a variable at one locality is independent of values of the variable at neighbouring localities. A variety of methods have consequently been developed to correct for the effects of spatial autocorrelation (partially reviewed by Miller *et al.*, 2007).

Methods of Measuring SAC

Defining Spatial Weights Matrix: Where the spatial relationships among spatial units are specified. There are several approaches to define spatial relations between two locations or spatial units; they can essentially be classified into two main groups: spatial contiguity approach and the distance based approach. Typical types of neighbouring matrices for spatial contiguity approach are: the rook, the bishop, the queen contiguity matrices *W*, and for the distance approach, the k-nearest neighbours or the critical cut-off neighbourhood matrices (J. LeSage: 1999).

Contiguity Matrix: Represents n×n symmetric matrix, where wij = 1, when i and j are neighbours and 0 when they are not. By convention, the diagonal elements are set to zero. We often report our collection of weights w_{ij} as a spatial proximity matrix (also called spatial connectivity or spatial weight matrixes (Cliff and Ord 1981) The (i, j) th element of a spatial proximity matrix W, denoted w_{ij}, quantifies the spatial dependence between regions i and j, and collectively, the w_{ij} define a neighbourhood structure over the entire area Perhaps the simplest neighbourhood definition is provided by the binary connectivity matrix. To express the degree of proximity between observations in space The Researcher may attribute a value of one if the observations are nearby (neighbours) and zero otherwise.

$$w_{ij} = \begin{cases} 1, & \textit{if region i and j share a boundary} \\ 0, & \textit{otherwise} \end{cases}$$

Tests of Spatial Autocorrelation: Spatial association, also referred to as spatial autocorrelation, corresponds to situations where observations or spatial units are non-independent over space, that is, nearby spatial units are associated in some way (Cressie: 1993). Such association can be identified in a number of ways, using a scatter-plot where each value is plotted against the mean of neighbouring areas the Moran's scatter plot, or using a spatial autocorrelation statistic such as Moran's I or Geary's C. Moran's I is a measure of global spatial autocorrelation, while Geary's C is more sensitive to local spatial autocorrelation (Carvalho: 2008).Both of these statistics require the choice of a spatial weights or contiguity matrix, usually denoted by the letter *W* that represents the topology or spatial arrangement of the data and represents our understanding of spatial association among all areas units [M. Fischer: 2006]. Spatial autocorrelation analysis was conducted by using Open GeoDa software (GeoDa Center for Geospatial Analysis and Computation, Arizona State University, AZ, USA) and used to identify the spatial clustering of the Distribution of TB Cases in the Study area. Test for spatial autocorrelation are designed to quantify the extent of clustering and to allow for statistical inference. The null hypothesis (under the normality and independence assumptions) is given by:

H_0 : no spatial autocorrelation (H0 : $\rho = 0$)

H_1: spatial autocorrelation (spatial dependence) ($H_1: \rho \neq 0$)

Global Measures of Spatial Autocorrelation: Global spatial autocorrelation is a measure of the overall clustering of the data. Global indices of spatial autocorrelation have been used to evaluate the degree to which similar observations tend to occur near each other (Rogerson, 1999; Waller & Gotway, 2004 and Jackson & Waller, 2005). Global spatial autocorrelation is a measure of the overall clustering of the data which provides one correlation statistic to summarize the whole area. But if there is no global autocorrelation or no clustering in the whole area, one can look for clusters at a local level using the measure known as local spatial autocorrelation. Moran's I and Geary's C is the measures of global spatial autocorrelation.

Moran's I: Moran's I tests for global spatial autocorrelation for continuous data. It is based on the cross-products of deviations of observations from their mean and is calculated by accounting for location of the observations. Moran's I deals with the correlation of values of a single variable. In Pearson's r, the denominator is the product of the standard deviations of the two variables, whereas in Moran's I there is only one variable involved. Moran's I formula is:

$$I = \frac{n}{S_0} \frac{\sum_{i}^{n}\sum_{j}^{n}(x_i - \bar{x})(x_j - \bar{x})w_{ij}}{\sum_{i}^{n}(x_i - \bar{x})^2}$$

Range of Moran's I: Moran's I usually takes values in the interval [–1, +1], although values lower than –1 or higher than +1 may occasionally be obtained. Positive spatial autocorrelation occurs when similar values occur near one another. Negative spatial autocorrelation occurs when dissimilar values occur near one another and approximately zero when the observations are independently arranged in space.

Geary's C: Computation of Geary's C results in a value within the range of 0 to +2. With zero being a strong positive spatial autocorrelation, through to 2, this represents a strong negative spatial autocorrelation. Calculation is similar to *Moran's I*, For Moran, the cross-product is based on the deviations from the mean for the two location values while for Geary, and the cross-product uses the actual values themselves at each location. Geary's C statistic (Geary; 1954) is based on the deviations in responses of each observation with one another and is given as,

$$C=\frac{n-1}{2S_0}\frac{\sum_i\sum_j w_{ij}(x_i-x_j)^2}{\sum_i (x_i-\bar{x})^2}$$

Interpretation of these values is very different; essentially the opposite of Moran I. Geary's C varies on a scale from 0 to 2, C of approximately

1 indicate no autocorrelation /random,

0 indicate perfect positive autocorrelation clustered,

2 indicate perfect negative autocorrelation /dispersed.

Local Indicators of Spatial Autocorrelation: Spatial autocorrelation can also be measured at the local level to evaluate the extent of autocorrelation within local neighborhoods. Local measures captures the many local spatial variation and spatial dependency while global measurements provide only one set of values that represent the extent of spatial autocorrelation across the entire study area (Mueller *et al.*, 2008).

Getis and Ord's Local (Gi (d) Statistic: Local G statistics or Getis-Ord Gi* is used to find out the spatial association of the high and low values of the feature. The output that the Getis Ord Gi* provides is a Z-scores. So, Unlike Global and Local Moran's I, Getis Ord Gi* does not need further Z-score calculation. The $G_i(d)$ local statistic Dirk et al., (2008) is an indicator of local clustering that measures the „concentration‟ of a spatially distributed attribute variable. The $G_i(d)$ statistic helped us identify „hot spots‟ in spatial data. The hypothesis of Local G-Statistics or Getis-Ord Gi*

H_0: There is no clustering of high or low values within the specified distance of location i

H_1: There is clustering of high or low values within the specified distance of point i.

Local G-Statistics or Getis-Ord Gi* can be written as follows (Getis and Ord, 1996):

$$Gi^*=\frac{\sum_{j=1}^{n} w_{ij}(d)(x_j-\bar{x})}{S_1\sqrt{\frac{n\sum_{j=1}^{n} w_{ij}{}^2-(\sum_{j=1}^{n} w_{ij})^2}{n-1}}}$$

In the case of Gi* statistic high positive values indicate a clustering of high values, while highly negative values reveal clusters of low values. Under the null hypothesis, the expected value of 0 indicates that no clustering is occurring at the specified spatial lag d. Wulder and Boots (1998) have noted that the Getis-Ord Gi statistic gives information about both the degree of clustering and the average values of the cluster. As a consequence, the Gi cannot differentiate between a lack of autocorrelation, and a cluster of average.

Poisson Regression Model for spatial data: In the situation where the distribution of the TB is clustered then very popular model is fitting Poisson regression model when assumed the equality of mean and variance. However, in the Poisson model equality of the variance of TB distribution and the mean is assumed which is too restrictive for overdispersed data where the variance in the data is higher than the expected one from the model. In such cases fitting overdispersed Poisson regression and negative binomial regression models are more relevant. It is known that Poisson regression model is appropriate and is useful when the outcome is a count type with large counts of rare events (McCullough and Nelder, 1989).

In a Poisson regression model, observed counts Y_i assumed to have a Poisson distribution, with expected values depending on predictor variables, $X=(x_{1,}x_2, x_3, \ldots,).T$he possible values of Y are the nonnegative integers: 0, 1, 2, 3, and so on. It is assumed that large counts are rare event. This creates the Poisson regression model with parameter λ_i .The probability that the number of cases takes on the entity is y_i and given by:

$$P(Y_i = y_i) = \frac{\lambda_i^{y_i} e^{-\lambda i}}{y_i!}, y_i = 0, 1, 2, 3...$$

Can obtain a simple linear model of the form:

$$\lambda_i = X_i^T \beta$$

Where $X_i^T \beta$ is the usual linear combination of predictors for case i.

Negative Binomial (NB) distribution: Negative binomial regression models do not assume an equal mean and variance and particularly correct for overdispersion in the data, which is when the variance is greater than the mean (Osgood, 2000).The negative binomial regression model is more flexible than the Poisson model and is frequently used to study count data with overdispersion (Hilbe, 2007; Hoffman, 2004). In fact, the negative binomial regression model is in many ways equivalent to the Poisson regression model because the negative binomial model could be viewed as a Poisson-gamma mixture model (Hilbe, 2007). However, the difference is that the negative binomial regression model has a free dispersion parameter. In other words, the Poisson regression model can be considered as a negative binomial regression model with an ancillary or heterogeneity parameter value of zero (Hilbe, 2007). In the negative binomial regression model, a random term reflecting unexplained between-subject differences is included (Gardner *et al*., 1995), that is, the negative binomial regression adds an overdispersion parameter to estimate the possible deviation of the variance from the expected value under Poisson regression. Therefore, using the negative binomial regression to model count data with a Poisson distribution has the consequence of generating more conservative estimates of standard errors and may modify parameter estimates (Hilbe, 2007).

The density of the negative binomial distribution is defined by:

$$f(y_i, \lambda_i : \phi) = \frac{\Gamma(y_i + \phi)}{\Gamma(\phi)\Gamma(y_i + 1)} \left(\frac{\phi}{\lambda_i + \phi}\right)^{\phi} \left(\frac{\lambda_i}{\lambda_i + \phi}\right)^{y_i}$$

Where: $\Gamma(.)$ is Gamma function ϕ is the dispersion parameter.

Test of overdispersion: Deviance and Pearson's Chi-square divided by degree of freedom are used to probing overdispersion in Poisson regression. Values greater than 1 indicate overdispersion that is the variance greater than the mean. We can test the significance of overdispersion with a Likelihood Ratio Test which follows chi square (χ^2 $(1-2\alpha,1)$) distributed with 1 degree of freedom based on Poisson and Negative Binomial distributions. The null hypotheses of this test assume equality of the mean and the variance imposed by the Poisson distribution against the alternative that the variance exceeds the mean. The negative binomial distribution $\mathrm{var}(y_i) = \lambda_i(1+\lambda_i\phi^{-1})$ the negative binomial distribution reduces to the Poisson when ϕ^{-1}=0.The Likelihood Ratio Test statistic for this hypothesis is given as:

LR =–2[LL (Poisson) – LL (negative binomial)]

Reject H 0: if LR > $\chi^2(1-2\alpha,1)$

RESULTS

Descriptive Statistics: The All-out number of TB cases in Ethiopia from February 2018 to January 2019 was 157,548.The most elevated dispersion of TB cases was seen in Oromia (61416) by percent (38.98%) 0r (65.18 New Cases per 100,000 populace) while the least was in Harar (981) by percent (0.62%) or (1.04 New Cases per 100,000 populace). In this investigation, the analyst tracked down that the yearly occurrence pace of TB was 167.2 New Cases per 100,000 populaces in Ethiopia. From Fig 1: the Moran disperse plot, the presence of a Negative spatial autocorrelation across the District was checked since the majority of the perceptions were situated in the II quadrant. There were likewise some theoretical exceptions with information that were far off from the mean.

From Table 1: grouping of comparative qualities (either high or low), though a negative Moran I coefficient shows a bunching of unique qualities or District with low qualities is encircled by neighbours with high qualities or The other way around The worldwide Moran's record measurement for the TB Cases was - 0.57590901 (p-esteem =0.0074), demonstrating the presence of critical Negative spatial autocorrelation in TB Cases over the entire Area in Ethiopia. The Moran Scatterplot is a delineation of the connection between the upsides of the picked property (TB Cases) at every District and the worth of a similar characteristic (TB Cases) at adjoining Area.

Table 2 demonstrate that, Tigrai, Amhara, SNNPRS and Addis Ababa Locales are Positive spatial relationship (noticed is more noteworthy than anticipated). The remainder of the Areas show Negative spatial connection (since noticed not exactly anticipated).

Local *Gi Test for Spatial Autocorrelation:** To address for various examinations when utilizing Gi*, importance levels were changed by Getis and Ord's standards Note that the nearby bunch map (Fig. 1) upholds the consequence of Table 3.The meaning of the Gi* measurement is surveyed by normalized Z esteem. A positive and factual huge Z an incentive for the Gi* measurement shows spatial grouping of high qualities. A negative and measurable critical Z an incentive for the Gi* measurement shows spatial grouping of low qualities (Getis and Ord's, 1992)

Table 3 shows that the H-H (High-High), L-L (Low-Low), L-H (Low-High) and H-L (High-

Low) groups that are huge at 5%. In the LISA examination, areas situated in Affar, Benishangulgumuz and Gambela Locale showed a High-High sort of relationship, implying that these regions had a High TB Cases and the encompassing regions had likewise High TB case. While the other Locale showed a High-low or low-high kind of relationship, implying that these regions had a High TB Cases and the encompassing regions had additionally low TB case and the other way around.

Overdispersed Poisson Relapse model is one model that can be utilized to defeat over scattering on Poisson relapse. The examination for an overdispersed Poisson relapse model was begun by testing the importance and relationship of each informative variable could have with the reliant variable.

Table 4 shows that at 95% importance level the specialist Saw that the p-worth of all boundaries is more modest than 0.05. With the goal that the boundaries β1, β2, β3 and β4 huge impact on the model. Then, at that point Poisson relapse models created by

$$log\lambda = 5.311 + 0.000055x_1 + 0.04457x_2 + 1.303x_3 \ + 2.785x_4$$

Indicator factors that impact the quantity of TB cases in Ethiopia was Populace Density(x1), HIV Pervasiveness (x2), Smoke Commonness (x3) and Liquor Predominance (x4).

Negative Binomial Regression Analysis Model

Another technique for dissecting the connection between Illustrative factors and reaction variable is the negative binomial relapse. For every Indicator Factors, one can decide the huge impact on the reaction variable.

The outcomes displayed in Table 5 show the presence of positive connection among reliant and logical factors (TB case stacking), with an importance level of 95%, it tends to be seen that the p-upsides of all boundaries are more modest than 0.05. The negative binomial relapse model was created as follows:

$$log\lambda = 5.533 + 0.0000122x_1 + 0.0433x_2 + 1.221x_3 + 2.829x_4$$

Results of test of overdispersion: For the Poisson model, the Pearson Chi-square and deviance values divided by the degrees of freedom are sufficiently larger than 1. This is a possible indication that the fit is overdispersed; however, this is also justified by applying a formal statistical test of dispersion.

From Table 6 we observed that to carry out the LR test for significance of overdispersion, test of hypothesis is given by:

H_0: $\phi = 1$

H_1: $\phi > 1$

The result obtained from -2(Log-likelihood (Poisson) – Log-likelihood (negative binomial)) is 28413.94, which corresponds to P-value <0.001. Hence, we reject H_0 and conclude that the mean and variance are not equal as a result the assumption of standard Poisson regression model has to be abandoned. This is shows that presence of significant overdispersion.

Comparisons of an overdispersed Poisson and negative binomial regression models:

Comparisons of Overdispersed Poisson regression model and negative binomial regression was conducted to determine a better model used in modelling the number of cases of Tuberculosis in each Region in Ethiopia. The criterion for selection of the best model used is AIC. Best model is the model that has the smallest AIC value.

Based on AIC, BIC and Log likelihood values in Table 7, the smallest AIC, BIC and Large value of Log likelihood is a negative binomial regression model. Then the best model for the number of cases of Tuberculosis is obtained from the negative binomial regression model. This suggests that the negative binomial regression model is more appropriate in the case overdispersion Poisson regression.

Table 1: ***Results of Global Moran's I and Geary's C Statistics***

Assumption	Coefficient	Observed	Expected	Variance	Z	P
Normality	Moran's *I*	-0.57590901	-0.1	0.03817971	-2.4356	0.0074
Normality	Geary's *C*	2.6437818	1.00	0.3037139	2.9827	0.0014
Randomization	Moran's *I*	-0.57590901	-0.100	0.02696425	-2.8982	0.001
Randomization	Geary's *C*	2.6437818	1.000	0.2558659	3.2497	0.0006

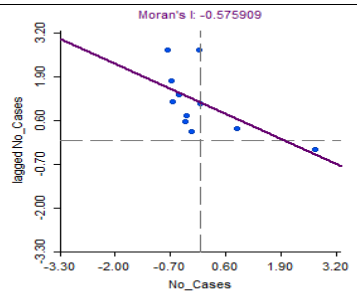

Fig 1: Global Moran's I Scatter Plot for TB Cases.

Table 2: Results of Local Moran's I Test

ID	Region	Observed	Expected	Variance	Z	Pr(z>0)
1	Tigrai	-0.05983202	-0.1	0.25117656	0.040147578	0.04680599
2	Affar	-0.26341115	-0.1	0.12453783	-0.463053343	0.6783369
3	Amhara	0.34749322	-0.1	0.01245378	0.0168048342	0.01023903
4	Oromia	-0.81361380	-0.1	0.05418299	-3.065714930	0.9989142
5	Somalia	-0.22958717	-0.1	0.16675074	-0.317342408	0.6245081
6	Benishangul	-1.35537598	-0.1	0.25117656	-2.504864657	0.9938751
7	SNNPRS	0.01618987	-0.1	0.25117656	0.0231834848	0.04083331
8	Gambela	-0.73228564	-0.1	0.25117656	-1.261606071	0.8964547
9	Harar	-2.30177316	-0.1	0.50445400	-3.100000668	0.9990324
10	A.A	-0.09692356	-0.1	0.50445400	0.004331494	0.04982720
11	Dire Dewa	-0.84587971	-0.1	0.25117656	-1.488261475	0.9316590

Table 3: Results of Local Gi* Test

ID	Region	Observed	Z	P
1	Tigrai	0.3763398	0.3763398	0.0352
2	Affar	1.7530338	1.7530338	0.0401
3	Amhara	1.1565190	1.1565190	0.0123
4	Oromia	-0.1470700	-0.1470700	0.4404
5	Somalia	1.0455393	1.0455393	0.1469
6	Benishangul	2.6255490	2.6255490	0.0043
7	SNNPRS	1.6436051	1.6436051	0.0051
8	Gambela	1.9746307	1.9746307	0.0244
9	Harar	2.6963991	2.6963991	0.9964
10	A.A	2.6794797	2.6794797	0.0099
11	Dire Dewa	1.6765755	1.6765755	0.9525

Table 4: Results of over dispersed Poisson model

Parameter	Estimate	Deviance	Standard error	Chi-square	Pr>Chi-square
Intercept	5.311	28504	0.021	63960.82	<0.001
Population Density	0.0000055	28532	0.000001033	28.35	<0.001
HIV prevalence	0.04457	80102	0.002446	332.03	<0.001
Smoke Prevalence	1.303	126962	0.005088	65583.51	<0.001
Alcohol Prevalence	2.785	58760	0.01626	29336.57	<0.001

Table 5: Results of NB regression model.

P>Chi-square	Parameter	Estimate	Deviance	Standard error	Chi-Square
<0.001	**Intercept**	5.533	11.265	0.5302	108.90
<0.001	**Population Density**	0.0000122	11.343	0.00003927	0.097
<0.001	**HIV Prevalence**	0.04334	28.884	0.09045	0.229
<0.001	**Smoke Prevalence**	1.221	70.950	0.1484	67.69
<0.001	**Alcohol Prevalence**	2.829	33.721	0.06059	2180.04

Table 6: Results of over dispersion of fit test.

Criterion	Estimate	Poisson model	Negative binomial model
Deviance	Value of Statistic/DF	4750.615	1.88
Pearson Chi-square	Value of Statistic /DF	4536.52	1.74
Log-likelihood	Value of Statistic	-14311.6	-104.6288

Table 7: The choosing of best model

Model	AIC	BIC	LL
Overdispersed Poisson	28633.2	28635.19	-14311.6
Negative Binomial	221.2577	223.645	-104.6288

DISCUSSION

In this investigation, the specialist tracked down that the yearly occurrence pace of advised TB was 167.2 New Cases per 100,000 populace, which is extremely high when the analyst contrasted and the WHO TB 2018 report for Ethiopia (151 New cases for each 100,000 populace) (WHO: 2018). The occurrence of TB revealed in this examination was higher contrasted for certain other African nations, for example, Egypt, Ghana, and Sudan (Kyu et al., 2016). The high rate pace of TB in Ethiopia presents a genuine test for the public TB control endeavours to accomplish the worldwide End-TB targets (WHO: 2015).

It was feasible to distinguish that the Appropriation of TB in a neighbourhood not impacts the Dissemination of the adjoining area, which was recognized by a worldwide Moran list -0.57590901(p=0.0074),which addressed a Negative Spatial autocorrelation among neighbourhoods and this Negative Spatial autocorrelation show factual importance. Thusly, the dangers for transmission of TB in a locale can't be impacted by a TB scourge in adjoining regions (Queiroga et al., 2016).

This examination showed that the spatial dissemination of TB in the investigation region was non-irregular and grouped with a measurable importance (P=0.0074) in an Area. The spatial investigation recognized three critical spatial groups for a high rate of TB Steady outcomes were accounted for by different examinations (Patrick et al.,2004)).The considers showed that TB is circulated in the investigation region not haphazardly but rather in bunches in a spatial example.

The examination was to explore the spatial dispersion of TB Cases in Ethiopia and to evaluate the job of medical services access and social elements on spatial bunching of informed TB. The consequence of this examination showed that TB was spatially bunched (Scattered) in Ethiopia and there were huge connections between frequency of told TB and admittance to medical care and information about TB. (Alene et al, 2017).

This investigation Tracking down that Distinctive informative factors such HIV and populace thickness were utilized to see their huge impact on TB by utilizing Poisson relapse model. The outcome shows that positive affiliation is seen between HIV, populace thickness and TB. This examination can concur with Study done' by (Thomas; 1988), done on a raised TB occurrence happened in upper copepod Massachusetts.

This Finding additionally demonstrate that Worldwide Moran's I, Geary's C and Moran disperse plot are utilized in deciding circulation of TB. These were utilized in recognizing spaces of problem area for giving solid consideration in checking and to decrease TB conveyance. The qualities for Worldwide Moran's I shows that the presence of critical TB grouping Ethiopia. This Outcome is concurred with (Habte; 2011) done on spatial examples of TB in North Shoa Zone.

CONCLUSIONS

This investigation utilized ARCGIS and spatial examinations which have been applied to numerous epidemiological explores to break down and all the more obviously show the spatial examples of TB in Ethiopia. Both worldwide and neighbourhood procedures were applied with an aim to uncover spatial attributes of the illness. These incorporated Moran's I strategy as a worldwide recognition instrument, and Nearby Moran (LISA) as neighbourhood location apparatuses.

The consequences of the examination propose that there was huge disparate grouping (Negative spatial autocorrelation) among adjoining Areas in Ethiopia and critical nearby bunching of Tuberculosis happens among Districts inside adjoining Locales. This outcome gives helpful data on the predominant epidemiological circumstance of tuberculosis in the examination region. The discoveries, as far as the presence of problem areas of TB in Ethiopia, can help the common wellbeing officials to heighten their therapeutic measures in the distinguished regions and to give future systems for more powerful control of the sickness.

Acronomy

AIC: Akaike Information Criterion, SBIC: Bayesian Schwarz Information Criterion, AIDS: Acquired Immunodeficiency Syndrome, EPHI: Ethiopian Public Health Institute, FMOH: Federal Ministry of Health, LISA Local Indicator of Spatial Association, GIS: Geographic Information System, WHO: World Health Organization.

Acknowledgement

My special and heartfelt gratitude goes to Geremew Muleta (PhD scholar) my advisor, for his guidance, support, continual help, and comments from proposal development up to the final work of this thesis and will always be grateful for that. I would like to thank also my Co-Advisor Demeke Kifle (M.Sc.) who gave me all the necessary advices and valuable comments. Finally, I am highly grateful to all my friends for their continual encouragements towards success of my work.

Ethical Consideration: The Research Ethics Review Board of Jimma University has provided an ethical clearance for this study. The data was obtained from Ethiopia Public Health Institute and Central Statistical Agency based on official letter written by Jimma University.

Authors' Contribution: Getahun Dejene Conceived and designed the study, analysed and interpreted data and wrote the initial draft of the paper. Geremu Muleta offered overall advice on academic guidance and reviewed the final paper for important intellectual content.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. Alene, K. A., Viney, K., McBryde, E. S., & Clements, A. C. (2017). Spatiotemporal transmission and socio-climatic factors related to pediatric tuberculosis in north-western Ethiopia. *Geospatial health*.
2. Ali, H., Zeynudin, A., Mekonnen, A., Abera, S., & Ali, S. (2012). Smear positive pulmonary tuberculosis (PTB) prevalence amongst patients at Agaro teaching health center, south west Ethiopia. *Ethiopian journal of health sciences*, *22*(1), 71-76.
3. Cazabon, D., Alsdurf, H., Satyanarayana, S., Nathavitharana, R., Subbaraman, R., Daftary, A., & Pai, M. (2017). Quality of tuberculosis care in high burden countries: the urgent need to address gaps in the care cascade. *International Journal of Infectious Diseases*, *56*, 111-116.

4. Clements, A. C., Lwambo, N. J., Blair, L., Nyandindi, U., Kaatano, G., Kinung'hi, S., ... & Brooker, S. (2006). Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Tropical medicine & international health*, *11*(4), 490-503.
5. Club, T. B. (2016). Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pp. 267-281 in 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971, eds. BN Petrov and F. Csáki. Budapest: Akadémiai Kiadó. Bankevich, A., S. Nurk, D. Antipov, AA Gurevich, M. Dvorkin, AS Kulikov, VM Lesin, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single. *Systematics and Evolution of the Arundinoideae and Micrairoideae (Poaceae)*, *18*(1), 139.
6. Dangisso, M. H., Datiko, D. G., & Lindtjørn, B. (2015). Accessibility to tuberculosis control services and tuberculosis programme performance in southern Ethiopia. *Global health action*, *8*(1), 29443.
7. Derbie, A., & Mekonnen, D. (2015). Prevalence of smear positive pulmonary tuberculosis among patients visiting Bahir Dar health center, Northwest Ethiopia. *International Journal of Public Health and Epidemiology*, *4*(6), 179-182.
8. Elliott P, W. (2004). Spatial epidemiology: current approach sand future challenges. *EnvironHealthPerspect*, *112*(9), 998.
9. Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological bulletin*, *118*(3), 392.
10. Getnet, F., Demissie, M., Assefa, N., Mengistie, B., & Worku, A. (2017). Delay in diagnosis of pulmonary tuberculosis in low-and middle-income settings: systematic review and meta-analysis. *BMC pulmonary medicine*, *17*(1), 202.
11. Gideon, S. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.
12. Golub, J. E., Bur, S., Cronin, W. A., Gange, S., Baruch, N., Comstock, G. W., & Chaisson, R. E. (2005). Patient and health care system delays in pulmonary tuberculosis diagnosis in a low-incidence state. *The International journal of tuberculosis and lung disease*, *9*(9), 992-998.
13. Haase, I., Olson, S., Behr, M. A., Wanyeki, I., Thibert, L., Scott, A., ... & Schwartzman, K. (2007). Use of geographic and genotyping tools to characterise tuberculosis transmission in Montreal. *The International Journal of Tuberculosis and Lung Disease*, *11*(6), 632-638.
14. Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.
15. Jamison, D. T., Breman, J. G., Measham, A. R., Alleyne, G., Claeson, M., Evans, D. B., & Musgrove, P. (Eds.). (2006). *Disease control priorities in developing countries*. The World Bank.

16. Jia, Z. W., Jia, X. W., Liu, Y. X., Dye, C., Chen, F., Chen, C. S., ... & Liu, H. L. (2008). Spatial analysis of tuberculosis cases in migrants and permanent residents, Beijing, 2000–2006. *Emerging infectious diseases*, *14*(9), 1413.
17. Keshavjee, S., Dowdy, D., & Swaminathan, S. (2015). Stopping the body count: a comprehensive approach to move towards zero tuberculosis deaths. *The Lancet*, *386*(10010), e46-e47.
18. Kulldorff, M., & Nagarwalla, N. (1995). Spatial disease clusters: detection and inference. *Statistics in medicine*, *14*(8), 799-810.
19. Lakew, Y., Benedict, S., & Haile, D. (2015). Social determinants of HIV infection, hotspot areas and subpopulation groups in Ethiopia: evidence from the National Demographic and Health Survey in 2011. *BMJ open*, *5*(11), e008669.
20. Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*, *15*(3), 209-225.
21. Liestøl, K., Tretli, S., Tverdal, A., & Mæhlen, J. (2009). Tuberculin status, socioeconomic differences and differences in all-cause mortality: experience from Norwegian cohorts born 1910–49. *International journal of epidemiology*, *38*(2), 427-434.
22. Maciel, E. L. N., Pan, W., Dietze, R., Peres, R. L., Vinhas, S. A., Ribeiro, F. K., ... & Golub, J. E. (2010). Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil. *The international journal of tuberculosis and lung disease*, *14*(11), 1395-1402.
23. McBryde, E. S., & Denholm, J. T. (2012). Risk of active tuberculosis in immigrants: effects of age, region of origin and time since arrival in a low-exposure setting. *Medical journal of Australia*, *197*(8), 458-461.
24. Mekonnen, A. (2014). Smear-positive pulmonary tuberculosis and AFB examination practices according to the standard checklist of WHO's tuberculosis laboratory assessment tool in three governmental hospitals, Eastern Ethiopia. *BMC research notes*, *7*(1), 295.
25. Middelkoop, K., Bekker, L. G., Morrow, C., Zwane, E., & Wood, R. (2009). Childhood tuberculosis infection and disease: a spatial and temporal transmission analysis in a South African township. *South African medical journal*, *99*(10).
26. Moonan, P. K., Bayona, M., Quitugua, T. N., Oppong, J., Dunbar, D., Jost, K. C., ... & Weis, S. E. (2004). Using GIS technology to identify areas of tuberculosis transmission and incidence. *International Journal of Health Geographics*, *3*(1), 23.
27. Munch, Z., Van Lill, S. W. P., Booysen, C. N., Zietsman, H. L., Enarson, D. A., & Beyers, N. (2003). Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. *International journal of tuberculosis and lung disease*, *7*(3), 271-277.
28. Nemera, A. (2016). *Integration of the Ethiopian Climate Resilient Green Economy (CRGE) Strategy into the Growth and Transformation Plan (GTP-II) Process: Implications to the Most Vulnerable Social Groups* (Doctoral dissertation, Addis Ababa University).

29. Nunes, C. (2007). Tuberculosis incidence in Portugal: spatiotemporal clustering. *International journal of health geographics*, *6*(1), 30.
30. Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates. *Journal of quantitative criminology*, *16*(1), 21-43.
31. Ostfeld, R. S., Glass, G. E., & Keesing, F. (2005). Recent developments in remote sensing for environmental monitoring. *Trends in Ecology & Evolution*, *6*(20), 328-336.
32. Pai, M., & Temesgen, Z. (2017). Mind the gap: Time to address implementation gaps in tuberculosis diagnosis and treatment. *Journal of clinical tuberculosis and other mycobacterial diseases*, *6*, 14.
33. Pedigo, A., & Aldrich, T. (2011). Neighborhood disparities in stroke and myocardial infarction mortality: a GIS and spatial scan statistics approach. *BMC public health*, *11*(1), 644.
34. Pronyk, P. M., Kahn, K., & Tollman, S. M. (2007). Using health and demographic surveillance to understand the burden of disease in populations: The case of tuberculosis in rural South Africa1. *Scandinavian journal of public health*, *35*(69_suppl), 45-51.
35. Queiroga, R. P. F. D., Sá, L. D. D., Nogueira, J. D. A., Lima, E. R. V. D., Silva, A. C. O., Pinheiro, P. G. O. D., & Braga, J. U. (2012). Spatial distribution of tuberculosis and relationship with living conditions in an urban area of Campina Grande-2004 to 2007. *Revista brasileira de epidemiologia*, *15*, 222-232.
36. Randremanana, R. V., Sabatier, P., Rakotomanana, F., Randriamanantena, A., & Richard, V. (2009). Spatial clustering of pulmonary tuberculosis and impact of the care factors in Antananarivo City. *Tropical Medicine & International Health*, *14*(4), 429-437.
37. Sasson, C., Cudnik, M. T., Nassel, A., Semple, H., Magid, D. J., Sayre, M., ... & Columbus Study Group). (2012). Identifying high-risk geographic areas for cardiac arrest using three methods for cluster analysis. *Academic Emergency Medicine*, *19*(2), 139-146.
38. Shaweno, D., Trauer, J. M., Denholm, J. T., & McBryde, E. S. (2017). A novel Bayesian geospatial method for estimating tuberculosis incidence reveals many missed TB cases in Ethiopia. *BMC infectious diseases*, *17*(1), 662.
39. Tadesse, H. (2011). *Statistical Analysis of Spatial Distribution of Tuberculosis in North Shoa Zone, Ethiopia* (Doctoral dissertation, Addis Abeba university).
40. Theron, G., Jenkins, H. E., Cobelens, F., Abubakar, I., Khan, A. J., Cohen, T., & Dowdy, D. W. (2015). Data for action: collection and use of local data to end tuberculosis. *The Lancet*, *386*(10010), 2324-2333.
41. Van Soolingen, D. (2001). Molecular epidemiology of tuberculosis and other mycobacterial infections: main methodologies and achievements. *Journal of internal medicine*, *249*(1), 1-26.
42. Verma, A., Schwartzman, K., Behr, M. A., Zwerling, A., Allard, R., Rochefort, C. M., & Buckeridge, D. L. (2014). Accuracy of prospective space–time surveillance in detecting tuberculosis