SPATIAL ANALYSIS AND MAPPING OF INFANT MORTALITY IN KENYA ON THE BASIS OF DEMGRAPHIC AND HEALTH SURVEY DATA

BY

MUTWIRI ROBERT MATHENGE

SCHOOL OF MATHEMATICS

COLLEGE OF BILOGICAL AND PHYSICAL SCIENCES

UNIVERSITY OF NAIROBI

A PROJECT SUBMITTED TO THE SCHOOL OF MATHEMATICS IN PARTIAL FULFILLMENT FOR THE AWARD OF MASTER OF SCIENCE DEGREE IN SOCIAL STATISTICS

August 16, 2010

NAIROBI – KENYA

Declaration by Student

I the undersigned declare that this project is my original work and to the best of my knowledge has not been submitted for the award of degree in any other University.

MUTWIRI ROBERT MATHENGE

REG NO: I56/72440/08

Signature..... Date..30-07-2010

Declaration by supervisors

This project has been submitted for examination with my approval as a supervisor.

DR.THOMAS. O.ACHIA
SCHOOL OF MATHEMATICS
UNIVERSITY OF NAIROBI
Signature.....
Date..30/07/2010

MRS. ANN WANG'OMBE
SCHOOL OF MAHEMAICS
UNIVERSITY OF NAIROBI
Signature.....Wangombe.....
Date..30-07-2010

Acknowledgements

I would like to thank my senior supervisor Dr Thomas O. Achia and Mrs. Ann Wang'ombe for there invaluable support, guidance and encouragement throughout the project and the entire graduate programme.

To my course mates, Morris Mungai, Bonface Kimathi, Lilian and Eunice, thank you to some it all, without your wise counsel, encourangement, discussions and concerns my studies would have been the same.

Special thanks to my father and mother,Henry and Margaret , my brothers and a sister Betty for support and encourangement throughout the study. Lastly but not least to the Almighty God for protection, strength, guidance, wisdom and courage to face the studies without fear.

Dedication

To grandfather Zachary M'Murithi who wished to be great scholar but was forced out of school to Herd cattle and My Parents for taking care of me.

Abstract

This study set out to examine and map the spatial variation of infant mortality in Kenya. We used data from Demographic and Health Survey (DHS) database to explore spatial variation. Generalized linear mixed model(GLMM) with Enumeration Areas (EA) specific random effects was used to assess the effects of geographical heterogeneity and other covariates. The model based Geostatistical methods were used to quantify the spatial variations of the observations using the variograms and fitted the exponential and matern parametric models to the sample variograms. Then utilizing the fitted variogram function, Trans-gaussian kriging was performed infant mortality rates based on both models and produced smooth maps.

Generalized linear mixed model (GLMM) showed significant geographical heterogeneity in infant mortality. However, moran's I statistic showed spatial autocorrelation unaccounted for by GLMM. Modeling the correlation between people as a decreasing function of the spatial distance between them, Geostatistical models gave information no only on the magnitude but also on the scale of spatial variation. The socioeconomic status and infant mortality varied significantly across districts in Kenya. EA indicators better explained spatial variation of mortality when measured across a continuous space rather than within administrative areas.

The resulting map broadly agreed with the the previous studies on the variation of risk in the country, and further showed marked variation even at local level. High risk areas were in Nyanza regions, while low risk areas are in Central of the country. The maps provided an initial description of the geographic variation of IMR in Kenya, and might help in the choice and design of interventions, which is crucial for reducing the child mortality by two thirds by 2015 .

ABBREVIATIONS

EA : Enumeration Area

HIV : Human Virus

MDG : Millennium Development Goals

GDP : Gross domestic Product

GLMM : Generalized Linear Mixed Model

IMR : Infant mortality rate

KDHS :Kenya Demographic and Health Survey

UN : United Nations

UNICEF : United Nations International Children's Fund

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Background to the Study

More than 10,000 newborn babies die every day as a result of health problems that can either be prevented or treated. Every year it is estimated that under-nutrition contributes to the deaths of about 5.6 million children under the age of five; 146 million children in developing countries are underweight and at an increased risk of early death. [27]. Infant mortality, (the probability of dying between birth and exactly one year of age, expressed per 1,000 live births) is considered as a standard indicator of population health used throughout the world; the rates of infant mortality can reflect levels of social and economic development, levels of care and the effectiveness of preventive programs as well as post birth services to both mothers and their children. [28]. It is one of the components of United Nations human development index [37] hence its description is very vital for planning public health strategies in sub Saharan Africa [26]. In Kenya, approximately eight out of every 100 live births die before their fifth birthday, representing a huge wastage of potential manpower. [21]. According to [6], about 97% of newborn infants are expected to survive through the first five years. They intimated that reduction in this survival probability in any society is due to the operation of social,

economic, biological, and environmental forces. To reduce the present national level of infant mortality rate by two-thirds by 2015 is one of the Millennium Development Goals (MDG).

In Kenya, public spending on health care has been and continues to rise without corresponding improvements in health outcomes. During 2006, the health budget alone as a percentage of GDP and total Government expenditure accounted for 7.92%, and approximately 2 per cent respectively. Despite increased spending in health, there has not been a significant improvement in the health indicators in the country. Indeed, most of the health indicators have been worsening for a period of over two decades as shown by the trend of infant mortality rate. The trend has been 119 in 1969, 88 in 1979 and further declined to 66 per 1000 in 1989, although there was a sudden substantial increase in the year 1999 as it was 77.3. [21]. To draw feasible interventions programmes to enhance further reduction of infant mortality rate in Kenya, it is of most important to understand the direction and magnitude of socioeconomic, biological and demographic factors that are affecting child mortality in Kenya.

Previous studies in Kenya have mainly focused on the contribution of individual and household factors in explaining infant and child mortality differences in the country. Such studies have also found infant and child mortality differentials at aggregated levels of region while neglecting the critical influence of community-level variables and small area variations (spatial autocorrelation) [14, 7, 12, 8]. In the face of dwindling economic resources and global economic meltdown, spatial analysis and mapping of child mortality will help improve our understanding of the mortality situation in the country and subsequently in the design and implementation of policies and interventions to lower disparities and achieve uniform decline through out the country. This can be done by spatial analysis and mapping of the important factors so that government interventions should focus on those key areas that will have direct and immediate impact on child health instead of engaging in multifaceted approach with no visible impact.

1.1.1 Statement of the Problem

Weak health information infrastructure that does not take into account the spatial variability of mortality gives insufficient support for planning and resource allocation. [8] noted that for Kenya to attain the Millennium development goals, spatial inequalities among the ethnic groups and provinces need to be addressed. Most studies in Kenya have consistently reported regional differences in infant and child mortality. However, there has been no effort to model and map the geographical differences in mortality. Limited statistical work has been carried out from the data arising from the demographic and health surveys in Kenya. Previous works have concentrated on investigating the determinants of infant and child mortality.

Thus, the current study entails finding a suitable model that relate child survival and its determinants to its geographical location in Kenya. The Trans-gaussian kriging model is utilized in predicting the spatial distribution of infant mortality and its determinants.

1.1.2 Objective of the study

The main objective of this study is to explore the spatial variation of infant mortality in the Kenya and to identify some of the factors behind the variation.The study will also highlight any evidence of mortality concentration among geographic localities in Kenya.

Specific Objectives

i . To investigate whether there is spatial autocorrelation in the data.

ii .To model the geographical trend of infant mortality.

iii . To predict and produce smooth maps of infant mortality in Kenya.

1.1.3 Significance of the Study

Though infant mortality in Kenya is one of the highest in sub Saharan Africa standard, there is a dearth of study which integrates geographical background in conceptualization and in the right perspectives. Most of the existing studies give overemphasis to maternal background and child related factors. As any intervention programme aims toward' providing cushion to child survival can't be individual oriented, there seems not much relevance of the findings of the existing studies on infant and child mortality in Kenya. We have no idea on how geographical location and inequality in living standards affect infant survival.

Taking into account the aforesaid discussion and also consideration of the fact that intervention programmes for ensuring child survival target geographical locations and recognizing that Kenya is geographically, socially, culturally, economically and demographically heterogeneous, we strongly felt the need to look at the contribution of geographical location to distribution of infant mortality in Kenya. This study shall not only fill the gap in research at least in the context of Kenya but also provide a vital direction for policy formulation and implementation. We are also not aware of any other study that was designed to test for geographical location on infant mortality in Kenya. In most of the available studies regions are considered in place of geographical location and this is not the ideal way of dealing with geographical background.

Chapter 2

LITERATURE REVIEW

The literature focused on two key areas which were considered important in analysis and mapping of infant mortality. These includes studies on determinants of infant mortality and spatial and analysis and mapping of mortality.

2.1 Determinants of Infant Mortality

There is a huge literature on the determinants of infant and child mortality across the globe. Most studies in childhood mortality have focused on socioeconomic and demographic factors (for example [30, 31, 7, 10, 35, 8]), demographic (for example [36]), biological for example [34, 14], environmental factors (for example [12] through the use of censuses and surveys data.

[22] followed mosley and Chen framework and studied the trends and determinants of childhood mortality in the late 1980's and 1990's. They attributed the increase in infant and child mortality to the increase in HIV prevalence in the period, 1980-1990's.

Regarding the association between socioeconomic status and child mortality, [9] found spatial variations in Malaysia. [30] noted that education can contribute to child survival by making

women more likely to marry and enter motherhood later and have fewer children, utilize prenatal care and immunize their children. Similar findings have been reported elsewhere. [31].

[34] studied biological and behavioral influences of infant mortality of Malaysian infants. They found the influence of breast feeding on infant mortality. [17] reviewed the literature on the causes of infant and child mortality differentials in East Africa. He noted that infection, malnutrition and diarrhea were the major killers of infants and children, while the mother's level of education or lack of it appeared to be the central variable. [18] examined the effects of parental education and demographic factors on infant mortality in Lesotho. He found the effects of parental education on infant survival. [32] studied the effect of access to health care on infant mortality in Indonesia. She found that adding a maternity clinic to a village decreases the odds of infant mortality by almost 15%, compared to the risk of none before while an additional doctor reduces the odds by about 1.7%. [6] proposed a framework for studying the determinants of infant and child survival in developing countries, which considered the five proximate determinants, namely, maternal factors (age, parity, birth interval), environmental contamination (air,food, water), nutrient deficiency; injury (accidental, intentional) and personal illness. They noted that over 97% of newborn infants are expected to survive through the first five years. They further intimated that the reduction in this survival probability in any society is due to the operation of social, economic, biological and environmental forces.

In Kenya, [22] reported an inverse relationship between mother's education level and economic status (wealth index) and child mortality. While for the relationship between urban/rural residence and child mortality, urban areas showed higher mortality risks than rural, but when adjusted for HIV prevalence, child mortality was lower in urban areas. [7], followed Mosley and Chen framework and noted that a birth interval of less than 19 months was associated with an almost 50% increase in the risk of death while an interval of 36 or more months

reduces the risk by about 32%. Adjusting for a number of socioeconomic factors [8] found that child survival was better for those who were of birth order 2-3, birth interval more than 2 years, not outcomes of multiple births, living in wealthier households, had a access to drinking water and sanitation facilities, and users of low polluting fuels as their main source of cooking. Maternal age, maternal education and gender of the child had no significant association with child mortality. [13] examined the causes and risk factors affecting infant morbidity and mortality in Pakistan and found that the different determinants of infant morbidity and mortality include age, sex, multiple births, and mode of delivery, gestational age, and weight at birth, parity of mother, vaccination, maternal education, age, birth spacing and socioeconomic conditions. Breast feeding is an important determinant which lowers the rate of infection related morbidities. [10] studied the impact of community well-being on infant mortality in Orissa, India. They found the influence of community well being on infant mortality. In the study of [20] on the determinants of child mortality in rural Punjab, she found the existence of death clustering in child mortality. Other determinants of infant and child mortality had supported the significant effect of some biological and demographic predictors on the the phenomenon. [25, 14, 7]

2.2 Spatial Analysis and Mapping of Infant Mortality

Spatial differences in child mortality have been reported in Africa [19, 33, 2, 15, 5]. Risk factors contributing to this geographical variation include bio-demographic and socio-economic factors, the prevalence of infectious diseases and the variability in the quality of child health care [17, 16]. [19]used simple regression analysis and found spatial variation of infant mortality by a number of demographic and socioeconomic variables. [33] studied spatial variation in child mortality in ten West African countries and come up with findings suggesting that places closer to coastal areas and urban centers negates child survival. [9] studied spatial patterns

and association between mortality and prosperity in Malaysia and found spatial variation of mortality and prosperity across districts. [2] used Bayesian geostatistical models to study the spatial patterns of infant mortality and malaria endemicity in Mali. They intimated that accounting for spatial correlation resulted in more precise estimates of the standard error and widens the confidence limits of the estimated log odds ratios. [1] reviewed three procedures for fitting spatial generalized linear mixed model. [15] proposed a spatial generalized linear mixed model to describe the variation in the prevalence of malaria among a sample of village resident children in Gambia. They noted that the estimates of the presumed spatial autocorrelated village level random effects showed strong evidence of spatial dependence. [38] studied spatial association between malaria pandemic and mortality. They developed a multivariate spatial regression model for estimating the risk of mortality associated with malaria across Ogun state in Nigeria. They further noted that taking account of spatial autocorrelation in the data widens the confidence interval and improves the model.

2.3 Overview of Literature Review

There is a general consensus in literature that the determinants of infant mortality are not static but vary from one place to another across Kenya suggesting they are spatial. This is true for studies which employ both direct and indirect techniques to estimate infant and child mortality. As observed in most studies, household's income has significant effect on children survival prospects. Higher mortality rates are experienced in low income households as opposed to their affluent counterparts.

The mother's level of education is strongly linked to child survival. Higher levels of educational attainment are generally associated with lower mortality rates, since education exposes mothers to information about better nutrition, use of contraceptives to space births, and

knowledge about childhood illnesses and treatment. Larger differences have been found to exist between the mortality of children of women who have attained secondary education and above and those with primary level of education or less.

Differentials by urban/rural residence have commonly been observed, with urban areas having more advantages and therefore better child survival prospects. Differentials by region of residence and ethnicity have also been observed with Nyanza province inhabited mostly by Luo ethnic group having the highest infant mortality rates while Central province inhabited by Kikuyu community has the least infant mortality rates.

As concerns the demographic variables, the patterns of mortality by maternal age and birth order are typically U-shaped. Children born to both relatively old and young women have higher mortality rates than others; the interpretation of the effect of maternal age at birth on infant mortality must be biological, that is, it depends on reproductive maturity. Moreover, first and higher order births also have higher mortality rates since the birth order reflects the components of the child's biological endowments. As for the child's gender, it is widely believed that male mortality is higher due to biological disadvantages.

Chapter 3

METHODOLOGY

3.1 The Data

The data used in spatial analysis and mapping of infant mortality is taken from the 2003 Demographic and Health survey (DHS) for Kenya. The 2003 DHS covered 8,195 women aged 15-49 and 3578 men aged 15-54 selected from 400 Enumeration areas (EA's) throughout Kenya. The survey covered both rural and urban populations. The survey collected detailed information relating to demographic, child health care and GIS coordinates of EA's.

The demographic and health surveys utilized a two-stage sample design. The first stage involved selecting sample points from a national master sample frame maintained by Central Bureau of Statistics (CBS) the fourth National Sample Survey and Evaluation Programme (NASSEP) IV. In 2003, a total of 400 EA's 129 in urban and 271 in rural areas were selected. From these samples, the desired sample of households was selected using systematic sampling methods. There was small number of households selected in the North Eastern province because of difficulties in traveling and interviewing in the sparsely populated largely nomadic areas in that province. Households were also oversampled in Rural areas due to heterogeneity

of sample proportions.

The Demographic and Health Survey did not collect coordinates of individual households as a result its not possible to apply Generalized linear mixed modeling approach to measure the spatial variation of infant mortality. We can overcome the absence of GIS coordinates at household level by aggregating information collected at individual level to proportions explaining significant risk and fitting spatial linear model. To achieve this, we will use Generalized Linear Mixed Model with EA specific random effects on socioeconomic, demographic and biological variables to assess the geographical heterogeneity on infant mortality. While Spatial aggregation of point referenced data is increasingly used, there are some limitations on there use. Aggregation of point referenced data into spatial units such as census tracts or regions are more reflective of data collection and/ or modeling convenience rather than homogeneous or cohesive regions in the real world. The sdmistrative units are modifiable and contain artifacts related to degree of spatial aggregation or placement of boundaries. This may lead to loss of information and biased conclusions, thus not the best approach [42].

Continuous variables were categorized before starting the analysis. Age of the mother at birth was categorized into three age groups less than 20, 20-35 and 35 and above. The breakpoints were selected because of the need to cover the different reproductive trajectories: adolescence, young adults and adults. Age of the mother at first birth was categorized into two groups less than 19 and 19 or more, to capture the prevalence of teenage pregnancy. Breastfeeding status was categorized into ever and never while birth order was categorized into three groups: first order, 2-4 birth order and 5+ birth order. Preceding birth interval was grouped into two groups: less than 2 years and 2 or more years. Categorized variables were further edited by combining of their groups in one or two groups either because of the small number of observations in those categories or to make the analysis and the interpretation more meaningful. Also in ethnicity variable, ethnic groups with less than 50 observations

like Embu, Kisii, Kuria, Mijikenda/Swahili, Taita/Tavate and Turkana were all grouped into 'others'group. In the case of religion of the mother, those with no religion are combined with those whose religion is other than Christianity and Islam. For birth size variable, category of those being considered as larger or very large than average as large. Also the category, smaller or very small than average were combined into small birth size. This categorization has been used by [8].

3.1.1 Generalized Linear Mixed Model

In the first section of this study we apply Generalized linear mixed model with EA-specific random effects to assess the effects of geographical heterogeneity and sociodemographic factors on infant mortality on KDHS data (2003). With this approach the dependent variable is child survival status which is a binary outcome taking value 1 if child is alive and 0 otherwise. The independent variables used are mothers age at first birth, mothers work place, mothers education level, fathers education level, ethnicity, birth size, birth order and interval breast feeding status which were found to be associated with infant mortality.

Generalized linear mixed model is prototypically a logistic regression with random effects. It is used to assess the effects of categorical variables on the dependent variable. In this analysis we will use GLMM to determine the categories of covariates that explain the highest risk of infant mortality for aggregation and use in spatial analysis. The first category of each variable is used as a reference category against which other categories are compared in the original data.

Let $s_i : i = 1, \ldots, n$ be n spatial locations where the response Y_{ij} is observed. At each location s_i, the response Y_{ij} is associated with a covariate vector X. Then the Y_{ij} is conditionally bernoulli distributed which follows an exponential distribution function with the link function

given by

$$log(\frac{p_i}{1-p_i}) = X'(s)\beta + \alpha Z \tag{3.1.1}$$

X and Z are $n \times p$ and $n \times q$ vector of known covariates, β is $1 \times p$ vector of unknown parameters and α is a $1 \times q$ vector of unknown parameters The Likelihood function is given by

$$\int \prod_{j=1}^{n=1} f_{ij}(y_{ij}|\alpha_i, \beta, \Phi) f(\alpha|D) d\alpha_i \tag{3.1.2}$$

which was approximated by laplace technique using glmer function in lme4 package of R 2.11.0

3.2 Analysis of Spatial Autocorrelation

In the analysis of the spatial autocorrelation, the application of contiguity measure is required. Contiguity has a rather broad definition depending on the the research question; however, most analysis in spatial autocorrelation adhere to a common definition of neighborhood relations such as rook's case, bishop's case or queen's case.

Rook's case contiguity considers neighborhood of four locations adjacent to each EA. Meanwhile, bishop only considers the diagonal of the relationship and queen's or king's case considers a neighborhood of eight EA, the probability of infant death has been measured and spatially analyzed using the queen's neighborhood structure.

To measure the autocorrelation, the moran's I correlation coefficient, was employed. This measure was used to evaluate the spatial pattern and examine the existence of the spatial autocorrelation among the mean probabilities of infant mortality in the different EA's. Moran's I is one of the oldest indicators of the spatial autocorrelation, which has been used in many studies [9]. To examine the assumption of independency, the morans I formula proposed by

Moran 1950 is given by;

$$I = \frac{N \sum_i \sum_j W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_i \sum_j W_{ij}) \sum_j (x_i - \bar{x})^2} \tag{3.2.1}$$

where N is the number of EA's, x_i is the probability of infant death at location i, x_j is the probability of infant deaths in location j, $\bar{x}$ is mean probability of infant mortality given by $\bar{x} = \frac{\sum x_i}{N}$, and w_{ij} is the spatial weights applied between location i and j.

Like the correlation coefficient, the values of moran's I range from 1, which indicates a strong positive spatial autocorrelation, to 0 which equals to random pattern and -1 which indicates strong negative spatial autocorrelation.

We used the moran's I correlation coefficient to test the hypothesis:

H_0 : No spatial autocorrelation (No spatial clustering of Infant deaths)

H_1 : There is spatial autocorrelation (The is clustering of infant mortality)

Moran's I spatial correlation coefficient was calculated for the probability of infant death in kenya using queens structure of contiguity.

$$Z = \frac{I - E(I)}{\sqrt{var(I)}}, \tag{3.2.2}$$

$$E[I_i] = \left(\Sigma_j \frac{w_{ij}}{m_2}\right) E[z_i z_j],$$

where the expectation term is $E[z_i z_j] = \frac{-m_2}{(n-1)}$, $E[I_i] = \frac{-w_i}{(n-1)}$ and where z_i and z_j a pair of coordinate points, w_{ij} is the spatial weight matrix at location i and j, w_i is the spatial weight matrix at location i, m_2. It is evidently a ratio of quadratic forms in Y that provides the idea for obtaining approximate first and second moments through the delta method. Moran shows under the null model where the Y_i are independently and identically distributed, I is asymptotically normally distributed with

$$E(I) = \frac{-1}{n-1}$$

and

$$var(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2 S_0^2}$$

where $S_0 = \Sigma_{i \neq j} W_{ij}$, is the sum of the elements in the weights matrix so that if this matrix is row standardized to 1 then $S_0 = n$, $S_1 = \frac{1}{2}\Sigma_{i \neq j}(W_{ij} + W_{ji})^2$ and $S_2 = \Sigma_k(\Sigma_j(W_{kj} + \Sigma_i W_{ik})^2$, w_{ij} and w_{ji} is the spatial weight of the link between i and j.

Local spatial clustering

Local Indicators of Spatial Associations (LISA's) measure the degree of spatial dependence to allow for the effects of neighborhood based on each cluster's associated value (in this case, mortality), where neighborhood is defined according to some measure of proximity or contiguity. The main purpose of such indexes is to provide a local measure of similarity between each region's associated value and those of nearby regions. [41] proposed the local Moran's I_i statistic to test for local autocorrelation. Local spatial clusters, sometimes referred to as hot spots, may be identified as those locations or sets of contiguous locations for which the local Moran's I_i is significant. He stated that the indication of local patterns of spatial association may be in line with a global indication, although this is not necessarily the case. It is quite possible that the local pattern is an aberration that the global indicator would not pick up, or it may be that a few local patterns run in the opposite direction of the global spatial trend. Local values that are very different from the mean would indicate locations that contribute more than their expected share to the global statistic. These may be outliers or high leverage points and thus would invite closer scouting. Moran's I_i serves two purposes or provide two interpretations: First, it may be interpreted as indicator of hot spots. Second, it may be used to assess the influence of individual locations on the magnitude of the global Moran statistics and to identify outliers. The [41] defined I_i as:

$$I_i(d) = \frac{(x - \bar{x})}{\frac{1}{n}\Sigma_i^n (x_i - \bar{x})^2} \Sigma_{j=1, j \neq 1}^2 W_{ij}(d)(x_i - \bar{x}), \tag{3.2.3}$$

where $w_{ij}(d)$ is the weight matrix given a local neighborhood search of radius d. He showed

that the sum of all possible local moran values is equal to the global moran's statistic. The expected value of local moran statistic is $E(I_i) = -\frac{\sum_j W_{ij}}{(n-1)}$ and variance is $Var(I_i) = \frac{w_{i2}(n-b_2)}{n-1} + \frac{2w_{i(kh)}(2b_2-n)}{(n-1)(n-2)} + \frac{w_i^2}{(n-1)^2}$, where $b_2 = \frac{m_4}{m_2^2}$ and $m_r = \frac{\sum_i Z_i^r}{n}$ is the r-th moment of z, $w_i(2) = \sum(w_{ij}^2) j \neq i$ and $2w_{i(kh)} = \sum_k \sum_j w_{ik} w_{ih}$, $k \neq i$ and $h \neq i$, $Z(I_i) = \frac{(1-E(I_i))}{\sqrt{var(I_i)}}$.

3.3 Variogram Analysis and modeling

To test the hypothesis of spatial dependency, we fitted two empirical variograms, exponential variogram model and Matern variogram model, dependent variable being the proportion of infant deaths per EA. The independent variables considered in the analysis are; proportion of large/small births, proportion of average births size, proportion of breast feed infants, proportion of infants never breast feed, proportion of mothers < 4 years at first birth, proportion of mothers 19+ at first birth, proportion of infants born < 2 years to preceding birth, proportion of infants born 2+ years after the preceding birth, proportion of infants with < 4 living siblings, proportion of infants with more than 4 siblings, proportion of fathers with primary education, proportion of infants with secondary+ education, longitude and latitude. The probability of infant death at location i depended s on a set of variables X. We view $y(s)$ as a random variable where s denotes the coordinates s_i and s_j such that

$$Y(s) = x(s) + \epsilon(s) \tag{3.3.1}$$

where X(s) is some function of s, known as the drift or trend and $\epsilon(s)$ is a random variable with variance defined by

$$\begin{aligned} Var[Y(s_i) - Y(s_j)] &= E\{[Y(s_i) - Y(s_j)]^2\} \\ &= 2\gamma(h) \end{aligned} \tag{3.3.2}$$

The variogram function is given as:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{(s_i,s_j)\in N(h)} [Y(s_i) - Y(s_j)]^2 \tag{3.3.3}$$

where $Y(s_i)$ is the i^{th} observation at location s_i and $Y(s_j)$ is the is the j^{th} observation at location s_j, N(h) is the set of pairs of points such that $|s_i - s_j| = h$, (s_i, s_j) a pair of coordinates and $|N(h)|$ is the number of pairs in the set.

3.4 Parametric models

In order to carry out kriging for optimal linear prediction, we fitted two parametric models over the estimated empirical variogram. Several possible choices of the parametric models such as the linear model, exponential model, Matern model, rational quadratic model, wave model and spherical model are available [39].

i. Matern model

The matern model is defined [39] as

$$\hat{\gamma}(h) = \begin{cases} \tau^2 + \sigma^2[1 - \frac{(2\sqrt{v h\phi})^v}{2^{v-1}\sqrt{v}} K_v(2\sqrt{v h\phi}] & \text{if} h > 0 \\ 0 & \text{Otherwise} \end{cases} \tag{3.4.1}$$

where τ^2 denotes the nugget effect that measures the microscale variation which may have resulted from discontinuity process of the mortality rates among the EA'S, $\tau^2+\sigma^2$ is the sill that quantifies the availability of the mortality rates in long distances. If $\sigma^2 = 0$, $\hat{\gamma}(h)$ is a constant for all $|h| > 0$ that indicates no spatial correlation of the mortality rates among the EA's; Φ is the range that can be interpreted as the distance beyond which the infant mortality rate is correlated.

ii. Exponential Model

In addition to the matern model, for comparison purposes we fitted also the exponential

model on the empirical variograms. The exponential model is defined [39] as

$$\hat{\gamma}(h,\phi) = \begin{cases} 0 & h = 0 \\ \tau^2 + \sigma^2\{1 - exp^{(\frac{-|h|}{\phi})}\}, & h \neq 0 \end{cases} \tag{3.4.2}$$

where the parameters in the exponential model have the same interpretation as that in the matern model. However the correlation decays to zero in the exponential model.

The trend in our data is model using a linear function 3.3.1. the linear component of the model is non constant function. In order to fit a spatial trend model to our data which has unknown parameters $\hat{\theta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p, \tau^2, \sigma^2, \Phi)$. We shall use the maximum likelihood method to estimate these parameters. The log likelihood function to be maximized is:

$$\begin{aligned} L(\beta, \tau^2, \sigma^2, \Phi) = -0.5\{nlog(2\pi) + log\{(|\sigma^2 R(\Phi) + \tau^2 I)|\} \\ + (y - X\beta)'(\sigma^2 R(\Phi) + \tau^2 I)^{-1}(y - X\beta)\} \end{aligned} \tag{3.4.3}$$

3.5 Model Assumptions

In the above model, we make the following assumptions;

i. S is a stationary gaussian process with $E[S(s)] = 0$ and $cov\{S_i, S_j\} = \sigma^2\rho(s_i - s_j)$;

ii. Conditional on S, the Y_i: $i = 1, \cdots, n$ are mutually independent gaussian random variables, with expectation $\mu(x_i) + S(x_i)$ and variance τ^2;

iii. The spatial correlation between location i and j is a function of distance between them.

iv. The predicton function $Y(s_0)$ of location s_o is is a linear function of S.

3.6 Trans-gaussian Kriging

To predict the probability of infant deaths at the unsampled locations, we used logit transform of the proportion proportions of infants death in given by

$$y_i^* = h(p_i) = log(\frac{p_i}{1-p_i})$$

with inverse transform

$$p_i = h^{-1}(y_i^*) = \{1 + exp^{-y_i^*}\}^{-1}$$

and used the linear gaussian model to analyze the data y^*-scale,to obtain predictors for $\hat{s}(x)$ and associated prediction variance V(x) at each location. The target prediction on the original scale is $Y(s) = h^{-1}\{s(x)\}$ for an arbitrary location x. We followed Tylor series expansion of Y about $\hat{s}(x)$ and writing $g(.) = h^{-1}(.)$ the dependence on location x is given by;

$$Y(s_0) \approx g'(\hat{s}) + (s - \hat{s})g'(\hat{s}) + 0.5V(x)g''((\hat{s})) \tag{3.6.1}$$

Hence the trans-gaussian predictor for $Y(s_0)$,

$$\hat{Y}(s_0) = \{1 + exp^{-\hat{s}(x)}\}^{-1} - 0.5V(x)exp^{-\hat{s}(x)}\{1 - exp^{-\hat{s}(x)}\}\{1 + exp^{-\hat{s}(x)}\}^{-3} \tag{3.6.2}$$

Trans-gaussian kriging provides an optimal way of estimating the value of Y at any given location (EA). We used these estimates to form smooth maps of infant mortality rates of the region under study.

3.7 Data Analysis and Model selection

3.7.1 Data Analysis

In order to analyse and map infant mortality, three levels of analysis we carried out three levels of anaysis: univariate to determinate the variables associated with child survival status,

generalized linear mixed model fit to determine the risk on infants survival and geostatistical analysis to model spatial dependency, predict and produce smooth maps of infant mortality.

All the analysis were carried using R v2.11.0 statistical software and GeoDa spatial analysis software. The R packages functions used in this analysis are implemented in geoR , maptools, Rcolorbrewer and spdep.

3.7.2 Model selection

i. AKaike Information Criterion (AIC)

A satisfactory compromise between goodness of fit and parsimony can be achieved by applying the Akaike Information Criterion (AIC). For a given set of data the variable part of the AIC is estimated by

$$AIC = -2logL(\hat{\beta}, \hat{\theta}, \hat{\sigma}^2, \hat{\tau}^2) + 2(p+1) \tag{3.7.1}$$

where p is the number of parameters in the model. The information criterion leads to the selection of the model with the smallest value of (3.7.1), with p introducing a penalty term for models with more parameters.

3.8 Results

Fitting the generalized linear mixed model on the data, we assume that p is the probability of infant will die before its first birthday given a set of fixed explanatory variables by taking EA specific random effects to assess the effects of geographical heterogeneity. Then $\frac{p}{1-p}$ is the odds of infant dying before its first birthday and $ln\frac{p}{1-p}$ is the log odds of p. The observations

Table 3.1: Association between IMR and sociodemographic factors

Variable	**Values of $\chi 2$**	**df**	**Significance at 5% level**
Breast feeding	1203.00	1	<0.0001
Birth Size	29.93	5	<0.0001
Preceding Birth interval	33.16	1	<0.0001
Birth order	9.90	2	0.003
Maternal age at first birth	10.07	1	0.001
Partners education level	29.83	6	<0.0001
Region	42.88	7	<0.0001
Ethnicity of mother	85.52	8	<0.0001
Mothers education level	30.65	5	<0.0001
Mother works place	5.74	1	0.017
District of residence	160.71	68	<0.0001

are child survival status within one year of birth that is

$$y_i = \begin{cases} 1, & \text{if child is alive,} \\ 0, & \text{Otherwise.} \end{cases}$$

Since the dependent variable is a log odds of each parameter is also on the log odds of p scale. The fixed effects parameters predicts the log odds of infant death. This part of output shows the distribution of the deviance residuals for individual cases used in the model. We exponentiate the coefficients and interpret them as the odds ratio. In the final model [Table 3.2], we concluded only on variables that were significantly associated with child survival based on χ^2 tests in the univariate analysis (are shown in Table 3.1). The univariate results suggest that factors such as breast feeding, ethnicity, sex of the child, size of the child at birth, region

Table 3.2: Variables in the Final GLMM for factors associated with IMR

Variable		Estimate	odds	Std. Err	Pr(> |z|)
	(Intercept)	0.77838	2.1780	0.45547	0.087455 .
Birth Size	Small		1.00		
	Average	0.66829	1.951	0.18787	0.000375 ***
	Large	0.34281	1.4080	0.20657	0.097009 .
Ethnicity	Kikuyu		1.00		
	Kalenjin	−0.5686	0.5663	0.37242	0.126814
	Kamba	−0.5663	0.5676	0.38942	0.145884
	Luo	−0.6709	0.5112	0.32770	0.040625 *
	Luhya	−1.7022	0.1823	0.31890	9.40e-08 ***
	Maasai	0.5163	1.6758	0.67943	0.447289
	Meru	−0.4026	0.6686	0.59160	0.496193
	Somali	−0.6077	0.5446	0.40223	0.130820
	Other	−0.4891	0.6132	0.33855	0.148578
Breast feeding	Never		1.00		
	Ever	1.33783	3.811	0.22696	3.76e-09 ***
Birth Interval	<2 years		1.00		
	2 | years	0.57908	1.7844	0.15179	0.000136 ***
No Living children	<4		1.00		
	4-7	0.4721	1.6033	0.1822	0.009563 **
	8 |	−0.0224	0.9779	0.2955	0.939631
Paternal education	No education		1.00		
	Primary	0.2052	1.2277	0.2580	0.426463
	Secondary	0.6790	1.9719	0.2907	0.019516 *
	Higher	1.5766	4.8383	0.5100	0.001991 **
	σ^2	0.426		0.65269	
AIC	1669				

Table 3.3: Full Model and Reduced Model fit information

Model	AIC	BIC	LogLik	Deviance
Full Model	1678	1830	-813.9	1628
Reduced model	1669	1785	-815.6	1631*

of the child, preceding birth interval and birth order, maternal and paternal education level and the number of living children were significantly associated with infant mortality in Kenya.

Table 3.2 represents the estimates of fixed effects coefficients of the best fit model, significance probability and relative odd ratios which are calculated for each categorical variable. The regression coefficient for birth size was 0.6683 which implies that adjusting for the effects of birth size, an average birth has a positive impact on infant mortality. It was significant at 5% level. The odd ratio for birth size is 1.951 which indicates that an an average birth has 1.951 higher chances of surviving as compared to a small birth. Ethicity is the most important determinant of the outcome in Kenya. Infant deaths is more likely for the Luo [odds=0.5112, 95% confidence level] and Luhya [odds=0.1823,95% confidence level] compared to Kikuyu infants. The estimated regression coefficient and the odds ratio of Ever breast feed infants was 1.3378, which have positive significant effect at 5% level on infant survival. Birth interval of 2 or more years reduces the risk of infants death by a factor of 1.7844 as compared to infant born less than two years after the preceeding birth. While children born to mothers with 4-7 surviving children have lower risk of death comapared to children born to mothers with less than 4 surviving children [odds=0.4721, 95 confidence interval]. Fathers level of education significantly impacts on infants survival with children born to fathers who have no education at a higher risk of dying as compared to their counterparts with secondary or higher education [odds=1.9719, 95% confidence interval].

Taking EA, Table 3.3 shows results of the the reduced model fit based on the rule of thumb,"the smaller the AIC the better the model". The estimate of the random effects was

0.426 [standard deviation=0.65269].

3.9 Descriptive Results of the selected categories of variables

Out of 4 727 single live births that took place between exactly one and five years before the survey, there were 368 deaths before the first birth date giving IMR of 77.8 per 1000 live birth. Table 1 below shows that most of the interviewed (79.7 %) have at least attended the primary school compared to (83.34%) of the fathers. As shown in the table about three-fifths of live-births were to mothers with primary education in the rural areas compared tO (25.53%) in urban areas. In rural areas, most of the births took place at home (67.6%), while 74.47% were in urban areas. The number of male and female babies in the study sample is almost the same in both areas. Most of the respondents, in both urban and rural areas, reported that their average birth size. The percentage of teenage deliveries is about 20% in both areas, while for those who gave birth at age of 35 and above was almost doubled (13.2%) in rural compared to urban areas. The infant mortality rate is calculated from the information drawn from questions asked in the birth history section of the women's questionnaire. In the birth history, for each live birth, information is collected on each sex, month and year of birth, survivorship status and current age or if the child had died, the age at death. The relationship between infant mortality and various socioeconomic and demographic factors are examined. In terms of levels and differentials of infant mortality, Nyanza province (128 1000) has the highest level of IMR while Central province (45.76) has the least one. Again, Luo children (158.19 per 1000) have the highest IMR while it is the least in the Kikuyu (37.08 per 1000) group. Urban (75.39 per 1000) and rural (78.69 per 1000) areas show almost the same level of IMR. Mothers with no education have the highest level of infant mortality than mothers who have primary, secondary or higher education in Kenya.

Table 3.4: Levels and differentials of Infant mortality rate in Kenya

Variable	No. births	IMR per. 1000	Variable	No. births	IMR per. 1000
Residence Type			**Ethnicity**		
Urban	1207	75.39	Kalenjin	490	83.67
Rural	3520	78.69	Kamba	434	57.60
Gender			Kikuyu	836	37.08
Male	2607	84.39	Luhya	778	82.26
Female	2120	69.81	Luo	531	158.19
Breast Feeding			Masai	148	67.57
Never	232	668.10	Meru	159	44.03
Ever	4455	46.02	Somali	486	113.17
Birth order			Others	865	59
First born	1201	64.95	**Religion**		
2nd to 3rd	1675	75.22	Catholic	2832	78.39
4th to 4 years	617	132.90	Protestant	774	90.91
Maternal Education			Muslems	137	58.39
No Education	955	92.15	No Religion	979	70.48
Primary	2735	84.095	**Fathers education**		
Secondary	834	51.56	No Education	733	102.3
Mothers age at birth			Primary	2132	88.2
Less than 19 years	2232	90.96	Secondary	1179	58.5
Above 19 years	2495	66.13	Higher	356	28.1
Age of the mother			**Wealth Index**		
Less than 20 years	249	92.37	Poorest	1183	92.14
20-29 years	2663	75.85	Poor	887	77.79
30-39 years	1455	72.85	Rich	868	80.65
40-49 years	360	102.78	Richest	1049	72.45
Type of Delivery			**Delivery place**		
Ceaserean	151	544.30	Home	2737	80.38
Normal	2593	57.43	Hospital/clinic	1990	74.37
			Total	**4727**	**77.8**

Chapter 4

RESULTS

4.1 Spatial Descriptives

We first explored variations in infant mortality by administrative districts as per 1999 census as shown in Figure 4.3. Infant mortality rates varied from 0 to 267 per 1000 live births across district in Kenya.The map of standard deviations of infant death rates showed spatial variations in infants deaths rates in Kenyan districts, with Samburu,Makueni, and Kericho districts showing the highest variations as shown if figure 4.1.

Moran's I, the classic indicator of spatial autocorrelation was calculated for infant mortality at 0.0123. The spatially autocorrelated random or E(I) was -0.0025, and the standard deviation was 0.0319. The significance test of I under the assumption of normality (Z-test) was 0.4653 (p=0.3209) and was not significant. Because the value of Moran's I was greater than the expected I, this test indicates that Infant deaths are clustered and not dispersed; however, the significance tests indicate that the difference between these two values is not significantly greater than what would be expected by chance alone.The moran scatter plot is shown in Figure 4.5.

The levels of infant mortality differed substantially within the country as seen from Figure 4.3. A striking example was found in many districts (e.g, Garissa, Tana River, Tharaka, and

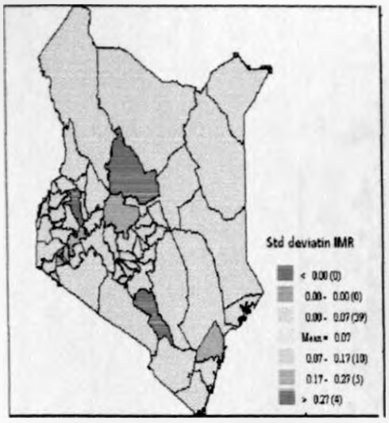


Figure 4.1: Infant Mortality

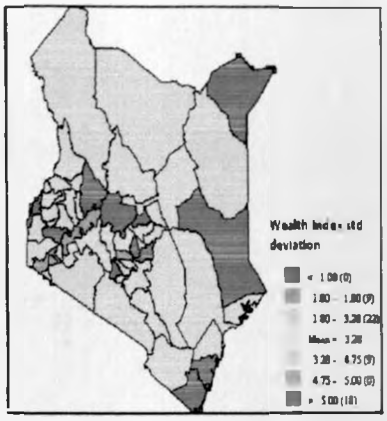


Figure 4.2: Wealth Index

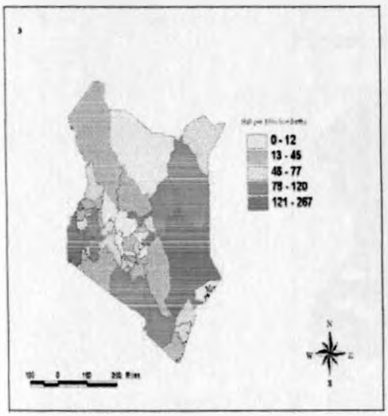


Figure 4.3: Distribution of IMR

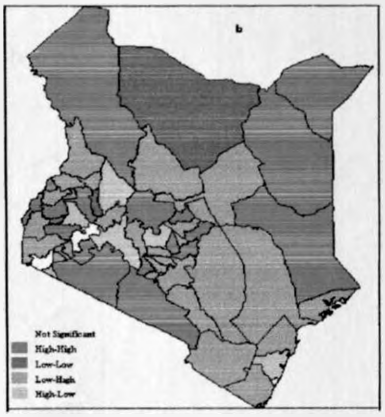


Figure 4.4: Association between IMR and wealth index

Nyanza region), which showed relatively high levels of mortality as shown in figure 4.3. As causes of death were not studied, explanations of this spatial pattern are not straightforward. Whether material or individual circumstances cause mortality differences may be debatable.

Figure 4.6 is the LISA map infant mortality in Kenya. The colored districts are significant with $p = 0.05$ for each district. The map indicated the three main patterns of spatial association, that is, clusters of high values surrounded by high values (HH), clusters of low values surrounded by low values (LL) and low values surrounded by high values (LH). It is worth noting that there is high infant mortality in districts with high socioeconomic inequalities and

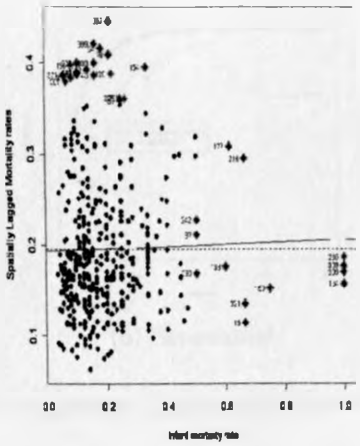


Figure 4.5: Moran Scatter plot

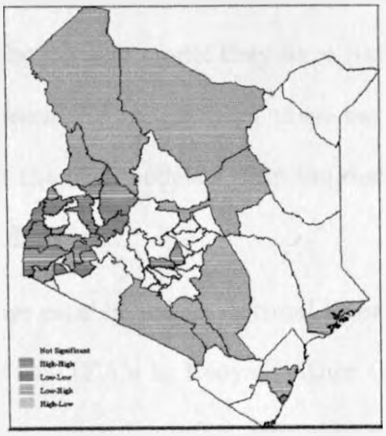
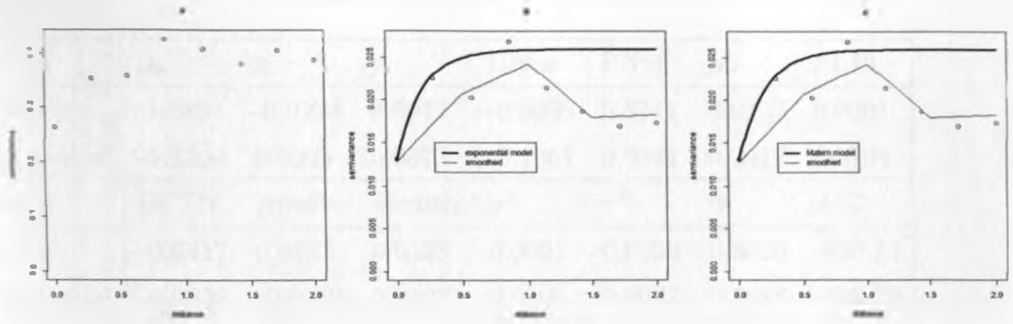


Figure 4.6: LISA of Infant Mortality

low infant mortality in districts with low socioeconomic status(wealth index) 4.4.

4.2 Variogram Models of infant mortality

The empirical variograms and their fitted matern and exponential curves are shown in figure 4.7, a, b and c for infant mortality rate in Kenya. These plots suggests that spatial correlation of infant mortality rates exists. If there were no spatial correlation among the



(a) Empirical (b) Exponential (c) Matern

Figure 4.7: Fitted empirical, Exponental and Matern Variograms

infant mortality rates, the variograms would be flat across the horizontal axis (distance).The estimate of θ in exponential model of the infant mortality rates of the EA's for Kenya are (0.0455,0.0318,0.0330) and by the matern model they were (0.3261,0.0625, 0.5199). Although there was large differences between the ranges from these two models, from figure 4.7 there seemed little difference between the two models within the distance interval for the estimated empirical variograms from our datasets.

Using the fitted variograms, we established an optimal linear prediction (kriging) model of infant mortality rates for all places (EA's in Kenya) within the study area. As an example, suppose that we are interested in estimating infant mortality rates for EA_p in Kisumu district of Nyanza province which was not in the dataset. The location of the EA is 34.5^o longitude and -0.1^o latitude south. The estimated infant mortality rate by Matern model is 288.15 and for Exponential model 194.94 per 1000 live births respectively which is less than that from Matern model.

Based on the results of kriging, we formed smooth maps for the study area. The results using both variogram models are shown in Figure 4.8 a, b, c and d for infant mortality. From these maps, we can see that the infant death retes were higher in Western Kenya (around Nyanza province) and lower in central Kenya (around central Rift valley and Central province). Both

Table 4.1: Spatial trend with Matern correlation structure

Model	β_0	x	y	L2yrs	PNbf	ph	PL19
Matern	-1.089	-0.0054	0.0042	-0.0938	0.3741	0.0157	0.0801
Exponential	-1.3304	0.0005	0.0073	-0.1051	0.3941	0.0412	0.0704
Model	BSNN	Away	Wealth	σ^2	τ^2	Φ	AIC
Matern	-0.0417	0.0128	0.0333	0.0001	0.0709	0.3629	-620.14
Exponential	-0.0359	-0.0129	0.0427	0.0100	0.0637	0.5688	-628.79

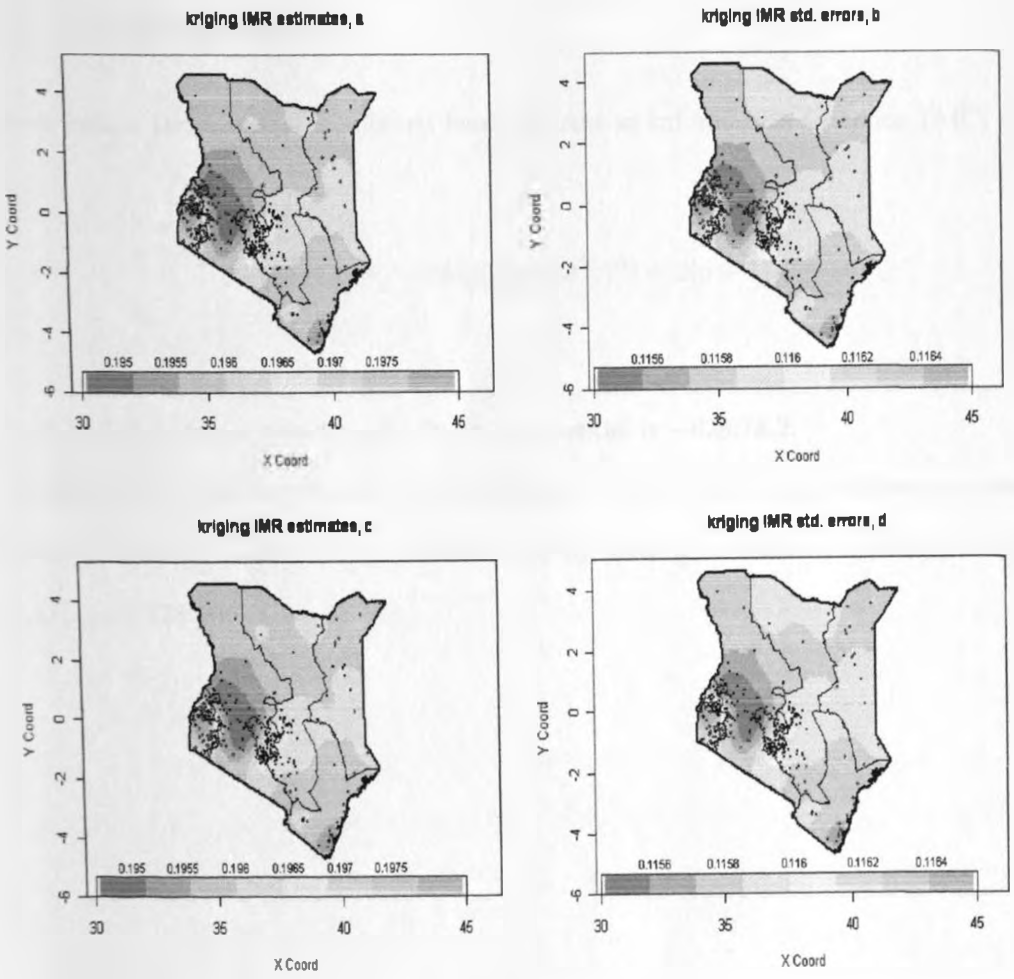


Figure 4.8: Spatial prediction of distribution of IMR in Kenya

Table 4.2: Model Selection

1			2	
Model	P	AIC	p	AIC
Exponential	13	-628.78	4	-623.86
Matern	13	-620.14	4	-713.99

Maternal and exponential models produced the same results.

4.2.1 Model selection

The best suface trend model is selected based by Akaike Information Criterion (AIC) defined as

$$AIC = -2logL(\hat{\beta}, \hat{\theta}, \hat{\sigma}^2, \hat{\tau}^2) + 2(p+1) \tag{4.2.1}$$

The AIC of the surface trend model fit by exponential is $-628.78.2$.

By comparing the AIC for the for four models, the best fitting model following exponential structure is model 1 with $AIC = -628.78$ and by matern correlation structure is Model 2 with $AIC = -713.99$.

Chapter 5

Discussion and Recommendation

5.1 Discussion

This study set out to identify the socio-demographic determinants of infant mortality as well as quantifying their impact with a view to identifying the relative importance of each variable and the geographical variation of infant Mortality in Kenya. The study has shown significant spatial variation in infant mortality in Kenya. The mean estimates of the residual smooth spatial effects (brown=high risk of infant death, purple=Low risk of infant death) are shown in the left panel of figure 4.8 a. In addition, the predicted variance of probability of infant death map in the right panel of Figure 4.8 b indicates significance of the spatial effects (Brown/Purple=significantly negative/positive effects, light blue=non-significant). In Nyanza, clustering of high infant mortality has been attributed to genetically linked diseases such as sickle cell anemia, HIV/AIDS infection, poor parenting and social as well as economic deprivation, [14, 8] and the prevalence of malaria [4].

Ethnic mortality differentials are largely attributed to differences in levels of socioeconomic development between areas inhabited by the various groups [8]. However, in Kenya, cultural differences in child health care practices and beliefs on disease causation and patterns of diseases could be more important [14]. For example, although the Kisii and the Luo ethnic

group inhabit Nyanza province, the risk of infant mortality has consistently been reported to be higher among the Luo group than the Kisii group [7, 14, 21, 8]. In another study, [40] found infant mortality to be high among the Luo children as compared to there counterpart from other communities, this suggests that there are some other unmeasured factors that explain clustering of infant mortality mortality rate in this community. The solution of the puzzle can be in the concentration of this tribe in Nyanza province, which at the same time showed the highest levels of IMR in the counrty. This province alone acccounts for about 30% of the National HIV burden in Kenya, which might be a major confounder in the analysis of infant mortality data. We did not have data to pursue this line of argument. The mystery of Luo and Nyanza needs more in-depth investigation. Other areas such as Nairobi have be shown to have high clustering of infants deaths this can be attributed to the fact that about 79% of the urban dwellers live in Kenya in slums which is associated with poor standards of living, pollution and prevalence of infectious diseases [29].while there is high clustering of infant deaths around Mt. Kenya in Mbere, Embu and Krinyaga districts. This can be attributed to infectious and waterborne diseases such as bilhazia. malaria etc due to the water logged soils which are good breeding grounds for mosquitoes that cause malaria and poor child care.

Districts of low infant mortality are restricted along the central Rift Vally region and North-west of central province. This low probability of child deaths in these region has been attributed to socio-economic advantage, urbanization and education ,[8, 14].

5.2 Study Limitations

Since our results depend on EA's sampled, the limitations of this study hinges primary on the sampled EA's. Only a few EA's were sampled in North eastern and upper Eastern region region due to large traveling distance and low population density in these areas. This may lead to a bias with regard to prediction at unsampled locations. Another limitation is the lack

of population density data for the sampled EA's which may confound the predicted results. Another serous limitation is the lack of HIV prevalence data for the infants, particularly since the experience in Kenya suggests that this infection is an important cause of reported infant mortality. Lack of effective treatment programmes using anti-retroviral drugs to prevent mother-to-child transmission is likely to result in increased HIV transmission. The viral load of HIV transmission and the influence of HIV may therefore cause infant mortality. HIV is thus a major confounder when considering the spatial variation of infant mortality rates thus studies might be strengthened by controlling HIV prevalence. This modeling extension can be the focus of further research when HIV prevalence data become available at child level.

5.3 Recommendation

Despite the limitations discussed above, we feel that this study fills a gap in knowledge of geographical variations of infant mortality in Kenya. The maps identify areas of increased risk and patterns which have important implications for health policy aimed at reducing all cause infant mortality by two thirds by 2015. Such a goal may take a comprehensive approach aimed at:

- A reduction in HIV incidence which affect infant mortality
- Targeted prevention of mother-to-child treatment programmes,
- Integrated management of infants illness and;
- Dedicated programmes aimed at improving the quality of child health care for areas where high childhood mortality have been found.

R scripts

```
data=read.csv("dfs.csv",header=TRUE)
names(data)
library(geoR)
attach(data)
data.geo=as.geodata(data,coords.col=2:3,data.col=4,
covar.col=c(5,6,7,8,9,10,11,12,13))
summary(data.geo)

##############################################
# Comparing empirical and theoretical
# variograms
##############################################
par(mfrow=c(2,2))
bin2 <- variog(data.geo,uvec = seq(0,2,l=8))
plot(bin2,ylim=c(0,0.027),main="b")
lines.variomodel(cov.model = "exp", cov.pars = c(0.013,0.2),
nugget = 0.013, max.dist = 2, lwd = 3)
smooth <- variog(data.geo, option = "smooth", max.dist = 2,
n.points = 5, kernel = "normal", band = 0.2)
lines(smooth, type ="l", lty = 1)
legend(0.3, 0.015, c( "exponential model", "smoothed"),
lty = c(1,1,2), lwd = c(3,1))
bin2 <- variog(data.geo,uvec = seq(0,2,l=8))
plot(bin2,ylim=c(0,0.027),main="c")
lines.variomodel(cov.model = "matern", cov.pars = c(0.013,0.2),
nugget = 0.013, max.dist = 2, kappa = .5, lwd = 3)
smooth <- variog(data.geo, option = "smooth", max.dist =2,
n.points = 5, kernel = "normal", band = 0.2)
lines(smooth, type ="l", lty = 1)
legend(0.3, 0.015, c( "Matern model", "smoothed"),
lty = c(1,1,2), lwd = c(3,1))

##########################################################
```

```
# PARAMETER ESTIMATION
#####################################################
#Exponential Model
ml1 <- likfit(data.geo,ini=c(0.015,0.22),nug = 0.013,lambda = 0.5,
 trend=~coords+L2yrs+pNbf+ph+pL19+BSNN+Away+wealth)
#Matern model
ml12<- likfit(data.geo,ini=c(0.015,0.22),nug = 10,lambda = 0.5,
 kappa = 0.5,trend=~coords+L2yrs+pNbf+ph+pL19+BSNN+Away+wealth)

################################################
# Creating the prediction grid
################################################
library(maptools)
kenya=readShapePoly("KEN_outline.shp")# Spatialpolygon object
#plot(kenya)
kenya1=kenya@polygons[[1]]@Polygons[[1]]@coords
str(kenya1)
class(kenya1)
#plot(kenya1)
############################################
# Regional boundaries
############################################
kenya2=readShapeLines("KEN-level_1.shp")
#############################################################
# Kriging
#############################################################
# defining the grid
pred.grid <- expand.grid(seq(34,42, l=100), seq(-5,5, l=100))
#############################################################################
#Color ordering
#############################################################################
library(RColorBrewer) # creates nice color schemes
nclr <- 8
plotclr <- brewer.pal(nclr,"BuPu")
#plotclr1 <- brewer.pal(nclr,"PuOr")
```

```
#plotclr <- plotclr[nclr:1] # reorder colors

################################################
#Spatial prediction of Infant Mortality in Kenya
################################################
kc <- krige.conv(data.geo, loc = pred.grid,
krige = krige.control(obj.m = ml1),borders=kenya1)
par(mfrow=c(1,2))
plot(kenya2)
image(kc, loc = pred.grid,col=plotclr,ylim=c(-6,5),
      x.leg=c(34,42),y.leg=c(-6,-5.5))
title("kriging IMR estimates")
plot (kenya2,add=TRUE)
points(data.geo$coords,type="p",cex=0.5)

image(kc,col=plotclr, loc = pred.grid,val=sqrt(kc$krige.var),
     ylim=c(-6,5), x.leg=c(34,42),y.leg=c(-6,-5.5))
title("kriging IMR std. errors")
plot (kenya2,add=TRUE)
points(data.geo$coords,type="p",cex=0.5)


library(spdep)
data.loc <- cbind(data$x, data$y)
# k nearest neighbors
k <- 4
nn <- knearneigh(data.loc, k, longlat=T)
data.neighbors.knn <- knn2nb(nn)
plot(data.neighbors.knn, data.loc, add=T, lwd=2, col="green")
# plotting second-order lags
col.lags <- nblag(data.neighbors.knn, 2)
#plot(col.lags[[2]], data.loc, add=T, lwd=2, col="red", lty=2)
# spatial weights matrix
w.cols <- 1:399
w.rows <- 1:399
```

```
w.mat.knn <- nb2mat(data.neighbors.knn, zero.policy=TRUE)
#image(w.cols, w.rows, w.mat.knn, col=brewer.pal(3,"BuPu"),
borders=kenya1)
moran.test(pii,nb2listw(data.neighbors.knn, style="W"))
moran.plot(pii,nb2listw(data.neighbors.knn),pch=19,
 ylab="Spatially Lagged Birth size",xlab="Birth size")
```

References

[1] Gemperli A. and Vounatsou P, (2003) Fitting generalized linear mixed models for point-referenced spatial data. *Journal of Modern Applied Statistical Methods*, 2 , 481-495

[2] A. Gemperli, P. Vounatsou, I. Kleinschmidt, M. Bagayoko, C. Lengeler and T. Smith. (2004) Spatial patterns of infant mortality in Mali: The effect of Malaria Endemicity. *American Journal of Epidemiology*, Vol. 159, pp. 64-72.

[3] Kleinschmidt I,Bagayoko M, Clarke GPY, Craig M, and Le Sueur D A spatial statistical approach to malaria mapping. *International Journal of Epidemiology* (2000),29(2): 355-361.

[4] Lawrence N. Kazembe, Christopher C. Appleton, Immo Kleinschmidt.(2007) Spatial analysis of the relationship between early childhood mortality and malaria endemicity in Malawi. *Geospatial Health* 2(1), pp. 41-50

[5] Liva Montana, Melissa Neumann and Vinod Mishra. (2007) Spatial Modeling of HIV prevalence in Kenya. DHS Working Papers, No. 27

[6] W. Henry Mosley and Lincoln C. Chen (1984) An Analytical Framework for the Study of Child Survival in Developing Countries. Population and Development Review, Vol. 10, Supplement: Child Survival: Strategies for Research (1984), pp. 25 – 45. Published by: Population Council.

[7] D. W.R Omariba, Roderic Beaujot and Fernando Rajulton (2007). Determinants of infant and child mortality in Kenya: an analysis controlling for frailty effects. *Population Research Policy Review* 26:299321 DOI $10.1007/s11113 - 007 - 9031 - z$.

[8] Hisham Elmahdi Mustafa, Lecturer, University of Khartoum, and Clifford Odimegwu (2007). Socioeconomic Determinants of Infant Mortality in Kenya: Analysis of Kenya DHS 2003. *Journal of Humanities and Social Sciences*, Volume 2, Issue 2, 2008

[9] Faisal G. Khamis, Abdul Aziz Jemaine, and Kamarulzaman Ibrahim. On spataial patterns and Association Between these Patterns of Mortality and Prosperity in Malaysia. *American Journal of Applied Sciences* 5(7): 881-890. 2008

[10] Barsharani Maharana and L.L. Singh (2009) Impact of Community Well-being on Infant Mortality in a Demographically Backward State in India, http://iussp2009.princeton.edu/download.aspx?submissionid=93413

[11] Elizabeth Zenger, (1993) Siblings' Neonatal Mortality Risks and Birth Spacing in Bangladesh Demography, Vol. 30, No. 3, pp. 477-488

[12] Mutunga C.J (2007) Environmental Determinants of Child Mortality in Kenya. World institute for Development Economics Research, United Nations University; Research Paper No. 2007/83

[13] Hafsa Habib, Maheen Lohani, Habibullah Khan, Muhammad Hussain Khan, (2009) Infant Morbidity Leading To Infant Mortality. *Gomal Journal of Medical Sciences*, Vol. 7, No. 2. pp

[14] Lawrence Ikamari, Sibling Mortality Correlation in Kenya. J*Journal of biosocial Scence* i. (2000) 32, 265-278.

[15] Peter , Rana Moyeed, Barry Rowlingson, and Madeleine Thomson, Childhood malaria in the Gambia: a case-study in model-based geostatistics. *Journal of Applied. Statistics.* (2002) 51, Part 4, pp. 493-506.

[16] Kandala Ngianga-Bakwin and Nyovani Madise, The Spatial Epidemiology of Childhood Diseases in Malawi and Zambia. *African Population Studies Supplement B* vol 19

[17] Stephen Kaduuli (1988) Infant and Child Mortality in Eastern Africa: Causes and Differentials. National Centre for Development Studies. Australian National University.

[18] Stephen Kaduuli (1989). Infant Mortality in Lesotho: Parental Education and Demographic Factors. http://ssrn.com/author=808607.

[19] Ezekiel Kalipeni (1993): "Determinants of Infant Mortality in Malawi: A Spatial Perspective." *Social Scienc Medical.* Vol. 37, No. 2. pp. 183-198.

[20] Monica Das Gupta (1990), Death Clustering, Mothers' Education and the Determinants of Child Mortality in Rural Punjab, India. Population Studies, Vol. 44, No. 3 (Nov., 1990), pp. 489-505 http://www.jstor.org/stable/2174464. Accessed: 11/12/2009 08:49.

[21] Central Bureau of Statistics (CBS) [Kenya], Ministry of Health (MOH), and ORC Macro. (2004): Kenya Demographic and Health Survey 2003. Calverton, Maryland: CBS, MOH, and ORC Macro.

[22] Hill, K., G. Bicego and M. Mahy. (2001): "Childhood Mortality in Kenya: An Examination of Trends and Determinants in the Late 1980s to Mid 1990s." http://www.jhsph.edu/popcenter/publications/pdf/WP01-01.pdf (accessed 19/07/2009).

[23] Ngianga-Bakwin Kandala, Chen Ji, Nigel Stallard, Saverio Stranges, and Francesco P. Cappuccio. *Spatial Analysis of Risk Factors for Childhood Morbidity in Nigeria.*, American. Journal. Tropical. Medicine. Hygiene., 77(4), 2007; pp. 770778

[24] Ngianga-Bakwin Kandala. Bayesian geo-additive modeling of childhood morbidity in Malawi. *Applied Stochastic Models In Business And Industry Appl. Stochastic Models Bus. Ind.*, 2006; 22:139-154

[25] Gyimah, S. O. (2002): "Ethnicity and infant mortality in Sub-Saharan Africa: The case of Ghana." Population Studies Centre, University of Western Ontario, Canada. http://www.ssc.uwo.ca/sociology/popstudies/dp/dp02-10.pdf (accessed 15/10/2008).

[26] McElroy, P.D., F.O. Ter Kuile, A.W. Hightower, W.A. Hawlley, P.A. Phillips-Howard, A.J. Oloo, A.A. Lal, and B.L. Nahlen. (2001): "All cause mortality among young children in western Kenya ". *American Journal of Tropical Medicine and Hygiene.* 64:18 – 27.

[27] UNICEF, (2006): State of World's Children 2006.

[28] Fukuda. Y, Keiko.N And Takehito . T (2004). "Wide range of socioeconomic factors associated with mortality among cities in Japan." *Health Promotion International*,19(2)

[29] Kyobutungi. C, Abdhalah A.Z, Ezeh.A and Yazoume' 'Ye'. The burden of disease profile of residents of Nairobi's slums: Results from a Demographic Surveillance System. *Population Health Metrics* 2008, 6:1 http://www.pophealthmetrics.com/content/6/1/1.

[30] Hobcraft, J. (1993): "Women's education, child welfare and child survival: a review of the evidence" . *Health Transition Review*; 3(2):159 – 173.

[31] Pavalavalli. G and B.M. Ramesh (1997). "Maternal Education and the Utilization of Maternal and Child Health Services in India" . *International Institute for Population Sciences Mumbai*, India Macro International Inc. Calverton, Maryland, U.S.A.

[32] Elizabeth Frankenberg(1995): "The effects of access to health care on infant mortality in Indonesia." *Health Transition Review* 5, 143 – 163.

[33] Balk D, Pullum T, Storeygard A, Neuman M, (2004). A spatial analysis of childhood mortality in West Africa. Pop Space Place 10, 175 – 216.

[34] Butz W.M, and J.P Habicth (1982). Biological and Behavioural Influences On The Mortality of Malaysian Infants, The Agency For International Development.

[35] Uddin .J, Hossan. Z, and Mohammad .O.U, (2009). Child Mortality in a Developing Country: A Statistical Analysis. *Journal of Applied Quantitative Methods*, 3(4) : 270−283

[36] H.C. Quamrul, Rafiqul. I and Kamal Hossain (2010). Effects of Demographic Characteristics on Neonatal, Post neonatal, Infant and Child Mortality. *Current Research Journal of Biological Sciences* 2(2): 132-138.

[37] Human development index. *United nations*

[38] B.M. Dansu and O.E Asiribo. Spatial Association between Malaria Pandemic and Mortality. *Data Science Journal*,6, 143 – 153.

[39] Cressie, N. A. C. (1993). Statistics for Spatial Data. *Wiley, New York*, revised edn.

[40] Moïsi J. C, Gatakaa H, Noor M. A, Williams T. N, Bauni E, Tsofa B, Orin S Levinel, J , Scott G. A (2010) "Geographic access to care is not a determinant of child mortality in a rural Kenyan setting with high health facility density." . *BMC Public Health*,http://www.biomedcentral.com/1471 – 2458/10/142

[41] Anselin, L. (1995). Local indicators of spatial association: LISA. Geographical Analysis 27(2), 93116.

[42] Neil Wrigley, Tim Holt, David Steel, and Mark Tranmer *Analysing, Modelling, and resolving the ecological fallacy*, in Longley and Batty. Spatial Analysis: Modelling in a GIS Environment *Wiley, 1996* pp. 24-25