

https://dspace.org/

Contextual influence, frailty and the spatial patterns of child mortality in Nigeria

Salau, Sulaiman

2011-07-07
http://hdl.handle.net/10539/10278

Downloaded from a DSpace repository. https://dspace.org/



University of the Witwatersrand, Johannesburg

CONTEXTUAL INFLUENCE, FRAILTY AND THE SPATIAL PATTERNS OF CHILD MORTALITY IN NIGERIA

By

Sulaiman Salau

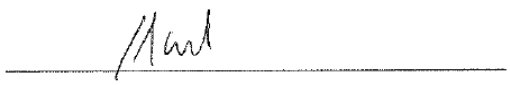
Supervisor: Professor Jacky Galpin
School of Statistics and Actuarial Science

A research report submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science.

February 2010.

DECLARATION

I declare that this research report is my own, unaided work. It is being submitted for the Degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.



(Signature of candidate)

February, 2010

ABSTRACT

The main aim of the project is to investigate statistical models that account for the influence of contextual factors and frailty on Child Mortality (CM) and to investigate the spatial patterns of CM in Nigeria. Using data from the Nigerian Demographic and Health Survey, results from a descriptive investigation of clustering showed that clustering of child mortality exists at the household, community and state levels and these need to be taken into account in the multivariate analysis by the inclusion of frailty effects at the relevant levels. A total of 8 models were evaluated using geo-additive survival models and the results in Chapter 4 reveal that the inclusion of frailty terms as well as the inclusion of contextual variables at the community level lead to an improvement in the model fit, thereby suggesting the importance of contextual and frailty effects. A higher share of state level variability in the data was due to the structured spatial effect. Although, the spatial patterns were found to be insignificant, they point to very interesting patterns in child mortality variations.

ACKNOWLEDGEMENT

I would like to express my gratitude to my supervisor Professor Jacky Galpin, whose suggestions, assistance, encouragement and comments has this study possible. My sincere thanks go to my family and friends for their support. Finally, I will like to thank the Demographic and Health Surveys program (www.measuredhs.com) for providing the survey data used in this study.

TABLE OF CONTENTS

DECLARATION i
ABSTRACT ii
ACKNOWLEDGEMENT iii
TABLE OF CONTENTS iv
LIST OF FIGURES v
LIST OF TABLES vi
LIST OF ABBREVIATIONS vii
Chapter One: Introduction 1
1.2 Background of the study area 6
1.3 Problem statement 8
1.4 Aims and Objectives of the Study 8
1.5. Organization of the Report 9
Chapter Two: Literature Review 10
2.1 Conceptual Framework 10
2.2 Pathway of Influence 12
2.2.1 Proximate Determinants 12
2.2.2 Household level factors 13
2.2.3 Community level factors 14
2.3 Child Mortality and its differentials in Nigeria 15
Chapter Three: Data and Methods 18
3.1 Data Sources 18
3.2 Statistical Methods 20
3.2.1 The Geo-additive Discrete-Time Survival Model 25
Chapter Four: Results 33
4.1 Unit of analysis and outcome 33
4.2 Variable selection 34
4.2.1 The selection and construction of community-level variables 34
4.2.2 Final Data set 36
4.3 Descriptive summaries 37
4.3.1 Results of the survival analysis 41
4.3.2 Investigation of clustering of deaths 43
4.4 Multivariate analysis 46
4.4.1 Modelling Strategy and Model Comparison Approach 47
4.4.2 Sensitivity analysis 50
4.4.3 Interpretation of categorical covariates (fixed effect) 51
4.4.4 Interpretation of non-linear effects 55
4.4.5 Interpretation of the spatial effect 56
4.5 Determinants of Infant mortality 58
Chapter Five: Summary and Conclusions 60
5.1 Summary 60
5.2 Recommendations 61
5.3 Limitations of the Study /Suggestions for future research 61
REFERENCES 63
APPENDICES 71

LIST OF FIGURES

Figure 1: Map of Nigeria showing the 37 spatial units considered.. 7
Figure 2: Schematic presentation of the conceptual framework for the study of CM 11
Figure 3: Hierarchical structure of the dataset 34
Figure 4: Maps depicting the nature of spatially explicit variables considered..... 35
Figure 5: Kaplan-Meier Survival Curves for community level covariates..... 42
Figure 6: Results from Spatial autocorrelation for U5M 45
Figure 7: Non-linear effects of metrical covariates 56
Figure 8: Maps of the posterior mean of spatial effects..... 57
Figure 9: Non-linear and spatial effects for IM 59

LIST OF TABLES

Table 1: Community level contextual factors to be considered in the study 19
Table 2: Community level variables from factor analysis 36
Table 3: Descriptive statistics of child level variables 38
Table 4: Descriptive statistics of mother level variables 39
Table 5: Descriptive statistics of household variables 40
Table 6: Descriptive statistics of community variables 41
Table 7: Distribution of births and deaths in households 43
Table 8: Distribution of births and deaths in communities 44
Table 9: Creation of child-period dataset from the original child-level data set 46
Table 10: Models considered 47
Table 11: Results from Models 1 -7b – Model fit and Variance components of random and non-linear effects 49
Table 12: Sensitivity to choice of hyperparameter values for Model 6 50
Table 13: Posterior summaries for child level effects models 1-7b 52
Table 14: Posterior summaries for mother level effects models 1-7b 53
Table 15: Posterior summaries for household effects models 1-7b 54
Table 16: Posterior summaries for community effects models 1-7b 55
Table 17: Posterior summaries for community effects model 6 - IM 58

LIST OF ABBREVIATIONS

AIC	Akaike's Information Criterion
CI	Credible Interval
CM	Child Mortality
DCW	Digital Chart of the World
DFID	Department for International Development
DHS	Demographic and Health Survey
DIC	Deviance Information Criterion
EA	Enumeration Areas
ESDA	Exploratory Spatial Data Analysis
FCT	Federal Capital Territory
GIS	Geographic Information System
GNP	Gross National Product
GPS	Global Position System
GPW	Gridded Population of the World
GRF	Gaussian Random Field
IM	Infant Mortality
K-M	Kaplan-Meier
LGA	Local Government Area
LISA	Local Indicators of Spatial Association
M-H	Metropolis-Hastings
MARA	Mapping Malaria Risk in Africa
MAUP	Modifiable Area Unit Problem
MCMC	Markov Chain Monte Carlo
MICS	Multiple Indicator Cluster Survey
MRF	Markov Random Field
NDHS	Nigeria Demographic and Health Survey
NIMA	National Imagery and Mapping Agency
NPC	Nigerian Population Commission
PPS	Probability Proportional to Size
TFR	Total Fertility Rate
U5M	Under five Mortality
UN	United Nations
WFS	World Fertility Survey
WHO	World Health Organization

Chapter One: Introduction

Childhood mortality (CM) remains a major public health issue in developing countries where it is estimated that over 10 million preventable child deaths occur yearly (World Health Organization [WHO], 2005). A high level of childhood mortality leads to high fertility through physiological, replacement and insurance effects (Preston, 1978; Montgomery and Cohen, 1998) resulting in rapid population growth - a situation that can hamper development. Indeed, mortality in the childhood years has been identified as an important indicator of a population's public health and socio-economic conditions (Masuy-Stroobant and Gourbin, 1995), and reduced childhood mortality not only offers opportunities for improving living conditions, but also has an effect on life expectancy. Of high priority in the developing world is the reduction of under-five mortality rates to two thirds of their 1990 levels by the year 2015 (United Nations [UN], 2000), with national governments as well as the international community supporting various research and intervention initiatives geared towards the attainment of the goal.

CM is usually monitored using two indicators: Infant mortality (IM) - (death between birth and first year of life) and Under-five mortality (U5M) - (death between birth and the fifth year of life). CM as used in this study refers to U5M except where otherwise stated. The focus on U5M is based on the fact that IM is a rare and noisy event and a large sample is often required for its' modelling (Mosley and Chen, 1984). Moreover, available statistics reveal that the national U5M in Nigeria is worse than IM (NPC [Nigerian Population Commission] and ORC Macro, 2004).

In spite of substantial reduction in CM rates experienced in most developing countries (Hill and Pebley, 1989), statistics reveal that the sub-Saharan Africa region has the greatest proportion (about 45%) of the global annual incidence of child deaths (UN, 2005). The pace of CM reduction within the African continent has also not been uniform, and gains in mortality reduction earlier experienced in some African countries have either begun to stagnate or reverse, raising fears that the millennium development target on CM reduction may not be met by the target date (Rutstein, 2000; UN, 2005; WHO, 2005). The situation is true for Nigeria, where CM rates are high, with evidence of substantial geographical variation across the country (NPC, 1998; NPC and ORC Macro, 2004). This

calls for an in-depth examination of the trends and patterns of CM rates, as well as their association with other factors, and the identification of high risk sub-groups for more effective targeting and CM reduction.

Most CM studies in Africa use data from surveys such as the Demographic and Health Survey (DHS). Such surveys use a stratified multistage sampling design, resulting in a hierarchically structured data set. The primary sampling unit is regions within a country, with communities being selected within each region. Within the community, households are selected, and limited information collected on the children and their parents. This sample design results in data that are correlated within households, within communities, and within regions, which needs to be taken into account when analyzing the data (Chromy and Abeyasekera, 2003).

The existing CM studies in Nigeria have primarily focused on the influence of a few individual and household factors in explaining CM differentials in the country (see for example: Iyun, 1992; Ahonsi, 1995; Adebayo, Fahrmeir and Klasen, 2004). The causes of CM are however multifaceted, often involving a number of factors operating at various levels and in complex ways. Despite evidence that factors at other contextual levels (arbitrarily or administratively defined geographical units such as household, community or region) may affect child survival (Mosley and Chen, 1984; Sastry, 1996; Root, 1997; Curtis and Hossein, 1998), there is a general scarcity of studies examining the influence of context-level factors on CM in developing countries. This can mainly be attributed to the non-availability of adequately measured contextual factors and the difficulty in incorporating them with routinely collected data when they are available (Sastry, 1996; Curtis and Hossein, 1998).

Statistical techniques mostly employed in the analysis of CM data include logistic regression (used when the dependent variable is binary) and Poisson regression (used when interest lies in modelling death rates) (Fahrmeir and Tutz, 2001). Standard specifications of these methods usually assume that the observations are independent (ignoring the hierarchical structure of the data utilized), and that heterogeneity (differentials in mortality) in the population under study can be explained by the set of

measured covariates included in the model (thereby neglecting the influence of unobserved heterogeneity). Failure to account for the clustering of CM risk and the influence of omitted covariates may yield inconsistent and inefficient estimates which in turn, may lead to invalid or wrong conclusions (Hobcraft, McDonald, and Rutstein, 1985; Guo and Rodriguez, 1992; Sastry, 1997a). Conventional specifications of the models also do not take into account time varying covariates (such as breastfeeding), non-linear effects of certain covariates (such as mother's age) and the censoring of observations. In CM studies the censoring type is right-censoring, occurring when a child has not run all the risks of death (by virtue of being less than the age of interest) and does not experience death during the period of interest. In using the logistic and Poisson regression models, data on recent births (children who are not yet five in the case of U5M) are usually excluded to alleviate bias caused by censoring. This approach leads to a loss of data that may carry valuable information.

The other set of techniques frequently used is survival analysis, appropriate when survival times and survival status data are available. The Cox proportional hazard (Cox, 1972) is given by:

$$h(t_i \mid X_i) = h_o(t_i)\exp(\beta' X_i)$$

where t_i is the time to death or censoring of child i, h_o is the baseline hazard, X_i is a vector of covariates and β is a vector of parameters. Survival analysis can accommodate censored observations in addition to modelling time varying and non-linear effects (Fahrmeir and Tutz, 2001).

In time-to-event studies with hierarchical structured datasets, there are two main approaches in modelling the data. The first is the fixed-effects method, which applies to situations were one is looking at specific treatments, and these are the only treatments of interest. In this case, variations in CM are explained entirely by the covariates included in the model, that is to say, unobserved heterogeneity is treated as a fixed parameter and is modelled as one of covariates. In modelling community specific heterogeneity, for example, one community is picked as a baseline community and a set of indicator

covariates is included for all other communities to estimate the community-specific differences.

The second approach is to use random-effects methods (Vaupel, Manton and Stallard, 1979), which are appropriate when one is using a sample of respondents and assumes that these are a random sample from a wider population. These models are often referred to as frailty models, a term used in the CM literature to explain the situation where children in certain groups are more susceptible to death than others, perhaps due to group-specific factors which are mostly unmeasured, immeasurable or unknown, or that result from a survey design which imposes a correlation of mortality risks among children belonging to the same group. The random-effects method assumes that the context-level frailty is distributed over the population according to some distribution function. In other words, random effects models incorporate frailty into the model estimates as an uncorrelated error component, and the frailty effects are considered as resulting from random sampling from a certain distribution function (whose mean and variance can be estimated).

To illustrate this model, let t_{ij} be the time to death or censoring for child j in cluster i. Let Z be a vector of child and cluster-specific covariates.

In the fixed-effect approach, the hazard rate for the j^{th} child from community i is modelled as:

$$h_{ij}\left(t_{ij} \mid Z_{ij}\right) = h_o\left(t_{ij}\right)\exp\left\{\beta' Z_{ij} + \theta' X\right\}$$

where β is a vector of regression coefficients for the covariates Z_{ij}, θ is a vector of unknown parameters, $X = \left\{X_1, \mathrm{K}, X_k\right\}$ and X_i=1 if child j is in community i, 0 otherwise. This fixed-effect approach therefore implies that there are i communities and the last one is the base category.

In the frailty setting, the hazard rate for the j^{th} child from community i is

$$\begin{aligned} h_{ij}\left(t_{ij} \mid Z_{ij}\right) &= h_o\left(t_{ij}\right)\exp\left\{\beta^{'} Z_{ij} + u_i\right\} \\ &= h_o\left(t_{ij}\right)\exp(u_i)\exp\left\{\beta^{'} Z_{ij}\right\} \\ &= h_o\left(t_{ij}\right)\xi_i \exp\left\{\beta^{'} Z_{ij}\right\} \end{aligned}$$

where $\xi_i = \exp(u_i)$ and the u_i's are assumed to be iid $\sim N(0, \sigma^2)$, and represent the cluster-specific frailty effect designed to capture differences among the clusters. The above approach can be extended to more than one nested level.

In general, CM studies that have considered frailty effects can be broadly classified into two groups. The first group consists of those that consider frailty at the family or community level (or both) as unstructured random components of the model and often ignore frailty at higher contextual levels (such as regions) (see, for example, Curtis, Diamond and McDonald, 1993; Madise and Diamond, 1995; Guo and Rodriguez, 1992; Guo, 1993; Sastry, 1997a-c; Curtis and Steele, 1996). The second group consists of more recent studies that consider frailty mostly at the contextual levels of community or region (Banerjee, Wall and Carlin, 2003, Gemperli, Vounatsou, Kleinschmidt, Bagayoko, Lengeler, and Smith, 2004, Kandala, Fahrmeir and Klasen, 2002; Kandala, Magadi and Madise, 2004; Adebayo *et al.*, 2004; and Adebayo and Farhmeir, 2005). These studies use both unstructured random effects (which assume that the frailty components at the contextual level of enquiry are independent), and spatially structured frailty effects (which take into account the fact that geographical locations at close proximity are more likely to be similar to each other than those far apart). In these studies, Bayesian methods generally allow for specification and estimation of all factors (including the frailty terms) in a single framework and also allow for the incorporation of empirical information (for example, previous knowledge about the modelling of the baseline effect) in the model. Bayesian modelling also helps alleviate the problem of sparse data in contextual studies through the use of smoothing techniques which involves borrowing strength from neighbouring areas in order to obtain more reliable estimates.

Determination of the magnitude of the unobserved regional/state effects is important since this is the level at which most health policies implemented at the lower geographical levels are made, and because state/region level covariates are often not included in studies. It is possible that the frailty effects observed at the contextual levels may be attributable, in part, to the frailty effect at lower levels. The literature does not show that any child survival study has explored frailty effects at the family, community and state levels simultaneously while exploring possible spatially structured frailty at the point-location (households or communities) level or aerial/lattice (state, region) level.

Increasing availability of geographically referenced data and remotely sensed geographic information, coupled with the recent advances in Geographic Information Systems (GIS) now makes it possible to measure certain contextual variables and integrate them with routinely collected survey data. Advances in spatial statistics analysis also facilitates the modelling of CM data using appropriate statistical techniques that take into account frailty at multiple levels for researchers to understand which level of heterogeneity (observed or unobserved) plays a greater role in the child's risk of death. The results from such analyses, when combined with mapping, provide a good way of visualizing mortality disparities, thereby facilitating the identification of areas where the situation warrants immediate action, and in the subsequent allocation of resources and interventions for meaningful and uniform reduction in CM. Employing appropriate statistical techniques with GIS capabilities, this research thus seeks to understand the determinants of CM and its differentials in the 37 states of Nigeria using a combination of data from the 2003 Nigerian Demographic and Health Survey (NDHS) and other contextual sources, such as information on malaria prevalence from the Mapping Malaria Risk in Africa (MARA) database.

1.2 Background of the study area

The West African country of Nigeria, with an estimated population of 145 million and an annual population growth rate of 2.4%, is the largest country in Africa and the tenth most populous country worldwide. Administratively, Nigeria is made up of 6 geopolitical zones consisting of 36 states and a Federal Capital Territory (FCT) at Abuja (see Figure

1). These make up the 37 spatial units considered in this report, all of which will be referred to as states for convenience. The states are further divided into 774 Local Government Areas (LGA). The country has diverse climatic and topographic conditions, and is also ethnically, culturally and religiously diverse (NPC, 1998; NPC and ORC Macro, 2004).



Figure 1: Map of Nigeria showing the 37 spatial units considered. Source: (NPC and ORC Macro, 2004).

Statistics indicate that about 45% of Nigeria's total population is less than age 15, with about 20% (24 million) under age five. In 2003, the total fertility rate (TFR)[1] for the country was 5.7 (NPC and ORC Macro, 2004). Nigeria is endowed with abundant natural resources, but the country's Gross National Product (GNP) per capita of $320 and an estimated 70% of the population living below the poverty level of US$1 per day makes it

[1] TFR is the average number of children a woman is expected to have if she experienced the current age-specific fertility levels for the whole of her reproductive life.

one the poorest countries in the world (World Bank, 2004; UNICEF, 2002). The phenomenon of poverty in the country, although widespread, is more concentrated in the rural areas and in the northern states due to differential accessibility and availability of government services (Department for International Development [DFID], 2000). In 2002, only 38% of Nigeria's population had access to adequate sanitation, while about one third of the population lacked access to safe water (World Bank, 2004). In terms of literacy, the percentage of adult females 15 years and above who were literate increased from 54.3% in 1999 to 59.4% in 2002 but these figures were still short of the male literacy rate which improved from 71.0% to 74.4% in the same period[2]. The primary school net attendance ratio, between 1996 and 2003 was also higher for males (64%) than for females (57%)[3].

1.3 Problem statement

In explaining child mortality differentials in Nigeria, three related issues are yet to be addressed: 1) assessing the influence of measurable community level variables on child survival 2) accounting for frailty (differing variances) in child survival at multiple levels and 3) describing the spatial patterns of child mortality risk across the country. Addressing these issues in the context of appropriate statistical modelling constitutes the main focus of this study.

1.4 Aims and Objectives of the Study

The main aim of the project is to account for the influence of contextual factors and frailty on CM in Nigeria and to investigate the spatial patterns of CM in the country.

The objectives are to:

- Evaluate the contribution of community level contextual factors to CM
- Determine the sources of frailty (household, cluster/community, and states) that are important in explaining CM differentials.
- Evaluate the effect of frailty terms on model estimates and model fit.

[2] World Development Indicators database, April 2005 (accessed: 17th, July 2005)
[3] http://www.childinfo.org (accessed: 17th, July 2005)

- Examine the potential bias incurred when the spatial dependence in the data is ignored and study the spatial pattern of CM risk across the 37 states of Nigeria with the aid of maps depicting the geographical differentials.
- Examine the differences between the models in order to understand the implications of using the wrong model.

1.5. Organization of the Report

The remainder of the research report is organized as follows. In Chapter Two, a literature review of CM related issues will be considered. An overview of the study area is also given in this chapter. Chapter Three contains a discussion on the data sources; an introduction to the statistical methods employed in the study i.e. the Bayesian model and the computational approach. The results of the study are presented in Chapter Four. Finally, conclusions are drawn in Chapter Five and various recommendations and suggestions for future research are given.

Chapter Two: Literature Review

There are two main aspects of the literature reviewed in this chapter. The first aspect deals with the CM conceptual frameworks, modelling frameworks and what has been done in Nigeria, while the second aspect is a summary section pulling together what various authors have contributed to the topic.

2.1 Conceptual Framework

A number of conceptual frameworks have been developed for the study of child health and survival. Models developed by demographers and economists (such as Schultz, 1984), lay great premium on the role of demographic and socioeconomic variables in determining mortality, while epidemiologists (such as Venkatacharya, 1985) place emphasis on the role of biomedical factors in morbidity studies. The two most referenced frameworks are those of Mosley and Chen (1984) and Schultz (1984).

Mosley and Chen's (1984) framework, developed for the study of IM and CM in developing countries, is considered to be the most comprehensive and systematic model developed and is the most referenced in the literature relating to child survival (Ruzicka, 1989; Masuy-Stroobant, 2002). Mosley and Chen (1984) employed a multidisciplinary approach, incorporating social and medical science research methodologies in the development of their model, which has both mortality and morbidity as outcomes. According to the framework, three sets of socio-economic factors operate through a set of five intermediate/proximate determinants namely; maternal fertility factors, environmental contamination, nutrient deficiency, injury and personal illness control to influence the level of IM and CM in a society. The socioeconomic variables include variables at individual, household and contextual levels.

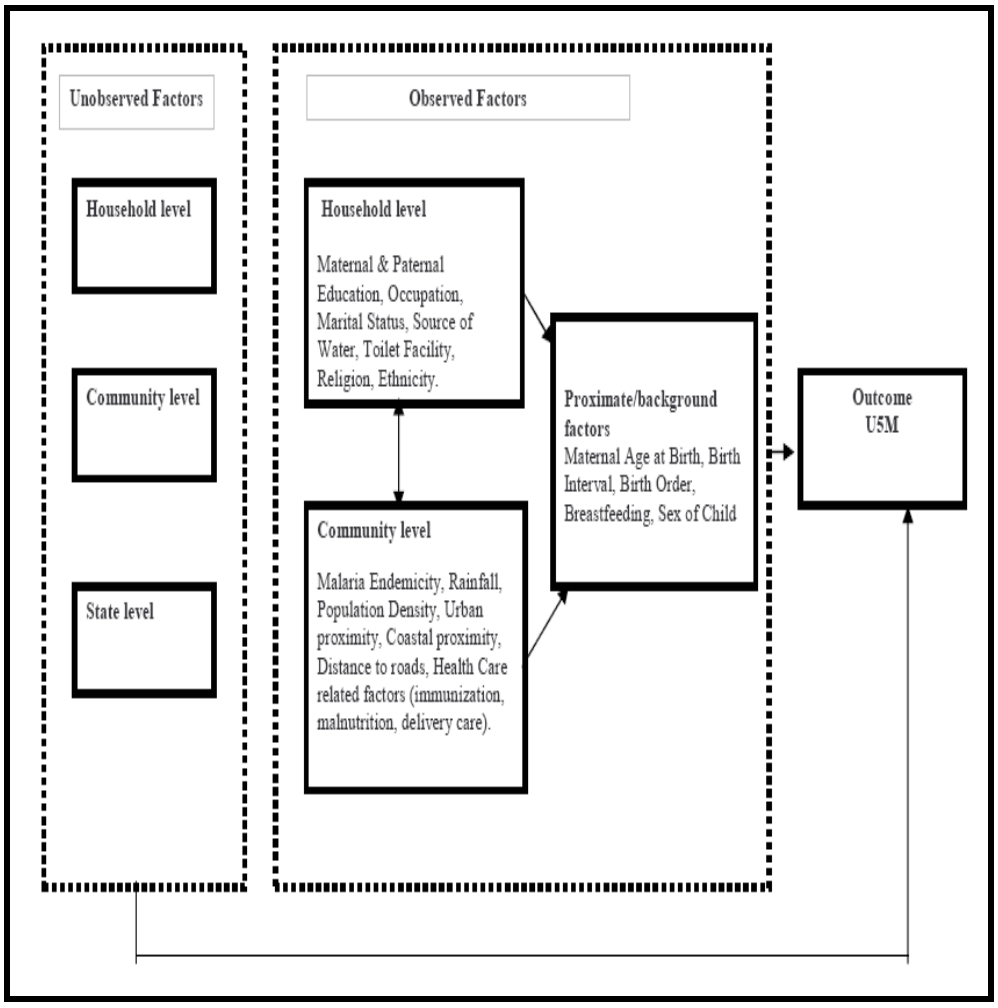


Figure 2: Schematic presentation of the conceptual framework for the study of Child mortality adapted from Mosley and Chen (1984) and Schultz (1984).

Schultz (1984) also makes a clear distinction between endogenous and exogenous causes of CM, and provides an additional mechanism for studying the unobserved influence on child survival. The framework proposed for this study is thus based on an adaptation of Mosley and Chen's (1984) and Schultz (1984) frameworks (see Figure 2). The dependent variable in the model above is U5M. As in the Mosley and Chen (1984) and Schultz (1984) frameworks, the representation above assumes that the context-level factors operate through the proximate determinants (which are mainly individual level attributes) to influence mortality. Variables such as nutrition, immunization or other health care factors, which appear under the classical proximate determinant category in Mosley and Chen's (1984) framework, are captured under the community level factors in Figure 2.

This has been done so that knowledge about the possible influence of the variables can be gained rather than speculated, in view of the fact that information on the variables often exists for a limited number of children. The unobserved factors at the household, community and state levels represent those other variables that are seldom captured and whose influence can be deduced based on the strength of the random term included in the model at the various levels.

2.2 Pathway of Influence

The following section reviews issues associated with some of the variables in the framework above.

2.2.1 Proximate Determinants

The proximate determinants as enumerated in Figure 2 above consist mainly of the demographic and biological characteristics of the mother and her child. Starting with the mother's characteristics, maternal age at the time of child's birth is known to exhibit a u-shaped relationship with child mortality; with mortality risk higher for children of younger and older women[4] (Hobcraft, McDonald and Rustein, 1984). The higher mortality among children of younger women can be attributed to their biologically immature reproductive system which results in their offspring having low birth weight, while the depletion of the maternal resources which progresses with age, makes the children of older women more susceptible to higher mortality. Studies have shown increased mortality risk among children born after short birth intervals, citing maternal resources depletion, competition amongst siblings, and increased transmission of disease due to crowding as the major factors (Hobcraft, McDonald and Rustein, 1985; Palloni and Millman, 1986).

Turning to the child's own attributes, male children generally experience higher mortality than female children primarily due to biological reasons. Higher female mortality is associated with cultural values especially in societies with strong male-child preference, in which case, biased allocation of health, nutrition and other resources in favour of the male -child explains the sex differential (D'Souza and Chen, 1980; Das Gupta, 1987).

[4] Usually less than 15 and greater than 35 years old respectively.

2.2.2 Household level factors

Mother's education has been described as the single most important determinant of child mortality (Caldwell, 1979). The education of mothers exhibits an inverse relationship with child survival, such that children of educated mothers experience lower mortality relative to children of uneducated women, with the relationship persisting even after controlling for other variables (Caldwell, 1979). Although, there is unanimity regarding the importance of maternal education on child survival, there is no agreement regarding the pathway through which mother's education influences mortality. Studies have shown that education equips the woman with the necessary knowledge and power which enables her to among other things, break away from harmful traditional practices, provide better domestic child care, participate better in child decision making and effectively utilize modern medical facilities in a timely manner (Caldwell, 1979; Hobcraft, McDonald and Rustein, 1985; Cleland and Van Ginneken, 1988). Others argue that the observed relationship between maternal education and child mortality may be as a result of certain independent/external factors such as access to toilet facilities and water, husband's education, fertility behaviour, breastfeeding and education of others in the community (Behrman and Wolfe, 1987; Tulasidhar, 1993; Desai and Alva, 1998; Adetunji, 1995).

Mother's occupation has a mixed impact with child survival. Mother's work may reduce the time she spends breastfeeding and in taking care of her child which may lead to increased mortality (Peterson, Yusof, DaVanzo and Habicht, 1986), but may also contribute to improved survival since working mothers who are educated may be better informed about immunization and child care trends. Father's education is often ignored in child mortality studies, but fathers in the developing world tend to make decisions regarding fertility, contraception and use of health care services, thus, decisions regarding child health and survival may also depend on the father and his level of education (Kuate-Defo and Diallo, 2002). With regards to toilet and water, studies have shown that childhood mortality is lower in households with piped water and flush toilets and the impact of these factors is more pronounced as the child gets older and has more frequent contact with the environment[5] (Balk, Pullum, Storeygard, Greenwell and Neuman, 2003).

[5] This is the stage of physical development, where the child does a lot of crawling and is more vulnerable to the effects of dirty environment.

Cultural disparities also exist in child mortality rates and have been captured in the literature using variables such as religion and ethnicity. Child mortality is often higher for children from Moslem and Traditionalist backgrounds than for Christian children, and cultural beliefs/attitudes about diseases and child care as well as the low status of women in certain religion explain the differentials (Caldwell and Caldwell, 1993; Gregson, Zhuwau, Anderson and Chandiwana, 1999; Ogunjuyigbe, 2004).

2.2.3 Community level factors

Mortality differentials by type of place of residence constitute the focus of a lot of studies. It has been highlighted that child mortality in urban areas is lower than for rural areas and the phenomenon may be attributed to the greater availability and accessibility of medical care facilities, public infrastructure such as safe water supply, as well as better income and education opportunities present in the urban areas. The study by Sastry (1997c) has however found that the observed mortality differentials by place of residence can be attributed to the role played by community level variables.

Previous research has shown a strong relationship between community level factors and child mortality. Typically, mortality risks are greater for children living in areas with: high HIV prevalence, low immunization coverage, high incidence of drought and food shortages (Adetunji, 2000; Hill, Bicego and Mahy, 2001; Curtis and Hossein, 1998; Balk *et al.*, 2003)

Population density has a U-shaped relationship with child mortality, with children resident in low and high-density areas at an elevated risk of dying (Balk *et al.*, 2003). High population density means an increased possibility of disease transmission and a greater competition for food, conditions which may lead to death. Low population density on the other hand means reduced access to health care and overall socioeconomic factors, implying a greater risk of mortality. With regards proximity to the coast (a proxy to easy access to markets), it has been shown that the risk of childhood death increases the further one resides from the coast (Balk *et al.*, 2003).

2.3 Child Mortality and its differentials in Nigeria

Nigeria, like other developing countries lacks accurate and comprehensive data[6] on the status and causes of childhood mortality. Available information however suggests that childhood mortality has declined over the years. For example, U5M rates declined from 290 in 1960 to 198 in 2003, while IM dropped from 165 to 98 in the same period (UNICEF, 2005). The mortality decline noticed especially in the late 1970's and early 1980's have been largely credited to the public health programmes initiated by the international community particularly in the area of immunization against the childhood killer diseases. Most childhood deaths have been attributed to pneumonia, malaria, measles, acute respiratory illness and diarrhoea – disease conditions that are preventable or treatable using low-cost interventions (NPC and ORC Macro, 2004; POLICY Project, 2002). Despite the earlier gains recorded in CM reduction, Nigeria currently occupies the 13th position amongst the countries in the world with the highest U5M rates (UNICEF, 2005), a position which suggests that more needs to be done in the area of child survival.

The pace of mortality decline within Nigeria has also not been uniform and consequently, CM rates exhibit wide geographic variation. The geographical pattern is however hard to discern for the whole country since available studies are highly localized[7]. The reports from the 1991 census and the 3 rounds of DHS conducted in the country, paint broad regional variations in CM rates. For example, the 1991 Nigeria census recorded the lowest IM rate (57/1000) for the southwest region and the highest (99/1000) for the northwest region (NPC, 1998). A similar pattern was reported in the 2003 Nigeria Demographic and Health Survey (NPC and ORC Macro, 2004) where the lowest under-five mortality rate of 103/1000 was reported for the South East, and the highest rate of 269/1000 was reported for the North-West region of the country. (NPC, 1998; NPC and ORC Macro, 2004). These studies used simple descriptive statistics and employed cross tabulations to show differential mortality patterns stratified by covariates such as the

[6] Detailed information on child health and survival in Nigeria has come from nationally representative surveys such as the DHS, MICS and WFS mostly conducted by international organizations.

[7] Most studies deal with selected geographical units such as a few regions, states or communities (Adedoyin and Watts, 1989; Iyun, 1992; Adetunji, 1995; Ahonsi, 1995; Ogunjuyigbe, 2004), and only occasionally consider the country as a whole (NPC, 1998; NPC and ORC Macro, 2000 and 2004; Adebayo *et al.*, 2004; Adebayo and Fahrmeir, 2005).

child's sex, mother's education and place of residence. State-wise variations in CM rates were reported in the 1991 census report, but the analysis did not include socio-economic factors to account for the observed variations and variations at levels lower than the state were not considered.

Regarding the determinants of child mortality, studies have been highly localized (dealing with specific areas such as regions, states or localities), and only occasionally applying to the country as a whole. Amongst the local studies are research by Adetunji (1995), Iyun (1992), Ogunjuyigbe (2004), Adedoyin and Watts (1989), Owa and Osinaike (1998), Feyisetan, Asa and Ebigbola (1997) and Lawoyin (2001). Important factors that affect child mortality documented in these studies include place of residence, education, tradition, toilet facility, water supply, access to medical and antenatal care.

Amongst the few recent country-wide studies that have dealt with the issue of child survival in Nigeria are the descriptive reports of the 1991 census (NPC, 1998) and those from the 1990, 1999 and 2003 NDHS (NPC, 1991, 2000; NPC and ORC Macro, 2004) as well as results from more recent systematic assessments (Adebayo *et al.*, 2004; Adebayo and Fahrmeir, 2005; and Kneib, 2005). The reports from the 1991 census and the 3 rounds of DHS conducted in the country, paint broad regional variations in child mortality rates and use simple cross tabulations to show differential mortality patterns by variables such as women's education, child's sex and place of residence State-wise variations in child mortality rates were also reported in the 1991 census but the analysis did not include socio-economic factors to account for the observed variations.

Turning to the studies that investigated the determinants of child mortality in a more detailed fashion, Adebayo *et al.* (2004) using data from the 1999 NDHS investigated the spatial distribution of IM (neonatal and post neonatal mortality) enquiring whether the determinants of a child's death differed in the different age groups considered. Their results from a geo-additive modelling (details of which are discussed in Chapter 3); show that spatial variation and the determinants of mortality differed considerably for the two age groups studied. Improved maternal education, being Christian, not being first born, being a singleton birth, and having assistance at birth significantly reduced the risk of

neo-natal mortality but the effect of the variables were less for post-neonatal mortality. Location effects influencing neonatal mortality also appeared to be negatively correlated with effects influencing post-neonatal mortality and speculations about the spatial differentials found tied to crowding, poverty, poor health service and geography which were not explicitly included in the models (Adebayo *et al.*, 2004).

Using the same dataset, Adebayo and Fahrmeir (2005) analyzed child mortality in Nigeria with flexible geo-additive discrete-time survival models which allows for the measurement of small-area spatial effects simultaneously with possibly non-linear or time-varying effects of other covariates (details of model are discussed in Chapter 3). Their results revealed mother's age (22-35 years), birth delivery assistance, hospital delivery and high preceding birth interval to be associated with lower child mortality risk. Distinct spatial patterns were also observed in their analysis, with significant high mortality associated with 4 of the 37 states, while lower mortality was associated with 6 of the 37 states (mostly northern states). The spatial variations were interpreted in terms of variables which were not captured in their analysis, including: disease environment, ethnicity/religion, topography, drought and malaria (Adebayo and Fahrmeir, 2005).

Apart from the works of Adebayo *et al.* (2004) and Adebayo and Fahrmeir (2005), most existing studies in Nigeria have ignored the spatial component of the dataset used and have not included frailty terms in their statistical model to take care of clustering or unobserved heterogeneity at any spatial unit. More recent systematic assessments of the determinants of CM by Adebayo *et al.* (2004) and Adebayo and Fahrmeir (2005), using data from the DHS, employed spatial statistical techniques which allow for the simultaneous measurement of small-area spatial effects and the effect of other covariates. In both studies, the effects of variables which were not captured in their analysis, including: disease environment, ethnicity/religion, topography, drought, crowding, poverty, poor health service and malaria were given as possible explanations for the resulting geographical variations observed. This suggests that where possible, the omitted factors should be included in statistical modelling of CM data. Only a few studies in Nigeria have included frailty terms in their statistical models to account for clustering or unobserved heterogeneity at any level, while most studies fail to properly consider the spatial component of their dataset (Adebayo *et al.*, 2004).

Chapter Three: Data and Methods

3.1 Data Sources

The study utilizes secondary data from multiple sources, with the majority of the data coming from the 2003 NDHS[8]. The NDHS was jointly conducted by the NPC and ORC Macro International USA. The 2003 NDHS is a nationally representative sample of urban and rural areas in which a 2-stage sampling design was employed. In the first stage of sampling, 365 Enumerator Areas (EAs) or clusters were randomly selected over the country, with probability proportional to size (PPS) of population from a list of EAs developed from the 1991 population census, where the measure of size is the number of households in the EA. In the second stage, a systematic random sample[9] of 7,864 households was selected from the chosen EAs. All females between 15 and 49 years and males in the 15-59 age groups who were permanent residents or visitors in the selected households on the night before the survey were eligible for interview (NPC and ORC Macro, 2004).

Using structured questionnaires administered to the eligible women, detailed information pertinent to all live births that had occurred to the chosen women in the 5 years before the survey was collected in addition to a complete birth history. A host of other demographic and health related information were also collected in the survey including the child's date of birth, birth weight, sex, survival status and age at death for deceased children. Data on parental education and occupation, type of place of residence, and household wealth, in addition to a host of other health and socio-economic factors were also obtained.

[8] DHS data are considered to be the most detailed source of demographic and health related information available in most developing countries where vital registration systems are virtually non-existent. Despite the heaping of reported ages at death and under-enumeration/reporting inherent in some of them, the surveys are also considered to be high quality sources for mortality data (Bicego and Ahmad, 1996; Curtis, 1995).

[9] The procedure involves first selecting a starting household at random from the household listing, and then selecting every k^{th} household - k is the sampling interval calculated as $k=N/n$ (where N is the total number of households and n is the number of households to be selected).

A child-based dataset consisting of information on 6029 children born in the five years preceding the survey was constructed using the data from the different survey questionnaires. The 2003 NDHS also collected location information (longitude and latitude coordinates) for each survey cluster using handheld Global Positioning System (GPS) devices, to aid easy linkage of the dataset to other geographically referenced data sources and to facilitate small area mortality studies.

Community (cluster) level measures of health are constructed from the DHS dataset using information such as incidence of illness (fever/cough and diarrhoea) in the previous two weeks, immunization and health facility use. Due to the fact that the 2003 NDHS did not collect information on some variables intended for use in the current analysis, supplementary information has been obtained from other sources (see Table 1). Values of all geographic variables have been obtained through the use of GIS software (Arc View GIS, ESRI (2003)) to each of the 2003 NDHS cluster locations, using the cluster GPS database provided by Macro International, to obtain a cluster-level dataset. The cluster-level file has then been linked to the child-based data file using the cluster identification number common to both datasets, to obtain an integrated child-level dataset consisting of all data variables relevant to the analysis.

Table 1: Community level contextual factors to be considered in the study

Variable	Source	Description
Population Density	CIESIN: Gridded Population of the World (GPW) v. 3 www.beta.sedac.ciesin.columbia.edu/gpw	Population Density per EA
Coastal proximity	National Imagery and Mapping Agency (NIMA) Digital Chart of the World (DCW)-derived continent boundary	Distance (Euclidean) to nearest point on the coastline
Malaria Endemicity	Mapping Malaria Risk in Africa (MARA) http://www.mara.org.za/lite/information.htm.	Malaria Endemicity per EA
Distance to roads	National Imagery and Mapping Agency (NIMA) Digital Chart of the World (DCW)	Distance (Euclidean) to nearest point on the road

3.2 Statistical Methods

There are three main stages of analysis considered in this project. The first stage involves preliminary analysis relating to the variables to be used and their relation to survival. The second stage involves an investigation of the correlation and spatial correlation structure of mortality, and finally, the last stage of analysis involves fitting more complex models to the data. The complex models have the ability to take into account the discrete time nature of the data and other special features of the dataset and have the ability to incorporate covariates at different levels in the form of fixed or random effects.

Summary statistics and Kaplan Meier curves

Summary statistics (means, standard deviations and percentages) are used to examine how varied the surveyed children are with respect to the covariates.

The study investigates the survival of a child following a 60 months exposure period (from birth to age 5). The time until the death of the child is thus the main outcome of interest. There are two key problems with this kind of data. Firstly, there is the issue of skewness of the survival times which arises due to some children having very long survival times and others having comparatively short survival times. This often implies that normality assumptions are violated and the data cannot be analyzed using conventional statistical techniques. The second is the problem of censoring. The children who are not yet five years old in the case of U5M and those who have not observed death at the end of the interview period are considered censored. The type of censoring in this case is known as right censoring which simply means that some children stop being observed before their deaths are observed, but each child is at least observed for some of the period, and thus, some information is collected about each child's survival.

The analysis of right-censored time to event data of this nature falls under the umbrella of survival analysis whose main goals include the estimation of the survivor and hazard functions, comparison of survival curves and the investigation of the effect of explanatory variables on survival times. Survival functions can be estimated either parametrically or non-parametrically. Parametric analysis is employed when the survival

times fit a theoretical density function such as the Weibull, Gompetz, Exponential, Lognormal or Gamma distribution in which case, parametric maximum likelihood estimation is used in modelling the survival function (Lee, 1980; Klein and Moeschberger, 1997). Nonparametric methods make no assumption about the functional form of the survival function, but instead, they use the information contained in the duration variable thus letting the data set speak for itself and as a result reducing the chances of misspecification of the true functional form of the survival function.

The Kaplan-Meier (K-M) method (Kaplan and Meier, 1958) is the most widely used non-parametric method of estimating the survival function $S(t)$, the probability that a child survives longer than time *t*. This method utilizes information from both the fully observed as well as the right-censored children.

The K-M estimate at time *t* is given by:

$$S(t) = \prod_{j|t_j \le t} \left(\frac{n_j - d_j}{n_j}\right)$$

where n_j is the number of children at risk of death at time t_j and d_j is the number of deaths at time t_j.

The K-M is based on several assumptions namely,

- the sample is chosen randomly and independently from a larger population,
- the deaths occurred at the times specified,
- the survival probabilities are the same for children interviewed early and late in the study,
- censored children have the same survival prospects as uncensored children,
- time to censoring and survival times are independent.

The survivor function is usually presented as a K-M curve which is a plot of probability of survival $S(t)$ on the vertical axis against survival time *t* on the horizontal axis. Vertical drops indicate times at which an event (in this case, death) was observed, while censored times are indicated by short vertical lines. The survival probability at a certain time, median survival time, mean survival time and other quantiles are summary statistics

that can be extracted from the survival curve. It is also often of interest to ascertain whether the survival curve of one group of children is different from another. As an example, we may want to know if male children live longer than females. This type of comparison can be achieved visually by comparing the survival curves or through statistical tests. The log-rank test (Mantel, 1966; Peto and Peto, 1972) is the most common method used in statistically comparing the overall difference between the survival curves for two or more groups.

The log-rank statistic tests the null hypothesis that at any time point, the survival functions for all groups are equal, against the alternative hypothesis that at least one survival function is different from the others for some time periods.

In other words, for *g* groups, the log-rank statistic tests:

$H_0 : S_1(t) = S_2(t) = ... = S_g(t)$ for all $t \le \tau$ against

H_1 : at least one of the $S_g(t)'s$ is different for some $t \le \tau$ (where τ is the largest time during which each group has at least one child at risk).

The log-rank statistic is given by:

$$\chi^2 = \sum_g \frac{(O_g - E_g)^2}{E_g}$$

where O_g is the observed number of deaths in each group, and E_g is the expected number of deaths in each group *g* assuming a null hypothesis of no difference in survival between the groups. O_g and E_g are calculated for each time when an event occurs. The log rank test is based on the same assumptions as the K-M given above, and under H_0, the log–rank statistic is χ^2 with $G - 1$ degrees of freedom, where *G* is the number of groups being compared. The decision to reject the null hypothesis is made using chi-square tables with the appropriate degrees of freedom.

Methods used for investigating the correlation and spatial correlation structure

Tests for correlation at various contextual levels will be used to detect the presence of spatial association in the data so that the appropriate frailty terms can be incorporated in the multivariate modelling in order to eliminate potential bias that will otherwise be present if such frailty terms are not included. To this end, cross tabulations are used to study the distribution of births and deaths and reveal the possible clustering of mortality at the household and community levels.

At the state level, Exploratory Spatial Data Analysis (ESDA) techniques of Global and Local indicators of Spatial Autocorrelation (LISA) (Cliff and Ord, 1981; Anselin, 1995; Ord and Getis, 1995) are employed to determine the extent of spatial association and presence of spatial clusters of U5M rates.

The Moran's I statistic (Moran, 1950) is a single global measure that tests for spatial association of a phenomenon.

The Moran's I is defined as:

$$I = \sum_i \sum_j w_{ij} (x_i - \mu)(x_j - \mu) / \sum_i (x_i - \mu)^2$$

where w_{ij} represents the spatial weight matrix elements, x_i is the measure of U5M rate in state i, and x_j is the measure of U5M rate in neighbouring state j, and μ is the average U5M rate for the country. A spatial weight matrix can be defined either by contiguity (where states share common boundaries) or by distance (where state centroids are within certain distance criteria). Contiguity-based weight matrices include Rook Contiguity (which uses only common boundaries to define neighbours) and Queen Contiguity (which uses all common points or borders). Distance-based weight matrices include distance bands and k nearest neighbours. For contiguity-based matrices, the matrix elements can broadly be defined according the following criteria: $w_{ij} = 1$ if states *i* and *j* are adjacent and zero otherwise, The matrix elements for distance-based matrices on the

other hand can be defined according the following criteria: $w_{ij}(d) = 1$ if state *j* is within distance *d* from state *i* and zero otherwise.

The Moran's I, like the Pearson's correlation coefficient, assumes values between -1 and +1. A value of +1 indicates strong positive autocorrelation; a value of -1 indicates strong negative autocorrelation, while a value of 0 indicates a random distribution of U5M rates. The significance of Moran's I is obtained using the permutation testing approach.

Anselin (1995) describes the LISA for each state *i*, and uses this to provide a value of spatial association for each state under consideration.

The LISA for state *i* is defined as:

$$I_i = z_i \sum w_{ij} z_j, \text{where } z_i = \frac{x_i - \bar{x}}{SD(x)}$$

LISA allows for identification of four different types of spatial clusters:

- High-High Cluster: States with high values of U5M surrounded by states that have high values of U5M (positive association – Hot spot) ,
- Low-Low Cluster: States with low values of U5M surrounded by states that have low values of U5M (positive association – Cold spot) ,
- Low-High Clusters: States with low values of U5M surrounded by states that have high values of U5M (negative association - spatial outliers) and
- High-Low Cluster – States with high values of U5M surrounded by states that have low values of U5M (negative association - spatial outliers).

Multivariate statistical model

The K-M curves and the log-rank test described above provide univariate analyses useful in assessing whether a covariate affects survival and are most suitable for descriptive purposes. They are particularly handy when the predictor variables are categorical and do not work easily with continuous[10] predictors such as age of mother at birth. However, they do not allow us to say how survival of a group is affected with the influence of other covariates included in the model. The Cox model (Cox, 1972) is commonly employed in analysing survival data in a multivariate way, allowing the effect of a set of covariates on survival time to be assessed. The Cox model also handles censored data, categorical and continuous variables as well as variables that change over time, all of which may influence survival. The Cox model also allows for frailty to be included at various levels, but its assumption that time is measured on a continuous scale makes it inappropriate for the current data. In the DHS surveys, the survival times of children are measured discretely in months which results in a lot of tied events, and which causes problems when continuous time models are used. A discrete formulation of time is therefore more appropriate than the Cox approach since tied events are not a problem with the discrete-time approach.

A standard discrete-time multilevel hazard model (Goldstein, 1995) is the first choice for this type of data. Such techniques have, however, been found to be inappropriate in cases where it is assumed that frailty at some level follows strong spatial patterns (Chaix, Merlo, Subramanian, Lynch and Chauvin, 2005).

The modelling framework for the study needs to take into account the special features of the dataset, whilst ensuring that the aims of the study are met. The Bayesian geo-additive discrete-time survival model described below can accommodate all the features of the dataset namely, presence of censored observations, non-linear and time varying covariates, frailty and spatial dependence. One major advantage of the Bayesian framework is that it allows for the inclusion of prior knowledge about the parameters

[10] Continuous covariates have to be arbitrarily divided into quartiles or other biologically meaningful groups and then treated as categorical covariates. This often leads to the loss of information contained in such variables.

along with information contained in the data to produce more robust results. The fact that the modelling process in the Bayesian framework does not reply on asymptotic theory also makes it possible to work with small sample sizes (Congdon, 2003).

3.2.1 The Geo-additive Discrete-Time Survival Model

The modelling details given below are derived from the works of Berger, Fahrmeir and Klasen (2002), Adebayo and Fahrmeir (2005), Hennerfeind, Brezger and Fahrmeir (2006), Knieb (2005) and Fahrmeir and Tutz (2001).

Consider the survival times $T \in \{1,...,k=60\}$ in months, where $T=t$ denotes death of a child in month t and k is the last observation in the interval. Let x_{it} be a vector of covariates observed up to month t. The discrete-time conditional probability of death in month t given that the child survived up to month t, is given by:

$$\lambda\left(t \mid x_{it}\right) = P\left(T = t \mid T \geq t, x_{it}\right) \qquad (1)$$

In a right–censored survival dataset such as ours, it is assumed that each child's survival information is captured as $\left(t_i, \delta_i\right)$, where t_i is the observed lifetime or time until death for child i, and δ_i is a censoring indicator with a value of 1 if child i is alive and 0 if the child is dead. For ease of analysis, the discrete-time survival model is often represented in the form of a logistic regression model by defining binary event indicators y_{it}, $t = 1,...,T$

$$y_{it} = \begin{cases} 1 & \text{if } t = t_i \text{ and } \delta_i = 1 \\ 0 & \textit{Otherwise} \end{cases} \qquad (2)$$

The equation in (1) is thus written as a binary response model given by

$$P\left(y_{it} \mid x_{it}\right) = h\left(\eta_{it}\right) \qquad (3)$$

where h is the response or link function, and η_{it} is a vector of covariates. Equation (2) can be treated as a probit, logit or multinomial function, with logit models being easier to estimate and interpret (Crook, Knorr-Held and Hemingway, 2003; Adebayo and Fahmeir, 2005).

An expression for the logit model is:

$$P\left(y_{it}=1\,|\,\eta_{it}\right)=\frac{e^{\eta_{it}}}{1+e^{\eta_{it}}} \qquad (4)$$

with a partially linear predictor

$$\eta_{it}=g_o\left(t\right)+x_{it}^{'}\gamma \qquad (5)$$

where $g_o\left(t\right)$, $t=1,2,....$ is the baseline hazard effect and γ are fixed effect parameters. Equations (4) and (5) may be represented as:

$$\frac{P(y_{it}=1\,|\,x_{it})}{P(y_{it}=0\,|\,x_{it})}=\exp\left(g_o\left(t\right)\right)\exp\left(x_{it}^{'}\gamma\right) \qquad (6)$$

Equation (6) can be regarded as the basic form of a semi-parametric survival model, where the baseline hazard $g_o\left(t\right), t=1,2,...$ is an unknown, usually non-linear function of t to be estimated from the data.

Incorporating the fixed effects, time-varying covariates, and spatial effect yields a geo-additive representation for equation (6) given by the expression:

$$\eta_{it}(t)=g_o\left(t\right)+\sum g_j(u_{ij})+\sum f_j(x_{ij})+f_{spat}\left(s_i\right)+b_i+c_i+x_i^{'}\gamma \qquad (7)$$

where $g_o\left(t\right)$ is the baseline function of time, $g_j(t)$ is the time-varying effects of covariates u_j, $f_j\left(x_{ij}\right)$ are non-linear effects of continuous covariates, $f_{spat}\left(s_i\right)$ is the effect of the state/district $s_i\in\left\{1,...,S\right\}$, b_i and c_i represent the cluster and household-specific frailty effects respectively, while x_{ij} are the fixed effect covariates, and γ is the vector of parameters. The spatial term, $f_{spat}\left(s_i\right)$ may be further split into a spatially-correlated (structured) and an uncorrelated effect. That is, $f_{spat}\left(s_i\right)=f_{str}\left(s_i\right)+f_{unstr}\left(s_i\right)$.

3.2.1.1 Prior distributions for covariate effects

The unknown model parameters γ and functions g_o, g_j, f_j and f_{spat} in equation (7) are considered random variables in a Bayesian framework and must be supplemented with suitable prior distributions for inference purposes. The choice of prior generally depends on the type of the covariate and a vast amount of literature exists detailing the treatment of covariates and prior specifications (including: Gelman, Carlin, Stern and Rubin; 1995, Leonard and Hsu; 1999, Carlin and Louis; 2000 and Bernardo and Smith; 2000). To this end, the specification of priors and hyper-parameters for each group of covariates follow the works of Berger *et al*. (2002); Adebayo and Fahrmeir (2005) and Hennerfeind *et al.* (2006), and are as follows:

3.2.1.2 Priors for fixed effects

In the absence of any prior knowledge about the covariates, independent diffuse priors (uninformative priors) $p(\gamma_j) \alpha$ *constant* are the most popular choice for modelling fixed effects.

3.2.1.3 Priors for continuous and time varying effects

The continuous and time varying effects in equation (7) are often assumed to vary smoothly and are modelled using the Bayesian Penalised Splines [P-splines] (Eilers and Marx, 1996; Lang and Brezger, 2004). In this approach, the function $f_j(x_j)$ is approximated by polynomial splines of degree *q*, i.e.

$$f_j(x_j) = \sum \tau_{jm} \tau_m(x_j)$$

where τ_m is the m^{th} basis function and $\tau = (\tau_1, \tau_2, ..., \tau_m)$ is a vector of regression coefficients.

3.2.1.4 Priors for Unstructured frailty

All uncorrelated random effects which include the group random effects (family and community random effects) as well as the unstructured spatial state effect are assumed to be independent and identically distributed $(i.i.d)$ Gaussian. The family random effect is modelled as: c_i : $N\left(0,\tau_c^{\,2}\right)$, the community random effect is modelled as: b_i : $N\left(0,\tau_b^{\,2}\right)$ and the unstructured spatial effect is modelled as: f_{unst} : $N\left(0,\tau_{unst}^{\,2}\right)$.

3.2.1.5 Priors for the spatially structured frailty

Spatial data is generally of two types: the point-location data which is based on measurements taken at exact locations in space (e.g. from exact longitude and latitude coordinates of a households or community) and aerial/lattice data which based on data gathered by artificially defined sites (usually administratively defined locations such as county, state, region). Structured spatial effects $f_{str}(s)$ are estimated either based on Markov random field (MRF) priors for lattice data or Gaussian random field (GRF) priors in the point-location data). Since we are interested in how the phenomenon of U5M varies over states (which are by nature lattice structures), the MRF prior which deals with lattice data is the preferred approach and is discussed below.

The MRF prior was proposed by Besag, York and Mollie (1991) for the correlated spatial effects. The MRF prior introduces a structure based on neighbourhood (areas are neighbours if they share a common boundary) and the mean effect of a phenomenon under consideration is taken as the mean of the effects of the neighbouring areas.

Let $f_{str}(s_i)=\beta_{js}$ be the structured spatial effect in equation (7), then the MRF prior is given by

$$\beta_{js} \mid \beta_{js'}, s_i \neq s_i^{'}, \tau_j^2 : \; N\left(\frac{1}{N_s}\sum \beta_{js'}, \frac{\tau_j^2}{N_s}\right)$$

where $s_i^{'} \in \partial_s$ denotes the set of neighbours of state *s*, N_s = number of neighbours and τ_j^2 is the variance parameter that controls for spatial smoothness.

3.2.1.6 Hyperparameters

In a fully Bayesian analysis, the variance parameters are also considered unknowns and are estimated by assigning priors to them (also called hyperparameters); thus, allowing for the simultaneous estimation of the variance parameter and the corresponding unknown functions. The hyperparameter is commonly assumed to be inversely gamma distributed (IG (*a*, *b*), with the scale parameter $a > 0$ and shape parameter $b > 0$ and *a* and *b* chosen such that the prior is weakly informative).

The values of *a* and *b* reflect different degrees of uncertainty about the variance parameter. A common choice for the hyperparameters are $a=1$ and a small value for *b*. An example is: $a = 1$, $b = 0.005$.This yields a flat distribution which is similar to a situation of no prior knowledge on the parameter space. Another common choice for the priors involves specifying equal scale and shape parameters (that is: $a=b$). An example of this is $a=b=0.001$, which yields a weakly informative but proper prior closely approximating the Jeffrey's non-informative prior and works better in sparse data situations. Crook *et al.* (2003) notes that decreasing the value of the shape paramater *b* corresponds to a lower prior guess of the size of the variance, since the inverse gamma distribution has its mode at $b/(a+1)$. Finally, in the Bayesian framework, it is assumed that all priors for parameters are mutually independent (Bolstad, 2004).

3.2.1.7 Inference / Estimation

Inference for the posterior distribution of the model parameters is fully Bayesian and is based on the Markov Chain Monte Carlo (MCMC) simulation technique. The MCMC simulation basically involves generating samples from the posterior distribution of the unknown parameters. Two major algorithms used for producing fully Bayesian estimates are the Gibbs Sampler and the Metropolis-Hastings (M-H) algorithm. The Gibbs Sampler simulates new values for a parameter based on the conditional distribution of that parameter. After each iteration step, new values are used to replace the old ones. This

process is then repeated until the estimates converge. Using the M-H approach on the other hand involves generating estimate values from a proposed distribution and then comparing the values to those from the previous iteration step using posterior probabilities. A decision is then made to either accept or reject the values based on the acceptance probability.

3.2.1.8 Model comparison

In Bayesian data analysis, model comparison and selection are employed for finding the "best" model, or subset of models, which describe the data, as well as for studying the sensitivity of results to prior specification (Vaida, Ghosh and Liu, 2008). The Deviance information Criterion (DIC) (Spiegelhalter, Best, Carlin and van der Linde, 2002) has been developed for comparing the fit and complexity of hierarchical models in the Bayesian setting. The DIC is an extension of the Akaike Information Criterion (AIC) and is based on the posterior distribution of the deviance statistic.

The DIC is calculated as:

$$DIC = \bar{D} + p_D \qquad (8)$$

In the above expression, $\overline{D} = E\left[D(\theta)\right]$ is the posterior mean of the deviance statistic $D(\theta)$ and represents a measure of the model fit to the data.

The deviance statistic is given by:

$$D(\theta) = -2\log\left(f(y|\theta)\right) + c$$

where $f(y|\theta)$ is the likelihood function for the observed data vector y given the parameter vector θ, and c is a constant.

In equation (8), p_D is the effective number of parameters in the model (a measure of model complexity) and is calculated as:

$$p_D = \overline{D} - D(\bar{\theta})$$

Here, $D(\bar{\theta})$ is the deviance evaluated at $\bar{\theta}$ - the posterior means of the parameters of interest.

When the DIC is used for model comparison, models with smaller values of DIC are preferred as they indicate a better fit and lower complexity, and while there is no standard for comparing DICs, the differences in the DIC values of two or more competing models are important (Spiegelhalter *et al*, 2002).

Burnham and Anderson (2002), in the case of the AIC, proposed as a rule of thumb that AIC differences within 1-2 units of the best model suggest similar support for both models (models cannot be differentiated), models with AIC differences of between 3-7 from the best model can be weakly differentiated and differences of more than 7 units is regarded as strong evidence in favour of the model with the smaller DIC. Spiegelhalter *et al.* (2002) suggest that the rule of thumb works reasonably well for the DIC. The major advantage of the DIC is that it can be easily calculated from output of the MCMC simulation (Spiegelhalter *et al*, 2002).

Another goodness-of-fit measure based on the DIC is the R^2_{DIC} which Miaou, Song and Mallick (2003) defined as:

$$R^2_{DIC} = 1 - \frac{DIC_{model} - DIC_{ref}}{DIC_{max} - DIC_{ref}}$$

The R^2_{DIC} attempts to standardize the DIC in the same way as the traditional R^2 (Miaou *et al*., 2003). In the above expression, DIC_{model} is the DIC value for the model under evaluation, DIC_{max} is the DIC value under a fixed one-parameter model and DIC_{ref} is a DIC value from a reference model (the best model) which can also be approximated as $DIC_{ref} = n$ (Miaou *et al*., 2003).

Chapter Four: Results

This chapter presents the results from the various analyses. The chapter begins by introducing the unit of analysis, the choice of variables and the level at which they are introduced as well as a discussion as to whether the variables should be modelled as fixed or random effects. Descriptive results are then presented as well as results from a multivariate analysis. Preliminary analyses, including univariate and bivariate analyses were performed using the statistical package SAS® Version 9.1.3 (SAS Institute Inc., 2002-2004). Multivariate analysis is then conducted and the models fitted are evaluated and compared, examining them as to goodness of fit or potential misfit, and then finally conclusions are drawn as to which model fits the data best. The multivariate analysis including the production of risk maps were implemented using BayesX Version 2.0 (Brezger, Kneib and Lang, 2005), while additional mapping was carried out in GeoDa version 0.9 (Anselin, 2003) and Arcview GIS version 3.3 (ESRI, 2002).

4.1 Unit of analysis and outcome

In this report, the individual (child) is the unit of analysis and the outcome variable is the risk of U5M (0–59 months). The overall aim is to assess the extent to which both measured and unmeasured factors at various level of aggregation (household, community and state) affect child survival.

The hierarchical structure of the data is depicted in Figure 3 and the definition of the various levels is:

- **Individual (child) level**: this is defined as the children under the age of five years who reside in the households. In this report, the individual child level is the lowest level and the unit of analysis.
- **Household level**: this is defined as the household in which the children live.
- **Community level**: is defined as a group of households in the same geographical area that share a common primary sampling unit within the DHS dataset.

- **State level**: In Nigeria, the state is the second tier of government after federal government. In the current analysis, each community belongs to one of the 37 distinct geographical locations that represent the states.

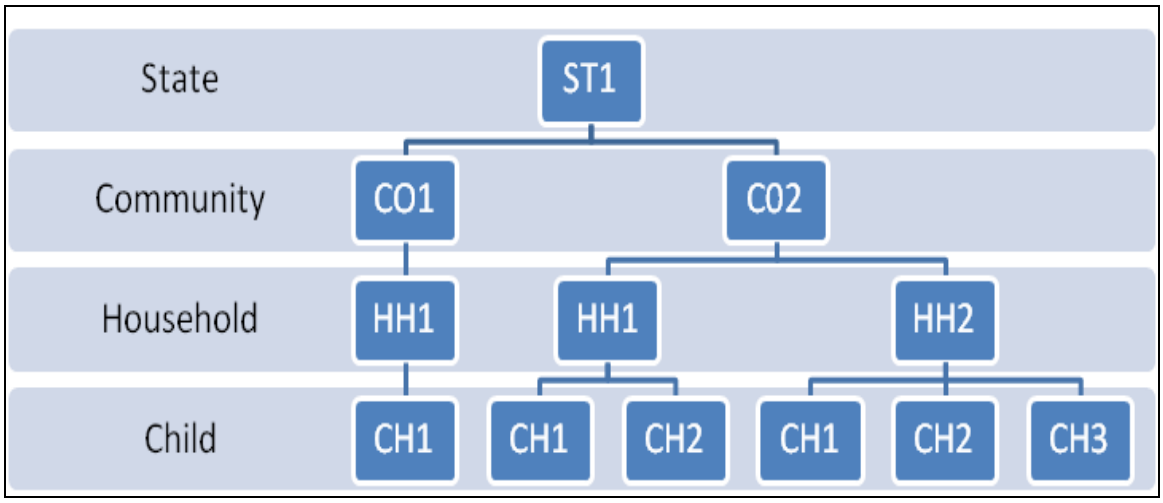


Figure 3: Hierarchical structure of the dataset

4.2 Variable selection

The selection of explanatory variables was guided by the Mosley and Chen (1984) conceptual framework and previous research on child mortality (including Sastry, 1997b; Desai and Alva, 1998; Kravdal, 2004 and NPC and ORC Macro, 2004). The full list of variables consists of socioeconomic and demographic factors at the individual, household, community and state levels and the full description of these variables is given in Appendix A.

4.2.1 The selection and construction of community-level variables

The community level characteristics considered in the present study fall into two groups. The first set of community measures consists of variables derived for each of survey clusters from the spatially explicit databases described in Chapter 3. These include population density, distance to road, distance to coast and malaria prevalence. These were obtained by overlaying the DHS cluster locations with the other data sources and extracting the mean pixel value for each of the covariate at the community level *(*see Figure 4*)*.

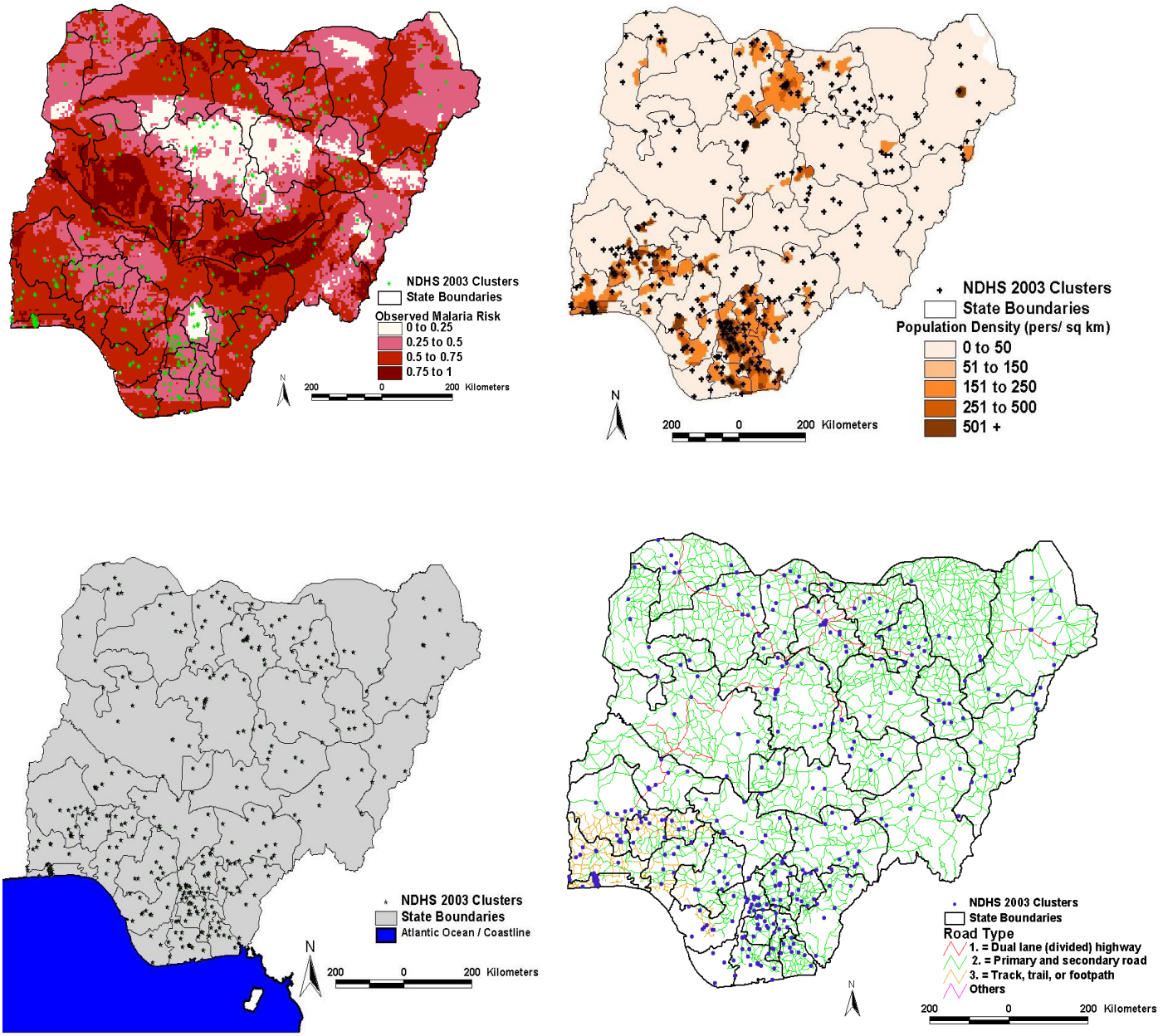


Figure 4 Maps depicting the nature of spatially explicit variables considered

The second set of community variables were based on the aggregation of individual measures from the 2003 NDHS dataset. These variables relate to the health, nutrition and socio-economic conditions in the communities. In order to minimize the number of variables used in the analysis, this set community level variables were grouped into areas such as socio-economic and community environment. Within each group, a principal component analysis was used to obtain summary scores that could be used as an index. These scores were then dichotomized into high and low (Table 2).

Table 2: Community level variables from factor analysis

			Factor 1	
Index	**Items**	**Factor Pattern**	**Eigen value**	**% Variance Explained**
Community Environment Index	Percent with access to clean water in community	0.55	3.03	61%
	Percent with access to hygienic toilet in community	0.76		
	Percent with access to finished floor in community	0.79		
	Percent with access to clean cooking fuel in community	0.88		
	Percent with access to electricity in community	0.86		
Community Health Service index	Percent of births delivered in medical facility	0.90	4.48	75%
	Percent of births with postnatal care	0.93		
	Percent of births with antenatal care	0.93		
	Per cent of births delivered by a skilled attendant	0.90		
	Percent of mothers who had at least one tetanus injection	0.91		
	Per cent of children 12-23 months fully vaccinated	0.54		
Community Child Deprivation Index	Percent with risky birth interval	0.35	1.34	45%
	Percent born to too young or too old women	0.76		
	Percent of children with high birth order	0.80		
Community Maternal Socioeconomic Index	Percent with at least secondary education	0.84	2.46	49%
	Percent White Collar job	0.56		
	Percent of single women or monogamous unions	0.56		
	Percent with access to at least one media type	0.71		
	Average composite score on Autonomy in community	0.79		

4.2.2 Final Data set

To overcome any potential problems with the analysis, data relating to twins was excluded and data concerning children who were not usual residents of the community in which they were sampled was removed so that community specific factors are not wrongly assigned to children who were not usual residents of the community in which they were sampled.

The occurrence of missing values was generally small and observations with missing values were assigned to the 'other' category. In order to avoid unstable categories due to small numbers, most of the categorical variables were re-categorized to be comparable to previous studies. The largest category for each categorical variable was assigned as the reference group.

4.3 Descriptive summaries

Table 3 presents descriptive statistics for the main child level variables of interest disaggregated by the child's survival status. The mean survival times *S(t)*, its standard error *SE S(t)* and the p-value that indicates if the survival times are significantly different for each group of covariate are included in Tables 3 through 6. The initial discussion is on the distribution of children by survival status and a summary of the findings from the survival analysis is given in section 4.3.1. The final dataset utilized for analysis had information on 5684 children of whom 752 (13%) had died before attaining age five. Table 3 reveals that the sample was almost equally distributed by child's sex and birth order. The majority of the children had no older siblings or were of preceding birth intervals of more than three years. A high proportion of the sampled children had no succeeding birth intervals by virtue of being last births. As reported by the mothers of the children, most of the children had average/larger birth weights. On the average, mother's age at child birth was 27 years. About two thirds of the children were delivered at homes, 60% did not receive professional prenatal care, 64% had a traditional birth attendant and only about 46% of the children had antenatal care at a health facility. Long labour at birth was the major problem that mothers had at the birth of the child (24%) and convulsions at birth were the least frequent problems (3%).

Taking the child's survival status into account, the percentage of children surviving was almost the same by gender and birth order. Children with preceding and succeeding birth intervals of less than 24 months had the worst survival. Fewer children also survived in the following groups, small/very small birth size, born to mothers of <18 years, delivered at homes, traditional birth assistance and home as source of prenatal care.

Table 3: Descriptive statistics of child level variables

Description	Dependent	# Alive	# Dead	Total	%Alive	Mean S(t)	SE S(t)	P-value
Gender	*Male*	2498	395	2893	86.35	41.73	0.29	0.3226
	Female	2434	357	2791	87.21	45.56	0.32	
Birth order	*First to third birth*	2573	377	2950	87.22	45.55	0.31	0.3038
	Fourth or higher birth	2359	375	2734	86.28	41.75	0.30	
Preceding Birth Interval	*No older siblings or > 36 months*	2532	301	2833	89.38	46.53	0.30	<.0001
	Less than 24 Months	872	206	1078	80.89	39.67	0.53	
	24 to 35 months	1528	245	1773	86.18	41.74	0.37	
Succeeding Birth Interval	*No younger sibling or > 36 months*	3676	349	4025	91.33	43.65	0.22	<.0001
	Less than 24 Months	439	260	699	62.8	35.47	0.84	
	24 to 35 months	817	143	960	85.1	42.58	0.44	
Size at Birth	Small/very small	718	178	896	80.13	29.61	0.44	<.0001
	Average or larger	4214	574	4788	88.01	45.97	0.24	
Mothers age at birth	<18 Years	398	88	486	81.89	43.23	0.86	0.0005
	18-34 Years	3704	517	4221	87.75	42.42	0.23	
	35 an older	830	147	977	84.95	41.06	0.53	
Place of delivery	*Homes/Others/Missing*	3149	586	3735	84.31	44.10	0.30	<.0001
	Health Facility	1783	166	1949	91.48	44.07	0.29	
Source of prenatal care	Skilled Birth Attendant	2101	147	2248	93.46	33.68	0.19	<.0001
	Traditional Birth Attendant/Other/None	2831	605	3436	82.39	43.75	0.31	
Birth Assistance	Trained Medical Personnel	1872	179	2051	91.27	43.98	0.29	<.0001
	Traditional Birth Attendant/Other/None	3060	573	3633	84.23	44.06	0.30	
Source of antenatal care	*Homes/Other/None*	2819	607	3426	82.28	43.69	0.31	<.0001
	Health Facility	2113	145	2258	93.58	33.73	0.18	
Long labour at birth?	*No*	3780	515	4295	88.01	42.55	0.23	<.0001
	Yes	1152	237	1389	82.94	43.28	0.51	
Excessive bleeding at birth?	*No*	4071	585	4656	87.44	45.68	0.24	0.0004
	Yes	861	167	1028	83.75	30.89	0.37	
Higher fever at birth?	*No*	4397	648	5045	87.16	45.52	0.24	0.0069
	Yes	535	104	639	83.72	40.46	0.68	
Convulsions at birth?	*No*	4819	717	5536	87.05	45.45	0.23	<.0001
	Yes	113	35	148	76.35	37.18	1.64	

Table 4: Descriptive statistics of mother level variables

Description	Dependent	# Alive	# Dead	Total	%Alive	Mean S(t)	SE S(t)	P-value
Mothers Highest Educational Level	*No education*	2428	462	2890	84.01	40.79	0.31	<.0001
	Primary	1191	184	1375	86.62	45.27	0.46	
	Secondary plus	1313	106	1419	92.53	44.53	0.33	
Mothers Occupation	No Work	1681	276	1957	85.9	44.68	0.41	0.0529
	White Collar Job	2069	288	2357	87.78	42.53	0.30	
	Agric and Others	1182	188	1370	86.28	41.80	0.42	
Type of Marital Union	*Monogamy/Never married*	3357	467	3824	87.79	42.38	0.24	0.004
	Polygamy	1575	285	1860	84.68	44.42	0.41	
Ethnicity	Hausa	1506	263	1769	85.13	44.54	0.42	<.0001
	Igbo	616	69	685	89.93	32.68	0.39	
	Yoruba	500	36	536	93.28	44.94	0.51	
	Fulani	406	86	492	82.52	30.91	0.53	
	Others	1904	298	2202	86.47	41.81	0.33	
Religion	Christian	1912	239	2151	88.89	42.91	0.31	<.0001
	Muslim	2928	487	3415	85.74	44.77	0.30	
	Traditionalist or Others/missing	92	26	118	77.97	37.75	1.85	
Media Exposure	No Media Exposure	1963	340	2303	85.24	41.28	0.34	0.007
	Exposed to at least one source	2969	412	3381	87.81	45.83	0.28	
Decision making index	No Decision	1875	321	2196	85.38	41.22	0.35	0.0047
	At least one decision	3057	431	3488	87.64	45.82	0.28	
Problem getting medical help	No problem	467	95	562	83.1	30.89	0.51	0.0044
	At least one problem	4465	657	5122	87.17	45.52	0.24	
Partners Occupation	No Work/No Partner	114	11	125	91.2	33.20	0.89	0.1553
	White Collar Job	1893	268	2161	87.6	42.35	0.32	
	Agric/Other	2925	473	3398	86.08	44.99	0.30	
Partners Highest Educational Level	*No education/Not married/Missing*	2006	390	2396	83.72	43.80	0.38	<.0001
	Primary	1183	190	1373	86.16	41.72	0.43	
	Secondary plus	1743	172	1915	91.02	43.87	0.30	

Looking at the mother level variables, Table 4 shows that almost half of the mothers surveyed did not have any education, about one third of the children were born to mothers who did not work and one third to mothers in polygamous marriages. Most of the children had a Muslim background, and about 53% belonged to the three major ethnic groups. Table 4 also indicates that Mothers of most of the children were exposed to at least one media source and made at least one decision that affected their lives. The

majority of the mothers however reported having at least one problem getting medical help. Among the children in the sample, only about 34% had fathers with secondary education or higher and majority of the fathers were employed. The results in Table 4 also suggest that children whose mothers had secondary education, whose mothers had white collar jobs, whose mothers were in monogamous unions, those of Yoruba and Christian backgrounds as well as those whose fathers had secondary education had lower percentages of deaths.

An investigation of the descriptive statistics of household variables (Table 5) reveals that majority of the children (76%) lived in households with well/surface water as source of drinking water. The majority of the children lived in households with pit latrine toilets and in households that used high pollution fuels. The percentage of children surviving was least for children in households with well water, no toilet facility, with natural floor as well as those in households using high pollution fuels.

Table 5: Descriptive statistics of household variables

Description	Dependent	**# Alive**	**# Dead**	**Total**	**%Alive**	**Mean S(t)**	**SE S(t)**	**P-value**
Source of drinking water	Piped or Tap	799	96	895	89.27	43.27	0.46	0.0006
	Well or Surface	3726	615	4341	85.83	44.82	0.27	
	Others	407	41	448	90.85	22.24	0.28	
Type of toilet facility	Flush	532	30	562	94.66	45.59	0.44	<.0001
	Pit latrine	3078	473	3551	86.68	45.27	0.29	
	No facility or Others	1322	249	1571	84.15	40.74	0.42	
Flooring materials	Natural and rudimentary	1949	392	2341	83.26	43.57	0.39	<.0001
	Finished	2983	360	3343	89.23	43.08	0.25	
Type of Cooking Fuel	Cleaner Fuels	1007	83	1090	92.39	44.44	0.38	<.0001
	High Pollution Fuels	3925	669	4594	85.44	44.67	0.26	
Household Wealth Status	*Poorest*	1111	221	1332	83.41	40.41	0.47	<.0001
	Poorer	1022	220	1242	82.29	43.09	0.54	
	Middle	973	150	1123	86.64	42.01	0.46	
	Richer	971	102	1073	90.49	43.71	0.41	
	Richest	855	59	914	93.54	45.01	0.38	

The descriptive statistics for community variables (Table 6), reveal that majority of the children lived in the Northern part of the country and mostly in rural areas (65%). The highest number of deaths was associated with communities with low scores on the community level indexes as can be seen in Table 6.

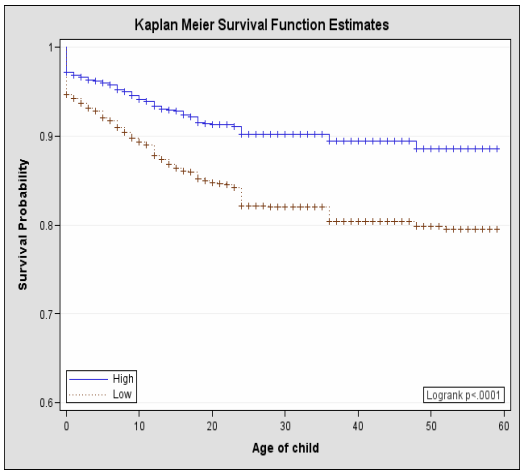
Table 6: Descriptive statistics of community variables

Description	Dependent	# Alive	# Dead	Total	%Alive	Mean S(t)	SE S(t)	P-value
Community environmental factors	*Low*	3037	572	3609	84.15	43.99	0.31	<.0001
	High	1895	180	2075	91.33	44.07	0.28	
Community health service index	*Low*	2776	543	3319	83.64	43.76	0.32	<.0001
	High	2156	209	2365	91.16	43.95	0.27	
Community child deprivation index	High	2057	240	2297	89.55	43.14	0.30	<.0001
	Low	2875	512	3387	84.88	44.45	0.31	
Community maternal socioeconomic index	*Low*	2857	533	3390	84.28	44.05	0.32	<.0001
	High	2075	219	2294	90.45	43.65	0.28	
Malaria prevalence	Low (0-35% reference category)	786	132	918	85.62	41.49	0.53	0.5091
	Medium (36–60%)	2429	358	2787	87.15	42.14	0.29	
	High Endemicity (>60%)	1717	262	1979	86.76	45.33	0.38	
Population density	<100 per sq km	1705	301	2006	85	44.50	0.40	0.0059
	100+ per sq km	3227	451	3678	87.74	42.38	0.25	
Distance to roads	< 1 km	2311	336	2647	87.31	42.25	0.29	0.22
	1+ km	2621	416	3037	86.3	45.05	0.32	
Coastal proximity	<500 km	2265	282	2547	88.93	42.88	0.29	<.0001
	500+ km	2667	470	3137	85.02	44.47	0.32	
Region	North Central	850	107	957	88.82	42.85	0.47	<.0001
	North East	1194	225	1419	84.14	40.93	0.44	
	North West	1470	258	1728	85.07	44.43	0.43	
	South East	438	50	488	89.75	32.64	0.46	
	South South	448	70	518	86.49	31.70	0.49	
	South West	532	42	574	92.68	44.64	0.51	
Type of Place of residence	Urban	1803	189	1992	90.51	43.72	0.30	<.0001
	Rural	3129	563	3692	84.75	44.27	0.30	

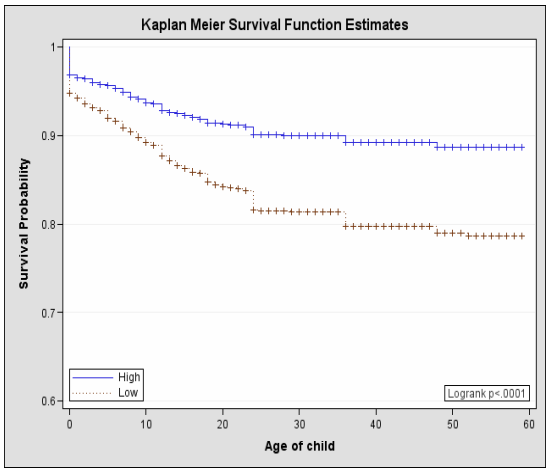
4.3.1 Results of the survival analysis

The results of survival analysis via the K–M method are displayed along with the summaries in Tables 3, 4, 5 and 6. In summary, there are significant differences in the survival times of children for most of the covariates considered. The variables not showing significant differences in survival times at the 5% level include: gender of child,

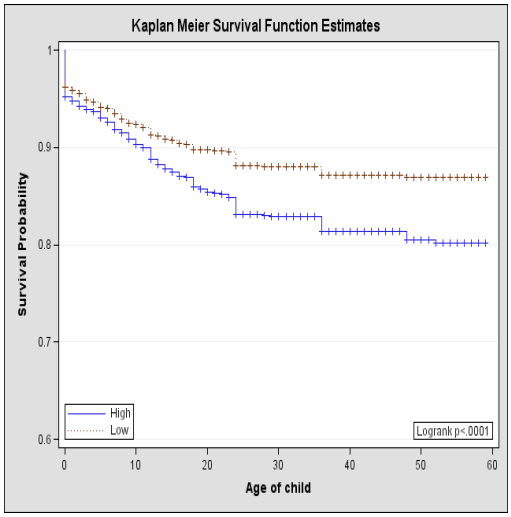
birth order, mother's occupation, partner's occupation, malaria prevalence and distance to roads.



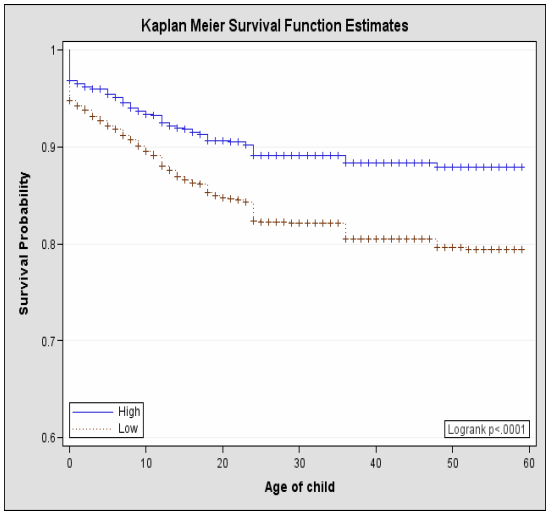
a) Community Environment Index



b) Community Health Service index



c) Community Child Deprivation Index



d) Community Maternal Socioeconomic Index

Figure 5: Kaplan-Meier Survival Curves for community level covariates

With special focus on the community level variables generated from factor analysis, the survival curves exhibited significant differences (Figures 5a-5d). In summary, children in communities with high community environment scores exhibited higher survival chances. Living in communities with access to good health service index was also associated with higher survival probabilities (Figure 5b). Low community child deprivation score is

significantly associated with greater survival probability and children in communities where maternal socioeconomic scores were high had better chances of survival than those in communities with low socioeconomic scores.

4.3.2 Investigation of clustering of deaths

The following section details the results of descriptive analyses conducted in order to establish if some correlation of deaths occurs as a result of children belonging to the same household, community and state. These tables were derived using the approach in Sastry (1997a and b). Table 7 shows the distribution of children and deaths per household from the 2003 NDHS There were 3215 households in the sample. A total of 752 deaths occurred to 635 families, while 2580 families never experienced a child death.

Table 7: Distribution of births and deaths in households

	Deaths in household	0	1	2	3	4	5	# families	% families	# children	%Children	#deaths	% deaths	#deaths/# children	(% deaths / % Children)*100
Children in Household	0														
	1	1342	123					1465	45.6	1465	25.8	123	16.4	0.08	63.5
	2	996	253	23				1272	39.6	2544	44.8	299	39.8	0.12	88.8
	3	170	110	32	3			315	9.8	945	16.6	183	24.3	0.19	146.4
	4	57	32	14	5	1		109	3.4	436	7.7	79	10.5	0.18	137
	5	12	14	7	1	2		36	1.1	180	3.2	39	5.2	0.22	163.8
	6	3	7	3				13	0.4	78	1.4	13	1.7	0.17	126
	7			1	1		2	4	0.1	28	0.5	15	2	0.54	404.9
	8		1					1	0.0	8	0.1	1	0.1	0.13	94.5
	# families	2580	540	80	10	3	2	3215		5684	100	752	100	0.13	
	# deaths	0	540	160	30	12	10	752							
	%deaths	0	71.81	21.28	3.99	1.6	1.33	100							

The number of children per household ranges from 1 to 8 per household, and there are on the average, 1.77 children per household. About 54% of the households have two or more children, and these children make up about 74% of the total children. Slightly over 28% of the deaths occurred to 3% of the households with two or more child deaths. Additionally, less than 1% of the households contributed three or more deaths; together they account for about 7% of the deaths (Table 7). Table 7 also shows that 46% of households have only 1 child, and that these households account for 16% of the deaths, giving a ratio of 0.63. However, the other 54% of children (who live in households with 2 or more children) accounted for 83.6% of the deaths, giving a ratio of 1.55. This is nearly 2½ times that in single child households, indicating that there is a clustering of deaths in larger households.

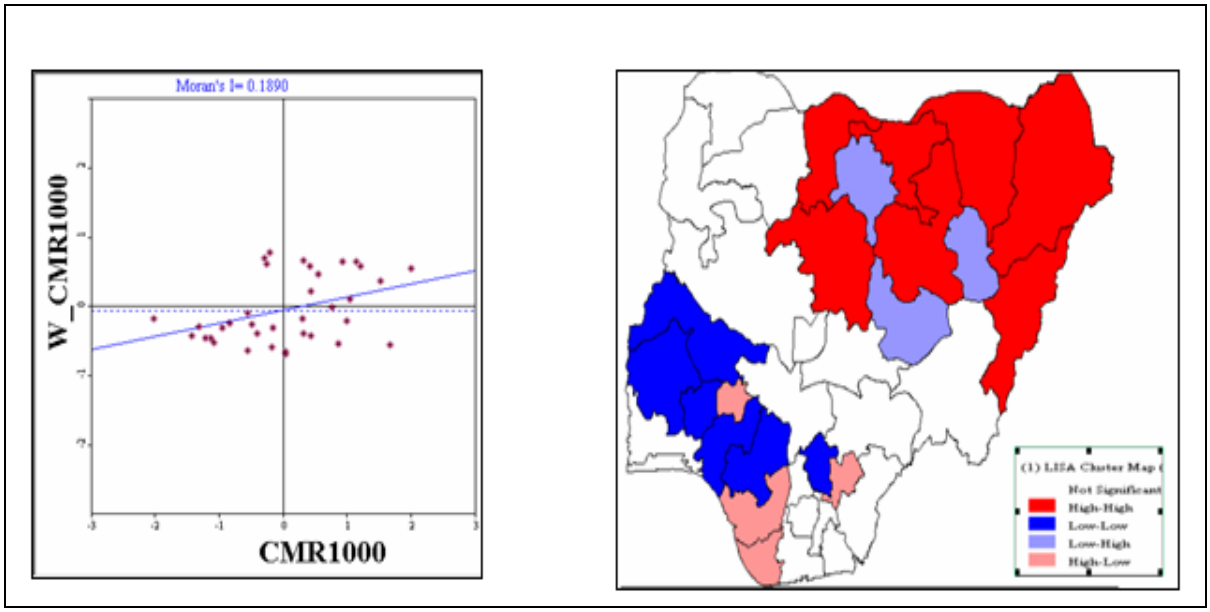
Looking at the distribution of births and deaths in communities (Table 8), there were a total of 752 deaths in the 361 communities in the dataset. A total of 112 communities did not experience any deaths, while 69% of the communities had experienced one or more deaths. Communities contributing two or more deaths make up 45% of the communities in the sample.

Table 8: Distribution of births and deaths in communities

		Deaths in Communities								# of communities	# Children	%Children	# Dead	% Dead	% of communities
	#	0	1	2	3	4	5	6	7-14						
Children in Communities	1-10	78	46	10	2	0	0	0	0	136	886	16	72	10	38
	11-20	31	35	33	9	11	2	2	1	124	1898	33	201	27	34
	21-46	3	7	14	16	17	7	12	25	101	2900	51	479	64	28
	# of communities	112	88	57	27	28	9	14	26	361	5684	100	752	100	100
	%of communities	31	24	16	7	8	2	4	7	100					
	# Dead	0	88	114	81	112	45	84	228	752					
	% Dead	0	12	15	11	15	6	11	30	100					
	# Children	1031	993	944	581	658	224	397	856	5684					
	% Children	18	17	17	10	12	4	7	15	100					

The results of the ESDA are displayed in Figures 6a and 6b. The nearest neighbour criterion was used in creating the weight matrix for this analysis. To this end, ten nearest neighbours which considered all lower number of neighbours was utilized. The Moran's I statistics computed for the whole study area gave a figure of 0.1890 (Figure 6a). This indicates a low positive spatial autocorrelation in U5M rates across the states in the country as a whole, implying that child mortality rates are not spatially randomly distributed.

The Moran's scatter plot map (Figures 6b) reveals that the hot spots for child mortality rates (areas of high mortality, surrounded by areas of similarly high mortality) are mostly found in the northern states (areas in red). Significant cold spots (areas of low mortality, surrounded by areas of similarly low mortality) are concentrated in the south-western part of the country (states coloured in blue). The majority of the states are devoid of spatial clustering (white areas). The map however reveals that Kano, Plateau and Gombe states are spatial outliers among the northern states. Specifically, these are states of low mortality surrounded by high mortality states.



a) Moran Scatter Plot **b) LISA cluster Map**

Figure 6: Results from Spatial autocorrelation for U5M

In summary, the result from this descriptive investigation of clustering in the preceding paragraphs suggests that clustering of child mortality does exist at the household, community and state levels, primarily because the majority of the units in the various levels (household, community or state) did not have any child deaths and only a few units in the different levels (household, community or state) account for the majority of the deaths in the sample. Therefore, clustering has to be taken into account in the multivariate analysis by the inclusion of frailty effects at the relevant levels.

4.4 Multivariate analysis

In order to implement the discrete time survival model described in Chapter 3, the data was restructured from a child level dataset (in which a child contributed one record), to a child period dataset (see Table 9 below). In the child period data, each child contributed one observation for each time period from birth until they died or were censored. For each child-month, the dependent variable (survival status) is coded 1 if the child died during that month and 0 otherwise. For example, a child who survives the first 3 months of life will have 3 records, while a child who dies at age 4 months will have 4 records. This resulted in a total of 142913 observations from the 5684 child based records.

Table 9: Creation of child-period dataset from the original child-level data set

Child level dataset				
Child ID	Duration (Months)	Survival Status	Gender	{Other variables …..}
001	4	0	1	
002	3	1	2	
Child-period dataset Illustration of a discrete time dataset				
Child ID	Discrete Time (Month)	Survival Status	Gender	{Other variables…..}
001	1	0	1	
001	2	0	1	
001	3	0	1	
001	4	0	1	
002	1	0	2	
002	2	0	2	
002	3	1	2	

4.4.1 Modelling Strategy and Model Comparison Approach

To study the determinants of child mortality and the extent of heterogeneity in mortality risk, several geo-additive survival models are estimated and compared. The models differ with respect to variable composition, treatment of covariates (whether as fixed or random), and inclusion of frailty term (see Table 10).

Table 10: Models considered

1. Child + Mother + HH variables
2. Model 1 + HH (random)
3. Model 2 + Community level variables
4. Model 3 + Community (random)
5. Model 4 + State (random)
6. Model 4 + State (spatial)
7. Model 4 + State (random) + State (spatial)
7b. Similar to model 7 but with a non-linear effect of mother's age at birth of child

Model 1 which is the simplest model consists of only covariate effects at the child, mother and household level. This model is the typical type of model considered in child mortality studies and does not include any random effects. Model 1 is then progressively expanded to include covariates and frailty effects at other levels. The full model (model 7) comprises of covariates at the child, mother, household and community levels as well as frailty effects at the household, community and state levels. In addition, model 7 splits the state level frailty effects into two so as to decide how much variation is spatially structured and how much is unstructured at the state level.

All models were estimated using BayesX version 2.0 (Brezger *et al.* 2005). For each model, 12,000 iterations were carried out, the first 2000[11] samples were discarded and every 10th observation thereafter was saved for parameter estimation. All models assumed non-linear effect of child's age, time-varying effect for breastfeeding: modelled via p-splines, and fixed effects of all other covariates. An additional model (7b) was also

[11] Convergence was monitored through autocorrelation functions and trace plots which are part of the output from the BayesX software and the plots showed evidence of good mixing behaviour and a minor autocorrelation.

considered. This model is similar to Model 7 except that the effect of mother's age is assumed to be continuous and is entered into the model as a non-linear effect in an attempt to assess the bias arising from modelling it as a fixed effect. Only main effects are considered for all covariates and interaction effects are not considered due to the number of covariates involved. Means, standard deviations and quantiles estimated from the posterior distributions are used to assess model fit for all models and credible intervals (CI) used to assess the significance of parameters. The DIC described in Chapter 3 was used to compare all the models and to explore the effect of adding covariates and frailty terms to Model 1.

The results for model fit and variance components are summarized in Table 11. Based on the DIC values, model 6 had the lowest DIC value and thus is the best model. Model 3 which incorporated child, mother, household and community level variables as well as household random effect had the second lowest DIC. Looking at the difference in DIC of other models relative to models 6, it can be concluded that models 2, 3, 4 and 7 can be weekly differentiated as they all have DIC difference of between 3-7 from the best model, while models 1, 5 and 7b cannot be supported (strong evidence in favour of the model 6 with the smaller DIC). The inclusion of random effects as well as community level variables to model 1 lead to increased model complexity but also to a substantial improvement in the DIC values, thereby suggesting the importance of contextual and frailty effects. Even though model 7 which incorporated spatial and random effects had a good fit, the proportion of total spatial variance attributed to the spatial clustering were 0.69 for Model 7 and 0.71 for Model 7b, indicating a higher share of spatial variability due to the structured spatial effect and further supporting model 6 as the preferred model. Finally, Model 7b which is a variant of Model 7 shows a higher DIC value compared to model 6, thereby supporting the inclusion of mother's age at birth as a categorical variable.

Table 11: Results from Models 1 -7b – Model fit and Variance components of random and non-linear effects

Estimation results for the DIC:	**Model1**	**Model2**	**Model3**	**Model4**	**Model5**	**Model6**	**Model7**	**Model7b**
Deviance	6023.45	5656.06	5578.84	5588.29	5632.76	5569.56	5619.01	5641.11
p_D	52.82	217.74	254.66	251.87	231.72	257.86	236.04	229.10
DIC	6129.09	6091.54	6088.17	6092.03	6096.19	**6085.28**	6091.10	6099.31
ΔDIC*	43.81	6.27	2.89	6.75	10.92	0.00	5.82	14.04
Variance components**								
Household effects		0.4689 (0.1119-0.8674)	0.5857 (0.3273-0.9513)	0.5302 (0.2298-0.9753)	0.4538 (0.1572-0.7993)	0.5447 (0.2621-1.1477)	0.4636 (0.184-0.8011)	0.4283 (0.1042-0.811)
Community effects				0.043 (0.0005-0.1875)	0.0409 (0.0006-0.1874)	0.0677 (0.0019-0.2003)	0.054 (0.0007-0.1931)	0.0457 (0.0009-0.1709)
State (Random)					0.0105 (0.0006-0.0484)		0.0108 (0.0005-0.049)	0.0125 (0.0005-0.0575)
State (Spatial)						0.0355 (0.1842-0.0009)	0.0241 (0.1451-0.0005)	0.0304 (0.1567-0.0007)
Age of child	16.9249 (8.9219-31.1239)	19.5834 (9.2194-40.37)	15.9612 (8.3957-28.7482)	16.2055 (8.5326-30.3106)	16.1456 (8.5572-30.3934)	16.0995 (8.4066-30.0385)	15.9213 (8.2426-30.3994)	15.8296 (8.4734-28.5438)
Breastfeeding	1.1889 (0.0646-6.4279)	0.5653 (0.0308-2.6391)	1.0291 (0.0711-5.1032)	0.8754 (0.066-3.7927)	0.8844 (0.0585-3.8003)	0.8041 (0.0645-3.5284)	0.9761 (0.0641-4.6433)	0.7774 (0.059-3.5333)
Mother's age at birth								0.009 (0.0007-0.0427)

*Difference of the best model against others

**CI in Parenthesis

4.4.2 Sensitivity analysis

The performance of the models in a Bayesian framework can be sensitive to the choice of the variance components priors, and this may arise due to small sample sizes (Gelman, 2006). Although results are insensitive to the choice of *a* and *b* for moderate to large data sets, a sensitivity analysis is recommend for checking the changes models with respect to changes in the hyperparameters (Hennerfeind *et al.* 2006). The sensitivity analysis was carried out with the same set of covariates as in model 6 and involved changing the prior distributions for the variance components using the following values (a=1,b=0.005) – almost diffuse prior, (a=1,b=0.00005) and (a=0.00005,b=0.00005). These values reflect different degrees of uncertainty about the variance components and details of hyperparameters are provided in Section 3.2.1.6.

Table 12: Sensitivity to choice of hyperparameter values for Model 6

	Hyperparameters			
	a=0.001, b=0.001*	a=1,b=0.005	a=1,b=0.00005	a=0.00005,b=0.00005
Model Fit				
Deviance	5569.56	5720.87	6006.45	5652.47
p_D	257.86	197.09	70.21	224.07
DIC	6085.28	6115.05	6146.88	6100.61
Random effects**				
Household	0.54466 (0.26211-1.14767)	0.34531 (0.08037-0.79865)	0.01308 (0.00002-0.07451)	0.43002 (0.09526-0.82882)
Community	0.0677 (0.00185-0.2003)	0.02399 (0.00185-0.11718)	0.00016 (0.00002-0.00116)	0.03857 (0.00006-0.16322)
State- Unstructured				
State- Structured	0.03551 (0.1842-0.00087)	0.0116 (0.05276-0.00135)	0.00023 (0.00122-0.00001)	0.02362 (0.18899-0.00003)

* Default values

** variance components - posterior mean and 95% CI in parenthesis

As can be seen in Table 12, the choice of hyper-parameter does affect the estimates. The benchmark model (model 6) which used priors: a=b=0.001 had the lowest DIC and can be considered the best model. Decreasing the value of b while maintaining a=1 resulted in a decrease in the size of the variance effects. The choice of a=b=0.001 is however considered appropriate for the current exercise since the DIC was lowest for this model.

4.4.3 Interpretation of categorical covariates (fixed effects)

The focus of the discussion from this point on will be the results of Model 6 which was the best model according to the DIC criterion. Comparisons will be drawn to other models where necessary. The parameter estimates obtained from the models are shown in Tables 13 through 16. Statistical significance of the effects was assessed at the 0.05 level by evaluating whether the 95% CI of the posterior distribution contained zero (0). An effect is therefore significant and marked with asterisk (*) if its 95% CI does not include zero. In general, if the sign of an effect is positive, it implies that there is a higher risk of mortality for children in that group relative to the reference category. As can be seen from Tables 13 through 16, the coefficients for the fixed effects are generally of the same magnitude and direction (had the same signs) and the same set of covariates were statistically significant across the models.

A close look at the posterior estimates for the child level effects in Table 13 reveals that mortality is significantly higher for children with preceding birth intervals of up to 35 months relative to those with no older siblings or with intervals of more than 35 months. The results for model 6 also suggests that, those with succeeding birth intervals of up to 24 months have a higher mortality risk compared to those with no younger siblings and with succeeding birth intervals of more than 24 months. Children with small sizes at birth have a higher chance of dying compared to those with birth sizes of average to large. Having long labour at birth as well as having convulsions at birth also significantly increases the risk of the child dying before the age of 5 years.

Mother's secondary education significantly reduces the mortality of children as can be seen from Table 14. Although not statistically significant, children born to mothers whose partners are in white collar jobs as well as those whose partners have secondary education have a lower mortality risk.

Table 13: Posterior summaries for child level effects models 1-7b

Description	Dependent	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model7b
Constant		-6.374*	-6.298*	-6.304*	-6.252*	-6.233*	-6.209*	-6.136*	-6.11*
Gender	*Male*								
	Female	0.013	0.014	0.006	0.007	0.007	0.009	0.011	0.000
Birth order	*First to third birth*								
	Fourth or higher birth	-0.026	-0.042	-0.05	-0.047	-0.04	-0.053	-0.049	-0.039
Preceding Birth Interval	*No older siblings or > 36 months*								
	Less than 24 Months	0.142*	0.15*	0.154*	0.14*	0.152	0.17*	0.162*	0.145
	24 to 35 months	0.139*	0.154*	0.155*	0.15*	0.154*	0.156*	0.162*	0.173*
Succeeding Birth Interval	*No younger sibling or > 36 months*								
	Less than 24 Months	0.575*	0.631*	0.645*	0.616*	0.633*	0.653*	0.644*	0.641*
	24 to 35 months	-0.162*	-0.18*	-0.177*	-0.178*	-0.174*	-0.183*	-0.176*	-0.171*
Size at Birth	Small/very small	0.16*	0.185*	0.209*	0.203*	0.213*	0.217*	0.211*	0.206*
	Average or larger								
Mothers age at birth	<18 Years	0.076	0.064	0.053	0.022	0.041	0.044	0.043	
	18-34 Years								
	35 an older	0.136	0.158	0.159	0.176	0.158	0.166	0.164	
Place of delivery	*Homes/Others/Missing*								
	Health Facility	-0.14	-0.135	-0.125	-0.115	-0.111	-0.113	-0.11	-0.151
Source of prenatal care	Skilled Birth Attendant	0.151	0.153	0.182	0.123	0.199	0.197	0.181	0.227
	Traditional Birth Attendant/Other/None								
Birth Assistance	Trained Medical Personnel	-0.034	-0.054	-0.04	-0.044	-0.056	-0.055	-0.049	-0.003
	Traditional Birth Attendant/Other/None								
Source of antenatal care	*Homes/Other/None*								
	Health Facility	-0.23	-0.219	-0.241	-0.205	-0.256	-0.249	-0.232	-0.283
Long Labour at birth	*No*								
	Yes	0.215*	0.231*	0.225*	0.214*	0.237*	0.229*	0.231*	0.242*
Excessive bleeding at birth	*No*								
	Yes	0.143	0.165*	0.157	0.148*	0.155*	0.166	0.167*	0.173
Higher fever at birth	*No*								
	Yes	0.124	0.155	0.165*	0.167*	0.152	0.163	0.168	0.169
Convulsions at birth	*No*								
	Yes	0.147	0.165	0.196	0.141	0.205	0.208*	0.225	0.217
Any problem at birth?	*No problem*								
	At least one problem	-0.176	-0.197	-0.193	-0.176	-0.193	-0.195	-0.195*	-0.22*

* Significant at 0.05% (i.e. 95% CI does not include 0)
Reference categories appear in italics

Table 14: Posterior summaries for mother level effects models 1-7b

Description	Dependent	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model7b
Mothers Highest Educational Level	*No education*								
	Primary	0.137	0.156	0.181*	0.16	0.178*	0.158	0.171*	0.17
	Secondary plus	-0.24*	-0.244*	-0.318*	-0.274	-0.293*	-0.277*	-0.281*	-0.274*
Mothers Occupation	No Work	0.033	0.037	0.047	0.048	0.05	0.05	0.039	0.054
	White Collar Job								
	Agric and Others	-0.019	-0.026	-0.029	-0.017	-0.009	-0.016	-0.012	-0.023
Type of Marital Union	*Monogamy/Never married*								
	Polygamy	0.019	0.022	0.042	0.039	0.036	0.043	0.036	0.047
Ethnicity	Hausa	0.096	0.115	0.012	0.048	0.016	0.01	0.018	0.025
	Igbo	-0.339*	-0.418*	-0.39	-0.484	-0.425	-0.43	-0.441	-0.475
	Yoruba	-0.13	-0.103	0.11	0.096	0.128	0.156	0.148	0.15
	Fulani	0.276*	0.316*	0.23	0.276	0.225	0.227	0.244	0.24
	Others								
Religion	Christian	0.095	0.065	0.01	-0.053	-0.004	-0.032	-0.038	-0.015
	Muslim								
	Traditionalist or Others/missing	-0.077	0.008	0.034	0.127	0.039	0.099	0.104	0.073
Media Exposure	No Media Exposure	-0.058	-0.048	-0.042	-0.042	-0.042	-0.04	-0.04	-0.038
	Exposed to at least one source								
Decision making index	No Decision	0.032	0.042	0.051	0.043	0.045	0.045	0.048	0.046
	At least one decision								
Problem getting medical help	No problem	0.057	0.052	0.055	0.053	0.043	0.045	0.049	0.049
	At least one problem								
Partners Occupation	No Work/No Partner	0.245	0.356	0.343	0.349	0.351	0.359	0.387	0.324
	White Collar Job	-0.055	-0.093	-0.079	-0.079	-0.081	-0.082	-0.094	-0.098
	Agric/Other								
Partners Highest Educational Level	*No education/Not married/Missing*								
	Primary	0.005	0.008	-0.005	0.006	0.0004	0.005	-0.002	-0.053
	Secondary plus	-0.1	-0.118	-0.104	-0.103	-0.106	-0.118	-0.103	-0.085

* Significant at 0.05% (i.e. 95% CI does not include 0)
Reference categories appear in italics

Table 15: Posterior summaries for household effects models 1-7b

Description	Dependent	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model7b
Source of drinking water	Piped or Tap	0.223*	0.195	0.219	0.208	0.221	0.213	0.223	0.18
	Well or Surface								
	Others	-0.24	-0.2	-0.231	-0.209	-0.232	-0.206	-0.22	-0.262
Type of toilet facility	Flush	-0.552*	-0.47*	-0.608*	-0.528*	-0.496*	-0.56*	-0.572*	-0.401*
	Pit latrine								
	No facility or Others	0.319*	0.269	0.351*	0.305*	0.28*	0.312*	0.327*	0.232
Flooring materials	Natural and Rudimentary	-0.024	-0.046	-0.023	-0.022	-0.035	-0.037	-0.033	-0.04
	Finished								
Type of Cooking Fuel	Cleaner Fuels	0.025	0.042	0.038	0.084	0.048	0.031	0.051	0.068
	High Pollution Fuels								
Household Wealth Status	*Poorest*								
	Poorer	0.219*	0.243*	0.217	0.239	0.234	0.227	0.22	0.237
	Middle	0.052	0.062	0.042	0.075	0.047	0.039	0.058	0.077
	Richer	-0.336*	-0.363*	-0.358*	-0.349*	-0.349*	-0.363*	-0.352*	-0.333*
	Richest	-0.023	-0.091	0.021	-0.09	-0.052	-0.038	-0.048	-0.11

* Significant at 0.05% (i.e. 95% CI does not include 0)
Reference categories appear in italics

Compared to children who live in households with a pit latrine, those who live in households with flush toilets have a significantly lower mortality risk, while those in households with no toilet facilities have significantly higher mortality chances as can be seen in Table 15 above.

The community variables did not generally yield statistically significant results, however Table 16 suggests that living in urban areas, living in the South-western part of the country and living in communities with high health service index are all associated with lower mortality.

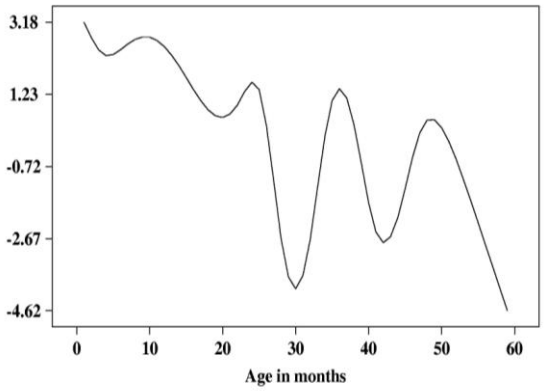
Table 16: Posterior summaries for community effects models 1-7b

Description	Dependent	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model7b
Community environmental factors	*Low*								
	High			-0.026	-0.023	-0.015	-0.006	-0.017	-0.011
Community Health service index	*Low*								
	High			-0.064	-0.056	-0.068	-0.063	-0.07	-0.062
Community Child deprivation index	High			-0.044	-0.046	-0.04	-0.04	-0.041	-0.046
	Low								
Community Maternal socioeconomic index	*Low*								
	High			0.123	0.111	0.119	0.11	0.117	0.115
Malaria Prevalence	Low (0-35% reference category)			0.116	0.127	0.127	0.142	0.138	0.122
	Medium (36–60%)								
	High Endemicity (>60%)			-0.002	-0.001	0.002	-0.007	-0.006	0.002
Population Density	<100 per sq km			-0.013	-0.01	-0.018	-0.004	-0.007	-0.013
	100+ per sq km								
Distance to roads	< 1 km			-0.071	-0.063	-0.068	-0.06	-0.062	-0.07
	1+ km								
Region	North Central			-0.015	-0.048	-0.036	-0.05	-0.03	-0.053
	North East			0.04	0.002	0.04	-0.027	0.013	0.015
	North West								
	South East			-0.02	0.092	0.026	0.009	0.037	0.052
	South South			0.227	0.215	0.197	0.252	0.209	0.217
	South West			-0.246	-0.239	-0.239	-0.19	-0.241	-0.248
Type of Place of residence	Urban			-0.076	-0.077	-0.087	-0.084	-0.083	-0.067
	Rural								

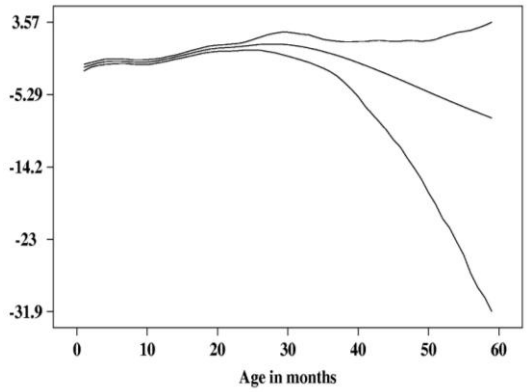
* Significant at 0.05% (i.e. 95% CI does not include 0)
Reference categories appear in italics

4.4.4 Interpretation of non-linear effects

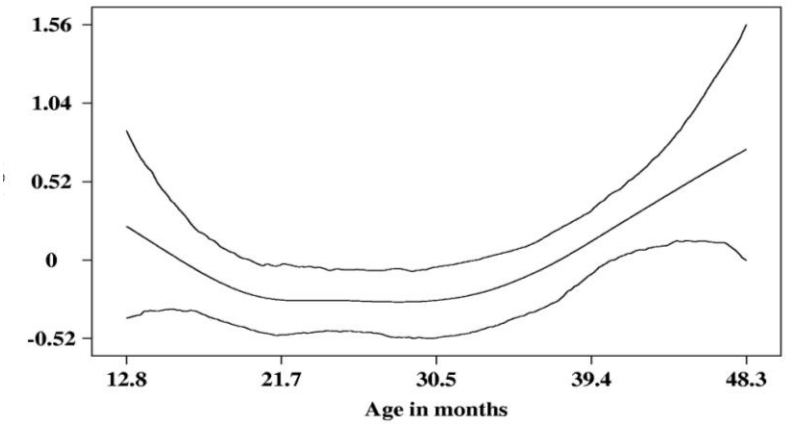
The results for smooth effects of continuous covariates modelled and fitted using penalized splines are displayed for models in Figure 7. In general, the effect of age of child shows a high risk of child death shortly after birth, and an overall decline in deaths as the child grows older (Figure 7a). The heaps appearing at various ages in the curve may be due to the heaping of survival times while the troughs may result from much smaller number of deaths being recorded between these time points. The modelling approach considered here ensures that the heaping has little effect on the estimation of the fixed effect covariates.



a) Effect of child's age – Model 6



b) Effect of breastfeeding – Model 6



c) Effect of Mother's age – Model 7b

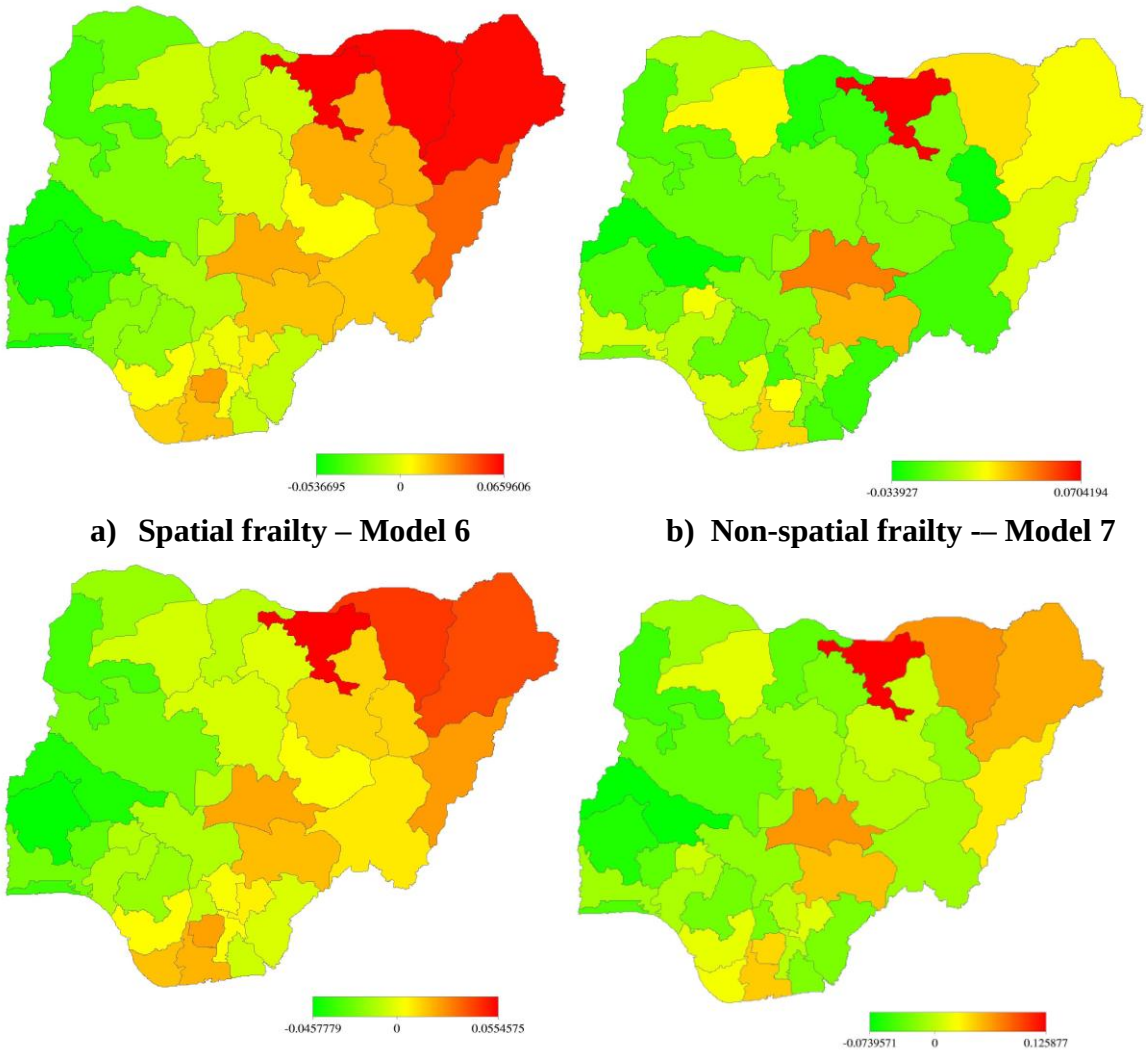
Figure 7: Non-linear effects of metrical covariates – Posterior Mean (Centre line) together with 95% CI (*CI not shown for Figure 7a for sake of clarity*).

Turning to the effect of breastfeeding, it can be observed from Figure 7b that mortality risk is reduced in the early ages, while its effect at the older ages (beyond 30 months) is insignificant. The effect of mother's age at birth of child is almost U-shaped with a higher risk of child deaths attributable to younger and older women (Figure 7c).

4.4.5 Interpretation of the spatial effect

Models 6 and 7 considered the spatial effects of state of residence on child mortality. Model 6 which incorporated only the structured spatial effect is superior in terms of the DIC to model 7 which considers both structured and unstructured spatial effects. Although the results did not show any major hot-spots or cold spots of child mortality,

the spatial pattern from model 6 (Figure 8a) points to the fact that once other variables have been taken into account, mortality risk tends to be higher in the North-Eastern parts of the country (Yobe, Borno, and Jigawa states) and lower in the South-western parts of the country (Lagos, Ogun, and Oyo States amongst others). The results from the LISA cluster (Figure 6b) as well those from Figures 8a-d suggest a concentration of mortality in the North-Eastern part of the country. However, the overall implication of the spatial effect is that although mortality risk exhibits spatial patterns, the spatial variations are probably explained by the covariates considered.



a) Spatial frailty – Model 6 **b) Non-spatial frailty –– Model 7**

c) Spatial frailty – Model 7 **d) Total spatial effects – Model 7**

Figure 8: Maps of the posterior mean of spatial effects

4.5 Determinants of Infant mortality

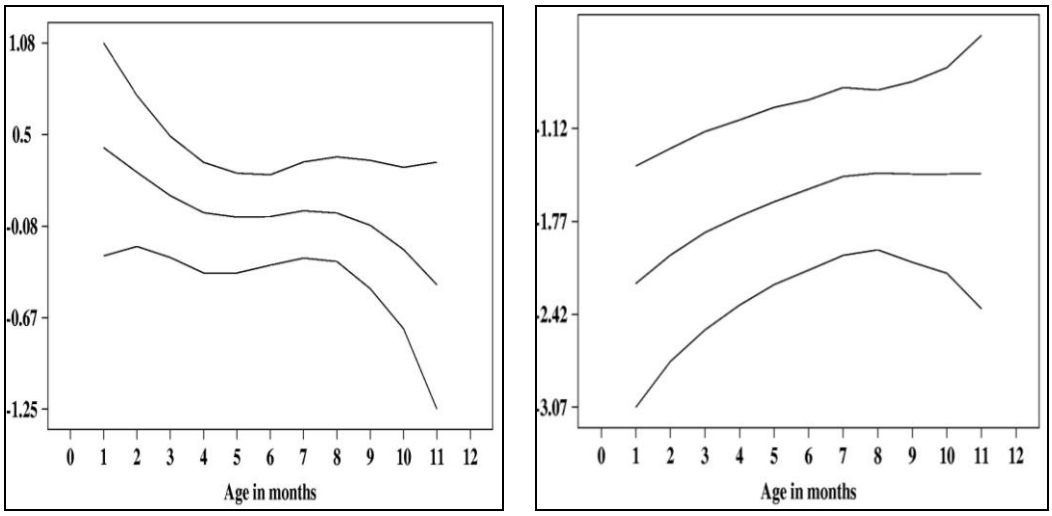
Since the risk factors associated IM and U5M can be very different, a separate analysis of the best fitting model (model 6) was fitted to the data on IM. In the revised dataset for IM analysis, all deaths after 11 months were considered censored and the resulting child-period dataset had 52,065 observations using this approach. Table 17 gives the posterior summaries for the community level variables considered and similar to the results for U5M, it can be observed that the results are not statistically significant but also suggest that living in urban areas, living in the South-western part of the country and living in communities with high health service index lowers IM risk.

Table 17: Posterior summaries for community effects model 6 - IM

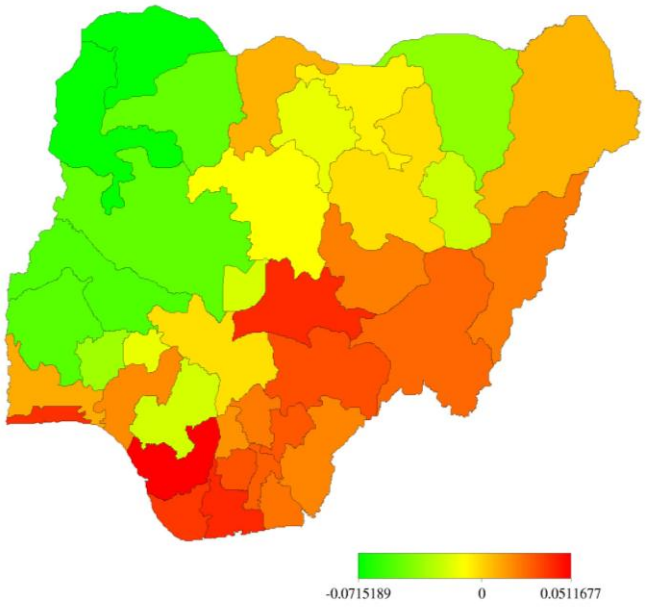
Description	**Dependent**	Model6 - IM
Community environmental factors	Low	
	High	0.06
Community Health service index	Low	
	High	-0.154
Community Child deprivation index	High	0.1
	Low	
Community Maternal socioeconomic index	Low	
	High	0.028
Malaria Prevalence	Low (0-35% reference category)	0.134
	Medium (36–60%)	
	High Endemicity (>60%)	0.139
Population Density	<100 per sq km	-0.015
	100+ per sq km	
Distance to roads	< 1 km	-0.12
	1+ km	
Region	North Central	0.102
	North East	0.127
	North West	
	South East	-0.352
	South South	0.216
	South West	-0.291
Type of Place of residence	Urban	-0.022
	Rural	

The results for smooth effects of continuous covariates on IM fitted using penalized splines are displayed in Figure 9. The effect of age of child shows a high risk of child death shortly after birth, and an overall decline in deaths as the child grows older (Figure

9a). The effect of breastfeeding on IM is such that mortality risk is reduced in the early ages and increases almost linearly with the child's age (Figure 9b). There is a shift in spatial patterns of IM when compared to the results from U5M. The map in Figure 9c shows that the risk of IM tends to be higher in the southern parts of the country. The difference in the pattern is an indication that the modelling of mortality at the childhood ages should take into account the various definitions of childhood mortality.



a) Effect of child's age – IM **b) Effect of breastfeeding – IM**



c) Structured spatial effect – IM

Figure 9: Non-linear: a & b - Posterior Mean (Centre line) together with 95% CI and spatial effects for IM (c)

Chapter Five: Summary and Conclusions

5.1 Summary

The main aim of the project was to account for the influence of contextual factors and frailty on CM and to investigate the spatial patterns of CM in Nigeria. Chapter 1 outlined the problem, and as well as the demographic and statistical issues and also set out the aims and objectives of the study. A literature review was undertaking in Chapter 2, while Chapter 3 listed the data used in the study, defined possible models, and discussed some of the model issues.

The analysis carried out in Chapter 4 examined the effect of community level factors on child mortality as well as the spatial patterns associated with child mortality risk in Nigeria. The results of survival analysis via K-M method revealed that there were significant differences in the survival times of children for most of the covariates considered and the only variables not showing significant differences in survival times were gender of child, birth order, mother's occupation, partner's occupation, malaria prevalence, population density and distance to roads. Results from a descriptive investigation of clustering showed that clustering of child mortality exists at the household, community and states levels and these need to be taken into account in the multivariate analysis by the inclusion of frailty effects at the relevant levels.

All the covariates considered were included into the geo-additive survival models. A total of 8 models were evaluated and the results in Chapter 4 revealed that most of the community level factors considered had no significant effect on child mortality once the household and individual level factors had been taken into account. The results also suggest that the inclusion of frailty terms as well as the inclusion of contextual variables at the community level lead to an improvement in the DIC values thereby suggesting the importance of contextual and frailty effects. A higher share of state level variability in the data was due to the structured spatial effect. The spatial patterns were also found to be insignificant although, they point to very interesting patterns in child mortality variations

in the country. The analysis however indicates that the child and household level factors play an important role in child mortality reduction.

5.2 Recommendations

The importance of correct model choice (particularly with respect to fixed, random and spatial components) has been demonstrated, and the quality of model fit should always be investigated before conclusions are drawn and policies formulated. The findings from this study are preliminary but we give recommendations as follows. Policy programs should focus on the education of women on the need to practice child spacing. Policy makers should develop strategies to narrow the wealth gap in the country. There should be an overall improvement in the area of service delivery with more houses connected to clean, affordable and regular pipe borne water systems. The tools used in the present analysis can also be beneficial in other ways. For example, the mortality cold spots could be studied closely to find out why the areas exhibit different conditions from their immediate neighbours. This would help in devising targeted intervention which will be more effective in child mortality reduction.

5.3 Limitations of the Study /Suggestions for future research

This study faces the following limitations:

1. Due to the cross-sectional nature of the data, the covariates may not reflect the socio-economic and ecological conditions of the child at the time of death
2. The methodology used is vulnerable to various biases due to factors such as migration.
3. The dichotomization of some community level variables may have resulted in the loss of information. Alternative specifications, such as the direct use of the component scores or the categorization of such scores into more than 2 levels are worth considering.

4. There are uncertainties related to the Modifiable Areal Unit Problem[12] (MAUP) (Heywood, 1998).

The literature suggests analysis of spatial effects at multiple levels as a means of alleviating problems related to MAUP. Therefore, to check the sensitivity the choice of geographical unit in measuring spatial effects, a geo-statistical (kriging) model with cluster as the spatial unit of analysis could be explored in addition to the lattice model (state level model), which is the main focus of this work.

[12] MAUP arises when artificial units of spatial reporting (for example states) are used in reporting highly localized spatial occurrences, thereby resulting to misleading spatial patterns

REFERENCES

Adebayo, S.B. and Fahrmeir, L. (2005). Analyzing child mortality in Nigeria with geo-additive discrete-time survival models. *Statistics in Medicine*, 24(5): 709-728.

Adebayo, S.B., Fahrmeir, L. and Klasen, S. (2004). Analyzing Infant Mortality with Geo-additive Categorical Regression Models: A Case Study for Nigeria. *Economics and Human Biology*, 2(2): 229-44.

Adedoyin, M., and S. Watts. (1989). Child Health and Child Care in Okele: an indigenous area of the city of Ilorin, Nigeria. *Social Science and Medicine*, 29(12): 1333-1341.

Adetunji, J.A. (1995). Infant Mortality and Mother's Education in Ondo State, Nigeria. *Social Science and Medicine*, 40(2): 253-263.

Adetunji, J.A. (2000). Trends in under-5 mortality rates and the HIV/AIDS epidemic. *Bulletin of the World Health Organization* , 78: 1200–1206.

Ahonsi, B.A. (1995). Age variation in the proximate determinants of child mortality in South-west Nigeria. *Journal of Biosocial Science*, 27(1): 19-30.

Anselin, L. (1995). Local indicators of spatial association –LISA. *Geographical Analysis*, 27: 93-115.

Anselin, L. (2003). *GeoDa 0.9 User's Guide.* Spatial Analysis Laboratory Urbana-Champaign, IL: University of Illinois.

Balk D., Pullum T., Storeygard A., Greenwell F. and Neuman M. (2003). *Spatial Analysis of Childhood Mortality in West Africa.* Calverton, Md.: MEASURE DHS+, ORC Macro (DHS geographic studies 1) HQ 766 W31 G46 #1 Macro International. 44p

Banerjee, S., Wall, M.M and Carlin, B. P. (2003). Frailty modeling for spatially correlated survival data with application to infant mortality in Minnesota. *Biostatistics*, 4: 123-142.

Behrman, J.R., and Wolfe, B.L. (1987). How Does Mother's Schooling Affect Family Health Nutrition Medical Care Usage, and Household Sanitation? *Journal of Econometrics*, 36: 195–204.

Berger, U., Fahrmeir, L., Klasen, S. (2002). *Dynamic Modelling of Child Mortality in Developing Countries: Application for Zambia*. SFB 386 Discussion Paper No. 299, University of Munich (available from http://epub.ub.uni-muenchen.de/1677/1/paper_299.pdf).

Bernardo, J.M. and Smith, A.F.M. (2000). *Bayesian Theory*. Chichester: Wiley.

Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* 43, 1-59.

Bicego, G. and Ahmad O. B. (1996). *Infant and child mortality*. Demographic and Health Surveys, Comparative Studies No. 20. Calverton, Maryland: Macro International Inc.

Bolstad W.M. (2004). *Introduction to Bayesian statistics*, 2nd Edition. New York: Wiley.

Brezger, A., Kneib, T. and Lang, S. (2005). *BayesX—Software for Bayesian Inference based on Markov Chain Monte Carlo simulation Techniques*. (Available from: http://www.stat.uni-muenchen.de/~bayesx/).

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information—Theoretic Approach*, 2nd Edition. New York, Springer.

Caldwell, J.C. (1979). Education as a Factor in Mortality Decline: An Examination of Nigeria data. *Population Studies*. 33(2): 395-413.

Caldwell, J.C. and Caldwell, P. (1993). Women's position and child mortality and morbidity in less developed countries. In N. Federici, K.O. Mason and S. Sogner (Eds.), *Women's position and demographic change* (pp. 122-139). New York: Oxford University Press.

Carlin, B.P. and Louis, T.A. (2000). *Bayes and Empirical Bayes methods for data analysis*, 2nd edition. New York, Chapman and Hall.

Chaix, B., Merlo, J., Subramanian, S. V., Lynch, J. and Chauvin, P. (2005). Comparison of a Spatial Perspective with the Multilevel Analytical Approach in Neighborhood Studies: The Case of Mental and Behavioral Disorders due to Psychoactive Substance Use in Malmo, Sweden, 2001. *American Journal of Epidemiology*, 162(2): 171 - 182.

Chromy, J.R. and Abeyasekera, S. (2003). Statistical analysis of survey data. In: *Household Sample Surveys in Developing and Transition Countries.* New York: United Nations Publication ST/ESA/STAT/SER.F/96, 2005, Chapter XIX, 388-417.

Cleland J.G. and van Ginneken, J.K. (1988). Maternal education and child survival in developing countries: the search for pathways of influence. *Social Science and Medicine*, 27(12): 1357-1368.

Cliff A. and Ord J.K. (1981). *Spatial Processes, Models and Applications*. London: Pion.

Congdon, P. (2003). *Applied Bayesian Modeling*. Wiley Series in Probability and Statistics. West Sussex, England: Wiley.

Crook, A., Knorr-Held. L. and Hemingway, H. (2003). Measuring spatial effects in time to event data: a case study using months from angiography to coronary artery bypass graft. *Statistics in Medicine*, 22: 2943-2961.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*, 34:187–220.

Curtis, S. L. (1995). *Assessment of the quality of data used for direct estimation of infant and child mortality in DHS-II Surveys*. Occasional Papers No. 3. Calverton, MD: Macro International Inc.

Curtis, S.L., Diamond, I. and McDonald J.W. (1993). Birth interval and family effects on post neonatal and mortality in Brazil. *Demography*, 30(1): 33-43.

Curtis, S. L. and Hossein, M. (1998). *The Effect of Aridity Zone on Child Nutritional Status*. West Africa Spatial Analysis Prototype Exploratory Analysis. Calverton, Maryland: Macro International Inc.

Curtis S. L. and Steele, F. (1996). Variations in familial neonatal mortality risks in four countries. *Journal of Biosocial Science*, 28: 141-159.

D'Souza, S. and Chen, L.C. (1980). Sex differentials in mortality in Bangladesh. *Population and Development Review*, 6: 257-70.

Das Gupta, M. (1987). Selective discrimination against female children in rural Punjab, India. *Population and Development Review*, 13: 77-100.

Desai, S. and Alva, S. (1998). Maternal Education and Child Health: Is There a Strong Causal Relationship? *Demography*, 35: 71-81.

DFID. (2000). *Nigeria: health briefing paper*. DFID HSRC, London. Available at: http://www.dfidhealthrc.org/shared/publications/Country_health/Nigeria.pdf

Eilers P.H.C. and Marx B.D. (1996), Flexible smoothing using B-splines and penalized likelihood, *Statistical. Science*, 11: 89–121.

ESRI (Environmental Systems Research Institute, Inc) 2002. *ArcView GIS* Version 3.3. Redlands, CA.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling based on Generalized Linear Models*. New York: Springer.

Feyisetan, B.J., Asa, S. and Ebigbola, J.A. (1997). Timing of birth and infant mortality in Nigeria. *Genus* 53 (3–4): 157–181.

Gemperli, A., Vounatsou P., Kleinschmidt I., Bagayoko M., Lengeler C. and Smith T. (2004). Spatial patterns of infant mortality in Mali; the effect of malaria endemicity. *American Journal of Epidemiology*, 159: 64-72.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1:515–534.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995). *Bayesian data analysis*. New York: Chapman and Hall

Goldstein, H. (1995). *Multilevel statistical models*. 2nd Edition. New York: Halstead Press.

Gregson, S., Zhuwau, T., Anderson, R. M., and Chandiwana, S. K. (1999). Apostles and Zionists: The influence of religion on demographic change in rural Zimbabwe. *Population Studies*, 53(2):179–193.

Guo, G. (1993). Use of sibling data to estimate family mortality effects in Guatemala. *Demography*, 30(1): 15-32.

Guo, G. and Rodríguez, G. (1992). Estimating a multivariate proportional hazards model for clustered data using EM algorithm, with an application to child survival in Guatemala. *Journal of American Statistical Association*, 87(420): 969-976.

Hennerfeind, A., Brezger, A. and Fahrmeir, L. (2006). Geo-additive Survival Models. *Journal of American Statistical Association*, 101(475): 1059-1064.

Heywood (1998). *Introduction to Geographical Information Systems.* New York: Addison Wesley Longman.

Hill, K. and Pebley, A. (1989). Child mortality in the developing world. *Population and Development Review*, 15(4): 657-683.

Hill, K., Bicego, G. and Mahy, M. (2001). *Child Mortality in Kenya: An examination of Trends and Determinants from the late 1980s to the mid-1990s*. Hopkins Population Center Working Paper.

Hobcraft, J.N., McDonald, J.W. and Rutstein, S.O. (1984). Socio-economic factors in infant and child mortality: a cross-national comparison. *Population Studies*, 38: 193–223.

Hobcraft, J. N., Mc Donald, J. W. and Rutstein, S. O. (1985). Demographic Determinants of Infant and Early Child Mortality: A Comparative Analysis. *Population Studies*, 39: 363-385.

http://www.childinfo.org (accessed: 17th, July 2005)

Iyun, B.F. (1992). Women's status and Childhood Mortality in two Contrasting Areas in South-western Nigeria: a Preliminary Analysis. *GeoJournal*, 26(1): 43-52.

Kandala, N. B., Magadi, M. A. and Madise, N. J. (2004). *An Investigation of District Spatial Variations of Childhood Diarrhoea and Fever Morbidity in Malawi*. S3RI Applications and Policy Working Papers, A04/14, Southampton University (available from: http://eprints.soton.ac.uk/12463/)

Kandala, N.B., Fahrmeir, L. and Klasen, S. (2002). *Geo-additive models of Childhood Undernutrition in three Sub-Saharan African Countries*. SFB 386 Discussion Paper No. 287, University of Munich (available from http://www.stat.uni-muenchen.de/sfb386/)

Kaplan, E. L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, 53: 457-481.

Klein, J.P. and Moeschberger (1997). *Survival analysis: techniques for censored and truncated data*. Springer.

Kneib, T. (2005). *Geo-additive hazard regression for interval censored survival times.* SFB 386 discussion paper 447, University of Munich (available from http://www.stat.uni-muenchen.de/sfb386/)

Kravdal, O. (2004). Child Mortality in India: the Community-Level Effect of Education. *Population Studies* 58: 177-92.

Kuate-Defo, B. and Diallo, K. (2002). Geography of child mortality clustering within African families. *Health and Place*, 8: 93-117.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13: 183-212.

Lawoyin, T.O. (2001). Risk factors for infant mortality in rural community in Nigeria. *Journal of Royal Society for Public Health*, 121(2): 114–118.

Lee, E.T. (1980). *Statistical methods for survival data analysis.* Lifetime Learning Publications.

Leonard, T. and Hsu, J.S.J. (1999). *Bayesian methods: an analysis for statisticians and interdisciplinary researchers*. Cambridge, New York.

Madise, N. and Diamond, I. (1995). Determinants of infant mortality in Malawi: An analysis to control for death clustering within families. *Journal of Biosocial Science*, 27(1): 95-106.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, 50 (3): 163-70.

Masuy-Stroobant, G. (2002). The determinants of infant mortality: how far are conceptual frameworks really modelled? In: Robert Franck (2002) *Explanatory Power of Models Bridging the Gap between Empirical and Theoretical Research in the Social Sciences*, Kluwer Academic Publishers Boston

Masuy-Stroobant, G. and Gourbin, C. (1995). Infant health and mortality indicators: their accuracy for monitoring the socio-economic development in the Europe of 1994. *European Journal of Population*, 11(1): 63-84.

Miaou, S.-P., Song, J., and Mallick, B. (2003). Roadway Traffic Crash Mapping: A Space-Time Modeling Approach. *Journal of Transportation and Statistics*, 6: 33–58.

Montgomery, M. R. and Cohen, B. (1998). *From death to birth: mortality decline and reproductive change*. Washington, D.C.: National Academic Press.

Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37:17-23.

Mosley, H. W. and Chen, L. C. (1984). An analytical framework for the study of child survival in developing countries. In *Child survival: Strategies for research,* ed. W. G. Mosley and L.C. Chen. New York: Population Council. 25.44.

National Population Commission [Nigeria] (1991). *Nigeria Demographic and Health Survey 1990*. Calverton, Maryland: National Population Commission and ORC Macro.

National Population Commission [Nigeria]. (1998). *1991 population census of the Federal Republic of Nigeria: Analytical report at the national level.* National Population Commission, Lagos [Nigeria].

National Population Commission [Nigeria] (2000). *Nigeria Demographic and Health Survey 1999*. Calverton, Maryland: National Population Commission and ORC Macro

National Population Commission [Nigeria] and ORC Macro. (2004). *Nigeria Demographic and Health Survey 2003.* Calverton, Maryland: National Population Commission and ORC Macro.

Ogunjuyigbe, P.O. (2004). Under-five mortality in Nigeria: perception and attitudes of the Yorubas towards the existence of "Abiku". *Demographic Research*, 11: 43-56.

Ord J.K. and Getis A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis*, 27: 286-306.

Owa J.A. and Osinaike A.I. (1998). Neonatal Morbidity and Mortality in Nigeria. *Indian Journal of Pediatrics*, 65: 441-449.

Palloni, A. and Millman, S. (1986). Effects of inter-birth intervals and breastfeeding on infant and early childhood mortality. *Population Studies,* 40: 215–236.

Peterson, C., Yusof K., DaVanzo J. and Habicht J.P. (1986). *Why were Infant and Child Mortality Rates Highest in the Poorest States of Peninsular Malaysia, 1941-75?* A Rand Note. Santa Monica, CA: Rand.

Peto R. and Peto J. (1972). Asymptotically efficient rank invariant procedures. *Journal of the Royal Statistical Society, Series A,* 135: 185–207

POLICY Project. (2002). *Child Survival in Nigeria; Situation, Response, and Prospects: Key Issues*. POLICY Project, USA.

Preston, S. H. (1978). *The effect of infant and child mortality on fertility*. New York: 16 Academic Press.

Root G. (1997). Population density and spatial differentials in child mortality in Zimbabwe. *Social Science and Medicine,* 44(3): 413–421.

Rutstein, S. (2000). Factors Associated with Trends in Infant and Child Mortality in Developing Countries During the 1990's. *Bulletin of the World Health Organization*, 78(10): 1256-1270.

Ruzicka, L., (1989). Problems and issues in the study of mortality differentials. In: Ruzicka, L., Wunsch, G. and Kane, P., Editors. *Differential Mortality: Methodological Issues and Biosocial Factors*, Clarendon Press, Oxford.

SAS Institute, Inc. (2000–2004). SAS version 9.1.3 software. SAS Institute, Inc. Cary, NC.

Sastry, N. (1996). Community characteristics, individual and household attributes, and child survival in Brazil. *Demography,* 33(2): 211-229.

Sastry, N. (1997a). A nested frailty model for survival data, with an application to the study of child survival in northeast Brazil. *Journal of the American Statistical Association*, 92: 426-435.

Sastry, N. (1997b). Family-level Clustering of Childhood Mortality Risk in Northeast Brazil. *Population Studies*, 51: 245-261.

Sastry, N. (1997c). What Explains Rural-Urban Differentials In Child Mortality In Brazil? *Social Science and Medicine,* 44:989-1002.

Schultz, T.P., (1984). Studying the Impact of Household Economic and Community Variables on Child Mortality. *Population and Development Review* 10 (Suppl.): 215-235.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society, Series B*, 64: 583–640.

Tulasidhar, V.B. (1993). Maternal Education, Female Labour Force Participation and Child Mortality: Evidence from the Indian Census. *Health Transition Review*, 3: 177-90.

UNICEF. (2002). *Nigeria: Information by Country*. Available from: http://www.unicef.org/infobycountry/nigeria_statistics.html (accessed 29 June 2004)

UNICEF. (2005). *The State of World's Children,* UNICEF, New York.

United Nations (2000). *Millennium Declaration.* New York: United Nations.

United Nations (2005). *Progress Towards the Millennium Development Goals 1990-2005.* New York: United Nations Department of Economic and Social Affairs Publication 2005.

Vaida F., Ghosh P. and Liu L. (2008): Mixed-Effects Models for Longitudinal HIV Virologic and Immunologic Data, In: Khattree R, Naik DN, editors. *Computational Methods in Biomedical Research,* Chapman and Hall/CRC Biostatistics Series.

Vaupel, J.W., Manton, K. and Stallard, E. (1979). Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality. *Demography*, 16(3): 439-454.

Venkatacharya, K. (1985). An Approach to the Study of Socio-Biological Determinants of Child Morbidity and Mortality. IUSSP Conference, Florence, Italy.

World Bank (2004). *World Development Indicators*. Washington D.C.: The World Bank.

World Development Indicators database, April 2005 (accessed: 17th, July 2005).

World Health Organization (2005). *World Health Report,* WHO HQ.

APPENDICES

APPENDIX A: list of variables used in the analysis

Level	Variable	**Description**
Child	Gender of child	*Male*
		Female
	Birth order	*First to third birth*
		Fourth or higher birth
	Preceding Birth Interval	*No older siblings or > 36 months*
		Less than 24 Months
		24 to 35 months
	Succeeding Birth Interval	*No younger sibling or > 36 months*
		Less than 24 Months
		24 to 35 months
	Size at Birth	Small/very small
		Average or larger
	Mothers age at birth	<18 Years
		18-34 Years
		35 an older
	Place of delivery	*Homes/Others/Missing*
		Health Facility
	Source of prenatal care	Skilled Birth Attendant
		Traditional Birth Attendant/Other/None
	Birth Assistance	Trained Medical Personnel
		Traditional Birth Attendant/Other/None
	Source of antenatal care	*Homes/Other/None*
		Health Facility
	Long Labour at birth	*No*
		Yes
	Excessive bleeding at birth	*No*
		Yes
	Higher fever at birth	*No*
		Yes
	Convulsions at birth	*No*
		Yes
	Any problem at birth?	*No problem*
		At least one problem

APPENDIX A: Continued

Level	Variable	**Description**
Mother	Mothers Highest Educational Level	*No education*
		Primary
		Secondary plus
	Mothers Occupation	No Work
		White Collar Job
		Agric and Others
	Type of Marital Union	*Monogamy/Never married*
		Polygamy
	Ethnicity	Hausa
		Igbo
		Yoruba
		Fulani
		Others
	Religion	Christian
		Muslim
		Traditionalist or Others/missing
	Media Exposure	No Media Exposure
		Exposed to at least one source
	Decision making index	No Decision
		At least one decision
	Problem getting medical help	No problem
		At least one problem
	Partners Occupation	No Work/No Partner
		White Collar Job
		Agric/Other
	Partners Highest Educational Level	*No education/Not married/Missing*
		Primary
		Secondary plus

Level	Variable	**Description**
Household	Source of drinking water	Piped or Tap
		Well or Surface
		Others
	Type of toilet facility	Flush
		Pit latrine
		No facility or Others
	Flooring materials	Natural and Rudimentary
		Finished
	Type of Cooking Fuel	Cleaner Fuels
		High Pollution Fuels
	Household Wealth Status	*Poorest*
		Poorer
		Middle
		Richer
		Richest

APPENDIX A: Continued

Level	Variable	**Dependent**
Community	Region	North Central North East *North West* South East South South South West
	Type of Place of residence	Urban *Rural*
	Malaria Prevalence	Low (0-35% reference category) *Medium (36–60%)* High Endemicity (>60%)
	Population Density	<100 per sq km *100+ per sq km*
	Distance to roads	< 1 km *1+ km*
	% with access to clean water in community	# of children in Households with Tap water / Total # of Children
	% with access to hygienic toilet in community	# of children in Households with Flush Toilet / Total # of Children
	% with access to finished floor in community	# of children in Households with Finished floor / Total # of Children
	% with access to clean cooking fuel in community	# of children in Households with cleaner fuel / Total # of Children
	% with access to electricity in community	# of children in Households Electricity / Total # of Children
	% of births delivered in medical facility	# of children delivered in medical facility / Total # of Children
	% of births with postnatal care	# of children in with postnatal care / Total # of Children
	% of births with antenatal care	# of children with antenatal care / Total # of Children
	% of births delivered by a skilled attendant	# of children delivered by a skilled attendant / Total # of Children
	% of mothers who had at least one tetanus injection	# of children whose mothers who had at least one tetanus injection / Total # of Children
	% of children 12-23 months fully vaccinated	# of children in Households with Tap water / Total # of Children
	% with risky birth interval	# of children in Households with Tap water / Total # of Children
	% born to too young or too old women	# of children born to mothers <18 and > 35 Years / Total # of Children
	% of children with high birth order	# of children with birth order greater than 3 / Total # of Children
	% with at least secondary education	# of children born to mother with at least secondary education / Total # of Children
	% White Collar job	# of children born to mothers with while collar jobs / Total # of Children
	% of single women or monogamous unions	# of children born to mothers in single and monogamous unions / Total # of Children
	% with access to at least one media type	# of children born to mothers who have access to radio, TV or newspaper/ Total # of Children