

Spatial Modelling of Fever Prevalence and Suspected Malaria Cases among Children in Ethiopia

Abstract: *Disease morbidity, mortality and speed of spread vary spatially. These have important implications for effective planning and targeting intervention strategies. The main aim of this study was to model the spatial dependence of both fever prevalence and suspected malaria cases among children and develop spatial regression models specifying the associated proximate risk factors. Data were obtained from 2011 EDHS collected for 144 districts at SNNP and Oromia regional states in Ethiopia. Spatial regression models were employed. The results showed that spatial lag model was found better to fit the data and explain the geographic variations of both childhood fever and suspected malaria rate data. From local and global spatial analysis, it was found that both childhood fever and suspected malaria prevalence rates in one district was directly affected by that of its neighbors. It was revealed from the study that altitude above sea level, access to piped water, proportion of children under five, vaccination coverage, wasting score, proportion of children born below average size and toilet availability were significant proximate determinants of childhood fever. And altitude, proportion of children born below average, vaccination coverage, stunting score, wasting score, proportion of children under five, mother education, and access to mass media were found to have significant effects on suspected malaria cases. It is suggested that any intervention to mitigate the health problem would use the fact that both fever and suspected malaria prevalence among children across the two regional states vary spatially.*

Keywords: Child Fever, Suspected Malaria, Spatial Dependence, Spatial model, Ethiopia

1. Introduction

Fever, Diarrhea and cough are a major cause of mortality among children in many developing countries, particularly in Sub-Saharan African countries. Few systematic studies of factors that influence the prevalence of fever, diarrhea and cough among young children were carried out in these countries (Kandala et al., 2006). In many of their attempts to improve public health, health planners depend on health care interventions that are based on a correct understanding of the diseases that affect childhood morbidity.

The spatial modeling and mapping of diseases is increasingly being undertaken to derive health metrics, guide intervention strategies, and advance epidemiological understanding. Spatial spillovers may cause to geographically dependent regions (Haining, 2003). Spatial modelling may improve our understanding of the epidemiology of the diseases for efficient and cost-effective control (Lawrence et al. 2007)

There is a need to understand the spatial relation ship and impact of spatial factors such as climate, water quality and management and education on human health. This will facilitate the work of decision makers in health sector to plan strategy to ensure safer living conditions to people without the fear of disease (Riedel et al., 2010). Mapping of variation in risk of childhood morbidity can help improve the targeting of scarce resources for public health interventions (Thomson et al., 1996). The our analysis studies spatial differentials of both childhood fever and suspected malaria cases among children in two Regional States of Ethiopia using data aggregated at districts/Woreda level for 2011 Ethiopian health and demographic survey data (EDHS, 2011).

The general objective of the study was to assess the spatial dependence of both childhood fever and suspected malaria prevalence and to develop spatial regression models specifying the associated proximate risk factors based on DHS data in Ethiopia.

The specific objectives are:

- To develop spatial models for fever and suspected malaria cases among children
- To estimate the spatial autocorrelation of fever prevalence and suspected malaria cases

across regions

- To identify the hot spot districts for childhood fever and suspected malaria prevalence rates

2. Methodology

Description of the Study Area and Population

SNNP Regional State: The Southern Nations Nationalities and Peoples Regional State (SNNP) is located in the Southern and Southwestern part of Ethiopia. Currently the region consists of 14 zones, 126 districts (weredas), 22 administration towns and 4 special districts (Woredas). According to central statistical agency report of 2008, the total population of the region is estimated to be 15,042,531 (CSA, 2008).

Oromia Regional State: This occupied the Central, Western, Southern, and Eastern part of Ethiopia. The region had 17 administrative zones, 245 districts (weredas), and 36 town administrations. Total population is estimated to be 27,158,471 (CSA, 2008).

Data Collection Procedures

This work is based on data available from the 2011 EDHS. The 2011 EDHS uses standard survey instruments to collect data on household members, such as working status and education of mother, sex of child, exposure to media, etc. It also collects household living conditions, such as housing characteristics and information on fertility, mortality and child health from mothers in reproductive ages (15-49). Individual data records were constructed for 10883 children in Ethiopia out of these children about 33.46% are taken from Oromia and SNNP Regional State (EDHS, 2011). Each record consists of disease information and the list of covariates that could affect the child's health with corresponding district where the children is come from. The EDHS data permit attributing childhood disease prevalence to three specific diseases, namely diarrhea, fever, and cough, observed in the last 2 weeks before the survey. We will provide some more information about the three diseases, which are used as response variables and information about the covariates considered in this study. In the 2011 EDHS, for each child under age 5, mothers were asked if the child had experienced an episode of diarrhea, a cough, accompanied by short, rapid breathing (symptoms of ARI),

or fever in the two weeks preceding the survey. Respondents were also asked if treatment was sought when the child was ill. The sample was selected in two stages. In the first stage, 647 clusters were selected from the list of enumeration areas (EA). EAs are developed at the Woreda level in Ethiopia context to mean district where children leaving. Spatially aggregated data across SNNP and Oromia Regional States on all variables were extracted and used to conduct the investigation. Information for the study was extracted from 2011 EDHS, and shape file map was obtained from Finance and Economic Development offices of SNNP and Oromia Regional State.

Variables Considered in the Study

The response variables are: rate of occurrence of childhood fever and rate of malaria drug intake rate by children who were ill of fever in a district. They are considered to be continuous variables.

The covariates were the baseline environmental, demographic and socioeconomic variables. These were altitude of district above sea level, proportion of households having access to piped water, proportion of children born below average size, vaccination coverage, stunting score, wasting score, mother education attainment, proportion of households having access to protected toilet, proportion of households having access to mass media. All rates and percentages were calculated at district level.

Explanatory Spatial Data Analysis (ESDA)

Exploratory Spatial Data Analysis (ESDA) is a set of techniques aimed at describing and visualizing spatial distributions, at identifying atypical localizations or spatial outliers, at detecting patterns of spatial association, clusters or hot spots, and at suggesting spatial regimes or other forms of spatial heterogeneity (Haining, 1990; Bailey and Gatrell, 1995; Anselin, 1998). These methods provide measures of global and local spatial autocorrelation. Spatial autocorrelation can be defined as the coincidence of value similarity with location similarity or dissimilarity. Therefore there is positive spatial autocorrelation when high or low values of a random variable tend to cluster in space and there is negative spatial autocorrelation when geographical areas tend to be surrounded by neighbors with very dissimilar values (Anselin, 2000). Spatial dependence (autocorrelation) is measured via spatial dependence

autocorrelation analysis: In this study, **Moran's I** and **Geary's C** statistics for spatial autocorrelation will be computed to ascertain evidence of disease clustering.

Methods of Measuring Spatial Autocorrelation

The first step that has to be made before one deal with spatial effects is the quantification of location aspect of spatial data based on location information from Cartesian space represented by latitude and longitude which allows as to measure distance between two points in location, and contiguity that reflecting the relative position in space of one region unit observation to other unit relaying on knowledge of the size and shape of observational unit depicted on map (Lesage, and Pace, 2009). Note that a "prior knowledge of the strength of the relationship between all pairs of places in the spatial system"; the weights are analogous to lag coefficients in autoregressive distributed lag time series models. Unlike in time series data where data points are ordered contemporaneously determining the order of observations in space is difficult as it is multi directional. Contiguity information is quantified as contiguity (spatial neighbors) matrix containing elements of 1 and 0.

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} \cdots & w_{1n} \\ w_{21} & w_{22} \cdots & w_{2n} \\ \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} \cdots & w_{nn} \end{pmatrix}$$

Where: n is number of districts or locations under study w_{ij} is element of $\mathbf{W}$ represents weight at local position of i^{th} and j^{th} districts which shares boundaries for $i, j = 1, 2, \dots, n$

To asses the nature and degree of spatial autocorrelation, it is necessary to represent the spatial arrangement of observations in order to get a sense of how close or distant they are apart from each other. To express the degree of proximity between observations in space we may attribute a value of one if the observations are nearby (neighbors) and zero otherwise. There are different other options for defining these weights, they may be based on neighborhood which have common boundary and based on distances between centroids (LeSage, 1998). In these cases, pairs of observations might be defined as neighborhoods or entering the actual distances as a measure of the degree of proximity.

To define the contiguity relation in terms of sets of neighbors of zones or sites having common boundary (Tobler, 1970), these are coded in the form of a spatial weight matrix $\mathbf{W}$, with a zero diagonal, and the off-diagonal non-zero elements often scaled to sum to unity in each row (standardized weights matrices), with typical elements:

$$W_{ij} = \frac{w_{ij}}{\sum_{j=1}^n w_{ij}} \quad (1)$$

where, non-standardized binary spatial weights matrix with typical elements is given by:

$$w_{ij} = \begin{cases} 1 & \text{if } i \text{ is linked to } j \\ 0 & \text{Otherwise} \end{cases}$$

The simplest definition of neighbors is the contiguity case. Here, we have used a queen contiguity definition which considers any two objects as neighbors if they share either a common edge or vertex (Anselin, 1988).

Tests of Spatial Autocorrelation:

The two most commonly used tests for spatial autocorrelation are Moran's I and Geary's C statistics. These tests indicate the degree of spatial association as reflected in the data set as a whole. While Moran's I is based on cross products to measure value association, Geary's C employs squared differences (Anselin, 1992).

Local Measures of Spatial Autocorrelation:

In order to observe if there is a local spatial cluster of high or low values, and identify the regions that contribute the most to the clustering (spatial autocorrelation), measures of local spatial autocorrelation such as Moran's Scatter Plot and local indicator of spatial association (LISA) statistics are used (Anselin, 1995).

Often our interest lies not only in determining whether the data as a whole exhibit spatial autocorrelation, but also, in identifying the specific observations that exhibit spatial autocorrelation with their neighbors.

Modelling Spatial Dependence

An important issue in empirical spatial analysis is how one can detect the presence of spatial effects, and moreover, how one can distinguish between spatial dependence as a nuisance

and a substantive spatial process (Anselin and Griffith, 1988). Once significant spatial dependence has been identified via global and local tests of spatial autocorrelation, the next step is to model this spatial autocorrelation via covariates. The spatial dependence (autocorrelation) can be incorporated in to the classical linear regression model in two distinct ways: as an additional predictor in the form of spatially lagged dependent variable (special lag model), or in the error structure (spatial error model).

Spatial lag model: is appropriate when the value of a dependent variable $\mathbf{Y}$ at a location is modeled as a function of the independent variables $\mathbf{X}$ in that location as well as the values of the dependent variable at the neighboring locations, that is, the spatial lag. A spatial lag is basically the weighted average of the dependent variable values at the neighboring locations (Anselin, 1988), included as an additional explanatory variable in the model as shown in equation (14) below. On the other hand, spatial lag model assumes that the autoregressive process occurs only in the response variable ("lagged-response model"), and thus includes a term $(\rho\mathbf{W})$ for the spatial autocorrelation in the response variable $\mathbf{Y}$, but also the standard term for the predictors and errors $(\mathbf{X}\beta+\varepsilon)$ as used in an ordinary least squares (OLS) regression. The spatial lag model takes the form:

$$\mathbf{Y} = \rho \mathbf{W}\mathbf{Y} + \mathbf{X}\beta + \varepsilon \quad (2)$$

which is equivalent to $\mathbf{Y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\beta + (\mathbf{I} - \rho\mathbf{W})^{-1}\varepsilon$

Parameter Estimation for the Spatial Lag Model

The parameter estimation of the spatial lag model is based on the maximum likelihood (ML) method. Ord (1975) gives the maximum likelihood methods for estimating the spatial lag and spatial error models. The logarithm of the determinate of the $(n \times n)$ asymmetric matrix $(\mathbf{I} - \rho\mathbf{W})$ does not tend to zero, it constraints the parameter values to their feasible range between the inverse of the smallest and largest eigenvalues of $\mathbf{W}$, since for positive autocorrelation, as $\rho \rightarrow 1$, $\ln|\mathbf{I} - \rho\mathbf{W}| \rightarrow -\infty$. The log likelihood functions for spatial lag models is

given by:

$$L(\beta, \sigma, \rho) = \ln |I - \rho W| - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \left((Y - \rho WY - X\beta)' (Y - \rho WY - X\beta) \right) \quad (3)$$

The first condition for the ML estimators yield nonlinear (in parameters) equations which are solved by numerical methods (Newton-Raphson iteration method). The ML estimate of ρ is obtained from a numerical optimization of the concentrated log-likelihood function (Anselin and Bera, 1998):

$$L_{lag}^c = -\frac{n}{2} \ln \left[\frac{(e_{ols} - \rho e_l)' (e_{ols} - \rho e_l)}{n} \right] + \sum_{i=1}^n \ln(1 - \rho \omega_i) \quad (4)$$

where e_{ols} and e_l are, respectively, the residuals from OLS regressions of $\mathbf{Y}$ on $\mathbf{X}$ and from $\mathbf{WY}$ on $\mathbf{X}$. Given the maximum likelihood estimate of ρ , the parameters β , and the error variance σ^2 are then easily computed.

Spatial Error Model: A spatial error is fundamentally a weighted average of the individual residuals of the neighboring locations (Anselin, 1992), which is added into the model as an additional explanatory variable as shown equation (5) below.

The spatial error model assumes that the autoregressive process occurs only in the error term and neither in response nor in predictor variables.

The spatial error model thus takes the form:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{v} \quad (5)$$

$$\mathbf{v} = \lambda \mathbf{W}\mathbf{v} + \varepsilon$$

where λ is the spatial auto regression coefficient

Parameter Estimation for the Spatial Error Model:

The maximum likelihood estimation for the spatial error model employs the error term into

log-likelihood function as follows:

$$L_{error} = \sum_{i=1}^n \ln(1 - \lambda\omega_i) - \frac{n}{2} \ln \left[\sigma^2 - \frac{(y - X\beta)'(I - \lambda W)'(y - X\beta)}{2\sigma^2} \right] \quad (6)$$

This function is then maximized over the parameter spaces. First $\hat{\lambda}$ is obtained and then β and σ^2 are estimated given the value of $\hat{\lambda}$

3. Results and Discussion

The main goal of this study were to model the spatial dependence of both childhood fever and suspected malaria prevalence rates and to develop spatial regression models specifying the associated proximate risk factors using 2011 Ethiopian Demographic and Health Survey data collected from 144 weredas/districts. Individual records were obtained for 3316 children of which 1719 were from Oromia and 1597 from SNNP Regional State.

Part One: Analysis of Childhood Fever Rate

The Moran's I and Geary's C coefficient, both being among the most widely implemented measures of spatial autocorrelation between neighboring districts, provide some insights regarding the global spatial autocorrelation in childhood fever prevalence rate.

Table 1: Moran's Statistics for Childhood Fever Prevalence Rate

Variable	Moran's I	Standardized Value	P-Value
FEVER RATE	0.5961	11.87598	0.001

Theoretical mean of Moran's stat is -0.00699 and St.dev 0.0508

Table 2: Geary's C Statistics for Childhood Fever Prevalence Rate

Variable	Geary's C	Standardized Value	P-Value
FEVER RATE	0.00426	-13.5399	0.0002

Theoretical mean of Geary's C stat is 1.00 and St.dev 0.07354

In this section, our focus is on their application to particular data analysis, the essen-

tial task being to seek for spatial dependence. Both Moran's I and Geary's C statistics are applied to test the presence of global spatial autocorrelation patterns in the distribution of childhood fever prevalence rate among the sampled districts of SNNP and Oromia regional states and the results are displayed in Table 1 and Table 2 above.

These statistics were computed to test the null hypothesis ($H_0 : \rho = 0$) of no significant spatial clustering of childhood fever in the entire study region ($\alpha = 0.05$). Based on the P-values of the reported Moran's I and Geary's C coefficients, we reject the null hypothesis of no spatial autocorrelation. Furthermore, the computed Z-statistic is positive for Moran's I and negative for Geary's C indicating the existence of significant positive spatial autocorrelation (clustering).

Significance map shown in Figure 2 visualizes districts that had significant local spatial

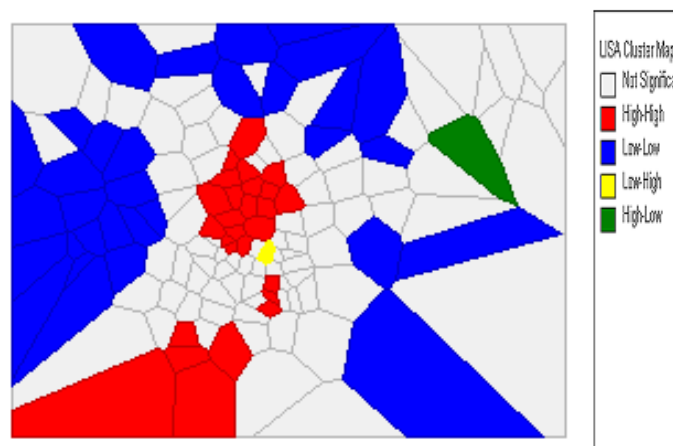


Figure 1: Cluster Map of Childhood Fever Prevalence Rate

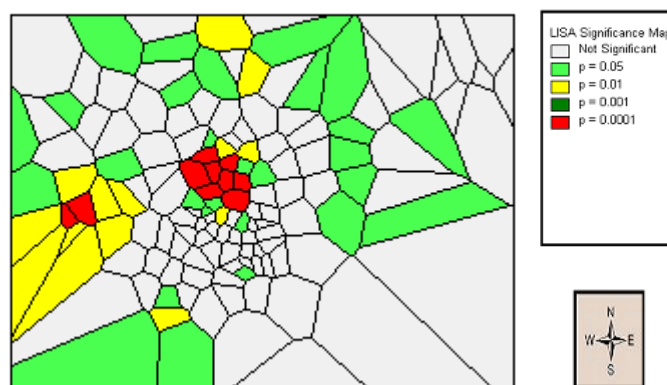


Figure 2: Significance Map of Childhood Fever Prevalence Rate

autocorrelation. The bright green, yellow and green colors in the significance map sig-

nifies evidence of significant local spatial clusters at 0.05, 0.01, and 0.001 significance levels respectively. The white color in the significance map indicates for districts which did not reject local spatial random patterns of childhood fever prevalence rate. Totally, 58 districts or 40.3% of the total 144 districts showed a significant local Moran's statistic for childhood fever prevalence rate at 5% level of significance.

The cluster map of childhood fever rate in Figure 1 shows both the locations significant local spatial autocorrelation from significance map (Figure 2). Red, blue, yellow and green colors in the cluster map indicates a district with high childhood fever prevalence rate was surrounded by districts with high childhood fever prevalence rate (HH), a district with low childhood fever rate was surrounded by districts with low childhood fever rate (LL), a district with low childhood fever rate was surrounded by districts with high childhood fever rate (LH), and a district with high childhood fever prevalence rate was surrounded by districts with low childhood fever prevalence rate (HL), respectively at 5% level of significance. White color in the cluster map indicates districts for which LMI didn't reject the null hypothesis of random distribution of childhood fever at 5% level of significance, i.e there exists a non significant local spatia autocorrelations patterns of childhood fever in these districts.

Among significant local spatial patterns of childhood fever prevalence rate, 96.8% of them had positive spatial autocorrelation at 5% level of significance and this result supports the evidence of positive spatial autocorrelation pattern in the distribution of childhood fever rates across two regional states obtained above in results of Moran's I and Geary's C tests for global spatial autocorrelation in Table 1 and Table 2.

Diagnostic for Spatial Dependence in Residuals of OLS regression of Childhood Fever Rate

The results of diagnostic tests are presented in Table 3 on which results of five methods that were used to assess the spatial dependence in the residuals of OLS model were displayed.

The test results show that both LM tests of the lag and Robust LM lag are significant, indicating the presence of spatial lag dependence of childhood fever prevalence rate, but both

Table 3: Test of Spatial Dependence from OLS regression Residuals of Fever Rate

Test	DF/MI	Value	
Moran's I	0.5961	11.8759	0.001*
LM (Lag)	1	127.23	0.000*
Robust LM (lag)	1	26.98	0.000*
LM (error)	1	136.12	0.06475
Robust LM (error)	1	8.89	0.08362

*significant at 5% level

LM error and Robust LM error tests were found to be insignificant. The robust measure for lag is significant whereas the robust measure for error is not, indicating less evidence for the importance of spatial error model. On the other hand the presence of spatial autocorrelation in the data violates the independence assumptions for the classical regression model and hence suggests explicit treatment with a spatial regression model.

Spatial Lag Model for Childhood Fever Rate

The Spatial Lag Model (SLM) can be fitted for the childhood fever rate. The results of maximum likelihood estimates of coefficients are presented in Table 4 below. The effect of individual explanatory variables on the geographical variations of childhood fever rate was tested by Z statistic in the control of spatial lag dependence. Estimated coefficient for spatial lag of childhood fever rate (ρ) was 0.8211, positive and significant at 5% level of significance indicating that childhood fever prevalence rate in one district depends directly on the childhood fever prevalence rate in the neighboring districts. Proportion of children born below average size (PCBAV SIZE) and proportion of children under five (PCU5) have positive and statistically significant effect on childhood fever prevalence rate whereas altitude, proportion of households having piped water (*WATER_100*), proportion of children ever vaccinated (*VACOV_100*), proportion of households having protected toilet (*TOILET_100*) and wasting score have negative and statistically significant coefficients at 5% level of significance. Stunting score, mother educational attainment (*EDAT_100*) and proportion of

households with access to mass media (*MEDIA*_100) have insignificant effect on childhood fever prevalence rate.

Table 4: Maximum Likelihood Estimate for Factors of Childhood Fever Rate in Spatial Lag Model

Variable	Coefficient	Std.Error	z-value	Probability
$W_FEVER(\rho)$	0.8211	0.05001	16.418	0.000*
CONSTANT	-0.4385	0.3211	-1.3649	0.0038*
ALTITUDE	-0.2105	0.0741	-2.8392	0.001*
WATER_100	-0.0019	0.00041	-4.6342	0.0045*
PCBA SIZE	0.011	0.0023	4.7826	0.001*
VACOV_100	-0.0061	0.00071	-8.451	0.004*
STUNT_SCORE	-0.0549	0.0312	-1.7596	0.067
WASTE_SCORE	-0.0582	0.0136	-4.2794	0.0401*
PCU5	0.0471	0.0043	10.949	0.003*
EDAT_100	-0.00239	0.00046	-5.1956	0.07301
TOILET_100	-0.001029	0.000135	-7.6199	0.046*
MEDIA_100	-0.0038	0.0012	-3.1667	0.105
Number of Observations	144	Degrees of Freedom	132	
R-squared	0.5904	Log likelihood	-85.0929	
S.E of regression	0.3331	Akaike Info Criterion	192.1858	
Sigma-square	0.112	Schwarz Criterion	224.854	

*significant at 5% level

Positive effect means that for a unit change in explanatory variable increase childhood fever prevalence rate in certain district by magnitude of estimate of parameter for that explanatory variable controlling the effect of neighbor districts, whereas negative effect means that for a unit change in explanatory variable, decrease childhood fever rate by magnitude of

estimate of parameter for the explanatory variable conditioning the neighbor districts infection rate effect and other variables constant.

Test for Spatial Correlations

As shown in Table 5 results for bi-variate Moran's I was 0.2468 and significant at 5% level of significance indicating a positive spatial correlation. That is, districts with high values for childhood fever prevalence were bounded by those districts having high values for suspected malaria prevalence rate and vice versa.

Table 5: Moran's I Statistics for Bi-Variate Spatial Autocorrelation Between Childhood Fever and Malaria Drug Intake Rates

Variable	Moran's I	Standardized Value	P-Value
MALARIA-S VS FEVER	0.2468	4.9861	0.001

Theoretical mean of Moran's stat is -0.00699 and standard deviation 0.0509

On average, low values for childhood fever and malaria suspected cases were in geographically close districts and similarly for high values rather randomly distributed in the two regions.

Part Two: Analysis of Malaria Drug Intake Rate

Explanatory Spatial Data Analysis of Suspected Malaria Rate

To test the spatial process in distribution of suspected malaria cases among children in the two regional states, Table 6 and Table 7 presents the results of Moran's I and Geary's C test respectively for testing the presence of global spatial autocorrelation. The reported P-values of both Moran's I and Geary's C statistics indicates that we can reject the null hypothesis of no spatial autocorrelation at 5% level of significance. In other words, the results of both tests indicates a positive spatial autocorrelation pattern of suspected malaria cases among childrens of the two regions.

Table 6: Moran's Statistics for Malaria Drug Intake Rate

Variable	Moran's I	Standardized Value	P-Value
MALARIA-S	0.4612	9.4626	0.001

Theoretical mean of Moran's stat is -0.00699 and St.dev 0.0495

Table 7: Geary's C Statistics for Malaria Drug Intake Rate

Variable	Geary's C	Standardized Value	P-Value
MALARIA-S	0.002527	-11.1266	0.0002

Theoretical mean of Geary's C stat is 1.00 and St.dev 0.0896

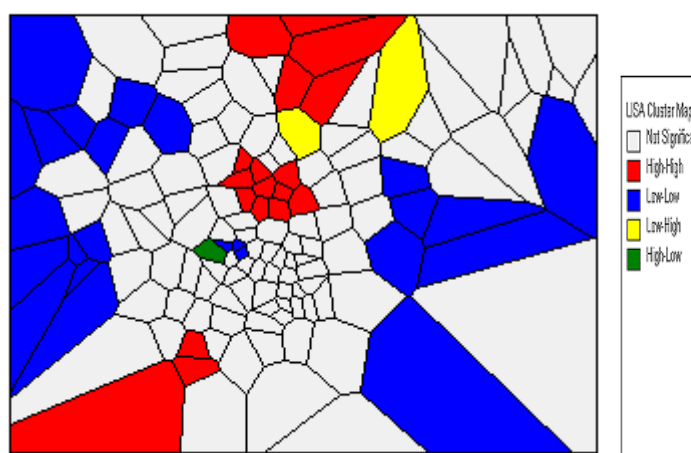


Figure 3: Cluster Map of Malaria Drug Intake Rate

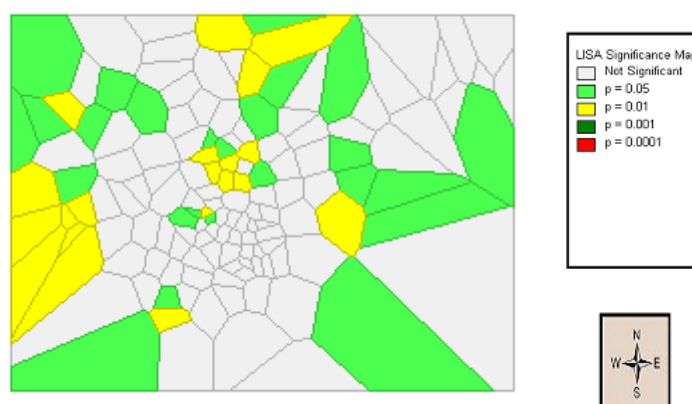


Figure 4: Significance Map of Malaria Drug Intake Rate

Figure 3 and Figure 4 presents cluster and significance maps of suspected malaria preva-

lence rate among the two regional states respectively. The Significance map (Figure 4) shows results of local Moran's I test for local spatial autocorrelation patterns of suspected malaria prevalence among children. There were about 30.55% (44) districts that had significant local spatial autocorrelation patterns of suspected malaria prevalence rate in the two regional states at 5% significance level, pointing out that there were presence of a spatial association of suspected malaria prevalence in many districts of the two regional states.

Fitting Spatial Regression Models for Suspected Malaria Rate

The most appropriate tests in identifying the form of spatial dependence in the data is Lagrange multiplier (LM) test.

Table 8: Diagnostics Test of Spatial Dependence in OLS regression Residuals Malaria Rate

Test	DF/MI	Value	Probability
Moran's I	0.4612	9.4626	0.001*
LM(Lag)	1	83.4876	0.000*
Robust LM(lag)	1	10.6311	0.0014*
LM(error)	1	80.86	0.000*
Robust LM (error)	1	3.065	0.029*

*significant at 5% level

The robust Lagrange multiplier tests, which provides a means of discriminating between the spatial lag or spatial error model, both significant at 5% significance level again suggest that both spatial lag and spatial error models can be used in modelling the suspected malaria prevalence rate. But the test for spatial lag model is significant even at 1% level of significance, hence it seems reasonable to fit spatial lag model (SLM) for suspected malaria prevalence rate among children in the two regional states

Spatial Lag Model for Suspecte Malaria Rate

In this case, we applied a statistical model which incorporates spatial dependence raised from spatial lag of the dependent variable, by adding the spatial lag of malaria drug intake

rate on the right hand side of the multiple linear regression equation to represent the direct influence of the neighboring districts. The maximum likelihood estimates, standard error, test statistic and p-values of factors analyzed in suspected malaria prevalence rate were displayed in Table 9.

Table 9: Maximum Likelihood Estimate for Factors of Malaria Drug Intake Rate in Spatial Lag Model

Variable	Coefficient	Std.Error	z-value	Probability
$W_MALARIA_S(\rho)$	0.7572	0.063	12.019	0.000*
CONSTANT	-0.5266	0.3012	-1.7484	0.046*
ALTITUDE	-0.158	0.0193	-8.1865	0.009*
WATER_100	-0.0056	0.00129	-4.33	0.107
PCBA SIZE	0.00201	0.000193	10.4144	0.004*
VACOV_100	-0.0061	0.000821	-7.4288	0.016*
STUNT_SCORE	-0.0109	0.0065	-1.6769	0.0303*
WASTE_SCORE	-0.0289	0.0048	-6.021	0.0451*
PCU5	0.05656	0.0131	4.3215	0.001*
EDAT_100	0.00426	0.00133	3.205	0.0419*
TOILET_100	-0.0201	0.00441	-4.557	0.091
MEDIA_100	0.0100	0.00239	4.1841	0.0410*
Number of Observations	144	Degrees of Freedom	132	
R-squared	0.4911	Log likelihood	-57.789	
S.E of regression	0.4064	Akaike Info Criterion	137.578	
Sigma-square	0.1651	Schwarz Criterion	170.246	
*significant at 5% level				

Estimated coefficient for spatial lag of suspected malaria prevalence rate (ρ) was 0.7572,

which was positive and significant at 5% level of significance indicates that suspected malaria prevalence rate in one district depends directly on the rate in its neighboring districts.

Proportion of children born below average (*PCBASIZE*) and proportion of children under five (*PCU5*) have positive and statistically significant effect on suspected malaria prevalence rate. The proportion of households having access to mass media (*MEDIA_100*), which was insignificant in childhood fever rate study, have positive and statistically significant effect on suspected malaria prevalence rate. This may be because of using mass media as a tool for promoting the severity of diseases and prevention or treatment mechanism by governments (since our study was based on reported malaria drug intake rate). The proportion of mother attained basic education (*EDAT_100*) also have positive and significant coefficient implying that on districts with high proportion of educated mother, the malaria drug intake rate will be high, since educated mothers are more aware of how to treat their ill children than those having poor educational background. The variables having negative and statistically significant effect on suspected malaria prevalence are proportion of children ever vaccinated (*VACOV_100*), stunting score *STUNT_SCORE*, wasting score (*WASTE_SCORE*) and altitude of a given district above sea level. We can interpret the coefficients of variables as a unit change in independent variable by its coefficient of estimate increase/decrease suspected malaria prevalence rate in a district (and its neighbor) by magnitude of coefficient estimate.

Discussion:

The study investigates the spatial pattern of both childhood fever and suspected malaria cases among childrenes of SNNP and Oromia regional states. A central feature of explanatory spatial data analysis is the use of formal statistical tests to asses the degree of spatial randomness observed in the data and spatial autocorrelation test is the most available tool in explanatory spatial data analysis for aggregated data.

The Moran's I and Geary's C statistics, being the most available measure of global spatial autocorrelation pattern in childhood fever and suspected malaria prevalence rate, showed that there was positive spatial autocorrelation patterns for both childhood fever and sus-

pected malaria prevalence.

The significance of the spatially lagged dependent variable (ρ) in both childhood fever and malaria suspected cases suggests that neighboring districts prevalence and suspected rates are important determinants of a districts infection possibility by fever and suspected malaria. These results indicate that the coefficient estimate and standard error of the OLS model assuming independent observations may display bias.

Around the center of the two regional states, we found a high-high spatial association of both childhood fever and suspected malaria prevalence rates, but a wider location for fever rate than suspected malaria rate. The bi-variate Moran's I test showed that childhood fever and suspected malaria prevalence rates are positively correlated (Table 5).

Among all the explanatory variables considered in the study, the spatial lag of the dependent variables (W_FEVER and $W_MALARIA_S$) have had the largest effect on the spatial distribution of childhood fever and suspected malaria rates, respectively, relative to the explanatory variables considered under the study. The result indicates geographically close districts had similar levels of disease rates and the result supports Tobler's law of geography, which states that geographically close districts with similar socio economic and demographic characteristics and vulnerability dimensions are more conducive grouping forces such as using of unprotected drinking water (Tobler, 2002)

4. Conclusions

The study reveals that there are positive spatial autocorrelation patterns for both childhood fever and malaria drug intake rates. Hot spot areas for the two childhood morbidities were found to be at the center of the regions. The bi-variate explanatory spatial data analysis result showed that childhood fever and malaria drug intake rates have correlated spatial patterns

From the analysis of measures of goodness of fit, the spatial lag model was found to be better in fitting the data and explain the geographical variations of the childhood fever and malaria drug intake rates. The presence of significant spatial lag parameter in the two child morbidities revealed that including coefficients for variables due to spatial multiplier bias in modelling the determinants of the two child morbidities is important.

Factors having significant influence on childhood fever prevalence rate are altitude, proportion of households having access to piped water, proportion of children born below average size, vaccination coverage, wasting score, proportion of children under five and proportion of households having access to protected toilet

Factors having significant influence on explanation of suspected malaria prevalence rate are proportion of children born below average size, altitude, vaccination coverage, stunting score, wasting score, proportion of children under five, mother education attainment, proportion of households having access to mass media.

The implication of the spatial dependence is that, in cases where the decisions on how to allocate funds for control interventions such as distribution of anti-malaria drugs, needs to have spatial dimension. Moreover, control interventions should not only focus on high fever and suspected malaria rate districts, but also the surrounding districts of high childhood fever and suspected malaria rates.

References

- [1] Anselin L.(2000): Spatial Externalities, Spatial Multipliers and Spatial Econometrics. Mimeo, Department of Agricultural and Consumer Economics. University of Illinois at Urbana-Champaign.
- [2] Anselin, L. (1998). Exploratory Spatial Data Analysis in a Geocomputational Environment. PP. 77-94 in Geocomputation, A Primer, edited by P.A. Longley, S. Brooks, B.Macmillan and R. McDonnell. New York: John Wiley.
- [3] Anselin L. (1995): *Local Indicators of Spatial Association -LISA*, Geographical Analysis, Ohio University Press Submitted 6(94), Revised Version 2.
- [4] Anselin L. (1992): Space-Stat Tutorial: A Workbook for Using SpaceStat in the analysis of Spatial Data, Typescript. University of Illinois at Urbana-Champaign, pp. 8-67.
- [5] Anselin L. (1988). Spatial Econometrics: Methods and Models, Kluwer Academic Publishers, Dordrecht: Online Google Books.

- [6] Anselin L. and Bera A. (1998): Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. Handbook of Applied Economic Statistics, pp. 237-289. Marcel Dekker, New York.
- [7] CSA(2008). The 2007 Population and Housing Census Analytical Report, Addis Abeba
- [8] EDHS (2011): Ethiopia Demographic and Health Survey Report, Central Statistical Agency, Ethiopia
- [9] Haining R. (1990): Spatial Data Analysis in the Social and Environmental Sciences, Cambridge: Cambridge University Press.
- [10] Kandala N., Magadi M., Madise NJ. (2006): An Investigation of District Spatial Variations of Childhood Diarrhoea and Fever Morbidity in Malawi. Soc Sci Med 2006, 62:1138-1152.
- [11] Lawrence N., Adamson S., Christopher.C.A, Immo K.(2007): Modelling the effect of malaria endemicity on spatial variations in childhood fever, diarrhoea and pneumonia in Malawi. International Journal of Health Geographics 2007, 6:33
- [12] Lesage, J., and Pace, K. (2009): Introduction to Spatial Econometrics. Boca Raton, Florida: CRC Press.
- [13] LeSage, J. (1998): Bayesian Estimation of Spatial Autoregressive Models. International Regional Science Review 20: 115-126.
- [14] Ord J. (1975): Estimation Methods for Models of Spatial Interaction, Journal of the American Statistical Association, 70: 120-126.
- [15] Riedel, N., Penelope, V., John, M., Laura, G., Victor, M., and Rick, W. (2010). Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modeling of the 2006 Zambia national malaria indicator survey, Malaria Journal, Vol.37: 1-2.
- [16] Thomson M., Connor S., Milligan P. and Flasse S.(1996): The Ecology of Malaria as Seen from Earth Observation Satellites. Annals of Tropical Medicine Parasit ,90 ,243-264
- [17] Tobler W. (1970): *A Computer Movie Simulating Urban Growth in the Detroit Region* Economic Geography 46 (2): 234-240.