Spatial Modelling of Fever Prevalence and Suspected Malaria Cases among Children

Aklilu Toma Shamenna[1] and Ayele Taye Goshu[2]

Hawassa University

[1,2]School of Mathematical and Statistical Sciences

Hawassa, Ethiopia

Email:aketomy@gmail.com; ayele_taye@yahoo.com

July, 2014

Abstract: *The purpose of this study was to model the spatial dependence of fever prevalence and suspected malaria cases among children along with their risk factors. Data were obtained from 2011 EDHS collected for 144 districts at SNNP and Oromia Regional States in Ethiopia. Spatial regression models were applied. The results showed that the spatial lag model better fitted to the data. Prevalence rate of each of the events in a district was shown to be affected by that of its neighbors status. It was revealed that altitude, access to piped water, proportion of children under five, vaccination coverage, child wasting score, proportion of children born below average size and toilet availability were significant risk factors of fever rate. Moreover, 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 the rate of suspected malaria cases. We recommend that any intervention to mitigate occurrence of malaria infection among children would take in to account nature of spatial variability and the related factors.*

Keywords: Suspected Malaria, Spatial Variability, Spatial Lag 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.

Kandala et al. (2008) conducted a study in Nigeria using 2003 Nigeria demographic and health survey data. The study provided evidence of consistent and highly significant spatial variation, at state level, in the risk of childhood diarrhoea, cough and fever in Nigeria. It was found in this investigation that the risks of childhood fever appeared to be reduced by urban residence, antenatal visits, hospital delivery, an educated mother, an educated mothers partner, a relatively high age (of the child) and a relatively old mother.

Ingrid et al. (2009) conducted a study in Adama, Ethiopia, by using small-scale spatial dependence. The main goal of the study was to identify foci of malaria transmission in urban communities. The results of the study indicated that proximity to vector breeding site, maximum and minimum temperature, rainfall were positively associated with malaria incidence.

Being a primary cause of poverty, some studies suggest that a better understanding of the relation ship between malaria and poverty is needed to enable the design of coherent and effective policies and tools to tackle the problem. Since poverty is related to socioeconomic factors, it is important to identify those factors that are related to the risk of malaria (Hay S. et al 2004).

Geographic spillover effects are important in modeling of disease incidence. As a significant neighbor effect indicator of spatially dependent variable, large value slope of Moran index indicated that the data under consideration is spatially dependent and employing OLS models will result in inconsistent and biased estimates due to spatial multiplier bias for the data under consideration (Asnakew et al., 2009).

Lawrence et al. (2007) modeled the effect of malaria endemicity on spatial variations in childhood fever, diarrhea and pneumonia in Malawi. The results of the study indicated that the risk of childhood fever increased with rural residence relative to urban residence, those who received Vitamin A relative to those who did not were at increased risk of fever, net owner ship and weight for age were associated with lower risk of fever. In this analysis, the risk of all three illnesses, i.e., fever, diarrhea and pneumonia was found to be associated with malaria endemicity, although this relationship was stronger with fever.

Smith T. (2003) argued that in the case of malaria, spatial correlation is present at both short and large scales, reflecting the transmission of malaria infection by the mosquitoes over space and the effects of environmental factors that determine mosquito survival over large areas in Africa.

A study done in northeastern Venzuela by Grillet et al.(2010) using local spatial statistics and geographically weighted regression (GWR) with objective to determine the spatial pattern of malaria incidence and persistence reported that GWR model greatly improved predictions of malaria risk compared with ordinary least squares (OLS) regression models.

The study using spatial lag model to analyze cholera incidence clustering in Ghana also indicated that the coefficient estimate and standard error using OLS model assuming independent observations may display substantial upward bias in estimating determinants of cholera (Frank., 2006).

The spatial modeling and mapping of diseases is increasingly being undertaken to derive health metrics, guide intervention strategies, and advance epidemiological understanding. Spatial modelling may improve our understanding of the epidemiology of the diseases for efficient and cost-effective control (Lawrence et al. 2007). 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 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 be15,042,531 (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 come from.

In the 2011 EDHS, for each child under age 5, mothers were asked if the child had experienced an episode of diarrhea, cough or fever in the two weeks preceding the survey. Respondents were also asked if treatment was sought when the child was ill. The 2011 EDHS sample was selected using a stratified, two-stage cluster design, and enumeration areas (EAs) were the sampling units for the first stage. Households comprised the second stage of sampling. A complete listing of households was carried out in each of the 624 selected EAs from September 2010 through January 2011. A representative sample of 17,817 households was selected (EDHS, 2011).

Spatially aggregated data across SNNP and Oromia Regional States on all variables were extracted and used to conduct the investigation. To include woreda/district information in the data file, percentages and rates are calculated for each woreda. The 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. For instance, percentage of households having access to piped water in a given woreda can be expressed in terms of the ratio of number

households responding 'YES' for access to piped water to total number of households included in the sample in that specific woreda/district.

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; 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 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 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.

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 options for defining these weights, they may be based on neighborhood which have common boundary and based on distances between centroids (LeSage, 1998).

The simplest definition of neighbors is the contiguity case. 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 **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:

$$W_{ij} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}} \tag{1}$$

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

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

Contiguity information is quantified as spatial neighbors matrix **W** containing elements of 1and 0 for w_{ij} where

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

where: n is number of districts or locations under study. The element w_{ij} of the spatial weight matrix **W** represents weight at local position of i^{th} and j^{th} districts which shares boundaries for $i, j = 1, 2 \ldots, n$.

In our analysis, we have used a **queen contiguity** definition of spatial weights matrix 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, Gearys 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 Bera, 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 (SLM): is appropriate when the value of a dependent variable **Y** at a location is modeled as a function of the independent variables **X** in that location as well as the values of the dependent variable **Y** 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 (2) 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 the term **(ρW)** for the spatial autocorrelation in the response variable **Y** and also the standard term for the predictors and errors **(Xβ+ε)** as used in an ordinary least squares (OLS) regression. The spatial lag model takes the form:

$$\mathbf{Y} = \rho\, \mathbf{WY} + \mathbf{X}\beta + \varepsilon \tag{2}$$

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

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 (nxn) asymmetric matrix **(I-ρW)** does not tend to zero, it constraints the parameter values to their feasible range between the inverse of the smallest and largest eigenvalues of **W**, since for positive autocorrelation, as $\rho \to 1$, $ln|I - \rho W| \to -\infty$. The log likelihood functions for spatial lag models is given by:

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

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):

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

where e_{ols} and e_l are, respectively, the residuals from OLS regressions of **Y** on **X** and from **WY** on **X** and ω_i's are the eigen values of the spatial weights matrix W. Given the maximum likelihood estimate of ρ, the parameters β, and the error variance σ^2 are then easily com-

puted.

Spatial Error Model (SEM): The spatial error model addresses the spatial autocorrelation (λ) existing in the residuals of the OLS models. The value of the dependent variable **Y** in a location is redefined as a function of the independent variables **X** and the regression residuals of the neighboring location, that is, the spatial error. 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. In this case, the usual OLS regression model **Y=X**β**+**ε , is complemented by a term (v) which represents the spatial structure (λW) in the spatially dependent error term (v).

The spatial error model thus takes the form:

$$\mathbf{Y} = X\beta + \mathbf{v} \tag{5}$$

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

where λ is the spatial auto regressive 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:

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

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 objective of this study is to assess the spatial dependence of both childhood fever and suspected malaria prevalence rates and to develop spatial regression models specifying the associated proximate risk factors in districts of SNNP and Oromia Regional States of Ethiopia by using 2011 Ethiopia demographic and health survey data collected for 144 weredas/districts. Individual data records were constructed for 3316 children in two regional States of which 1719 and 1597 records respectively for Oromia and SNNP Regional States (EDHS, 2011).

The analysis has been performed for 144 districts/woredas for which EDHS 2011 data have been collected of which 71 districts/woredas are in SNNP and 73 woredas in Oromia regional state. Statistical computations were carried out with GeoDA for spatial analysis, GIS for mapping and variable selection and SPSS for descriptive and correlation analysis.

In order to prevent that our findings are due to the formulation of spatial dependence imposed by the spatial weights matrix, we control the sensitivity of our results with respect to different distance and different order queen and rook contiguity matrix. Row standardized queen first order weight matrix was found to be reasonable to study spatial effect in these data set since in some part of the two regional states, shape file boundaries are not full.
Ethiopia DHS data was collected for only 144 woreda/districts in Oromia and SNNP Regional States, but there are a total of 406 woredas in the two regions. This makes iceland in map after cutting districts where data is not available.

[i]. Analysis of Childhood Fever Rate

Univariate 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 essential 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.

Significance map shown in Figure 2 visualizes districts that had significant local spatial autocorrelation. The bright green, yellow and green colors in the significance map signifies 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 districts for which we 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.

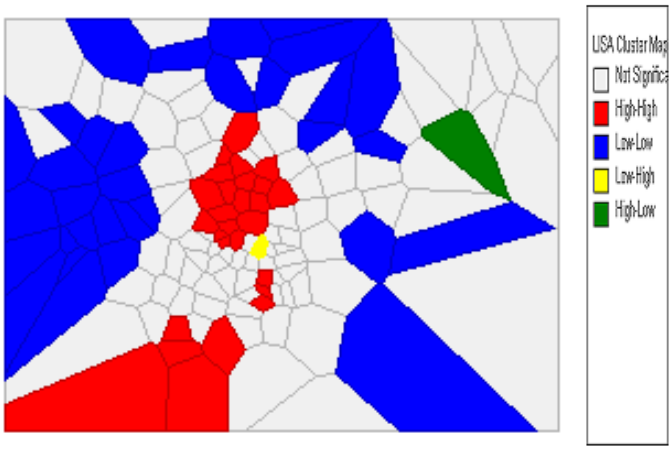

Figure 1: Cluster Map of Childhood Fever Prevalence Rate

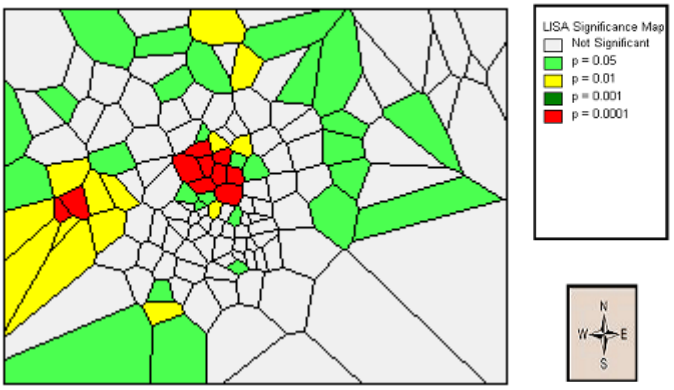

Figure 2: Significance Map of Childhood Fever Prevalence Rate

The cluster map of childhood fever rate in Figure 1 shows both locations having significant local spatial autocorrelation from significance map (Figure 2) as well as the type of spatial association (High-High (HH), Low-High (LH), Low-Low (LL) and High-Low (HL)). Red, blue, yellow and green colors in the cluster map indicates a district with high childhood fever prevalence rate surrounded by districts with high childhood fever prevalence rate (HH), a district with low childhood fever rate surrounded by districts with low childhood fever rate (LL), a district with low childhood fever rate surrounded by districts with high childhood fever rate (LH), and a district with high childhood fever prevalence rate 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 local Moran's index do not reject the null hypothesis of random distribution of childhood fever at 5% level of significance, i.e there exists a non significant local spatial autocorrelations patterns of child-

hood 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
Spatial clustering of childhood fever rate were indicated by the global and local spatial autocorrelation tests. The next step is to measure the relationships between fever rate and independent variables obtained at a neighborhood. When the data are spatially structured, OLS estimates can be biased and their significance is inflated as briefly. A diagnostic statistic indicating problems in OLS regression with spatial data is the degree of spatial non randomness of residuals and a common approach is to filter out or to treat the local spatial information as a 'noise.'

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 Lagrange multiplier (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 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 linear regression model and hence suggests explicit treatment with a spatial regression model.

Spatial Lag Model(SLM) for Childhood Fever Rate
The Spatial Lag Model (SLM) given in equation (2) can be fitted for the childhood fever

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

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) is equal to 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 where as 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 acess 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	Probablity
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 a unit change in explanatory variable increases childhood fever prevalence rate in certain district and its neighbors by magnitude of estimate of parameter for that explanatory variable fixing the effect of other variables constant, whereas negative effect means that a unit change in explanatory variable decreases childhood fever prevalence rate in certain district and its neighbors by magnitude of estimate of parameter for the explanatory variable controlling the effect of other variables constant.

Since altitude is fixed for a certain district and is fixed effect to childhood fever prevalence in certain district, the interpretation for its coefficient takes slightly different manner from other variables. Therefore coefficient can be interpreted as districts with high altitude are less likely to be infected by childhood fever whereas districts having lower altitude are more likely to be infected by childhood fever.

Test for Spatial Correlations

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

Table 5: Bi-Variate Moran's I for spatial autocorrelation between childhood fever and suspected malaria 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

That is, on average, both high and low values of childhood fever and suspected malaria rates were in geographically close districts.

[ii]. 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 results for testing the presence of global spatial autocorrelation respectively.

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

This result indicates that the null hypothesis of no spatial autocorrelation is rejected 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.

Figure 3 and Figure 4 presents cluster and significance maps of suspected malaria prevalence rate 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 in the two regional states. There are 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 is presence of a spatial autocorrelation of suspected malaria prevalence at distcrict level in the two regional states.

Anselin and Bera (1998) pointed out that in classical regression methods, the data are assumed to be randomly sampled from a homogeneous data generating process, but spatially-referenced data are often spatially clustered, and therefore are not randomly scattered in space and OLS model as econometric tool will produce spatially autocorrelated residuals resulting in biased estimates and all inferences based on the model may be incorrect.

The Moran's I test is a general test on spatial correlation without giving precise informa-

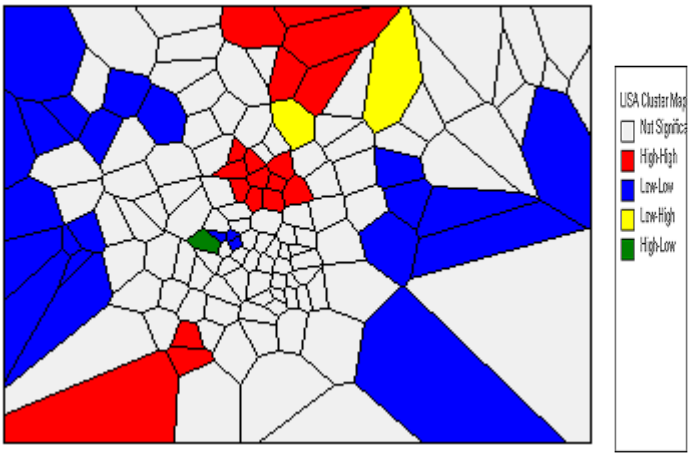

Figure 3: Cluster Map of Malaria Drug Intake Rate

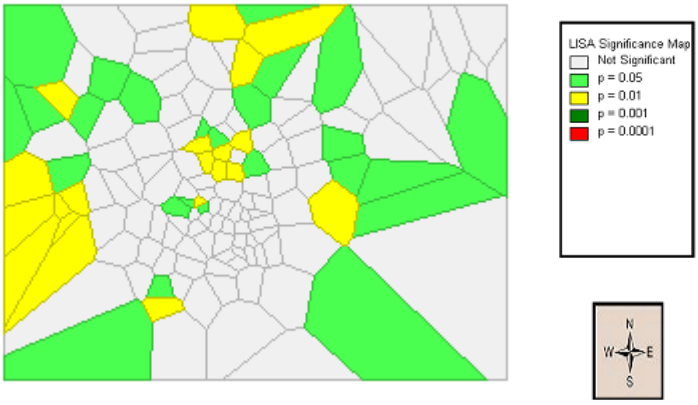

Figure 4: Significance Map of Malaria Drug Intake Rate

tion on the particular spatial structure and it provides evidence for a positive and significant spatial dependence in OLS residuals. In this case, the null hypothesis of no spatial autocorrelation is rejected (from Moran's I and Geary's C results above) indicating that OLS is inappropriate for the problem under consideration since the assumption of independent observation is violated.

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. The simple versions of LM test are powerful but not robust in local misspecification of the model, so the LM test for spatial lag dependence can be significant even if the form of the spatial dependence resembles spatial error dependence or vice versa. Thus, it is better to look at their robust part so as to come up with the correct

identification of the form of spatial dependence in the data.

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**

As suggested by Lagrange multiplier (lag) and Lagrange multiplier (error) tests which are both significant at 5% level of significance, both spatial lag and spatial error models are equally important to model determinants of suspected malaria prevalence rate.

The robust Lagrange multiplier tests, which provides a means of discriminating between the spatial lag and spatial error models, both significant at 5% significance level of significance 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 (see equation (2)). The maximum likelihood estimates, standard error, test statistic and p-values of factors analyzed in suspected malaria preva-

lence rate are displayed in Table 9.

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.

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

Variable	**Coefficient**	**Std.Error**	**z-value**	**Probablity**
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**

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 acess to mass media ($MEDIA\_100$), which was insignificant in childhood fever rate study, have positive and statistically significant effect on suspected malaria prevalence rate. The proportion of mother attained basic education ($EDAT\_100$) also have positive and significant coefficient.

The variables having negative and statistically significant effect on suspected malaria prevalence rate 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 children 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 suspected malaria prevalence rate. Again we have seen in explanatory spatial data analysis of the two morbidity studies considered in this study that these spatial patterns are not random. The autoregressive parameter estimate is positive and significant implying that there is spatial spillover in childhood fever and suspected malaria, the result is consistent with the result found in Malawi by Kandala. et al (2006).

The significance of the spatially lagged dependent variable (ρ) in both childhood fever and

suspected malaria cases suggests that neighboring districts fever prevalence and suspected malaria rates are important determinants of a districts infection possibility by fever and malaria. A significant ρ also indicate that the data under consideration is spatially dependent for the two morbidity studies and thus suggesting that employing OLS model will result in inconsistent estimates due to spatial multiplier bias for the data under consideration. Moreover, these results also indicate that the coefficient estimate and standard error of the OLS model which assumes independent observations may display bias (Anselin 2000).

In explanatory spatial data analysis, we found the spatial clustering of both childhood fever and suspected malaria prevalence rates among children is space. But these statistics fail to tell us the reason for the clustering of the diseases under consideration. Thus, to understand the factor that explain the spatial clustering of the two morbidities, the model based approach becomes necessary. The first step in model driving approach in spatial data is to check for the residuals of OLS regression for spatial autocorrelation.

The diagnostics results of spatial dependence in OLS regression residuals revealed that both spatial lag and spatial error models are equally important in modeling the suspected malaria rate. But based on the less significance of robust Lagrange multiplier (error) for residuals in spatial error term, as compared to that of spatial lag model, it is reasonable to take spatial lag model better than the spatial error model in fitting the suspected malaria rate data. The results of selecting spatial lag model for the two morbidity studies as a better fit to the data agrees with what Anselin (2000) pointed out, i.e spatial dependence cannot be handled by the result in inaccurate classical regression model.

Among all the explanatory variables considered in the study, the spatial lag of the dependent variables,that is, the spatial lag of childhood fever (W_FEVER) and spatially lag of suspected malaria rate ($W\_MALARIA - S$) have had the largest effect on the spatial distribution of childhood fever and suspected malaria rates relative to the other 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 character-

istics and vulnerability dimensions are more conductive grouping forces such as using of unprotected drinking water (Tobler, 1979).

The altitude of a given district above sea level, which acts as a function of factors such as maximum and minimum temperature of a given district, have negative and highly significant effect on both childhood fever and suspected malaria rate, showing that districts with low altitude are highly affected by childhood fever and suspected malaria. The result is consistent previous studies on malaria, on which the incidence of malaria is more high on those areas having altitude of less than 2500 meter above sea level (Guofa et al., 2003; Basel, 2008).

The coefficient estimate for the proportion of children born below average size ($PCBASIZE$) is significant and have positive effect on both childhood fever and suspected malaria rates of districts. This may be due to the fact that the greater proportion of children born below average size in district, the higher the rate of infection. Children who had low birth weight were more likely to suffer from disease than infants who had appropriate birth weight. This result is consistent with the study done by Lira et al (1996) in Brazil on which it was found that low birth infants experienced 32% more days with vomiting and 32% more days with childhood disease such as diarrhea.

Wasting score, being a general indicator of nutritional status of a child, have negative coefficient and is significantly associated with both childhood fever and suspected malaria rate where as stunting score have negative coefficient for the two childhood morbidities, but insignificant for childhood fever study. It is evidenced in Ngianga B. et al (2008) that both stunting and wasting scores are problems that occurred in children of under five most of the time and highly correlated with childhood morbidities. While under nutrition is always immediately related to either insufficient nutrient intake or the inability of the body to absorb nutrients (primarily due to illness), these are themselves caused by problems related to food security, care practices, and health care practices at the household level, which themselves are influenced by the socioeconomic and demographic situation of households and communities (Smith and Haddad, 1999).

The estimated coefficient for proportion of mothers who attained basic education ($EDAT\_100$) have positive and significant effect on the geographical variation of suspected malaria rate. This indicates that, since our study is based on the reported malaria drug intake rate, and mothers of a given child are assumed to be the major caregivers to child in home (relative to other family members), we conclude that in those districts having higher proportion of mother attained basic education, the malaria drug intake rate will be high. The result seems to be comparable with what Kazeen et al. (2008) found in their study in Nigeria. The results of Kazeen et al. (2008) have reiterated that most childhood fever are first treated at home by their mothers or care givers. Help was sought thereafter if the fever was unremitting despite the initial treatment or complication sets in.

The proportion of households having protected toilet is found to have negative and significant coefficient in the case of childhood fever prevalence rate, but is insignificant in the case of suspected malaria case study. These may be because of the reported fever among children may be related to other diseases such as for example diarrhea which is highly correlated with availablity of protected toilet in household as found in Girma et al. (2008).

For the proportion of households having access to mass media, its coefficient of estimate is positive and significant in the case of suspected malaria rates. This may be because of households with access to mass media are more aware of childhood disease and the prevention mechanisms. The same result is found in Shanne R and Hendrik R (1998) on which it is argued that households that have a radio may be better informed about children disease and prevention and are also able to afford the amenities necessary to prevent or to cope with disease in children.

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 to 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 controling 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

Anselin, L.(2000): Spatial Externalities, Spatial Multipliers and Spatial Econometrics. Mimeo, Department of Agricultural and Consumer Economics. University of Illinois at Urbana-Champaign.

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.

Anselin, L. (1995): Local Indicators of Spatial Association -LISA. Geographical Analysis, Ohio University Press Submitted 6(94), Revised Version 2.

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.

Anselin, L. (1988): Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht: Online Google Books.

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.

Asnakew, Y., Sucharita, G., Afework, H., Dereje, D. and Hrishikesh, P. (2009): Spatial Analysis of Malaria Incidence at the Village Level in Areas with Unstable Transmission in Ethiopia. International Journal of Health Geographics 2009.

Basel, P. (2008): Development of Bayesian Geostatistical Models with Applications in Malaria Epidemiology. Malaria Journal, vol. 9:4-6.

CSA (2008): The 2007 Population and Housing Census Analytical Report. Addis Abeba, Ethiopia.

EDHS (2011): Ethiopia Demographic and Health Survey Report. Central Statistical Aggency, Ethiopia.

Frank, B. (2006): Spatial Statistics for Epidemic Data, A Case of Cholera Epidemiology in Ashanti Region, Ghana,University of Twenty.

Girma, R., Wondwossen B., Bishaw D., and Tefera B.(2008): Environmental Determinants of Diarrhea among Under Five Children in Nekemte Town, Western Ethiopia.

Guofa, Z., Noboru, M. and Githeko, K. (2003): Association Between Climate Variability and Malaria Epidemics in the East African Highlands. Journal of Malaria, vol. 6:1-2.

Grillet, M., Roberto, B. and Marie, J. (2010): Disentangling the Effect of Local and Global Spatial Variation on a Mosquito-Borne Infection in a Neotropical Heterogeneous Environment. Tropical Journal of Medicine, Hyg.,82(2): 1-8.

Hay Si, Guerra, CA, Tatem, AJ, Noor, AM, Snow, RW: The global Distribution and Population at Risk of Malaria, Past, Present and Future. Lancet Infect Dis 2004, 4:327336.

Haining, R. (1990): Spatial Data Analysis in the Social and Environmental Sciences. Cambridge: Cambridge University Press.

Ingrid, P., Luisa, N., Wafaa, E., and Awash, T.: (2009): A Temporal-Spatial Analysis of Malaria Transmission in Adama, Ethiopia. The American Journal of Tropical Medicine and Hygiene, Vol. 86(6): 3-4.

Kandala, N., Stallard, N., Stranges, S. and Cappucio, F. (2008). Morbidity from Diarrhoea, Cough and Fever among Young Children in Nigeria. Annals of Tropical Medicine and Parasitology, Vol. 102, No. 5, 427445.

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.

Lawrence, N., Adamson, S., Christopher, C A., Immo K.(2007): Modelling the Effect of Malaria Endemicity on Spatial Variations in Childhood Fever, Diarrhea and Pneumonia in Malawi. International Journal of Health Geographics 2007, 6:33

Lesage, J., and Pace, K. (2009). Introduction to Spatial Econometrics. Boca Raton, Florida: CRC Press.

LeSage, J. (1998): Bayesian Estimation of Spatial Autoregressive Models. International Regional Science Review 20: 115-126.

Lira, P.,Ashworth, A. and Moris, S.(1996): Low Birth Weight and Morbidity from Diarrhea and Respiratory Infection in Northeast Brazil. J Prediater; 128 (4):497-504.

Ngianga, B., Kandala, L., Fahrmeir, Stephan, K. and Jan, P.(2008): Geo-Additive Models of Childhood Undernutrition in Three Sub-Saharan African Countries. Available on line at http://dx.doi.org/10.1002/psp.524 [Accessed on September 2009]

Ord, J. (1975): Estimation Methods for Models of Spatial Interaction. Journal of the American Statistical Association, 70: 120-126.

Smith, T. (2003): Development of Spatial Statistical Methods for Modeling Point- Referenced Spatial Data in Malaria Epidemiology. Journal of Malaria, Vol.7:1-25.

Smith, L. and Haddad, L. (1999): Explaining Child Malnutrition in Developing Countries. IFPRI Research Report No. 111, Washington DC.

Shanne, R. and Hendrik, R.(1998): Childhood Morbidity and Treatment Patterns. DHS Comparative Report; Macro International Incorporation; Cavelton, Mayland, United States of America.

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.

Tobler, W. (1979): Cellular Geography, Philosophy in Geography. Dordrecht, Reidel, pp. 379.

Tobler, W. (1970): A Computer Movie Simulating Urban Growth in the Detroit Region Economic Geography. 46 (2): 234-240.