SciencePG
Science Publishing Group

Geostatistics Analysis of Infant Mortality Rate in Ethiopia

Montasir Ahmed Osman Mohamed

Scientific Research and Development Center, Nawroz University, Kurdistan Region, Iraq

Email address:

a.montasir@yahoo.com

Abstract: In this paper, spatial statistical analysis of infant mortality rate in Ethiopia is addressed. The analysis investigated of a significance spatial autocorrelation attendance as well as an adapting of a generalized linear mixed model with spatial covariance structure. The results showed the distribution is much spatially associated. Some geographical, economical and healthy variables are used to estimate the model. Several examined variables have a significant effect in the model contrast to other have an insignificant impact. The results highlight the role of improving education to decline the risk of infant mortality rate. Male and children with extra weight are higher exposed and the risk is highly different from one zone to another.

Keywords: Spatial Statistics, Infant Mortality, Generalized Models, Mixed Models, Moran's I

1. Introduction

Infant mortality is the death of a child before completing the first year of age. The infant deaths number from every 1000 live births called infant mortality rate. This rate can be taken as an indicator to measure the health care and well-being of the society [3]. The most causes of infant mortality are birth defects, preterm, low birth weight, maternal complications of pregnancy and injuries such as suffocation [3, 6]. The infant mortality causes are significantly associated to structural factors like economic development, general living conditions, social wellbeing and the environment quality [4]. In 2005 the United Nations stated in the human development report the most powerful indicator to capture the divergence in the human development is child mortality [5].

The recent World Health Organization (WHO) reports showed 75% of under-five mortality happens within the first year of age. The risk of infant mortality in African countries is 55 per 1000 live births, and this is more than five times higher compared to European countries which the rate is 10 per 1000 live births [1]. Longitudinally, rates of infant mortality have clearly declined. In 1960, the estimated rate was 122 deaths per 1000 live births, while is 32 deaths per 1000 live births in 2015 [1, 2, 7].

When the analyzing data is collected in a geographical dimension, it is important to test for spatial dependence. If a spatial autocorrelation is detected, the locational attributes of units contain information about the variables. If this association is ignored, this might lead to biased estimators and false conclusion from the study. So, spatial statistical methods are necessary to be used to improve the precision of the results [8].

Nowadays, it is not uncommon for spatial statistical modelling to be used in medicine, biology, demography, environment and other fields because of the need for describing spatial variability in the data. The mixed generalized linear model is a useful tool to analyze spatial data [9, 10].

In this paper, spatial autocorrelation of the infant mortality rate in Ethiopia will be investigated. Moran's I and other related tests are the tools to examine the significance of spatial autocorrelation of infant mortality rate among the Ethiopian regions. Depending on the results of spatial autocorrelation, a generalized linear mixed model with spatial covariance structure will be adapted. Fifteen independent variables covered geographic, demographic and social domains are used in the model. For analyzing purpose, ArcGis, GeoDa and SAS software are used.

2. Spatial Autocorrelation

Spatial autocorrelation investigates in the term of what happens in a location is related or is not related to what happens in the neighboring locations, in addition to measure

of this relationship depending on the geographical data [11, 12, 13]. The required data should be points or polygons. The measures which used to scale spatial autocorrelation classified into two types, global and local measures of spatial autocorrelation. Global measures of spatial autocorrelation are measures apply to display in a single value the pattern of distribution for a single variable [14, 15]. The distribution pattern investigates in the randomness of data spread among the whole region. The investigation has to figure out one of two cases, there is clustering in data distribution or no clustering [16, 17]. Several measures are used to calculate these measures. The most common measure of spatial autocorrelation is Moran's I [11]. It uses the points or polygons of the regions as well as variable values to compute the value of spatial autocorrelation [15]. Its formula is:

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

Where: N is the number of observations (points or polygons), $\bar{x}$ is the mean of the variable values, x_i is the variable value at a particular location, x_j is the variable value at another location, W_{ij} is a weight indexing location of i relative to j and n is number of neighboring regions. Local measures are used to calculate the spatial autocorrelation between each sub-region and its neighboring which shared the borders. There are local versions of the Moran's I as follow [14]:

$$I_i = z_i \sum_{j=1}^{n} W_{ij} z_j \qquad (2)$$

Where: $z_i = \frac{x_i - \bar{x}}{SD_x}$, W_{ij} is a weight indexing location of i relative to j and n is the number of neighboring regions. These formulas are commonly used to test the spatial association in demography [8, 9], economics [18], diseases [19] and ecology [20].

3. Generalized Linear Mixed Model

Generalized Linear Mixed Models (GLMM) have attracted statisticians attention over the last decades. The word "Generalized" means non-normal distributions of the response variable, and the word "Mixed" means random effects in addition to the usual fixed effects of regression analysis [21, 22]. In other words, the GLMM are an extension of the generalized linear model in which the linear predictor contains random effects in addition to the fixed effects. The general formula of GLMM is:

$$y = g^{-1}(\eta) \qquad (3)$$

Where, y is the response is a vector, g is a linked function and η is the linear predictor function of fixed and random effects, then:

$$\eta = X\beta + Z\gamma + \varepsilon \qquad (4)$$

Where, X is a matrix of fixed effects, β is a vector of unknown parameters of fixed effects, Z is a matrix of random effects, γ is vector of unknown parameters of random effects assumed to be normally distributed with mean 0 and variance G, and ε is a vector of random errors term assumed to be uncorrelated to γ as well as normally distributed with mean 0 and variance R [23, 24, 25]. GLMM have become widely used in all fields regarding its flexibility to deal with different types of data and distributions.

4. Data and Analysis

The cross-sectional studies for Demographic and Health Surveys (DHS) are not uncommon in many countries over the world. The data in this paper is selected from the DHS conducted by the Ethiopian government in 2011 [26]. The survey objected to provide a demographic and health estimates for Ethiopia as a whole. The survey involved 596 randomly selected clusters and the total of 11654 households covered all administrative areas in the country. In this paper, the dependent variable is the dead of the child within the first year of age (binary variable) in addition to 15 independent variables potentially have an effect of infant mortality. The analysis of spatial association in the data and estimation of a generalized linear mixed model with spatial covariance structure will be done in this section.

To test the spatial autocorrelation in the data, Moran's I technique is used. The null hypothesis, in this case, is the spatial distribution of infant mortality rate is random, contrast to alternative hypothesis; infant mortality rate distribution is spatially correlated. The observed value for Moran's I is 0.141 with standard error 0.07, Z-value 2.149 and P-value 0.017. This value clearly pointed out there is a spatial autocorrelation in the infant mortality rate distribution in Ethiopia. Figure 1 and figure 2 below display the Moran's I scatter plot in addition to infant mortality rate map to show how the rates are distributed.
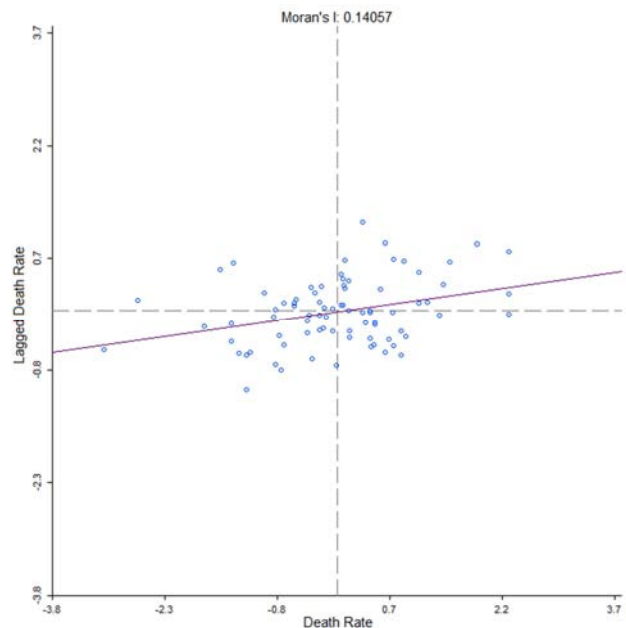

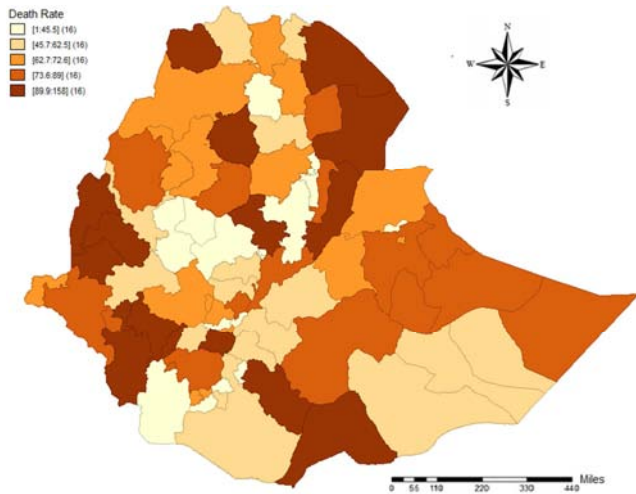*Figure 1. Moran's I scatter plot for infant mortality rate.*

Figure 2. Infant mortality rate quantile map.

Figure 1 reveals the significance of Moran's I plot slope which means significant spatial autocorrelation. The map in figure 2 exhibits the huge variation in mortality rate from area to another. The rates are classified into five groups each contain 16 areas. The first one has minimum rates from 1 to 45.5 infant deaths per 1000 live births; the others are 45.7–62.5, 62.7-72.6, 73.6-89 and 89.9-158. Some administrative areas have the same class in mortality rates in neighboring areas. On the other hand, some polygons are surrounded by a different class of mortality rates. In general, the classification of infant mortality rate by neighboring polygons is obviously attained.

Once, the spatial association is attendant in the data, the estimated model must be depends on this correlation. So, a generalized linear mixed model for the data is estimated. Death of the child is assigned as binary distributed variable, the link function is logit, the fixed effect variables are region, place of residence, respondent education, electricity availability, wealth index, twin of child, sex of child, size of child at birth, members of household, number of children less than five years, total children ever born, age of respondent at 1st birth, number of living children, hemoglobin level and husband/partner's age and the random effect is the cluster location with Gaussian spatial covariance structure (SP(GAU)).

Table 1. Significance of effects of the estimated model.

Effect	F-value	P-value
Region	0.68	0.740
Place of residence	0.03	0.858
Respondent education	0.91	0.436
Electricity availability	1.04	0.308
Wealth index	1.26	0.285
Child is twin	15.89	0.000
Sex of child	1.47	0.225
Size of child at birth	6.83	0.000
Members of household	1.38	0.240
Number of children <5	89.88	0.000
Total children ever born	177.42	0.000
Respondent age at 1st birth	4.90	0.027
Number of living children	128.57	0.000
Hemoglobin level	0.29	0.590
Husband/partner's age	8.75	0.003

The results in table 1 showed there are seven variables of examined effects have a significant impact on infant mortality rate in Ethiopia. These variables are the twin of the child, the size of the child at birth, the number of children less than five years, total children ever born, respondent age at 1st birth, the number of living children and husband/partner's age where the p-value is less than 0.05 for all these effects. On the other hand, the remaining variables in the model have an insignificant impact.

Table 2. Fixed effects estimates of the model.

Effect	Estimate	Std. Error	t-vaue	P-value	OR
Intercept	-2.002	1.508	-1.330	0.185	0.135
Region (ref. Dire Dawa)					
Tigray	0.523	0.448	1.170	0.244	1.686
Affar	0.611	0.446	1.370	0.171	1.842
Amhara	0.232	0.458	0.510	0.613	1.261
Oromiya	0.535	0.425	1.260	0.208	1.707
Somali	0.778	0.449	1.730	0.083	2.176
Benishangul-Gumuz	0.416	0.446	0.930	0.352	1.515
SNNP	0.355	0.437	0.810	0.416	1.426
Gambela	-0.032	0.509	-0.060	0.950	0.969
Harari	0.373	0.515	0.720	0.469	1.452
Addis Ababa	0.247	0.624	0.400	0.693	1.280
Place of residence (ref. rural)					
Urban	0.070	0.390	0.180	0.858	1.072
Respondent education (ref. Higher)					
No education	1.510	1.065	1.420	0.156	4.526
Primary	1.345	1.062	1.270	0.205	3.839
Secondary	1.071	1.136	0.940	0.346	2.917
Electricity availability (ref. yes)					
No	0.385	0.378	1.020	0.308	1.470
Wealth index (ref. Richest)					
Poorest	-0.114	0.415	-0.280	0.783	0.892
Poorer	0.130	0.424	0.310	0.759	1.139
Middle	-0.173	0.436	-0.400	0.692	0.841

Effect	Estimate	Std. Error	t-vaue	P-value	OR
Richer	-0.464	0.412	-1.130	0.260	0.629
Child is twin (ref. 2nd of multiple)					
Single birth	-1.624	0.526	-3.090	0.002	0.197
1st of multiple	0.634	0.680	0.930	0.351	1.885
Sex of child (ref. female)					
Male	0.193	0.159	1.210	0.225	1.212
Size of child at birth (ref. Very small)					
Very large	0.868	0.254	3.420	0.001	2.382
Larger than average	0.878	0.269	3.260	0.001	2.406
Average	0.064	0.235	0.270	0.786	1.066
Smaller than average	-0.131	0.369	-0.350	0.724	0.878
Members of household*	0.045	0.038	1.180	0.240	1.046
Number of children <5*	-1.147	0.121	-9.480	0.000	0.318
Total children ever born*	0.906	0.068	13.320	0.000	2.475
Respondent age at 1st birth*	0.053	0.024	2.210	0.027	1.054
Number of living children*	-1.067	0.094	-11.340	0.000	0.344
Hemoglobin level*	-0.002	0.004	-0.540	0.590	0.998
Husband/partner's age*	-0.034	0.012	-2.960	0.003	0.966

*Effects of continuous variables are assessed as one unit offsets from the mean.

Table 2 present the estimates and the odd ration of the model including spatial variability covariance structure (SP (GAU)). Based on these results, by reference to the region by Dire Dawa zone, the odds of a child to die within the first year is 86.6% higher for those children in Tigray zone (OR=1.686). In general, with the exception of Gambela zone, there is a higher risk of a child to die before completing the first year in all other regions compared to Dire Dawa (OR > 1). But in Gambela zone (OR=0.969) the probability is 3.1% times less than Dire Dawa. In term of the place of residence the risk of infant mortality in urban is 7.2% times higher than rural (OR=1.072). Respondent education is clearly effected the risk of infant mortality, where the chance of risk is more than four times if the respondent has no education, more than three times if the respondent has primary education, more than twice times if the respondent has secondary education referenced to higher education. Availability of electricity reduces the infant mortality risk, when by reference to yes (electricity available) odd ratio for no is 1.470. Wealth index has shown different directions, where the risk is higher in the case of the respondent is poorer (OR=1.139), lower in the case of the respondent is poorest (OR=0.892), middle (OR=0.841) and richer (OR=0.629) referenced by richest. A number of children in the pregnancy is clearly effects the infant mortality, when, if the child is a single there is 80.3% lower to die within the first year referenced to 2nd of multiple (OR=0.197) compared to 88.5% higher if the child is 1st of multiple (OR=1.885). With respect to sex of the child, males are higher exposed to infant mortality than female, where there is 21.2% times greater for male than female (OR=1.212). Regarding size of the child at birth, the relationship is obviously positive with infant mortality, where when the size is very large or larger than average the risk is more than double (OR=2.382/2.406) referenced to very small size compared to OR=1.066 and 0.878 for average and smaller than average size respectively. Members of household, total children ever born and respondent age at first birth have a positive effect in infant mortality when each additional unit over the mean increase the risk (OR=1.046, 2.475 and 1.054 respectively). On the other hand, the number of children less than five years, the number of living children, hemoglobin level and Husband/partner's age negatively effect in the infant mortality, where the risk is decreasing by each unit above the mean (OR=0.318, 0.344, 0.998 and 0.966 respectively).

5. Conclusion

From the results in the paper, we can conclude the following statements:

1) Infant mortality rate is spatially correlated among Ethiopian regions.
2) The results highlight the role of improving education to decline the risk of infant mortality rate.
3) Availability of electricity for the household increase the probability for the child to complete the first year of age.
4) Male are more exposed to the risk than female.
5) The size of the child at birth significantly affect infant mortality rate and the relationship is negative.
6) The infant mortality risk is highly different from one zone to another.

References

[1] Global Health Observatory (2016), Infant mortality, Available at http://www.who.int/gho/child_health/mortality/neonatal_infant_text/en/index.html. 18 January, 2017.

[2] WHO (2016), Under-five mortality, Available at http://www.who.int/gho/child_health/mortality/mortality_under_five_text/en/index.html. 15 January 2017.

[3] Center for Disease Control and Prevention (2016). "Infant Mortality," http://www.cdc.gov/reproductivehealth/maternalinfanthealth/infantmortality.htm.

[4] Daniel Reidpath and Pascale Allotey (2003). "Infant Mortality Rate as an Indicator of Population Health", Journal of Epidemiology and Community Health, 57, 344–346.

[5] United Nations (2005), Human Development Report, New York, UNDP, 4.

[6] WHO (2011). "Newborn death and illness", http://www.who.int/pmnch/media/press_materials/fs/fs_newborndealth_illness/en.

[7] World Bank (2016). "Mortality rate, Infant", http://data.worldbank.org/indicator/SP.DYN.IMRT.IN.

[8] Holmberg, H and HäggströmLundevaller E. (2015), "A test for robust detection of residual spatial autocorrelation with application to mortality rates in Sweden", Spatial Statistics, 14, 365-381.

[9] Dawit G. Ayele&Temesgen T. Zewotir (2016), "Childhood mortality spatial distribution in Ethiopia", Journal of Applied Statistics, 43(15), 2813-2828.

[10] LICHSTEIN, JEREMY W., SIMONS, THEODORE R., SHRINER, SUSAN A., AND FRANZREB, KATHLEEN E. (2002). "SPATIALAUTOCORRELATION AND AUTOREGRESSIVE MODELS IN ECOLOGY", Ecological Monographs, 72 (3), 445–463.

[11] Getis, Arthur (2008), "A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective", Geographical Analysis, 40, 297–309.

[12] Chou, Y. (1997), "Exploring Spatial Analysis in Geographic Information Systems", Onward Press, Santa Fe.

[13] Anselin, L. (1992), "SpaceStat Tutorial: A Workbook for Using SpaceStat in the analysis of Spatial Data", Typescript. University of Illinois at Urbana-Champaign, pp. 8-67.

[14] Anselin, L. (1995), "Local indicators of spatial association - LISA. Geographical Analysis", 27, 93-105.

[15] Moran, P. A. P. (1950), Notes on Continuous Stochastic Phenomena, Biometrika 37 (1): 17–23.

[16] Geary, R. C. (1954), The Contiguity Ratio and Statistical Mapping, The Incorporated Statistician, 5 (3): 115–145.

[17] Getis, A and Ord, J. K. (1992), The analysis of spatial association by use of distance statistics, Geographic Analtsis, 24 (3): 189-206.

[18] Melecky, Lukas (2015), "Spatial Autocorrelation Method for Local Analysis of The EU", Procedia Economics and Finance, 23, 1102 – 1109.

[19] Osman, Montasir A. (2016). "GIS Use to Analyze Distribution of Malaria Spread in Kassala State", 5th International Conference of Union of Arab Statisticians, Cairo. Feb 9-10.

[20] Dale, Mark (2002). "Spatial Autocorrelation and Statistical Tests in Ecology", Ecoscience, 9 (2), 162-167.

[21] McCullagh, P, &Nelder, J. A. (1989). Generalized Linear Models, 2nd ed. Chapman & Hall/CRC Press.

[22] Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. Journal of the American Statistical Association, 88 (421), pp 9-25.

[23] Stroup, W. W. (2012), Generalized Linear Mixed Models, CRC Press.

[24] Jiang, J. (2007), Linear and Generalized Linear Mixed Models and Their Applications, Springer.

[25] Fitzmaurice, G. M.; Laird, N. M.; Ware, J.. (2011), Applied Longitudinal Analysis (2nd ed.), John Wiley & Sons.

[26] CSA, Central Statistics Agency of Ethiopia and ORC Macro (2012): Ethiopia demographic and Health Survey 2011, Central Statistics Agency and ORC Macro, Addis Ababa and Calverton, MD.