Estimation of Mean Transition Time using Markov Model and Comparison of risk factors of malnutrition using Markov Regression to Generalized Estimating Equations and Random Effects Model in a Longitudinal Study

Thesis submitted to
The Tamil Nadu Dr. M.G.R. Medical University

For the degree of
Doctor of Philosophy

by
Visalakshi J

Department of Biostatistics
Christian Medical College

March, 2012

Table of Contents

INTRODUCTION

1. INTRODUCTION

1.1 Malnutrition:

The term 'Malnutrition' is used to refer to a number of diseases, each with a specific cause related to one or more nutrients (for example, protein, iodine or calcium) and each characterized by cellular imbalance between the supply of nutrients and energy on the one hand, and the body's demand for them to ensure growth maintenance, and specific functions, on the other. Malnutrition is an important indicator of child health. It is now recognized that 6.6 million out of 12.2 million deaths among children under-five – or 54% of young child mortality in developing countries – is associated with malnutrition (1). In addition to the human suffering, the loss in human potential translates into social and economic costs that no country can afford (1). Protein-Energy malnutrition (PEM) is one of the most serious health problems in resource poor countries where PEM accounts for more than 35% deaths of children aged less than five years and 11% of the total burden of disease (2). WHO estimates that 175 million children in the developing world are malnourished as indicated by low weight for age and 230 million are stunted as indicated by height for age (3). It is estimated that more than half of the young children in south Asia suffer from protein – energy malnutrition, which is about five times the prevalence in the Western hemisphere, at least three times the prevalence in the Middle East and more than twice that of east Asia (1). Protein-energy malnutrition (PEM) is one of the most serious health problems in Bangladesh (4). India has the highest percentages of undernourished children in the world (1). In 1990, only 53 developing countries had reliable national data on the prevalence of underweight in young children; however, by 1995, 97 countries had such data, and 95 countries also had data on stunting and wasting where, stunting refers to chronic malnutrition (1) (low height for age indices) and wasting is acute malnutrition as indicated by the child's weight for age indices (1).

1.2 Risk Factors associated with Malnutrition:

The etiology of childhood malnutrition is not very simple as it involves interactions of the biological, cultural and socioeconomic factors (4). In most South Asian countries, poverty, high population density, low status of women, poor antenatal care, high rates of low birth weight, unfavourable child caring practices, and poor access to child healthcare are the underlying contributors to the development of PEM (5). In some regions, such as sub-Saharan Africa and south Asia, stagnation of nutritional improvement combined with a rapid rise in population has resulted in an actual increase in the total number of malnourished children (1). Several factors are responsible for high undernutrition in India. Some of these are related to poverty and poor access to nutrition and health care and could be remedied within a short period. India has the dubious distinction of having a very high prevalence of low birth weight (LBW). In any community, under-five children are one of the most vulnerable groups for nutritional deficiencies, owing to many factors ranging from Low Birth Weight to maternal ill health to socio-economic and environmental factors (6). Estimates based on available data from institutional deliveries and smaller community – based studies suggested that nearly one-third of all Indian infants weigh less than 2.5 kg at birth. Studies carried out by Ghosh and co-workers in the seventies (7) have shown that LBW children have a low trajectory for growth in infancy and childhood. Internally displaced people deserve special care, as they constitute a large proportion World Wide. These groups of people are vulnerable to many health crises as they are triggered by many risk factors such as lack of shelter, poor personal hygiene and poor environmental sanitation, outbreaks of diseases, war, drought, famine and shortage of food. National surveys indicate that a third of the children from high income group who have not experienced any deprivations are undernourished. The high undernutrition rates among children appears to be mainly due to high low birthweight rates, poor infant and young child feeding and caring practices. Malnutrition is an important indicator of child health. A significant contributing factor to infant and child mortality, poor nutritional status during childhood also has implications for adult economic achievement and health (8). A study from South India suggested that the risk factors for severe malnutrition were found to be low mother's education, low family income, more among boys, use of firewood or coal for cooking and defecation within premises (9).

1.3 Generalized Estimating Equations:

Generalized Estimating equations (GEE) are extensions of linear regression analysis which are applied when there are repeated observations (responses) obtained from the same individual. As the response from each individual is obtained over time, these observations are not independent and hence the usual regression analysis is not applicable. There is considerable correlation as the responses are obtained from the same individual. Hence the GEE method accounts the correlation as a nuisance parameter and thereby included in the model as a covariate. The correlation structure is decided a prior which may be a difficult task especially when the outcome is a categorical variable. However, GEE is still robust for a wrong choice of correlation.

1.4 Random Effects Model or Multilevel Modeling (MLM):

Multilevel modeling (MLM) is being applied extensively over the past 10 years. In many situations, data is usually collected in a hierarchical manner. For example, data may be collected from individuals from the same household. Hence it is expected that there could be some correlation in the responses if they are obtained from the individuals from the same household and thereby the observations are not independent failing to the usual assumptions of linear regression analysis. Multilevel analysis accounts for the dependency and the correlation at each level. Multilevel analysis can also be applied if repeated observations are taken from the same individual. In such a situation, individuals are cluster and each unit of repeated measurement within that individual are the next level.

1.5 Markov Chain:

Mathematical models represent real world problems or phenomena by formal system (10). They offer several advantages over empirical studies, as well as a number of disadvantages. Among the advantages are the identification of the variables in a quantitative problem that have the most impact on the system, and in particular the ability to ask "What if?" questions of the model.

The Markov chain model has several attractive features that stem from the central assumption of the model: the probability of arriving in stat j, given that the process is in state i at time T, (known as a transition probability), is determined only by i and j. This strong assumption yield a homogeneous Markov Chain. Chronic diseases such as neurologic, cardiovascular and rheumatic disorders seem to obey the Markov assumption. The Markov model uses data on the probability of transition from one clinical status or disease state to another disease state to derive under certain assumptions, the duration of time spent in each disease state. It is possible to determine the future probability that a patient will be in each disease state from information on the current disease state.

Silverstein et al. demonstrated that the natural history of systemic lupus erythematosus in fact can be modeled as a homogeneous chain, and thus validate an assumption made by earlier users of the Markov model.

Multistate models can be used to describe changes in patient's health condition over time. In a study of chronic illness, these models classify patients into one of the finite number of distinct states at any given point during his or her follow-up. These states represent various health conditions, and the transition times correspond to the times at which these changes occur (11).

Another feature of Markov models is the natural division of a population into cohorts of different health states. With this allocation, and measure of QOL superimposed on the model, one can calculate state specific measures of health status or utility. These results can be incorporated into clinical decision analyses, using the Markov chain as a utility structure. Such modeling is particularly appropriate for cost-effectiveness analysis, wherein resources and health benefits are accrued incrementally, rather than summarized at the end of the model. The ideal population upon which to perform stochastic modeling is a true inception cohort, but such investigations are now just underway.

1.5.1 Mean first passage and Sojourn time:

Mean First Passage Time:

The Markov model uses data on the probability of transition from one clinical status or disease state to another disease state to derive under certain assumptions, the duration of time spent in each disease state. Since clinical data are often available on the likelihood of a patient's disease becoming more active or going into remission, it is often possible to calculate the duration of time in each disease state, and a life expectancy under the assumption of Markov Model, when one of the state (final state) is an absorbing state, such as death, where there is no possibility to transit from there. However, in situation such as malnutrition there is no such absorbing state. In such situation besides estimating the duration of time spent in each state, it is useful, to estimate the mean first passage time for a subject to transit from one state to another first time, for example, the mean first time for subject to transit from severe malnutrition state to normal state. This will help physicians to plan interventions effectively. Estimation of such time will help physicians and the epidemiologist to study the impact of staying in that state, for example, in severe malnutrition for some time and its impact on mental development in children.

Sojourn time:

Using the transition probabilities, it is imperative to estimate the expected duration stay in each state of disease. For example, how long a child is expected to be in severe malnutrition or in normal state. This can be done for both Markov models with absorbing and non-absorbing state. This will help the therapists to plan treatment options for specific duration to reverse the disease progression.

1.6 Markov Regression:

In usual regression the hypothesized variables are associated with the outcome after adjusting for other known risk factors and confounders. In Markov regression the concept of current disease state depends on the previous state of the disease. That is, when we model the duration of disease and the probability of transition from one state to another, the conditional probability concept or the hazard concept is incorporated in the regression analysis. Thus in chronic disease epidemiology, if the current state of the disease depends on the previous state, then it is appropriate to consider Markov regression. This also adjusts for other risk factors, confounders, besides the previous state information.

Several alternate approaches are available for examining longitudinal data on health states of a patient. Deciding which method to use often depends on the questions that need to be answered (11). A correlated ordinal model may suffice if one is interested in estimating the relation between the probability of being in a particular health state and time since diagnosis. If the objective is to predict time to event probabilities, a survival model can be implemented in which a time-dependent covariate may be used to express events that may affect a person's health condition. Multistate models are particularly useful for describing the complexities of a disease process (12). They are more applicable when interest lies in estimating the instantaneous rate of transition between various states, estimating the probability of transition from one state to another within a specific time period, or estimating the average period of single stay in a state (mean sojourn time) and also risk factors.

When patients are followed continuously and transition times are subject only to right censoring, a wide range of multistate modeling and estimation strategies are available (13). In realty, however, it is often not possible to observe patients continuously throughout the course of disease, especially when patients are assessed by a physician only at periodic clinic or home visits. In these cases, the exact times of state-to state transitions other than death are interval-censored. The transition is only known to have occurred within a bounded time interval, usually assessments. Under intermittent observation, the data available for an individual consist of the assessment times and the states that are occupied at each of these times. Data emerging from this observation scheme pose a variety of estimation on transition times and the fact that the number and timing of assessments may vary dramatically across patients. In such occasions, the Markov regression would be appropriate.

AIMS AND OBJECTIVES

2. AIMS AND OBJECTIVES

The main aim of the thesis is to find if malnutrition was associated with any of household factors from which the child was taken for the study such as type of fuel used for cooking, education of the mother and father, sex of the child, etc.

The objectives of the thesis are:

1. To estimate the first mean passage time which indicates the average time spent by a child to move from one state to another
2. To find risk factors of using GEE and Random effects model
3. To find risk factors of protein energy malnutrition using Markov regression with transition probabilities
4. To find the risk factors using Markov Regression with transition intensity rates
5. To compare the results obtained from GEE and Markov regression models using transition probabilities and transition intensities and to evaluate the coverage probabilities of the 95% CIs obtained using the above methods.

REVIEW OF LITERATURE

3. REVIEW OF LITERATURE

3.1 Definition of Malnutrition:

World Health Organization (WHO) defined the term 'malnutrition' as a condition that refers to a number of diseases each with a specific cause related to the cellular imbalance of one or more nutrients like protein, iodine and/or calcium and the body's demand to ensure growth, maintenance and proper functioning. Malnutrition can disable, maim or even kill. In the past few years, there is an economic growth and seems to be an improvement in food supplies, health conditions, availability of educational resources and social services but malnutrition seems to persisting virtually in all countries of the world (1).

3.1.1 Indices for measuring Malnutrition:

Growth assessment best defines the nutritional status of children. There are various anthropometric indices that are used to assess child's growth status. But the most widely used anthropometric index to determine nutritional status is Z scores (14). The standard indices of physical growth that describes the nutritional status of children is presented as height-for-age, weight-for-height, weight-for-age and BMI. Each of the above nutritional status is expressed in standard deviations from the median. Each index provides different information about growth and body composition which is used to assess the nutritional status. The height-for-age index is an indicator of linear growth retardation and cumulative growth deficits. Children whose height-for-age Z scores is below minus two standard deviations are considered short for their age (stunted) and are chronically malnourished. Children below minus three standard deviations from the median are considered to be severely stunted. The weight-for-height index measures body mass in relation to the body length and describes the current nutritional status. Children below -3SD are considered severely thin (wasted) and are acutely malnourished. Weight-for-age is a composite index of height-for-age and weight-for-height. It takes into account both acute and chronic malnutrition. Children below -3SD are considered to be severely underweight. An article from Chile had suggested BMI Z scores as an index of underweight. The classification using BMI Z scores have been used recently in clinical settings. Children whose BMI Z score was less than -3SD were considered as severely underweight (15).

3.2 Prevalence of Malnutrition:

In 1990, the WHO fact sheet reported that only 53 developing countries had reliable national data on the prevalence of underweight in young children; by 1995, 97 countries had such data, and 95 countries also had data on stunting and wasting. It is estimated that more than half of the young children in south Asia suffer from protein energy malnutrition, which is about five times the prevalence in the Western hemisphere, at least three times the prevalence in the Middle East and more than twice that of east Asia. Estimated for sub-Saharan Africa indicate that the prevalence is approximately 30%. In some regions, such as sub-Saharan Africa and south Asia, stagnation of nutritional improvement combined with a rapid rise in population has resulted in an actual increase in the total number of malnourished children. Currently, over one thirds of the world's malnourished children live in Asia (especially south Asia), followed by Africa and Latin America (1).

A study in 2009 reported that malnutrition was a contributing factor to infant mortality and also has implications for adult economic achievement and health. The article also reported that Pakistan has half of its children aged five years or less are stunted, over a third are underweight and a quarter of all births are low weight. These high levels of malnutrition contribute to about half of the 740,000 children deaths that occur every year in Pakistan (16).

Protein-Energy malnutrition (PEM) is one of the most serious health problems in Bangladesh too where PEM accounts for 35% of deaths of children aged less than five years and total burden of disease. There were earlier reports that were reported in the same article that, severely underweight children aged 6-59 months had more than eight fold increased mortality (4). Another study reported that India has the highest percentages of undernourished children in the world (17).

Malnutrition is also a significant problem in older children. It was indicated that there is very little known about the state of nutrition but some studies conducted in 1980s indicate that malnutrition is a significant problem with prevalence ranging from 47-70% in male school children in rural Pakistan. Within 7-10 year age, 36% were underweight, 39% stunted and 20% wasted. The prevalence of underweight in children between 5-7 years was 26% underweight, 32% stunted and 8% wasted (18).

The national nutrition survey in Bangladesh found that 29% of under-five children were moderately underweight and 12% were severely underweight according to weight-for-age Zscores. The prevalence of underweight decreased over the follow-up from the years 1987 to 2002 in a national school-based annual population surveys in 6 year old children (15).

The prevalence of under nutrition under 3 years of age in India as reported using the NFHS 2 data found that 45% of children were stunted, 47% underweight and 16% were wasted (19). In another study in children between 5-7 years of age in south India reported that 8.2% of children were severely malnourished as classified using weight-for-age percentiles. The prevalence of moderate malnutrition was 30% as reported in the study (9).

3.3 Risk Factors:

Most of the studies reviewed addressed many risk factors for the prevalence of malnutrition among children. A lot of studies reported that malnutrition was mainly a severe problem in resource poor developing countries.

A study done in Bangladesh reported that most South Asian countries, poverty, high population density, low status of women, poor antenatal care, high rates of low birth weight, unfavorable child caring practices, and poor access to child healthcare are the underlying contributors to the development of PEM. The study used children from 6 to 24 months who reported to Dhaka Hospital of the International Center for Diarrheal Disease. The results of the study showed that major risk factors were related to parental education, employment, income, child birth order and early feeding practices. The study showed that there were no significant differences between the age groups, area of residence and year of enrollment. Results showed that weight-for-length Z score was -2.71 (sd = 0.76) for those children who had <-3 SD weight-for-age Z scores as compared to those children in the >2.5 WAZ scores which was -0.55 (sd = 1.12). The results also showed that number of children in the family was also a contributing factor for less WAZ score. It was also found that severely-underweight children were more likely to have undernourished mothers (BMI < 18.5 kg/m^2). The other factor that influenced the risk of underweight was the 'education level' of mother. Mothers who completed <5 years of education were more likely to have undernourished children. Children who had shorter duration of predominant breastfeeding had 2.3 times odds of having <-3 WAZ scores as compared to children who had >2.5 WAZ scores. Children who came from families which had monthly income of <5000 had nearly 3 times the odds of having <-3 WAZ score. The other risk factors that contributed to higher risk of having <-3 Z scores were higher birth order, occupation of father (4).

A stratified multistage random sampling study in Uganda was done to assess the dietary and environmental factors influencing stunting and other poor nutritional status of children <30 months. The study consisted of 261 children where 70% were from rural areas and remaining 30% were from semi-urban. The study findings were that older children had higher incidence of stunting than younger children. The study showed that current breastfeeding and age of weaning were not associated with incidence of stunting and underweight. Children who had never consumed milk showed a higher incidence of underweight, children who were fed foods on low energy density of stunting and not underweight. The other findings that were included in the study were that age of the mother, parity, occupation and number of children had no influence on stunting and underweight. The study showed that better education of mother had less stunted children. None of the infants whose mothers' education was primary were stunted. Children from rural areas were more underweight than those from urban and low socio-economic families had more underweight children in urban areas. Unprotected source of water was also one of the important factor that contributed to more underweight. Type of fuel used for cooking was also another factor that contributed to the underweight among children. Families who used charcoal or paraffin had lower under weighed children as compared to those families that used firewood for cooking. The adjusted analysis from the same study showed families with very low economic status had 2.6 times the odds of children being underweight as compared to families that were mid-upper status. Families that used protected source of water were 21% less likely to have underweight children as compared to those families that did not use protected source of water to drink (20). In Sudan, (21) a total of 327 children from 200 families were enrolled into a study. According to WHO criteria, the prevalence of malnutrition was 56.1% with 30% mild, 13.1 % moderate, and 12.8% severely nourished. The study did not show that age of child, sex, lack of immunization and lack of breast feeding were not very influencing factors of malnutrition. However, another study from Kenya suggested that lack of immunization was a contributing factor to malnutrition among children. The prevalence of Kwashiorkor and marasmus in Uganda seemed to be increasing and led to 40% of deaths to malnutrition alone. Hence a study was conducted to compare the feeding practices, health facility utilization and socio-demographic factors of mothers or caretakers of malnourished children with those of well nourished. This study concluded that some socio-economic factors were associated with severe PEM. Some of those risk factors are early cessation of breastfeeding, failure to complete the primary course of immunization and coming from economically disadvantaged household, without any livestock and living in mud walled dwellings. However, type of occupation and education level of the mothers did not seem to have any significant influence on nutrition status. The other important finding of this study was that children from urban setting were associated with severe malnutrition and in fact all children from urban area were mostly from slum areas in Kampala.

Another study that was conducted in Brazil to find the risk factors for protein energy malnutrition in pre-school children found that birth weight of the child, presence of upper respiratory infections, gravida and parity were the most important to be considered. There were 233 children involved in the study. The Z scores for weight-for-age and weight-for-height were calculated from the anthropometric measurements of the infants. The children included in the study were <72 months of age. The children were classified malnourished if the weight-for-age and height-for-age Z score was less than -2 SD, and weight-for-height Z score less than -1 SD. The prevalence of malnutrition was found using three different indices which were Gomez, Waterlow and OMs. The prevalence of 2^{nd} degree malnutrition according to Gomez was 3.1% whereas, according to Waterlow and OMS the prevalence was 36.6% and 17.6% respectively.

The significant risk factors that were associated with weight-for-age were birth weight, number of pregnancies, birth order, body mass index of the mother and maternal weight. The significant factors associated with weight-for-height were birth weight, age of mother at birth of the child, The birth weight had shown a relative risk (RR) of 5.7 with 95% confidence interval to be 2.1-15.2. The factors associated with height-for-age were also birth weight, age of the mother at birth, maternal height, Body mass index of the mother, birthplace of mother and father. This study had birth weight as a factor that was important for weight-for-age, weight-for-height and height-for-age indices (22).

A case control study from Bangladesh was done to see the characteristics of children aged 6-24 months with or without severe underweight of those who reported to the Dhaka Hospital. There were 507 children with weight-for-age Z score (WAZ) <-3 and 500 children from the same communities with WAZ > -2.5. The study results were presented as parental or family factors that may be associated with the presence of severe underweight and other factors that are pertaining to the child that may be associated with the presence of severe underweight. The family factors that were associated were age of the mother, weight of the mother, height of the mother, body mass index of the mother, education of the mother (in years), education of the father, family income, number of children under five in the family and total number of children in the family. From the other factors, birth order of >=3 had an odds of 1.6 times of having WAZ <-3 as compared to birth order <3. If there was no predominant breastfeed for 4 months then there was 2.7 times the odds of having WAZ <-3 (underweight) as compared to a child who had breastfeed for complete 4 months. If BCG vaccination was not given, then that child had 4.6 times the odds of being underweight as compared to child who had BCG vaccine. The other important factors that were significant in the 'child factor' was using unsanitary latrine. These above factors were unadjusted factors. The adjusted factors that contributed for underweight were teen aged mother, education of the mother, predominant breastfeeding for 4 months, education of the father, monthly family income, undernourished mother, shorter mother and father's job category. If the mother was a teen aged person then there was 3 times the odds of having underweight child. If the mother was undernourished, then there was nearly 4 times the odds of having underweight baby. If the mother was illiterate or had less than five years of education, then there was nearly 3 times the odds of having underweight baby (4).

A cohort study was conducted was in south India in the year 1982 had also reported on the risk factors of malnutrition. This study had reported the factors associated with underweight at baseline measurement. The anthropometric data was collected for children between the ages 5 – 7 years for every six months for seven times after baseline. The children were graded as mild (70-80%), moderate (60-70%) or severely (<60%) malnourished based on weight-for-age percentiles. There were 2496 children at baseline. The overall prevalence of moderate malnutrition was 30% and severe was nearly 8%. The boys had higher prevalence of severe malnutrition. The adjusted ordinal logistic regression analysis results found that boys had 1.3 times the odds of having severe malnutrition as compare to girls. If the yearly family income was <=2000 rupees then there was 1.7 times the odds of having malnutrition as compared to family income being >8000 rupees per year. Illiterate or just literate mothers had 1.3 times the odds of having malnourished children as compared to mothers who had secondary or college education. The mothers who had primary or middle school education also had 1.5 times the odds of having malnourished children as compared to mothers who had secondary or college education. If the families used dung or fire wood for cooking, then there was 1.4 times the odds of having moderate or severe malnutrition as compared to families that used gas or kerosene for cooking. If the defecation was within the premises then there was more likely that the child had moderate or severe malnutrition as compared to defecation done in open fields. The other interesting finding was type of roof. If the type of roof was thatched, then there was a high odds of having moderate or severe malnourished child as compared to houses that had RCC or pukka type roofs (9).

3.4 Generalized Estimating Equations:

Need to perform Generalized Estimating Equations (GEE):

The generalized estimating equations (GEE) method, an extension of the quasi-likelihood approach, is being increasingly used to analyze longitudinal and other correlated data, especially when they are binary or in the form of counts consist of the age- and sex-standardized heights and data on the covariates gender and socioeconomic status) of 144 children in a sample of 54 randomly selected households in Mexico. The results presented as Odds ratio (OR) were compared using the logistic regression analysis. The result using the logistic regression analysis was found to be 9 whereas after adjusting for the correlation using GEE it was found to be 5.4 (23).

A study was done to present that if a longitudinal data was modeled using regression techniques that ignore correlation biased estimates of the regression parameter variances occur. This was illustrated using the childhood health intervention in Brazil which showed that standard errors differed substantially by about 50% on an average for the two models where the two models were logistic regression and GEE that incorporated within correlation. The “months” variable was less associated in the correlated model than the naïve model. The naïve model overestimated the standard errors (24).

A study explains the structure of GEE with count responses. The main challenge mainly in a longitudinal data is when data are correlated within subject such as that provided in longitudinal studies and also in which data are clustered within subgroups. Failure to incorporate correlation of responses can lead to incorrect estimation of regression model parameters especially when the correlations are very large. This incorrect estimation has been demonstrated using a study that collected data from a laboratory that involved assembling Lego objects over five consecutive sessions. The responses are not normally distributed because they consist of count of the number of trips out of the room. The other variables that were correlated to the outcome were the object that needs to be assembled, the size of the object. The data was analyzed using Poisson distribution with independent correlation structure, Poisson distribution with unstructured correlation and Poisson distribution with one dependent autoregressive correlation. (25).

Working Correlation structures and its bounds:

In a longitudinal data analysis using GEE, the variance is considered as nuisance parameter. In some situations, it is important to understand the structure of the variability the same as understanding the mean structure. Moreover, treating the variance structure as a nuisance parameter can lead to misleading conclusions (26).

A study reported the complication in using GEE especially for discrete outcomes. In GEE analysis, the dependence is modeled using the working correlation matrix that is estimated using the method of moments. An unbiased estimating equation that is optimal under some conditions is used to estimate the regression parameters. The process is iterated between estimation of the regression and the working correlation parameters until convergence. However, one has to be cautious on the bounds of the correlation parameters imposed by the nature of the random variables and inferences drawn from such bounds may be unreliable. As an illustration, there was a study to test the efficacy of self-help relapse-prevention booklets for smokers who had already achieved initial abstinence at baseline. The authors analyzed the data using GEE with logit link function with smoking status as the repeated outcome, group and time as the model factors. It was noted that the correlation bounds were outside the estimated correlation. Hence the recommended method was to select the correlation estimate within the correlation bounds (27).

Generalized linear model analyses of repeated measurements typically rely on simplifying mathematical models of the error covariance structure for testing the significance of differences in patterns of change across time. The robustness of the tests of significance depends, not only on the degree of agreement between the specified mathematical model and the actual population data structure, but also on the precision and robustness of the computational criteria for fitting the specified covariance structure to the data. GEE solutions utilizing the robust empirical sandwich estimator for modeling of the error structure were compared with general linear mixed model (GLMM) solutions that utilized the commonly employed restricted maximum likelihood (REML) procedure. Under the conditions considered, the GEE and GLMM procedures were identical in assuming that the data are normally distributed and that the variance-covariance structure of the data is the one specified by the user. The question addressed in this article concerns relative sensitivity of tests of significance for treatment effects to varying degrees of misspecification of the error covariance structure model when fitted by the alternative procedures. Simulated data that were subjected to Monte Carlo evaluation of actual Type I error and power of tests of the equal slopes hypothesis conformed to assumptions of ordinary linear model ANOVA for repeated measures except for autoregressive covariance structures and missing data due to dropouts. The actual within-groups correlation structures of the simulated repeated measurements ranged from AR(1) to compound symmetry in graded steps, whereas the GEE and GLMM formulations restricted the respective error structure models to be either AR(1), compound symmetry (CS), or unstructured (UN). The GEE-based tests utilizing empirical sandwich estimator criteria were documented to be relatively insensitive to misspecification of the covariance structure models, whereas GLMM tests which relied on restricted maximum likelihood (REML) were highly sensitive to relatively modest misspecification of the error correlation structure even though normality, variance homogeneity, and linearity were not an issue in the simulated data. Goodness-of-fit statistics were of little utility in identifying cases in which relatively minor misspecification of the GLMM error structure model resulted in inadequate alpha protection for tests of the equal slopes hypothesis. Both GEE and GLMM formulations that relied on unstructured (UN) error model specification produced non-conservative results regardless of the actual correlation structure of the repeated measurements. A random coefficients model produced robust tests with competitive power across all conditions examined (28).

Some authors have argued that Chaganty and Joe (2004, 2006) have argued that the GEE correlation structures are not correlations at all, but rather weighted matrices. Their claim is based on the supposition that the range of correlations for multivariable binary distributions – i.e., Bernoulli distributions – are based on the marginal means, which they believe preclude the working correlation from being the true correlation of the data (29, 30).

A study was conducted that compared the GEE and random coefficient analysis. These two techniques are the most commonly used techniques that adjust for the correlations between the responses when an individual is measured repeatedly over time. The data used for the comparison was from the Amsterdam Growth and Health study investigating the longitudinal relationship between lifestyle and health in adolescence and young adulthood. There were six measurements on 147 observations. The main hypothesis was to find the relationship between serum cholesterol levels and physical fitness at baseline, body fatness and smoking behavior classified as smoking or non-smoking and gender. The serum cholesterol levels were expressed in mmol/liter or categorized into upper and lower tertiles respectively. The results were compared using the continuous outcome and binary outcome between GEE and random coefficient model. The results from the study was that the GEE and random coefficient model for continuous outcome was similar but there was a difference in the standard errors when dichotomous outcome was considered for analysis (31).

Choice of Correlation Structure:

A study reported that although GEE is becoming popular in handling correlated response data in longitudinal studies, due to an attractive property that one can use some working correlation structure that may be wrong but the resulting regression coefficient estimate is still consistent and asymptotically normal. One such convenient choice is independence model which is treating the correlated responses as if they are independent. However, for time-varying covariates there is a dilemma in using independence correlation model as it may be very inefficient producing biased estimates. Hence this study proposed resampling methods like bootstrap methods to do the estimation and this is illustrated through an application to the Lung Health Study that investigated the effects of smoking cessation on lung function and on the symptom of chronic cough (32).

There was a study that compared several approaches to select the best working correlation structure and it was suggested that all approaches be used to select correlation structure and then decide on the best correlation and the problem spreads more when there is a small sample size (33).

Quasi Likelihood estimation:

Correlated response data are common in biomedical studies. Regression analysis based on the GEE is an increasingly important method for such data. However, there seem to be few model-selection criteria available in GEE. The well-known Akaike Information Criterion (AIC) cannot be directly applied since AIC is based on maximum likelihood estimation while GEE is non-likelihood based. The authors proposed a modification to AIC, where the likelihood is replaced by the quasi likelihood and a proper adjustment is made for the penalty term. Its performance is investigated through simulation studies. For illustration, the method is applied to a real data set (34).

Selecting an appropriate working correlation structure is pertinent to clustered data analysis using GEE because an inappropriate choice will lead to inefficient parameter estimation. A study investigated the well-known criterion of QIC for selecting a working correlation structure, and has found that performance of the QIC is deteriorated by a term that is theoretically independent of the correlation structures but has to be estimated with an error. This lead to propose a correlation information criterion (CIC) that substantially improves the QIC performance. Extensive simulation studies indicate that the CIC has remarkable improvement in selecting the correct correlation structures. They also illustrated findings using a data set from the Madras Longitudinal Schizophrenia (35).

The generalized estimating equation is a popular method for analyzing correlated response data. It is important to determine a proper working correlation matrix at the time of applying the generalized estimating equation since an improper selection sometimes results in inefficient parameter estimates. The authors proposed a criterion for the selection of an appropriate working correlation structure. The proposed criterion is based on a statistic to test the hypothesis that the covariance matrix equals a given matrix, and also measures the discrepancy between the covariance matrix estimator and the specified working covariance matrix. They evaluated the performance of the proposed criterion through simulation studies assuming that for each subject, the number of observations remains the same. The results revealed that when the proposed criterion was adopted, the proportion of selecting a true correlation structure was generally higher than that when other competing approaches were adopted (36).

The method of generalized estimating equations for regression modeling of clustered outcomes allows for specification of a working matrix that is intended to approximate the true correlation matrix of the observations. A study investigated the asymptotic relative efficiency of the GEE for the mean parameters when the correlation parameters are estimated by various methods. The asymptotic relative efficiency depends on three features of the analysis, namely (i) the discrepancy between the working correlation structure and the unobservable true correlation structure, (ii) the method by which the correlation parameters are estimated and (iii) the 'design', by which we refer to both the structures of the predictor matrices within clusters and distribution of cluster sizes. Analytical and numerical studies of realistic data-analysis scenarios show that choice of working covariance model has a substantial impact on regression estimator efficiency. Protection against avoidable loss of efficiency associated with covariance misspecification is obtained when a 'Gaussian estimation' pseudo likelihood procedure is used with an AR(1) structure (37).

The GEE technique is often used in longitudinal data modeling, where investigators are interested in population-averaged effects of covariates on responses of interest. GEE involves specifying a model relating covariates to outcomes and a plausible correlation structure between responses at different time periods. While GEE parameter estimates are consistent irrespective of the true underlying correlation structure, the method has some limitations that include challenges with model selection due to lack of absolute goodness-of-fit tests to aid comparisons among several plausible models. The quadratic inference functions (QIF) method extends the capabilities of GEE, while also addressing some GEE limitations (38).

There was a study that showed the difficulties with GEE particularly with logistic regression GEE analysis. There are also other authors who have considered these claims to be based more on semantics than on statistics especially for binary response models (39).

3.5 Random Effects Model:

When is a Random effects model applied:

A study reported the appropriate application, interpretation and compared with straightforward marginal models like GEE approaches. This study addressed the limits that needed to be placed on interpretation of the coefficients and inferences derived from random-effects models involving binary outcomes. The other issue is the diagnostic checks that are appropriate for evaluating whether such random effect models provide adequate fit to the data. These issues were addressed by means of an extended case study using data on adolescent smoking from a large cohort study. The authors also applied Bayesian estimation methods to fit discrete-mixture alternative to the standard logistic-normal model and posterior predictive checking was used to assess the model fit. The authors described surprising parallels in the parameter estimates from the logistic-normal and mixture models and used them to question the interpretability of the subject specific regression coefficients from the standard multilevel approach. Positive predictive checks suggested a serious lack of fit of both multilevel models. The authors expressed that lessons learnt from the case study provided guidance for further investigations (40).

A study reported the situation when random coefficient models must be applied. Regression models with random coefficients arise naturally in both frequentist and Bayesian approaches to estimation problems. They are becoming widely available in standard computer packages under the headings of generalized linear mixed models, hierarchical models, and multilevel models. I here argue that such models offer a more scientifically defensible framework for epidemiologic analysis than the fixed-effects models now prevalent in epidemiology. The argument invokes an antiparsimony which is that models should be rich enough to reflect the complexity of the relations under study. It also invokes the countervailing principle that you cannot estimate anything if you try to estimate everything (often used to justify parsimony). Regression with random coefficients offers a rational compromise between these principles as well as an alternative to analyses based on standard variable-selection algorithms and their attendant distortion of uncertainty assessments. These points are illustrated with an analysis of data on diet, nutrition, and breast cancer. Random effect models are also known as random coefficient models. The random coefficients models are also methods that adjust for the correlation in longitudinal data. These models like GEE can also be applied when there are repeated responses from the same individual where each individual is considered as a cluster. Random coefficients models provide the variance from which correlation termed as Intraclass correlation coefficient (ICC) is calculated. Maximum likelihood is the standard methods of estimation for linear mixed models. However, evaluation of likelihood is computationally difficult. The log likelihood is maximized using numerical integration (41).

Intraclass Correlation Coefficient (ICC):

A study was done that compared multilevel methods with the traditional methods where the outcome which was the number of alcohol-free weeks per patient during 1 year. The study had 2 level models. The comparison of results were based on a hypothetical observational PBRN (practice-based research network) study and simulated database to illustrate the findings. The data set consists of 500 patient-level observations. Patients were randomly sampled from 1 physician in each of 5 clinics (100 patients per physician), with 3 clinics located in an urban area and 2 in a rural setting. In a simple 2-level model, the sources of variance are within-groups and between-groups. Using a PBRN context with patients sampled from clinics, the total variation in patient outcomes can be partitioned into 2 variance components: *within-clinics variance* (ie, variance among patients in the same clinic) and *between-clinics variance* (ie, variance between patients in different clinics). When patients within groups are very similar to each other, we have less information than we would have from the same number of patients obtained in a simple random sample. An important measure that describes these dependencies in the data is the intraclass correlation coefficient (ICC); this statistic measures the extent to which individuals within the same group are more similar to each other than they are to individuals in different groups. The estimated ICC indicates that the ratio of the between-clinic variance to the total variance is about 55%, calculated as ICC = 1.76/(1.76 + 1.41), suggesting that patients within clinics are more similar to each other than to those at other clinics. Had the researchers ignored the hierarchical structure of the data and used traditional analytic approaches, they would have erroneously concluded that physician advice had little or no influence on patient alcohol consumption behavior. On the other hand, all the HLMs that assess the relationship between physician time advising patients on alcohol consumption and patient behavior lead to the conclusion that physician advice is effective, at least in some settings (42).

An adolescent study was done that included 8th and 10th grades with varying amounts of cigarette smoking experience. These were categorized into three categories such as < 6cigarettes, 6-99 and 100+ cigarettes. The outcome was physiological sensation change categorized into five ordered categories such as -2, -1, 0, 1 and 2. The ICC was found to be 0.44 suggesting that there was correlation within subjects. The study showed that physiological sensation diminishes as smoking level increases (43).

Estimation procedures for count data in random effect models:

In the social and health sciences, data are often structured hierarchically, with individuals nested within groups. This was presented using dyadic data. Dyadic data represent a special case of hierarchically clustered data, with individuals nested within dyads. Dyads constitute a special case of hierarchically structured data with variation at both the individual and dyadic level. Analyses of data from dyads pose several challenges due to the interdependence between members within dyads and issues related to small group sizes. Multilevel analytic techniques have been developed and applied to dyadic data in an attempt to resolve these issues. In this article, the authors described a set of analyses for modeling individual- and dyad-level influences on *binary* outcomes using SAS statistical software; and the authors discuss the benefits and limitations of such an approach. For illustrative purposes, the authors applied these techniques to estimate individual-dyad-level predictors of viral hepatitis C infection among heterosexual couples in East Harlem, New York City (44).

Least squares analyses (e.g., ANOVAs, linear regressions) of hierarchical data leads to Type-I error rates that depart severely from the nominal Type-I error rate assumed. Thus, when least squares methods are used to analyze hierarchical data coming from designs in which some groups are assigned to the treatment condition, and others to the control condition (i.e., the widely used groups nested under treatment experimental design), the Type-I error rate is seriously inflated, leading too often to the incorrect rejection of the null hypothesis (i.e., the incorrect conclusion of an effect of the treatment). To highlight the severity of the problem, a paper presented simulations showing how the Type-I error rate is affected under different conditions of intraclass correlation and sample size. For all simulations the Type-I error rate after application of the popular correction is also considered, and the limitations of this correction technique discussed. They concluded with suggestions on how one should collect and analyze data bearing a hierarchical structure (45).

A study reported that PQL, restricted maximum likelihood methods as approximate to Maximum likelihood methods are a failure in mixed models especially for binary data (46). This article provided a conceptual introduction to the issues surrounding the analysis of clustered (nested) data. It defines the intraclass correlation coefficient (ICC) and the design effect, and explains their effect on the standard error. When the ICC is greater than 0, then the design effect is greater than 1. In such a scenario, the standard error produced under the assumption of independence is underestimated. This increases the Type I error rate. This is illustrated based on the effect of non-independence on the standard error. The paper also shows that after accounting for the design effect, the decision about the statistical significance of the test statistic changes. When there is a failure to account for the clustered nature of the data, it was concluded that the difference between the two groups is statistically significant. However, once they adjusted the standard error for the design effect, the difference is no longer statistically significant (47).

Comparison of Fixed Effects, GEE and Random Effects model:

A study reported the differences between fixed effects, random effects and GEE analysis. This article reported the underlying assumptions to assess the covariate effects on the mean of continuous, dichotomous or count outcomes. This paper reports the structural differences and similarities of the random effects, the linear mixed model, the fixed effects and generalized estimating equations in a longitudinal data. Let the random draw from a population of interest be (Y_i, X_i), where I denotes the sampling unit$Y_i = \left(Y_{i1}, Y_{i2}, \ldots Y_{in_i}\right)'$ the time-ordered $n_i \times 1$ vector of responses and $X_i = \left(x_{i1}, x_{i2}, \ldots x_{in_i}\right)'$ an $n_i \times p$ matrix of explanatory variables with X_{ij} a $p \times 1$ vector associated with the response Y_{ij}. The conditional mean vector and covariance matrix are respectively, $\mu_i = E(Y_i|X_i)$ and $V_i = E[(Y_i - \mu_i)(Y_i - \mu_i)'|X_i]$. In the above notation, each component of the conditional mean $\mu_{ij} = E(Y_{ij}|X_i)$ is a function of all the covariates. The total number of observations in the sample is $N = \sum_{i=1}^{n} n_i$. Let g be a known link function such that $g\left(\mu_{ij}\right) = X'_{ij}\beta$ where $\beta = (\beta_1, \ldots \beta_p)'$ is a pX1 vector of unknown parameters. Whereas the mean $\mu_i$depends on β, the covariance matrix $\mathbf{V_i}$ may depend on β and perhaps additional parameters α so that the total number of parameters is $p+p_1$.

Marginal Model: This specifies only the conditional mean $\mu_i = E(Y_i|X_i)$ but treats the parameters in $\mathbf{V_i}$ as nuisance parameters. A distribution function in the exponential family usually suggests the form of mean and variance of Y_{ij}.

Random Effects model: In this model, correlation is induced through an unobserved heterogeneity ζ_i in the conditional mean specification n$\mu_{ij} = E(Y_{ij}|X_{ij}, \zeta_i)$. The random coefficient model will also fall under this umbrella allowing one to acknowledge dependencies at different levels of a hierarchy (48).

Recent advances in statistical software have led to the rapid diffusion of new methods for modeling longitudinal data. Multilevel (also known as hierarchical or random effects) models for binary outcomes have generally been based on a logistic–normal specification, by analogy with earlier work for normally distributed data. The appropriate application and interpretation of these models remains somewhat unclear, especially when compared with the computationally more straightforward semiparametric or 'marginal' modeling (GEE) approaches. In this paper we pose two interrelated questions. First, what limits should be placed on the interpretation of the coefficients and inferences derived from random-effect models involving binary outcomes? Second, what diagnostic checks are appropriate for evaluating whether such random-effect models provide adequate fits to the data? We address these questions by means of an extended case study using data on adolescent smoking from a large cohort study. Bayesian estimation methods are used to fit a discrete-mixture alternative to the standard logistic–normal model, and posterior predictive checking is used to assess model fit. Surprising parallels in the parameter estimates from the logistic–normal and mixture models are described and used to question the interpretability of the so called 'subject-specific' regression coefficients from the standard multilevel approach. Posterior predictive checks suggest a serious lack of fit of both multilevel models. The results do not provide final answers to the two questions posed, but we expect that lessons learned from the case study will provide general guidance for further investigation of these important issues (40).

Several approaches have been proposed to model binary outcomes that arise from longitudinal studies. Most of the approaches can be grouped into two classes: the population-averaged and subject-specific approaches. The generalized estimating equations (GEE) method is commonly used to estimate population-averaged effects, while random-effects logistic models can be used to estimate subject-specific effects. However, it is not clear to many epidemiologists how these two methods relate to one another or how these methods relate to more traditional stratified analysis and standard logistic models. The authors address these issues in the context of a longitudinal smoking prevention trial, the Midwestern Prevention Project. In particular, the authors compare results from stratified analysis, standard logistic models, conditional logistic models, the GEE models, and random-effects models by analyzing a binary outcome from two and seven repeated measurements, respectively. In the comparison, the authors focus on the interpretation of both time-varying and time-invariant covariates under different models. Implications of these methods for epidemiologic research have been discussed which found that both estimates for random effects and standard errors were larger than GEE model although the test statistic results were similar (49).

A study compared GEE to random effects model in genetic association analysis. The authors proposed a retrospective multilevel model (rMLM) approach to analyze sibship data by using genotypic information as the dependent variable. Simulated data sets were generated using the simulation of linkage and association (SIMLA) program. Then they compared rMLM to sib transmission/disequilibrium test (S-TDT), sibling disequilibrium test (SDT), conditional logistic regression (CLR) and generalized estimation equations (GEE) on the measures of power, type I error, estimation bias and standard error. The results indicated that rMLM was a valid test of association in the presence of linkage using sibship data. The advantages of rMLM became more evident when the data contained concordant sibships. Compared to GEE, rMLM had less underestimated odds ratio (50).

Diagnostic issues in Random effect models:

Commonly applied diagnostic procedures in random-coefficient (multilevel) analysis are based on an inspection of the residuals, motivated by established procedures for ordinary regression. The deficiencies of such procedures are discussed and an alternative based on simulation from the fitted model (parametric bootstrap) is proposed. Although computationally intensive, the method proposed requires little programming effort additional to implementing the model fitting procedure. It can be tailored for specific kinds of outliers. Some computationally less demanding alternatives are described (51).

3.6 Markov Chain:

Most of disease conditions exhibit a property that the present state of condition is due to the just previous condition. Hence a study in Taiwan used this property of Markov chain to assess the efficacy of screening non-insulin dependent diabetes mellitus. Non-insulin diabetes (NIDDM) was very common in Taiwan. It is important that this disease is detected in the asymptomatic phase. In order to project the above progression a five state illness-and-death Markov chain model was proposed to estimate these transition parameters. The annual incidence of asymptomatic NIDDM was 10.67 (8.26 – 13.79) per 1000. The progression rate from asymptomatic to symptomatic NIDDM was 2.27% implying that there was a high risk of dying from NIDDM for subjects with symptomatic NIDDM (52). This paper, however, is based on absorbing state where there is no possible transition from death state. Therefore there is a need for non-absorbing state model.

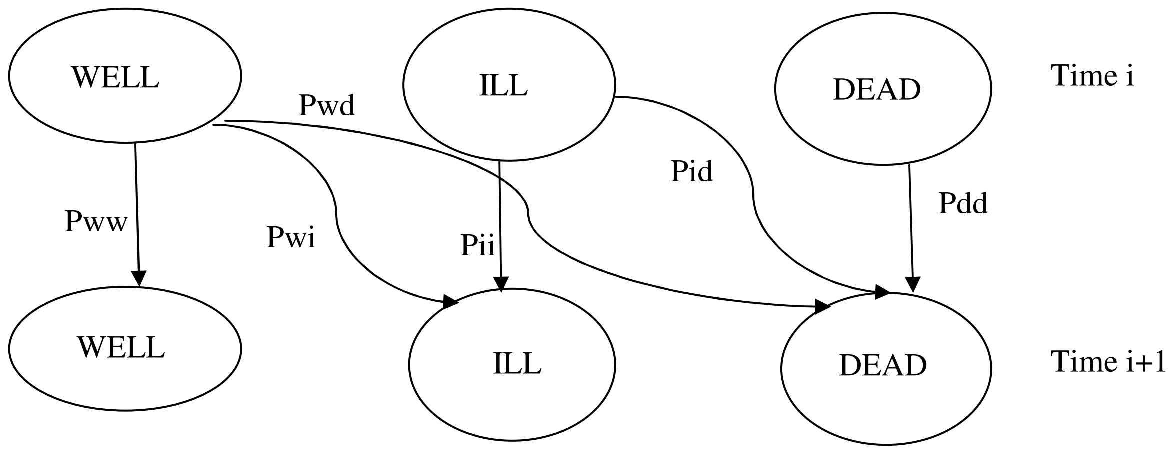
For many recurrent events, a change in state during a sufficiently narrow time span usually involves a move to next state only. For example, the person with no headache during one interval is more likely during the next span of time to have a slight headache than a severe headache. A Markov chain is a stochastic model that describes the probabilities of transition among sites of a system. The assumption that characterizes a Markov chain is that the transitional probability is completely determined by the present state of the system. A headache diary was completed by 177 female and 57 male headache patients. The patients were asked to record their level of headache during each 24-hr interval for 28 days. The main aim of the study was to see if the movement from time t to t+1 was different for males and females. This was tested using log linear model. The $G^2 = 25.17$ for 9 degrees of freedom suggesting that there was a difference in the transition probability for males and females (53). This study has not estimated the likely duration of stay in each state of headache transition which may be of clinical importance.

Another study had a hypothesis which stated “Should anticoagulation be withheld in a patient with an artificial heart valve and recent hemorrhagic cerebral infarction?” This problem was answered using Markov model. There are three major states of health which were well, disabled and dead. In addition to the above states there was also another additional state which was a temporary state of health, minor event state. The data was from the Framingham cardiovascular disease study. A transition probability matrix of the Markov process was obtained. The Markov model was also used to generate quality-adjusted life expectancy values (10).

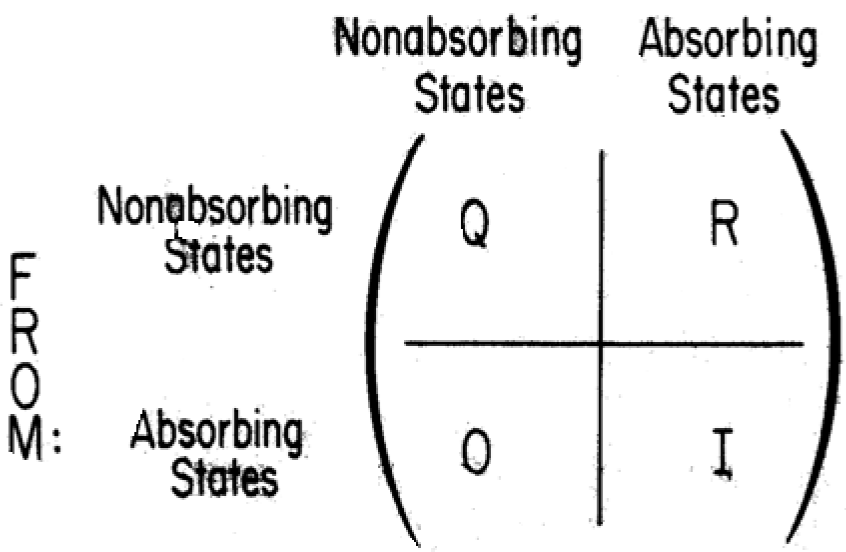
A study was conducted to describe the lifetime clinical course and costs of Crohn's disease in a 24-year population-based inception cohort of patients with Crohn's disease in Olmsted county. The disease states were defined by medical and surgical treatment. A Markov model analysis calculated time in each disease state and presented value of excess lifetime costs in comparison with an age- and sex matched cohort (54).

A Markov model of prognosis was evaluated by comparing the duration of disease activity states and life expectancy with Kaplan Meier curves for 98 patients with systemic lupus erythematous with 1080 patient years of observations. A four state homogenous Markov chain was constructed to determine the transition probabilities between the disease states. The proportion of the patient population in each disease state over time provided a convenient graphic summary of the natural history of SLE from which the Kaplan Meier survival curves were obtained. A Markov model yielded a clinically useful description of outcome for multistate disease (55).

Physician's estimate of prognosis under alternative treatment plans in a principal factor in therapeutic decision making. Methods of reporting prognosis such as five-year survivals, survival curves and quality adjusted life expectancy are crude estimates of natural history. The author describes a general purpose model of medical prognosis based on the Markov process and shows how this simple mathematical tool be used to generate detailed and accurate assessments of life expectancy and health status. Natural history of a chronic disease can be viewed for an individual patient as a sequence of particular states of health. An example described in the paper was if a patient was classified into one of three categories: WELL, ILL and DEAD. At any time i, the patient resides in one of the states. This is represented as shown in the figure:



The oval represents the patient being in one of the three ovals (upper row) and transitions that occur over a fixed time interval i are illustrated by the arrows from i to i+1. It is possible to leave WELL or ILL via a transition and hence are termed as Non-absorbing states where as once DEAD state is reached no possible transitions can be made and hence is an absorbing state. This paper has described the constant transition probabilities that are possible only with Markov chain model. The matrix formulation has four important sections.



The section labeled Q which reflects the probability of not being absorbed and the probability of being in the WELL state is sum of the four probabilities of the elements. Each element is subtracted from a corresponding element in another 2X2 matrix of ones on the diagonal and zeros elsewhere. Matrix N is the fundamental matrix of absorbing Markov chain and has its elements by column the expected time in each of absorption state given the starting state corresponding to the row of N matrix which is displayed in the figure below which represents the single step transition between and within absorbing and transient states (10).

Another study reported the mean sojourn time which is the time spent in the preclinical detectable phase for chronic disease like breast cancer that plays an important role in the design and assessment of screening programs. This paper developed two-parameter Markov chain model and the model was developed explicitly to estimate the preclinical incidence rate and the rate of transition from preclinical to clinical state without using control data. Using this method to the data from Swedish two county study of breast cancer screening in the age group 70-74, the mean sojourn time was found to be 2.3 with 95 percent confidence interval ranging from 2.1 – 2.5 which was close to the result based on the traditional method however, the 95 percent confidence interval was narrower using Markov model. The reason for greater precision of the latter is the fuller use of all temporal data since the continuous exact times to events are used in our method instead of grouping them as in the traditional method (56). This paper presented the sojourn time but did not explain the transition time from one clinical state to another.

Another study conducted in India used the systemic lupus erythematosus disease activity index (SLEDAI) score to find factors that predict the survival outcome for patients with systemic lupus erythematosus patients. The subsequent visits were abstracted from the case notes and the quantitative data from SLEDAI score was used to construct a Markov chain mathematical expression to predict life expectancy. The predicted life expectancy using Markov chain was

13.9 years and it was also found that proteinuria caused a 50% reduction in the life expectancy but increased disease activity at onset did not predispose to a poor outcome (57).

A study considered the estimation of the intensity and survival functions for a continuous time progressive three-state semi-Markov model with intermittently observed data. The estimator of the intensity function is defined nonparametrically as the maximum of a penalized likelihood. Thus the authors obtained smooth estimates of the intensity and survival functions. This approach also accommodated complex observation schemes such as truncation and interval censoring. The method is illustrated with a study of hemophiliacs infected by HIV. The intensity functions and the cumulative distribution functions for the time to infection and for the time to AIDS are estimated. Covariates can easily be incorporated into the model (58). This study explains for continuous time Markov model.

3.7 Markov Regression:

A study proposed an online binary classification procedure for cases when there is uncertainty about the model to use and parameters within a model change over time. The authors accounted for model uncertainty through dynamic model averaging, a dynamic extension of Bayesian model averaging in which posterior model probabilities may also change with time. The authors applied a state-space model to the parameters of each model and allowed the data-generating model to change over time according to a Markov chain. The study also proposed an algorithm that adjusts the level of forgetting in an online fashion using the posterior predictive distribution, and so accommodates various levels of change at different times. This method was applied to the data from children with appendicitis who receive either a traditional (open) appendectomy or a laparoscopic procedure. Factors associated with which children receive a particular type of procedure changed substantially over the 7 years of data collection, a feature that is not captured using standard regression modeling (59). However, this approach may be intensive in computing and may have problem in generalizing.

A study describes the application of a multi-state model to diabetic retinopathy under the assumption that a continuous time Markov process determines the transition times between disease stages. There are three transient states that represented the early stages of retinopathy and one final absorbing state that represented the irreversible stage of retinopathy. Using a model with covariables, the authors explored the effects of factors that influenced the onset, progression and regression of diabetic retinopathy among subjects with insulin-dependent diabetes mellitus. The study also had time – dependent covariables in the model assuming that the covariables remained constant between two observations. The authors also demonstrated survival curves from each stage of the disease and for any combination of the risk factors (60). This was applied for absorbing state Markov model.

Stroke is a leading cause of death worldwide. Stroke related disability manifestation by neurological impairment has resulted in poor quality. The ability to predict changes in functional disability over time would be conducive to the clinical management and rehabilitation of stroke patients. Functionality disability changes with time and varies from individual to individual. The approach used to model the data was multi-state Markov model adequately after adjusting for the known covariates. The data was obtained form a multi-center study that involved 111 patients with a first stroke. These patients were assessed after at six time points. The univariate analysis showed the transition rate from state 1 which was poor function to state 2 (moderate function) as 1.19 per month and from state 2 to state 3 (good function) was 0.43 per month. Age, size of infarct, sex had no effect on transition rates. Baseline functional status showed an effect on transition rates. The multivariable analysis showed that baseline functional status had an effect from state 2 to state 3 (HR: 0.14; 0.06 -0.25). Size of the infarct showed an effect on transition rate from state 1 to state 2 with hazard ratio (HR) 2.3 (1.6-3.0) (61).

A study on liver fibrosis evolution in HIV-HBV-coinfected patients treated with tenofovir disoproxil fumarate (TDF) was conducted. The effect of TDF on liver fibrosis in 148 HIV-HBV-coinfected patients was prospectively evaluated using Fibrometer scores and liver biopsies in a subset of patients. The mean change from baseline in Fibrometer score was modelled using a GEE and a homogeneous continuous-time Markov models were used to study risk factors for regression or progression of liver fibrosis. It was found that the median follow-up of patients treated with TDF was 29.5 months (25th-75th percentile 20.9-38.1). In patients with a baseline fibrosis score of F3-F4, Fibrometer score decreased with a triphasic shape (Fibrometer change at 12, 24 and 36 months after TDF initiation was -0.079, -0.069 and -0.102, respectively). Progression in fibrosis score over time was influenced by age, alcohol consumption, low CD4(+) T-cell count and HCV coinfection, whereas HDV coinfection and longer duration of HBV infection prevented fibrosis regression. No influence of antiretrovirals other than TDF was found (62). Continuous time Markov models, and the advantages over survival models need to be explored further in medical field.

An observational study on cancer patients, progression of performance status over time was described by a multistate model. There were four states with one absorbing state. The performance status at each clinic visit was based on PPS score. PPS score 70-100 (stable state), PPS score 40-60 (state 2), 10-30 PPS score (state 3) and deceased (state 4). The transition intensity rates were obtained for transition for all the states. These transition intensity rates were then used to estimate the transition probabilities by the end of 1 month and 6 months. It was found that a patient who was in the transitional state had 11% chance of being in the stable state at the end of 1 month, 5% chance of being in the end-of-life state at the end of 1 month, 24% chance of being dead at the end of this time There was 0.8% chance of death for a patient in the stable state at the end of 1 month where as 15% chance of death if the patient was in the stable state at the end of 6 months (63).

A longitudinal study that compared Markov model regression model, markov regression model with random effects and a mover-stayer model to find the risk factors for transition in Bacterial Vaginosis among women. The study showed that Markov regression model found a poor fit while Markov regression with random effects that accounted for additional unexplained heterogeneity had better fit to the data. The study found that transition models that accounted for additional heterogeneity provided an attractive approach for describing the effect of covariates on the natural history of BV (64).

There was a study that was conducted to find the potential effects of interventions on cervical cancer. The authors constructed a Markov state-transition model of a cohort of HIV positive women in Cameroon. They examined the potential impact, on cumulative cervical cancer mortality of four possible scenarios: when no HAART and no screening was present (NHNS), HAART and no screening (HNS), HAART and screening once on HAART initiation (HSHI) and HAART screening once at age 35 (HS35). The model projected that compared to NHNS, lifetime cumulative cervical cancer mortality approximately doubled with HNS (65).

A study was done to find an exploration of factors associated with suicidality so as to understand the mechanisms that lead to suicide. Two samples in Germany were examined via internet regarding suicidality, depression, alcohol abuse, adverse childhood experiences and parent-child relationships. A Graphical Markov Model was constructed from the first subsample, testing for main, quadratic and interaction effects. All effects in the model were cross-validated using the second subsample. Depression was found to be a strong predictor of suicidality; alcohol abuse was not a predictor. Both maternal and paternal love also predicted suicidality; the former had an indirect effect via depression and the latter a direct effect. Early experiences with violence showed both a direct and indirect association with suicidality. In addition to depression being a predictor for suicidality, various pathways connect suicidality with early childhood experiences (66).

Markov regression using multilevel modeling and continuous time Markov regression are other areas, which need to be explored in terms of challenges in using them in longitudinal studies, the advantages over GEE or Random effects have to be researched further.

SCOPE AND PLAN OF WORK

4. SCOPE AND PLAN OF WORK

Scope: Clinical data on the prognosis of diseases with a clinical history marked by exacerbations and remissions are best derived from longitudinal study for group patients from the onset of disease. In Systemic Lupus Erythematosus (SLE), the clinically pivotal issues are correlated with the transition between exacerbations and remissions. The usual survival analyses or incidence rates by subgroups will not address the impact of transition. A Markov model is a different method of analyzing and reporting the prognosis of multi state disease processes. Since clinical data are often available on the likelihood of patient's disease becoming more active or going into remission, it is often possible to estimate the duration of stay in each disease state under the assumption of Markov model. Discrete time Markov model is an approach when there is discrete transition of health status of patients over time. Discrete transition refers to the transition occurring in fixed points in time. A discrete time Markov model can approximate a continuous time Markov model by defining a cycle length of interest such as yearly or 6 monthly. Markov model uses the probability of transition from one clinical or disease state to another. The data will deal with Protein-energy Malnutrition (PEM) among children as this will help us to estimate the mean transition time which specifies, for example, what is the average time taken to transit from a severe state to normal state of malnutrition and also to estimate average time of stay in a particular state, so as to plan appropriate intervention. Also, using the transition probabilities one can model the risk factors using the regression models, conditioning the current state of the outcome with the immediate state of the outcome. Thus, the dependence of serial outcomes are expected to be within two time points as compared to Generalized Estimating Equation (GEE) which deals with all time points or, the Random effects models or Multi Level Modeling (MLM) which deals with identifying and excluding the exact amount of correlation within individuals. Thus the usual GEE or MLM approach is expected to have wider standard errors (SE) as compared to Markov Regression (MR) models. Therefore there is a scope to find more number of risk factors as significant as compared to GEE or MLM models. However, the wider SE implying wider Confidence Intervals (CI) has to be shown as a consistent criteria and this can be shown through simulations. Hence, the coverage probabilities that the CIs could have, based on both methods, will provide us to suggest a best method. If expectation is that the MR shows to be a better method then this could change the practice.

Plan of work: Protein Energy Malnutrition (PEM) study data will be used for the above mentioned scope. The malnutrition level will be categorized into 3 categories as Normal, Mild/Moderate and Severe at the each visit, which is every 6 months follow up data. Nutritional status will be calculated using EPIINFO software. The transition probability matrices will be established using this nutritional status. Using Chapman Kolmogrov equations, Stationary distribution of the transition probability matrix and solving system of equations we will find First Mean Passage Time. The 95% CI will be computed using 10,000 simulations from the above estimations. A priori specified risk factors will be associated with the ordinal outcome using GEE procedures. As this being a longitudinal study, Autoregressive (1) correlation structure will be used to adjust for the correlation structure. In MLM model, the exact amount of correlation will be found out for the levels and adjustment will be done accordingly. The diagnosis to check whether the current status of the outcome depends on the immediate or the previous states of the outcome will be done. That is, whether Y_t where Y (the state of the malnutrition at current time "t") depends on Y_{t-1} or Y_{t-2} or Y_{t-3} etc will be assessed. The ordinal Logistic regression cumulative odds models using GEE and MLM will be used to analyze the data. The MR analysis will be done using Transition Probability matrix and Transition Intensity Matrix. The significant risk factors of MR analyses will be compared to the results from the GEE and MLM analyses. The comparison will also be based on the length and coverage probabilities of the 95% CI for the risk factors, which will be obtained from the simulation findings.

MATERIALS AND METHODS

5. MATERIALS AND METHODS

5.1. Data:

During 1982, seven localities and 22 villages were selected for this study. These localities and villages were selected from Vellore town and KV Kuppam development block sampling frames respectively. All children aged 5-7 years were screened for signs of malnutrition by consultant pediatricians. The children from rural and urban areas of Vellore town were screened at baseline and followed up for every six months for 7 times. The anthropometric data were collected by two Anthropologists independently and care was taken to reduce intra and inter observer variability. Inter and intra observer variation was handled by standardizing the procedure.

5.2. Malnutrition classification:

Malnutrition was assessed based on the indicators which are BMI Z scores, Height-for-age. The BMI Z scores were classified as "normal" if the BMI Z scores were >-2 standard deviations, "moderate" when Z scores were between -2 and -3 standard deviations and, "severe" if the Z scores were <-3 standard deviations (67). EPIINFO software was used to compute Z scores for every follow-up and baseline anthropometric measurements.

5.3. Risk Factors:

The main hypothesized risk factors for the study were 'defecation practices at household level' (within the household; in the open fields), 'type of fuel used for cooking in the house' (firewood or cow dung or coal; gas or kerosene) and 'presence of a separate kitchen within the household premises' (yes; no). The other confounders that were seen important that have to be adjusted were sex of the child (male; female) and area of residence (rural; urban). Some other covariates that were also included for Generalized Estimating Equations and Markov Regression using transition probabilities are education of mother and father (illiterate or literate; primary or middle school; high school or above), consanguineous marriage of the parents whose children were included in the study (yes; no), type of roof (thatched; tiled; RCC or pukka), type of house (brick and cement; brick and mud; others) and birth order (1; 2; >=3), number of members in a family (<=4; 5-6; >6), type of floor (kucha; pukka).

5.4 Cumulative Incidence of Severe Malnutrition:

It is the percentage of children who have experienced new cases of severe malnutrition before the end of each year. In other words, it was calculated as the ratio of the number of children who were normal or moderate at baseline and became severely malnourished before the end of the first year to total number of children in the first year.

5.5. Generalized Estimating Equations:

5.5.1 Model of a Generalized Estimating Equation:

For a given outcome y_{it}, we have a $(p \times 1)$ vector of covariates X_{it} associated with our parameter vector β. We also have a $(q \times 1)$ vector of covariates Z_{it} associated with the random effect ν_i. consider the marginal expectation of the outcome (integrated over the distribution)

$$\mu_{it}^{PA} = E[E(\mathcal{Y}_{it}|\nu_i)]$$

so that the responses are characterized by $g\left(\mu_{it}^{PA}\right) = X_{it}\beta^{PA}$

$$V\ (\mathcal{Y}_{it}) = V\left(\mu_{it}^{PA}\right)a(\boldsymbol{\phi})$$

Thus, the marginal expectation is the average response for observations sharing the same covariates. Generalized Estimating Equations (GEE) are also known as Population averaged (PA) models as they indicate that the marginal outcome are averaged over the population of individuals and that the coefficients. β^{PA} have an interpretation in terms of the response averaged over the population. The limited information maximum quasi likelihood (LIMQL) estimating equation for generalized linear model (GLM) is

$$\Psi\ (\beta) = \left[\left\{ \sum_{i=1}^{n} \sum_{t=1}^{n_i} \frac{y_{it} - \mu_{it}}{a(\boldsymbol{\phi})\boldsymbol{V}\ (\mu_{it})} \left(\frac{\partial \mu}{\partial \eta}\right)_{it} x_{jit} \right\}_{j=1,\ldots,p} \right]_{p\times 1} = [0]_{p\times 1}$$

The above equation can be re-written in the matrix of panels are

$$\Psi\ (\beta) = \left[\left\{ \sum_{i=1}^{n} X_{ji}^{T} D\left(\frac{\partial \mu}{\partial \eta}\right) \quad [V(\mu_i)]^{-1} \left(\frac{y_i - \mu_i}{a(\boldsymbol{\phi})}\right) \right\}_{j=1,\ldots,p} \right]_{p\times 1}$$

where, D() is the diagonal matrix. $V(\mu_i)$ is also a diagonal matrix that can be decomposed into

$$V(\mu_i) = \left[D\big(V(\mu_{it})\big)^{1/2} I_{(n_i \times n_i)}\, D\,\big(V(\mu_{it})\big)^{1/2} \right]_{n_i \times n_i}$$

The above equation denotes that the estimating equation treats each observation within a pane as independent. When the marginal distribution of the outcome for which the expected value and variance functions are averaged over the panels, the above identity matrix is the within-panel correlation matrix. The GEE is a modification of LIMQL estimating equation where it is replacing the identity matrix with a more general correlation matrix since the variance of the correlated data does not have a diagonal form

$$V(\mu_i) = [D\big(V(\mu_{it})\big)^{\frac{1}{2}} R(\alpha)_{(niXni)} D\big(V(\mu_{it})\big)^{\frac{1}{2}}]_{niXni}$$

$R(\alpha)$ is the correlation matrix that is estimated through the parameter α.

Generalized Estimating equation (GEE) is an iterative procedure, using quasi-likelihood to estimate the regression coefficients. The estimated regression coefficients reflects the relationship between the longitudinal development of the underweight as classified using BMI-Z scores and the corresponding predictor variables such as defecation practices, presence of a separate kitchen in the household, mother's education, father's education, sex of the child, type of fuel used for cooking, etc.

Let the random draw from a population of interest be (Y_i, X_i), where I denotes the sampling unit $Y_i' = (Y_{i1}, Y_{i2}, \dots Y_{in_i})'$ the time-ordered $n_i \times 1$ vector of responses and $X_i = (x_{i1}, x_{i2}, \dots x_{in_i})'$ an $n_i \times p$ matrix of explanatory variables with X_{ij} a $p \times 1$ vector associated with the response Y_{ij}. The conditional mean vector and covariance matrix are respectively, $\mu_i = E(Y_i|X_i)$ and $V_i = E[(Y_i - \mu_i)(Y_i - \mu_i)'|X_i]$. In the above notation, each component of the conditional mean $\mu_{ij} = E(Y_{ij}|X_i)$ is a function of all the covariates. The total number of observations in the sample is $N = \sum_{i=1}^{n} n_i$. Let g be a known link function such that $g(\mu_{ij}) = X_{ij}'\beta$ where $\beta = (\beta_1, \dots \beta_p)'$ is a pX1 vector of unknown parameters. Whereas the mean $\mu_i$depends on β, the covariance matrix $\mathbf{V_i}$ may depend on β and perhaps additional parameters α so that the total number of parameters is $p+p_1$.

Marginal Model: This specifies only the conditional mean $\mu_i = E(Y_i|X_i)$ but treats the parameters in $\mathbf{V_i}$ as nuisance parameters. A distribution function in the exponential family usually suggests the form of mean and variance of Y_{ij}. The estimator of β has the same structural form as the generalized least square estimator. The methods of estimation of the variance $\mathbf{V_i = V_i(\alpha)}$ are different. The true variance is not known but even though it may be misspecified, the asymptotic variance of GEE estimator of β can be made robust. However, some loss of infeasible efficiency could result if the data do not support the correlation structure (68, 69). Only in this case does the GEE ensure consistent estimation of effects of covariates on the marginal expectation of the outcome. The quail-likelihood information criterion is a reasonable way of choosing working correlation matrix and for selecting variables (70).

5.5.2 Parameterizing the working correlation matrix:

The efficiency of the regression parameters are gained by choosing a within-panel correlation. Its not very easy to choose a working correlation structure (71). Here are several ways which might hypothesize the structure. They are:

1. **Independent structure**: With this structure the correlations between subsequent measurements are assumed to be zero. In other words, this correlation structure assumes independence of the observations:

	t_1	t_2	t_3	t_4	t_5	t_6	t_7
t_1	-	0	0	0	0	0	0
t_2	0	-	0	0	0	0	0
t_3	0	0	-	0	0	0	0
t_4	0	0	0	-	0	0	0
t_5	0	0	0	0	-	0	0
t_6	0	0	0	0	0	-	0
t_7	0	0	0	0	0	0	-

1. **Exchangeable correlation structure:** In this structure the correlations between subsequent measurements are assumed to be the same, irrespective of the length of the time interval.

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
t_1	-	ρ	ρ	ρ	ρ	ρ	ρ	ρ
t_2	ρ	-	ρ	ρ	ρ	ρ	ρ	ρ
t_3	ρ	ρ	-	ρ	ρ	ρ	ρ	ρ
…	…	…	…	…	…	…	…	…
t_8	ρ	ρ	ρ	ρ	ρ	ρ	ρ	-

2. ***m* – dependent** (**stationary**) **structure:** The correlations *t* measurements apart are equal, the correlations $t + 1$ measurements apart are assumed to be equal, and so on for $t = 1$ to $t = m$. Correlations more than '*m*' measurements apart are assumed to be zero. When, for instance, a '2-dependent correlation structure' is assumed, all correlations one measurement apart are assumed to be the same, all correlations two measurements apart are assumed to be the same, and the correlations more than two measurements apart are assumed to be zero.

	t_1	t_2	t_3	t_4	t_5	t_6
t_1	-	ρ_1	ρ_2	0	0	0
t_2	ρ_1	-	ρ_1	ρ_2	0	0
t_3	ρ_2	ρ_1	-	ρ_1	ρ_2	0
t_4	0	ρ_2	ρ_1	-	ρ_1	ρ_2
t_5	0	0	ρ_2	ρ_1	-	ρ_1
t_6	0	0	0	ρ_2	ρ_1	-

4. **Autoregressive correlation structure**: The correlations one measurement apart are assumed to ρ; correlations two measurements apart are assumed to be ρ^2; correlations *t* measurements apart are assumed to be ρ^t.

	t_1	t_2	t_3	t_4	t_5	t_6
t_1	-	ρ^1	ρ^2	ρ^3	ρ^4	ρ^5
t_2	ρ^1	-	ρ^1	ρ^2	ρ^3	ρ^4
t_3	ρ^2	ρ^1	-	ρ^1	ρ^2	ρ^3
t_4	ρ^3	ρ^2	ρ^1	-	ρ^1	ρ^2
t_5	ρ^4	ρ^3	ρ^2	ρ^1	-	ρ^1
t_6	ρ^5	ρ^4	ρ^3	ρ^2	ρ^1	-

5. Unstructured correlation structure: With this structure, all correlations are assumed to be different (72).

	t_1	t_2	t_3	t_4	t_5	t_6
t_1	-	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
t_2	ρ_1	-	ρ_6	ρ_7	ρ_8	ρ_9
t_3	ρ_2	ρ_6	-	ρ_{10}	ρ_{11}	ρ_{12}
t_4	ρ_3	ρ_7	ρ_{10}	-	ρ_{13}	ρ_{14}
t_5	ρ_4	ρ_8	ρ_{11}	ρ_{13}	-	ρ_{15}
t_6	ρ_5	ρ_9	ρ_{12}	ρ_{14}	ρ_{15}	-

5.5.3 Generalized Estimating Equations for ordinal response:

The ordinal response for Generalized Estimating Equations (GEE) is given as

$$Pr(y_{it} > s \mid x_{it}) = \frac{\exp(\beta_0 + \beta_i x_{it}) - k_s}{1 + \exp(\beta_0 + \beta_i x_{it}) - k_s}$$

where, y_{it} is the ordinal response for i^{th} individual at the t^{th} time point within each individual 'i'. s = 1,2,3 ordinal responses which are 1 – normal, 2 – moderate and 3 – severe . The link function used for ordinal regression is cumulative-log-log; k_s is the response category specific parameter.

5.5.3 Estimation:

The estimation procedure in GEE is an iterative process. It involves the following steps:

1. First a 'naïve' linear regression analysis is carried out, assuming the observations within subjects are independent.
2. Based on the residuals of this analysis, the parameters of the working correlation matrix are calculated.
3. The last step is to re-estimate the regression coefficients, correcting for the dependency of the observations.

The estimation process alternates between steps two and three, until the estimates of the regression coefficients and standard errors are stabilized.

In GEE analysis, the within – subject correlation structure is treated as a 'nuisance' variable (i.e. as a covariate). So, in principle, the way in which GEE analysis corrects for the dependency of observations within one subject is the way that has been shown in equation (which can be seen as an extension of equation.

$$Y_{it} = \beta_0 + \sum_{j=1}^{J} \beta_{ij} X_{itj} + \dots + \mathrm{CORR}_{\mathrm{it}} + \mathcal{E}_{it}$$

where Y_{it} are observations for subject i at time t, β_0 is the intercept, X_{itj} is the independent variable j for subject i at time t and $\mathrm{CORR}_{\mathrm{it}}$ is the working correlation structure, and $\mathcal{E}_{it}$ as the 'error' for subject *i* at time *t* (73).

Alternating Logistic Regression:

It has been argued that GEE with binary response logistic models may contain bias that cannot be eradicated from within standard GEE. A reason for the bias rests in the fact that the Pearson residuals, which we have been using to determine the various GEE correlation matrices are not appropriate when dealing with binary data and hence proposed a model termed *alternating logistic regression* which aims to ameliorate this bias, which clearly affects logistic GEE models (74). The alternating logistic regressions (ALR) algorithm models the association between pairs of responses with log odds ratios, instead of with correlations, as do standard GEE algorithms. The model is fit to determine the effect the predictors have on the pair-wise odds ratios. A results is that ALR is less restrictive with respect to the bounds on alpha than is standard GEE methodology. It is the ratio of the probability of success ($y=1$) to the probability of failure ($y=0$). Of a pair of responses, the odds that $y_{ij} = 1$, given that $y_{ik}=1$, is expressed as:

$$O(y_{ij};\ y_{ik}=1) = \frac{\Pr(y_{ij}=1, y_{ik}=1)}{\Pr(y_{ij}=0, y_{ik}=1)}$$

The odds that $y_{ij} = 1$, given that $y_{ij} = 0$ can be given as:

$$O(y_{ij};\ y_{ik}=0) = \frac{\Pr(y_{ij}=1, y_{ik}=0)}{\Pr(y_{ij}=0, y_{ik}=0)}$$

The odds ratio is the ratio of the two odds, which is then given as:

$OR(y_{ij}, y_{ik}) = \Psi_{ijk} =$

$[Pr(y_{ij}=1, y_{ik}=1)/Pr(y_{ij}=0, y_{ik}=1)]/[Pr(y_{ij}=1, y_{ik}=0)/Pr(y_{ij}=0, y_{ik}=0)]$

or, $OR(Y_{ij}, Y_{ik}) = \frac{Pr(y_{ij}=1, y_{ik}=1)/Pr(y_{ij}=0, y_{ik}=0)}{Pr(y_{ij}=1, y_{ik}=0)/Pr(y_{ij}=0, y_{ik}=1)}$

where *i* in above equation indicate a cluster, *j* is the first item of the pairs, and *k* is the second item of a pair. Alternating logistic regression (ALR) seeks to determine the correlations of every pair-wise comparison of odds ratios in the model. The logic of alternating logistic regressions is to simultaneously regress the response on the predictors, as well as modeling the association among the responses in terms of pair-wise odds ratios. The ALR algorithm iterates between a standard GEE logistic model in order to obtain coefficients, and a logistic regression of each response on the others within the same panel or cluster using an offset to update the odds ratio parameters. In other words, the algorithm alternates (hence the name) between a GEE-logistic model to obtain coefficients, and a standard logistic regression with an offset aimed at calculating pair-wise odds between members of the same panel, or $OR(Y_{ij}, Y_{ik})$. The algorithm may be constructed to initially estimate the log-odds ratios, subsequently converting them to odds ratios, or it can directly estimate odds and odds ratios.

The GEE logistic GEE model and alternating logistic regression fit well with respect to the differences in parameter estimates and robust standard errors. It is of interest to note that Twist

(2003, 2008) argues that logistic GEE models are prime examples (73) of the claim made by Liang and Zeger, developers of the GEE methods, that GEE analysis is robust against the wrong selection of correlation structure (75). However, such an argument is contrary to Diggle et al.'s (76) caveat that binary response models are not appropriate for GEE unless amended via ALR methods.

5.6. Random Effects Model or Multilevel Modeling (MLM):

The random effects model or multilevel structure includes correlation among observations within cluster, which in the present study are the households within which more than one child was obtained. Also, each child was followed up for seven time points after baseline. Hence there is correlation between the responses when the same child is followed over a period of time and/or the influence of two or more children sharing the same household environment (77). In random effects model, the residual variance is split into components that pertain to the different levels in the study. The two-level model with the grouping of children within the same household would include the residuals at the household and child level where as the one-level model which includes children with different follow-up time measured with the same child would include the residuals at the child level only (48, 78).

There are three levels in the present study which are follow-up times, children and household from where more than one child was obtained. The diagrammatic representation of the three levels is shown in figure 5.6

Figure 5.6: Schematic representation of cluster levels:

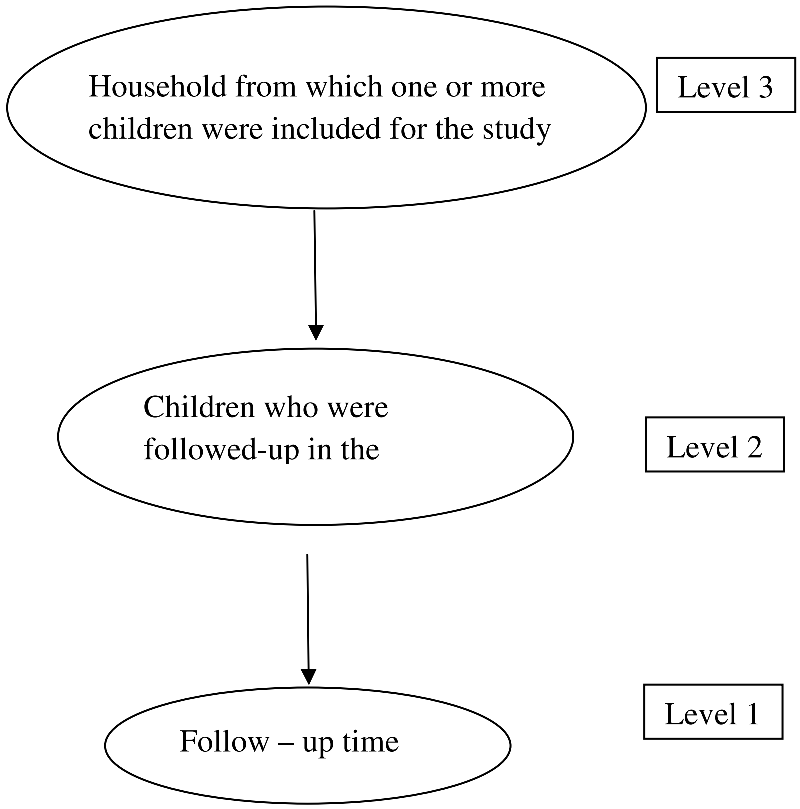


Figure 5.6 presented above represents the level of hierarchy with the lowest level being the 'follow-up time'. Each child's anthropometric measurements were recorded at baseline and every child's anthropometric measurements were again recorded for seven time points after every six months. Hence the next higher level above follow-up time is 'child level'. It was possible that if there were children who were in the age groups 5 – 7 years were included even if they belonged to the same household. Hence the next higher level is the 'household level'. Hence adjustments for the correlation in the responses need to be adjusted at each level. Hence random coefficients model with random intercept and random slope was considered at each level.

5.6.1 Micro and Macro level units:

The levels which are embedded within another level are also known as 'micro' levels where as the upper levels are referred to as 'macro' levels. In the present study, the macro level is the household and micro levels are children from the same household and follow-up of each child.

5.6.2 Aggregation:

A common procedure to address if certain risk factors are associated with macro level data is to aggregate the micro-level (lower level) data to macro-level data (higher level).

There can be three errors that happen with aggregation of data. The first potential error is the 'shifting of meaning'. A variable aggregated to macro level refers to macro unit and not micro units.

The second potential error with aggregation is the ecological fallacy. A correlation between macro-level variables cannot be used to make assertions about micro-level relations.

The third potential error is the neglect of the original data structure, especially when some kind of analysis of covariance is to be used.

5.6.3 The intraclass correlation (ICC):

The degree of resemblance between micro-units belonging to the same macro-unit can be expressed by the *intraclass correlation coefficient*. The term 'class' is conventionally used here and refers to the macro-units in the classification system under consideration. Let us assume a two-stage sampling design, and infinite populations at either level. The macro-units will also be referred to as *groups*.

5.6.4 Design effect:

It is the ratio of the variance obtained with the given sampling design to the variance obtained for a simple random sample from the same population, supposing that the total sample size is the same. A large design effect implies a relatively large variance. The design effect of a two-stage sample with equal group sizes is given by

Design effect = $1 + (n - 1)\,\rho_I$.

A relevant model here is the *random effect ANOVA* model. Indicating by Y_{ij} the outcome value observed for micro-unit i within macro-unit j, this model can be expressed as

$$Y_{ij} = \mu + U_j + R_{ij},$$

Where μ is the population grand mean, U_j is the specific effect of macro unit j, and R_{ij} is the residual effect from micro-unit i within this macro-unit. In other words, macro-unit j has the 'true mean' $\mu + U_j$, and each measurement of a micro-unit within this macro-unit deviates from this true mean by some value, called R_{ij}. Units differ randomly from one another, which is reflected by the fact that U_j is a random variable and the name 'random effects model. It is assumed that all variables are independent, the group effects U_j having population mean 0 and population variance σ^2 (the *population within-group variance*).

The total variance of Y_{ij} is then equal to the sum of these two variances,

$\mathrm{Var}(Y_{ij}) = \tau^2 = \sigma^2$

The number of micro-units within the j the macro-unit is denoted by n_j. The number of macro-units is N, and the total sample size is $M = \Sigma_j\, n_j$.

The intraclass correlation coefficient ρ_I can be defined as

$$\rho_1 = \frac{population\ variance\ between\ macro - units}{Total\ variance} = \frac{\tau^2}{\tau^2 + \sigma^2}$$

where, τ^2 is the between group or macro units' variance and σ^2 is the within group variance. It is the proportion of variance that is accounted for by the group level.

5.6.5 Within-group and between-group variance:

To disentangle the information contained in the data about the population between-group variance and the population within-group variance, we consider the *observed variance between group* and the *observed variance within groups.* These are defined in the following way. The mean of macro-unit *j* is denoted $\bar{Y}_{.j} = \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ij}$,

And the overall mean is $\bar{Y}_{..} = \frac{1}{M}\sum_{j=1}^{N}\sum_{i=1}^{n_j} Y_{ij} = \frac{1}{M}\sum_{j=1}^{N} n_j\,\bar{Y}_{.j}$.

The observed variance within group *j* is given by $S_j^2 = \frac{1}{n_j - 1}\sum_{i=1}^{n_j}\left(Y_{ij} - \bar{Y}_{.j}\right)^2$.

This number will vary from group to group. To have one parameter that expresses the within-group variability for all groups jointly, one uses the observed within-group variance, or pooled within-group variance. This is a weighted average of the variances within the various macro-units, defined as $S_{withinin}^2 = \frac{1}{M-N}\sum_{j=1}^{N}\sum_{i=1}^{n_j}\left(Y_{ij} - \bar{Y}_{.j}\right)^2 = \frac{1}{M-N}\sum_{j=1}^{N}\left(n_j - 1\right)S_j^2$.

The expected value of the observed within-group variance is exactly equal to the population within-group variance. For equal group sizes n_j, the observed between-group variance is defined as the variance between the group means,$S_{between}^2 = \frac{1}{(N-1)}\sum_{j=1}^{N}\left(\bar{Y}_{.j} - \bar{Y}_{..}\right)^2$

For unequal group sizes, the contributions of the various groups need to be weighted

$$S_{between}^2 = \frac{1}{\tilde{n}(N-1)}\sum_{j=1}^{N} n_j\left(\bar{Y}_{.j} - \bar{Y}_{..}\right)^2$$

In this formula, $\tilde{n}$ is defined by $\tilde{n} = \frac{1}{(N-1)}\left\{M - \frac{\sum_j n_j^2}{M}\right\} = \bar{n} - \frac{s^2(n_j)}{N\,\bar{n}}$,

where, $\bar{n} = M/N$ is the mean sample size and

$S^2(n_j) = \frac{1}{N-1} \sum_{j=1}^{N} (n_j - \bar{n})^2$ is the variance of the sample sizes.

The total observed variance is a combination of the within-group and the between-group variances, expressed as follows:

observed $total$ variance $= \frac{1}{M-1} \sum_{j=1}^{N} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_{..})^2 = \frac{M-N}{M-1} S^2_{within} + \frac{\bar{n}\,(N-1)}{M-1} S^2_{between}.$

The standard error of this estimator in the case where all group sizes are constant, $n_j = n$, is given by $S.E.(\hat{\rho}_I) = (1 - \rho_I)(1 + (n-1)\rho_I) \sqrt{\frac{2}{n\,(n-1)(N-1)}}$

5.6.6 Random Effects model for ordinal response:

The random effects model for ordinal response is the usual ordinal model which is *cumulative odds or proportional odds model*

The ordinal logistic regression model for random effects is given as:

$Pr(y_{ijt} > s \mid x_{ijt}) = \frac{\exp(\beta_0 + \beta_i x_{ijt} + Z_{ijt}\varepsilon_i) - k_s}{1 + \exp(\beta_0 + \beta_i x_{ijt} + Z_{ijt}\varepsilon_i) - k_s}$

where, y_{it} is the ordinal response for i^{th} individual at the t^{th} time point within each individual 'i' and within the jth household. Z_{it} is the covariates corresponding to random effects, ε_i

$\xi_i \sim N(0, \Sigma)$

's' = 1,2 and 3 represents the number of categories for ordinal response

where, 1 = normal, 2 – moderate and 3 – severe

k_s is the category specific parameter (43, 79, 80).

5.6.7. Intraclass correlation coefficient (ICC) for ordinal response:

Random intercept model means that each level is assumed to have different intercepts (risk) at baseline which needs to be taken into account. The intraclass correlation indicates the proportion of unexplained variance at the subject level; In other words, it reflects the magnitude of the between – subject variance. Reference is made to the threshold concept and the underlying latent response tendency that determines the observed ordinal response. For a logistic regression model, the latent response tendency, which is unobserved, is assumed to follow a standard logistic distribution, which has variance equal to $\pi^2/3$. Thus, for the logistic model assuming normally distributed random effects, the ICC equals $ICC = \frac{between\ cluster\ variance}{\left(\frac{\pi^2}{3}+between\ cluster\ variance\right)}$

5.6.8 Estimation:

Numerical integration was used to perform the integration over the random – effects distribution. Specifically, if the assumed distribution is normal, Gauss – Hermite quadrature can approximate the above integral to any practical degree of accuracy. Additionally, like the Laplace approximation, the numerical quadrature approach yields a deviance that can be readily used for likelihood-ratio tests. The integration is approximated by a summation on a specified number of quadrature points for each dimension of the integration. An issue with the quadrature approach is that it can involve summation over a large number of points, especially as the number of random – effects is increased (81). To address this, methods of adaptive quadrature have been developed that use a few number of points per dimension that are adapted to the location and dispersion of the distribution to be integrated (82).

5.6.9 Evaluation of Random intercept and/or random slope at each level:

The evaluation of random intercept and/or random slope is evaluated by taking the differences in the -2 log likelihood when that particular intercept or slope was included and excluded in the model. The -2 log likelihood was checked for sex of the child and area of residence in which the child lives.

5.7. Markov Chain:

A stochastic process is a collection of random variables indexed to time, t and the state, X. For example, we can write $\{X_t, t \in T\}$.

When T is finite, we refer it to as a countable stochastic process.

5.7.1 Definition of Markov Chain:

The stochastic process $\{X_t, t = 0, 1, 2, \ldots\}$ is called a Markov chain, if, for $j, k, j_1, \ldots j_{n-1} \mathcal{E} N$ (or any subset of i), $\Pr\{X_t = k \mid X_{t-1} = j, Xt_{-2} = j_1, \ldots, X_0 = j_{t-1}\} = \Pr\{X_t = k \mid X_{t-1} = j\} = p_{jk}$

In other words, conditioning on the history of the process up to stage n is equivalent to conditioning only on the most recent value Xt. When the present is given, *the future is independent of the past.*

The outcomes E_j (or the values j) are called the states of the Markov chain; if Xt has the outcome E_j (i.e. $X_t = j$), the process is said to be at state E_j (or simply at state j) at tth trial. To state j (or outcome E_j) there is no longer a fixed probability $\Pr\{Xt = j\}$ but to a pair of states (j, k) at the two successive trials (say, tth and $(t + 1)$st trials) there is a conditional probability p_{jk}. It is the probability of transition from the state j at nth trial to the state k at $(t + 1)$st trial. The transition probabilities p_{jk} are basic to the study of the structure of the Markov chain.

The transition probability may or may not be independent of t. If the transition probability p_{jk} is independent then Markov chain is said to be *homogeneous* (or to have *stationary transition probabilities*). If it is dependent on n (in which case it is denoted by $(^{t)}p_{ij})$, the chain is said to be non-homogeneous. In this thesis we confine to homogeneous chains.

The transition probability p_{jk} refers to the states (j, k) at two *successive* trials which may be tth and $(t + 1)$st trial then this transition is one-step and p_{jk} is called one-step (or unit step) transition probability. In the more general case, we are concerned with the pair of states (*j, k*) at two non-successive trials, say, state *j* at the *t*th trial and state *k* at the (*t* + *m*) the trial. The corresponding transition probability is then called *m*-step transition probability and is denoted by $p_{ij}^{(m)}$, i.e.

$p_{ij}^{(m)}$ = Pr $\{X_{t+m} = k \mid X_t = j\}$

The transition probabilities *pjk* satisfy

$P_{jk} > 0$, *and* $\sum_K pjk = 1$ for all *j*

These probabilities may be written in the matrix form $P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$

This is called the *transition probability matrix* or *matrix of transition probabilities* (t. p. m.) of the Markov chain. *P* is a *stochastic* or *Markov matrix*, i.e. a square matrix with non-negative elements and unit row sums (83).

5.7.2 Chapman-Kolmogorov equations in Markov chains:

All finite state space homogeneous Markov chains satisfy

$P_{m+k} = P_m P_k = P^{m+k}$ which implies that $P_m = P^m$ where P is the 1-step transition probability matrix. The matrix of m-step transition probabilities is the mth power of P, the matrix of one-setp transition probabilities.

If the number of steps is large, the transition probabilities are then called ergodic transition probabilities and are given be equilibrium probabilities $\delta = \lim_{n\to\infty} P^t$

Chapman Kolmogorov equations imply that $\delta = P\delta$ which provides a system of linear equations that are used to calculate the ergodic probabilities using the fact: $\sum_{i\in X} \delta_i = 1$; $\delta_i > 0$. A Markov chain with transition probability matrix P is said to have a stationary distribution δ if $\delta = P\delta$ and

$1\acute{\delta} = 1$. In the present study, the Markov chain is positive recurrent as starting at state i, the process will move to state j and return to state i. Also, the expected time for the process to re-enter state i is finite.

5.7.3 First Passage Times:

The first passage time from state i to state j is the number of transitions made by the chain in going from state i to state j for the first time.

Let $f_{ij}(t)$ be the probability that the first passage time from state i to state j is equal to t(=1,2….)

First passage time probabilities satisfy recursive relationship

$$f_{ij}(1) = p_{ij}$$

$$f_{ij}(2) = p_{ij}^{2} - f_{ij}(1)p_{ij}$$

$$\vdots$$

$$f_{ij}(t) = p_{ij}^{t} - \sum_{k=1}^{n-1} p_{ij}(t-k)$$

5.7.4 Mean First Passage time (MPT):

The expected first passage time from state i to state j is given as $\mu_{ij} = \sum_{t=1}^{\infty} t f_{ij}(t)$

The chain will move from state i to state j in one transition with probability p_{ij}.

For k≠j, the chain moves to state k with probability p_{ik}

On an average the number of transitions to visit j for the first time from i is $1 + \mu_{kj}$ transitions to move from state i to state j. Therefore, $\mu_{ij} = 1 + \sum_{k \neq j} \mu_{kj} p_{kj}$

The mean first passage times were calculated solving the above equations

Also, $\mu_{ii} = \frac{1}{\delta_i}$ which is the recurrent time for state i

5.7.5 Variance of the Mean First Passage Time:

The state i of a Markov chain is an absorbing state if pii = 1. A Markov chain is an absorbing Markov chain if and only if the following two conditions are satisfied:

1. The chain has at least one absorbing state
2. It is possible to go from any non absorbing state to an absorbing state.

The fundamental matrix for an absorbing Markov chain is defined as matrix Z , where

$Z = (In - Q)^{-1}$ where Q is such that P = $\begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$ where I is the identity matrix.

The fundamental matrix gives the expected number of visits to each state before absorption occurs. The mean first passage time Matrix M was obtained using M = (I – Z + EZ)D

where, Z is the fundamental matrix for P, D is the diagonal matrix with diagonal

elements dii = 1/δI (83). The variance of first passage time was obtained using the equation (84)

$Vi(fj) = Mi[f^2_j] - Mi[fj]^2$ where $M = [\mu ij]$ and $W = (Mi[fj^2]) = P[W - Wdg] - 2P[Z - EZ]D + E$

5.7.6 95% Confidence Interval for the first mean passage time:

The 95% confidence interval was obtained using Monte Carlo simulation of the first passage times. The first passage time was calculated for 1000 transitions using the transition probability. For any large stationary Markov chain, the cumulative probability is close to one. Hence the number of transitions in first passage time was decided based on the cumulative probability. These 1000 transitions' first passage time was then simulated for 10,000 times using Monte Carlo simulation. The 5th percentile and the 95th percentile were obtained as the 95% confidence interval for the mean passage time. This procedure was repeated for each cell in the transition probability matrix and the 95% confidence interval was obtained for cell in the transition probability matrix. The R program to calculate MPT is given below.

5.7.6 R Program to calculate the mean passage time:

```
###Overall Mean Passage Time###

rm(list=ls())

nstate = 3

N = matrix(c(8486,1213,85,1077,1954,482,207,556,1210), nstate, nstate, byrow=T)

P = N/rowSums(N) #transition probability matrix

# To find the limiting distribution

B = diag(nstate) - P + 1

delta = solve(t(B),rep(1,nstate)) #solving for solutions in a matrix

# Equation (2.3) of the reference

# Variances of first passage times in a Markov chain with applications to mixing times

#J. J. HUNTER. This is equivalent to a simplified formula obtained using fundamental matrix

# Reference: Finite Markov chains / by John G. Kemeny and J. Laurie Snell.

# Formula for mean of the first passage time: Theorem 4.4.7, eq(2), page 79

# Formula for variance of the first passage time: Theorem 4.5.3, page 83 (see W)

# Fundamental matrix = Z = (I-P-PI), (PI is the same as A in the book)

library(MASS)

I = diag(1,nstate)

e = matrix(1,nstate,1)

D = diag(1/delta,nstate)

PI = e%*%delta

Z = solve(diag(1,nstate)-P+PI)

E =  e%*%t(e)
```

```
M = (I - Z + E%*%diag(diag(Z),nstate))%*%D

M # Mean first passage time

# To obtain the second moment

M2=2*(Z%*%M - E%*%diag(diag(Z%*%M),nstate))+M%*%(2*diag(diag(Z),nstate)%*%D- I)

M2 # second moment of the first passage time

var.MFT = M2 - M*M # variance

sd.MFT = sqrt(var.MFT)

sd.MFT

# Function to compute mean passage time and its CI by Monte Carlo simulation

mpt = function(f){

# Computes the exact distribution of FPT

        for(i in 2:k){

                Q = Q%*%P

                term2 = 0

                Q1 = diag(nstate)

                for(j in (i-1):1)

                {

                        Q1 = Q1%*%P

                        term2 = term2 + f[j]*Q1[s,s]

                }

                f[i] = Q[r,s] - term2

        }
```

```
# This is needed for small probabilities (mostly in stationary distribution)
# The transtions more 100 had very high probabilities
# Setting the maximum value for the FPT because beyond that the probabilities are close to zero
    k0 = sum(cumsum(f) <= 0.99999)
    f0 = c(f[1:k0], 1-sum(f[1:k0]))
# Monte Carlo simulation -
# Simulating 1000 FPT from the exact distribution over (1, 2, ..., k0, k0+1) and finding its mean
# Repeating this m times
    m = 10000
    frs = rep(0,m)
    for(i in 1:m)
    {
        frs[i] = mean(sample(1:(k0+1),100,prob=f0,replace=TRUE))
    }
        frs[i]
    mean(frs) # compare this with M[r,s]
    var(frs)
    frs.LCL = quantile(frs,0.025)
    frs.UCL = quantile(frs,0.975)
    ppt = list(frs.LCL, mean(frs), frs.UCL)
}
```

```
# Call of function mpt to compute mean and CI for 9 combinations
# assuming that the probability of transitions more than 1000 would be extremely small
k = 1000
mpt.mean = matrix(0,nstate, nstate)
mpt.LCL = matrix(0,nstate, nstate)
mpt.UCL = matrix(0,nstate, nstate)
# Loop to go through all possible combinations of states
for(r in 1:nstate)
    {
        for(s in 1:nstate) {
                f = rep(0,k+1)
                f[1] = P[r,s]
                Q = P
                temp = mpt(f)
                mpt.LCL[r,s] = temp[[1]]
                mpt.mean[r,s] = temp[[2]]
                mpt.UCL[r,s] = temp[[3]]  }
}
}
mpt.LCL
mpt.UCL
```

5.7.7 Testing Hypothesis for Mean Passage Time:

In order to test the hypothesis that the MPT at various levels of malnutrition for risk factors such as rural and urban children are different, the Log linear model was used. The cell counts are modeled to see if there is any association between the variables. In other words, a log linear models that were fitted enabled us to find if the transition probabilities were different across the categories of the risk factor.

The saturated log linear model with only time t, time t+1 and was

$$\log(m_{ij}) = \mu + \lambda_i^y + \lambda_j^z + \lambda_{ij}^{yz}$$

$$\log(m_{ij}) = \mu + \lambda_i^y + \lambda_j^z + \lambda_{ij}^{yz} + \lambda_q^x + \lambda_{qi}^{xy} + \lambda_{qj}^{xz}$$

where, x is the risk factor being tested (x= 1,2,…c_1)

and state of malnutrition y = 1,2,3 at time t and z = 1,2,3 at time t+1

A particular risk is considered significant if the calculated deviance G^2 (which is the difference in likelihood value) was significant when that particular risk factor was included and excluded from the model.

The log linear model was assessed for sex of the child, area of residence from where the child was taken for the study and three hypothesized risk factors which are presence of a separate kitchen in the house, defecation and type of fuel used for cooking (53).

5.8. Markov Regression:

The main property of Markov chain is that the outcome in the future state is mainly affected by the present state and not past states. Markov regression is a regression technique which accounts for the present state to model the future state. Markov regression can be analyzed using modeling transition probabilities and modeling intensity rate between the states. Markov regression analysis was done only with those observations that had at least one transition. The remaining observations that had only one observation over time were excluded.

5.8.1 Markov Regression analysis using transition probabilities:

Markov regression using transition probabilities is a technique where the present state response is treated as additional covariates to the usual risk factors 'x_{it}'. Hence the transition model expresses the conditional probability as a function of both the risk factors and the present state response. These models are fit when observed times are equally spaced.

The ordinal logistic regression model is given as:

logit {Pr ($y_{i(t+1)}$ > s | x_{it}, $H_{i(t+1)}$) = $x_{i(t+1)}'\beta_q + \omega y_{it}$

where, history of a child 'i' $H_{i(t+1)}$ = {y_{ik}, k = 1,2….t}

The first order Markov chain is given as

$$\begin{pmatrix} \pi_{11} & \pi_{12} & \pi_{13} \\ \pi_{21} & \pi_{22} & \pi_{23} \\ \pi_{31} & \pi_{32} & \pi_{33} \end{pmatrix}$$

where, $\pi_{ab} = \Pr\left(y_{i(t+1)} > b \middle| y_{it} = a\right)$; a,b= 1,2,3

Each transition matrix sums upto 1. The transition models are modeled as functions of covariates

$x_{i(t+1)} = (1, x_{it1}, x_{it2}, x_{itp})$

A very general model uses separate logistic model for

Logit Pr($y_{i(t+1)} = 1|y_{it} = 1) = x_{\acute{\iota}(\acute{t}+1)}\beta_1$

Logit Pr($y_{i(t+1)} = 1|y_{it} = 2) = x_{\acute{\iota}(\acute{t}+1)}\beta_2$

Logit Pr($y_{i(t+1)} = 1|y_{it} = 3) = x_{\acute{\iota}(\acute{t}+1)}\beta_3$

which means that the effects of models differ depending on the present state of response. A more concise form of the model is

logit Pr($Y_{i(t+1)} = 1|Y_{it} = y_{it}) = x_{\acute{\iota}(\acute{t}+1)}\beta_0 + y_{it}x_{\acute{\iota}(\acute{t}+1)}\omega$ so that $\beta_1 = \beta_0 + \omega$

The above equation expresses the logistic model which includes as predictors the present state y_{it} as well as the interaction of y_{it} and risk factors. This enables us to test whether the particular risk factor have the same effect on the response probability whatever may be the state of y_{it} (may be y_{it} = 1, 2 or 3).

A first-order Markov model the contribution to the likelihood for the *i*th subject can be written as

$$L_i(\mathcal{Y}_{i1}, \ldots, \mathcal{Y}_{in_i},) = f(\mathcal{Y}_{i1}) \prod_{j=2}^{n_i} f(\mathcal{Y}_{it} \mid \mathcal{H}_{it})$$

In a Markov model of order *q*, the conditional distribution of Y_{it} is

$$f(\mathcal{Y}_{it} \mid \mathcal{H}_{it}) = f(\mathcal{Y}_{it} \mid \mathcal{Y}_{it-1}, \ldots, \mathcal{Y}_{it-q}),$$

so that the likelihood contribution for the *i*th subject becomes

$$f(\mathcal{Y}_{i1}, \ldots, \mathcal{Y}_{iq}) \prod_{j=q+1}^{n_1} f\left((\mathcal{Y}_{it} \mid \mathcal{Y}_{it-1}, \ldots, \mathcal{Y}_{it-q}\right)$$

In the logistic case, $f(\mathcal{Y}_{i1}, \ldots, \mathcal{Y}_{iq})$ is not determined from the GLM assumption about the conditional model, and the full likelihood is unavailable. An alternative is to estimate β and α by maximizing the conditional likelihood

$$\prod_{i=1}^{m} f\left(\mathcal{Y}_{iq+1}, \dots, \mathcal{Y}_{in_i} | \mathcal{Y}_{i1}, \dots, \mathcal{Y}_{iq}\right) = \prod_{i=1}^{m} \prod_{t=q+1}^{n_i} f(\mathcal{Y}_{it}|t).$$

when maximizing the above equation. There are two distinct cases to consider. In the first, $f_r(\mathcal{H}_{it}; \alpha, \beta) = \alpha_r f_r(\mathcal{H}_{it})$ so that $h\left(\mu_{it}^C\right) = x_{it}'\beta + \sum_{r=1}^{s} \alpha_r f_r(\mathcal{H}_{it})$. Here, $h\left(\mu_{it}^C\right)$ is a linear function of both β and $\alpha = (\alpha_1, \dots, \alpha_s)$ so that estimation proceeds as in generalized linear models (GLMs) for independent data. We simply regress Y_{it} on the $(p + s)$ – dimensional vector of extended explanatory variables $\left(x_{it}, f_1(\mathcal{H}_{it}), \dots, f_s(\mathcal{H}_{it})\right)$. The second case occurs when the functions of past responses include both α and β. To derive an estimation algorithm for this case, the derivative of the log conditional likelihood has the form

$$S^C(\delta) = \sum_{i=1}^{m} \sum_{t=q+1}^{n_i} \frac{\partial \mu_{ij}^C}{\partial \delta} v_{it}^{c^{-1}} \left(\mathcal{Y}_{it} - \mu_{it}^C\right) = 0,$$

where δ = (β, α). This equation is the conditional analogue of the GLM likelihood equation. The derivative $\partial \mu_{it} / \partial \delta$ is analogous to x_{it} but it can depend on α and β. Formulation of the estimation procedure as an iterative weighted least squares as follows. Let $\boldsymbol{Y}_i$ be the $(n_i - q)$-vector of responses for $j = q + 1, \dots, n_i$ and μ_{it}^C its expectation given $\mathcal{H}_{it}$. Let X_i^* be an $(n_i - q) \times (p + s)$ matrix with uth row $\partial \mu_{iq} / \partial \delta$ and $\boldsymbol{W_i} = \boldsymbol{diag}\left(\boldsymbol{1}/v_{ik+q}^c, u = 1, \dots, n_i - q\right)$ ***an*** $(n_i - q) \times (n_i - q)$ diagonal weighting matrix. Finally, let $Z_i = X_i^* \hat{\delta} + \left(Y_i - \hat{\mu}_i^C\right)$. Then, an updated $\hat{\delta}$ can be obtained by iteratively regressing $\boldsymbol{Z}$ on $\boldsymbol{X^*}$ using weights $\boldsymbol{W}$.

When the correct model is assumed for the conditional mean and variance, the solution $\hat{\delta}$ of $S^C(\delta)$ asymptotically, as m goes to infinity, follows a Gaussian distribution with mean equal to the true value, δ, and $(p + s) \times (p + s)$ variance matrix

$$V_{\delta} = \left(\sum_{i=1}^{m} X_i^{*\prime} W_i X_i^{*} \right)^{-1}$$

The variance $V_{\hat{\delta}}$ depends on β and α. A consistent estimate, $\hat{V}_{\hat{\delta}}$, is obtained by replacing β and α by their estimates $\hat{\beta}$ and $\hat{\alpha}$. Hence a 95% confidence interval for β_1 *is* $\hat{\beta}_1 \pm 2\sqrt{\hat{V}_{\hat{\delta}11}}$, *where* $\hat{V}_{\hat{\delta}11}$ is the element in the first row and column of $\hat{V}_{\hat{\delta}}$.

If the conditional mean is correctly specified and the conditional variance is not, we can still obtain consistent inferences about δ by using the robust variance, which here takes the form

$$V_R = \left(\sum_{i=1}^{m} X_i^{*\prime} W_i X_i^{*} \right)^{-1} \left(\sum_{i=1}^{m} X_i^{*\prime} W_i V_i W_i X_i^{*} \right) \left(\sum_{i=1}^{m} X_i^{*\prime} W_i X_i^{*} \right)^{-1}.$$

A consistent estimate $\hat{V}_R$ is obtained by replacing $V_i = Var(Y_i | \mathcal{H}_i)$ in the equation above by its estimate, $\left(Y_i - \hat{\mu}_i{}^C\right)\left(Y_i - \hat{\mu}_i{}^C\right)'$ (76).

5.8.2. Markov Regression using intensity rate:

Consider a process X = (x(t), t ≥ 0). A multistate process is a process that can take a finite number of states that is for any t, x(t) has values in {0,1,….k}. The law of multistate process can be specified by the transition probabilities Pr{x_t = j| x_{t-1} = k}.

Multistate models are governed by transition intensity functions. The transition intensity function represents the instantaneous incidence rate of moving from one state j to another state k at time t:

$$\lambda_{jk} = \lim_{\Delta t \to 0} \frac{P\{Y(t+\Delta t) = k | Y(t) = j}{\Delta t}, j \neq k$$

where Y(t) is the state occupied at time t. This transition intensity is the (j,k) entry of the transition intensity matrix denoted by Λ, the rows of which sum to zero. The model is time homogeneous, which means that $\lambda_{jk}(t) = \lambda_{jk}$ for all t. The diagonal entries of Λ are defined by the convention as: $\lambda_{ij}(t) = -\sum_{k \neq j} \lambda_{jk}(t)$

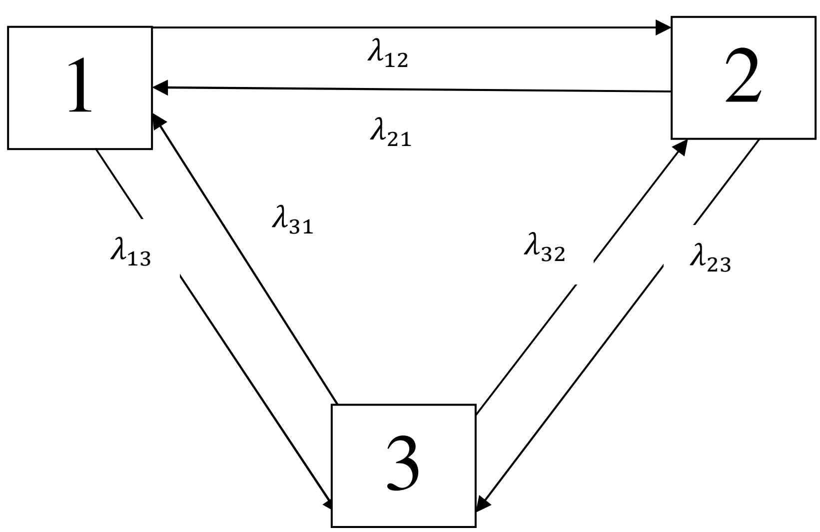
These above transition intensities can be used to calculate transition probabilities, probability of being in state j at time s and then in state k at time t:

$$p_{jk}(s,t) = P\{Y(t) = k|Y(s) = j\}, 0 \leq s < t$$

$p_{jk}(s,t)$ is not the actual time of transition from state j to stake k as the process has certainly entered other states between times s and t (Sutradhar et al., 2011). This probability is the (j,k) entry of the transition probability matrix denoted by P(s,t), the rows of which sum to 1. Based on the 3 state model in the study, the transition intensity matrix is:

$$\Lambda = \begin{bmatrix} -(\lambda_{12} + \lambda_{13}) & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & -(\lambda_{21} + \lambda_{23}) & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & -(\lambda_{31} + \lambda_{32}) \end{bmatrix}$$

The underlying 3-state model for examining progression among malnourished children (1-normal, 2 – moderate and 3 – severe)



The model includes other risk factors and hence the model is $\log(\lambda_{jk}) = \beta_{0jk} + \sum_{l=1}^{L} \beta_{ljk}$

where, l is the number of risk factors (l = 1,2,….L)

Suppose there are a random sample of n individuals at times $t_0, t_1, \ldots, t_m$.

Let n_{jkq} be the number of individuals in state j at t-1 and k at t and if we condition on the distribution of individuals among states at t_0, then the likelihood function for **θ** is

$$L(\boldsymbol{\theta}) = \prod_{q=1}^{m} \left\{ \prod_{j,k=1}^{r} p_{jk}(t_{q-1}, t_q)^{n_{jkq}} \right\}$$

when the time is homogeneous, $w_q = t_q - t_{q-1}$, q = 1,2,3,…m gives the log likelihood

$$L(\boldsymbol{\theta}) = \sum_{q=1}^{m} \sum_{j,k=1}^{r} n_{jkq} log p_{jk} w_q$$

The maximum likelihood estimate (MLE $\widehat{\boldsymbol{\theta}}$) is obtained by maximizing the above equation

Quasi-Newton procedure that use the first derivatives of log L(**θ**) which leads to faster convergence and an estimate of the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}$. This approach is made feasible by the provision of an efficient algorithm for the computation of P(t;**θ**) and its derivative with respect to **θ**.

$p(t; \boldsymbol{\theta}) = exp\{Q(\boldsymbol{\theta})t\}$ for a given **θ** is computed using a canonical decomposition. If, for the given **θ**, $Q(\boldsymbol{\theta})$ has distinct eigenvalues $d_1, \ldots, d_k$ and A is the $k \times k$ matrix whose jth column is a right eigenvector corresponding to d_j, then $Q = ADA^{-1}$, where $D = \text{diag}(d_1, \ldots, d_k)$. Then

$$P(t) = A \text{ diag } (e^{d1t}, \ldots, e^{dkt})A^{-1},$$

where the dependence of Q, $P(t)$, A and the d_j's on **θ** is suppressed for notational convenience. $Q(\boldsymbol{\theta})$ has distinct eigenvalues for almost all **θ** (Kalbfleisch at el., 1985).

5.8.3 Further Aspect of Estimation:

Implementing the algorithm for maximum likelihood estimation is as given below:

An initial estimate θ_0 of θ is usually obtained in an ad hoc way by examining the transition counts n_{jkq}. An alternative approach would involve a preliminary tabulation of the likelihood surface. There is an advantage to parameterizing the model by waiting $q_{jk} = \exp(\alpha_{jk}), j \neq k$. This is because the parameters α_{jk} can take any real value whereas $q_{jk} \geq 0$. This reparameterization avoids problems that can arise when an iteration results in parameter vectors outside the parameter space. It is possible to have $q_{jk} = 0$. When this happens, successive iterates of α_{jk} will typically become large and negative. In this situation it is useful to set the corresponding $q_{jk} = 0$ and fit the model, against using the α_{jk} 's, with one less parameter. The resulting estimate is then compared with one in which q_{jk} is taken as a small positive value. When the time w_t between successive observations are large, it is clear on intuitive grounds that not all parameters will be well estimated. The result is a likelihood surface for the q_{jk}'s or α_{jk}'s that has ridges defined by certain parameters that are imprecisely estimated. For example, if the w_l's are very large and the process is ergodic, then $p_{jk}(w_t) = \pi_j$, where $\pi' = (\pi_1, \ . \ . \ . \ , \pi_r)$ is the vector of equilibrium probabilities. In this case, a large number of individuals under study will allow precise estimation of π, but individual q_{jk}'s will be estimated imprecisely. If $w_l = w$ for all l, it may be possible to determine $\hat{\theta}$ in a relatively simple way. The empirical transition matrix $\tilde{p}_{jk\cdot}(w) = n_{jk} / n_{j\cdot\cdot}$, where $n_{jk\cdot} = \sum_{l=1}^{m} n_{jkl}$ *and* $n_{j\cdot\cdot} = \sum_{j=1}^{s} n_{jk}$, provides an estimate of $P(w)$. If the equation $\tilde{p}(w) = \exp(Qw)$ admits a solution $\hat{Q} = Q(\hat{\theta})$, then $\hat{\theta}$ is a MLE of θ.

5.8.4 Incorporating the covariates:

In most situations, there are measured covariates on each individual under study, and interest is on the relationship between these covariates and the intensities q_{jk} in the Markov model. One advantage of the methods described above is the generalization in a straightforward way to allow for the regression modeling of Q, though with many distinct covariate values in the sample, computations is too expensive to be easily implemented directly. Implementation may require that the covariates be grouped. Suppose that each individual has an associated vector of s covariates, $z' = (z_1, z_2, \ . \ . \ . \ . \ , z_s)$, where $z_1 = 1$. For given z, we suppose that the process is homogeneous Markov with transition intensity matrix $Q(z) = \left(q_{jk}\ (z)\right)$, Where $q_{jk}(z) = \exp\left(z'^{\beta_{jk}}\right)$, $\qquad j \neq k,$

And $q_{jk}\ (z) = -\sum_{j \neq i} q_{jk}(z)$. $\beta_{1jk}\ , \ . \ . \ . \ . \ , \beta_{sjk}'$ is a vector of s regression parameters relating the instantaneous rate of transitions from state j to state k to the covariates z. The algorithm requires a separate canonical decomposition of $Q_{(z)}$ for each of the r distinct covariate vectors z in the sample.

Let these be denoted by $z_h = (z_{1h}, \ . \ . \ . \ . \ , z_{sh})$ with $z_{1h} = 1$, and let $Q_h = Q(z_h) = (q_{jk}\ (z_h))$, $h = 1 \ , \ . \ . \ . \ . \ , \ r$,Let $n_{jkl}^{(h)}$ be the number of individuals with covariate values z_h that are in state j at t_{l-1} and k at t_l. The likelihood is then a product of terms where the hth term arises from data collected on a homogeneous model with intensity matrix Q_h. Thus the log-likelihood is

$$\log L(\theta) = \sum_{h-1}^{r} \sum_{l=1}^{m} \sum_{jk=1}^{s} n_{jkl}^{(h)} \log p_{jk}\ (w_l;\ z_h), \text{ where,} P_h(t) = (Q_h t) = \left(p_{jk}(t;\ z_h)\right).$$

The parameter θ is being used to indicate the vector of parameters in $\beta_{jk}\ (j \neq k)$ that are to be estimated (85).

5.9 Comparison of Markov regression models:

The comparison of Markov regression models was done considering binary outcome variable which was BMI and/or Height-for-age classified as “normal” and “moderate/severe” as the number of transitions from normal to severe was very small. The comparison for Markov regression with transition probabilities and intensity rates with GEE was done using only the five hypothesized variables. The comparison was done using simulations for 10,000 times. The log odds of the outcome (current state of malnutrition) was obtained using the mathematical model that involved the previous state of malnutrition and the five hypothesized variables. The regression coefficients were fixed to some value. Using these log odds values, the outcome was obtained using a binomial distribution. This was repeated for 10,000 times. The Markov regression with transition probabilities and GEE was performed for these obtained outcome at the current state. The coverage probability was calculated if the beta regression coefficients were included in the confidence limits. The average length of the confidence interval was also obtained for 10,000 simulations. The R program using height-for-age classification comparing GEE and Markov regression is provided below.

Program to compare GEE and Markov Regression using Simulations

```
rm(list=ls(all = TRUE))
library(foreign)
library(geepack)

datahtage <-
read.dta("C:\\Users\\keerthankavish\\Desktop\\PEM_NEW\\Chapters\\Merged_long_wide\\Mark
ov reg\\Diggle\\2012, Feb 10, htage_t1t2.dta"); # location and path where the file is stored
```

```
lengthmrar <- lengthdefmr <- lengthgeedef <- lengthargee <- 0

countmrdef <- countmrar <- countgeedef <- countgeear <- 0

p <- ob <- numeric(17556)

alpha <- 0.06 # estimates from the data for each risk factor

beta1 <- 0.64 # estimates from the data for each risk factor
beta2 <- 1.13 # estimates from the data for each risk factor
beta3 <- 274.01 # estimates from the data for each risk factor
sigma <- 200

y <- numeric(17556)

lengthdefmr <- lengthargee <- lengthgeedef <- lengthmrar <- numeric(10)

for (i in 1:10)

{

   for (j in 1:17556)

   {

  mu <- alpha+beta1*datahtage$defr[j]+beta3*datahtage$htagepbin[j] +
beta2*datahtage$arear[j]  # logistic model

 #Generating y's from the above model and also using binomial distribution

  ob[j] <- rnorm(1,mu,sigma)

  p[j] <- exp(ob[j]/1+exp(ob[j]))

  y[j] <- rbinom(1,1,p[j])

  }

length(datahtage$defr)

length(datahtage$htagepbin)

#performing logistic regression analysis with generated y's

estmr <- summary(glm(y~defr+htagepbin+arear,data=datahtage,family = binomial(link =
"logit")))

coefmatmr <- estmr$coefficients

#print(coefmatmr)

#performing gee with the generated y's

estgee <- summary(geeglm(y~defr+arear,id=idno,data=datahtage,family = binomial,corstr =
"ar1", std.err = "san.se"))
```

```
coefmatgee <- estgee$coefficients
#print(coefmatgee)
```

###Coverage probabilities and length of the confidence interval

```
###Markov regression coverage probabilities
lldefmr <- coefmatmr[2,1]-1.96*coefmatmr[2,2]
uldefmr <- coefmatmr[2,1]+1.96*coefmatmr[2,2]

###Markov regression length of the CI
lengthdefmr[i] <- (uldefmr - lldefmr)/2
if(lldefmr < beta1 & uldefmr > beta1)countmrdef = countmrdef+1

###Markov regression coverage probabilities
llarmr <- coefmatmr[3,1]-1.96*coefmatmr[3,2]
ularmr <- coefmatmr[3,1]+1.96*coefmatmr[3,2]

###Markov regression length of the CI
lengthmrar[i] <- (ularmr - llarmr)/2
if(llarmr < beta2 & ularmr > beta2)countmrar = countmrar+1

###Gee coverage probabilities
lldefgee <- coefmatgee[2,1]-1.96*coefmatgee[2,2]
uldefgee <- coefmatgee[2,1]+1.96*coefmatgee[2,2]

###Length of the CI
lengthgeedef[i] <- (uldefgee - lldefgee)/2
if(lldefgee < beta1 & uldefgee > beta1)countgeedef = countgeedef+1

###Gee coverage probabilities
llargee <- coefmatgee[3,1]-1.96*coefmatgee[3,2]
```

```
ulargee <- coefmatgee[3,1]+1.96*coefmatgee[3,2]###Length of the CI
lengthargee[i] <- (ulargee - llargee)/2
if(llargee < beta2 & ulargee > beta2)countgeear = countgeear+1
}

Countmrdef/10000
Countmrar/10000
Countgeedef/10000
Countgeear/10000

sum(lengthmrar)/10000
sum(lengthdefmr)/10000
sum(lengthgeedef)/10000
sum(lengthargee)/10000
```

RESULTS

6. RESULTS

6.1 Prevalence, Incidence and Cumulative Incidence of Malnutrition:

6.1.1 Prevalence, Incidence and Cumulative Incidence of Malnutrition using BMI classification:

The prevalence of malnutrition based on BMI classification at baseline is presented in table 6.1.1a. The prevalence of severe malnutrition at baseline was 22.5%, moderate malnutrition was 21.7%. Male children had higher prevalence of severe malnutrition (25%) than female children (19.9%). The prevalence of severe malnutrition in rural area was 16.5% while the prevalence in the urban area was higher (28%) (Table 6.1.1a).

Cumulative incidence of malnutrition based on BMI classification by area and sex of the child is presented in Table 6.1.1b. The table included malnutrition classification from follow-up 1 (every 6 months) to follow-up 7 (3.5 years). The cumulative incidence of severe malnutrition during the follow-ups was 11.6%. The cumulative incidence of severe malnutrition among male children was 14.7% which was higher than those of female children (8.4%) throughout the follow-up. The cumulative incidence of severe malnutrition was similar across rural (11.9%) and urban areas (11.3%) (Table 6.1.1b).

Table 6.1.1c presents the incidence density of severe malnutrition in each year of follow-up. The incidence density of severe malnutrition was calculated as children who transited from normal or moderate state at time t to time t+1. The incidence in the first year from baseline was found to be around 4%. It was also about 4.3% even in the second year from first year and increased to 6% from second to third year. The graphical representation is presented in figure 6.1.1 (Table 6.1.1c). The incidence and incidence density of malnutrition using BMI classification by age of the child, sex of the child and area of residence has been provided in the appendix (tables 1 - 3).

Table 6.1.1a: Prevalence of malnutrition using BMI classification at baseline by area and sex of the child

	Total	Prevalence of Malnutrition at Baseline					
		Normal		Moderate		Severe	
		n	%	n	%	n	%
Overall	2494	1391	55.8	542	21.7	561	22.5
Male	1271	670	52.7	283	22.3	318	25.0
Female	1223	721	59.0	259	21.2	243	19.9
Rural	1195	757	63.3	241	20.2	197	16.5
Male	594	359	60.4	125	21.0	110	18.5
Female	601	398	66.2	116	19.3	87	14.5
Urban	1299	634	48.8	301	23.2	364	28.0
Male	677	311	45.9	158	23.3	208	30.7
Female	622	323	51.9	143	23.0	156	25.1

Table 6.1.1b: Cumulative Incidence of Malnutrition by Area and Sex of the child using BMI classification for 3.5 years

	Total	Cumulative Incidence of Malnutrition during follow-up					
		Normal		Moderate		Severe	
		n	%	n	%	n	%
Overall	15629	10011	64.1	3801	24.3	1817	11.6
Male	8016	4798	59.9	2039	25.4	1179	14.7
Female	7613	5213	68.5	1762	23.1	638	8.4
Rural	7642	4812	63.0	1917	25.1	913	11.9
Male	3836	2266	59.1	995	25.9	575	15.0
Female	3806	2546	66.9	922	24.2	338	8.9
Urban	7987	5199	65.1	1884	23.6	904	11.3
Male	4180	2532	60.6	1044	25.0	604	14.4
Female	3807	2667	70.1	840	22.1	300	7.9

6.1.2 Prevalence, Incidence and Cumulative Incidence of Acute Malnutrition using Weight-for-age (underweight) classification:

The prevalence of acute malnutrition (underweight) by sex of the child and area of residence using weight-for-age is presented in Table 6.1.2a. The overall prevalence of severe underweight was 23% and the prevalence of severe underweight was higher in male children (26.8%) as compared to female children (19.1%). The rural and urban prevalence were similar (23.2% and 22.9% respectively) (Table 6.1.2a).

The cumulative incidence of overall acute malnutrition (underweight) as classified using weight-for-age is presented in Table 6.1.2b. The overall cumulative incidence of severe underweight was 10.8%. The cumulative incidence of severe underweight was higher among male children (15.5%) than female children (5.9%). The cumulative incidence of severe underweight among children was also higher in the rural area (13.3%) as compared to children from urban areas (8.4%) (Table 6.1.2b).

6.1.3 Prevalence, Incidence and Cumulative Incidence of Chronic Malnutrition (stunting) using Height-for-age classification:

The prevalence of chronic malnutrition (stunted) as classified using Height-for-Age is presented in Table 6.1.3a. The prevalence of severe chronic malnutrition was around 26%. The overall prevalence of severe chronic malnutrition was higher among boys (27.9%) than girls (23.5%). The prevalence of chronic malnutrition was higher in the rural than urban areas (33.2% vs 18.9%) (Table 6.1.3a).

The cumulative incidence of chronic malnutrition (stunted) was nearly 21%. The cumulative incidence for boys and girls was 24.1% and 16.9% respectively. The cumulative incidence of chronic malnutrition among children in rural areas was nearly 25% and for children from urban areas, the cumulative incidence was 16.1% (Table 6.1.3b). The incidence of malnutrition using height-for-age classification by age of the child, sex of the child and area of residence has been provided in the appendix (tables 4).

6.1.4. Prevalence, Incidence and Cumulative Incidence of Malnutrition (wasted) using Weight-for-height classification:

Table 6.1.4a presents the prevalence of malnutrition (wasting) as classified by weight-for-height by sex of child and area. The prevalence of wasting was 4.3% and it was 5.1% in boys as compared to 3.4% among girls. The prevalence of wasting among children was lower in rural areas (1.9%) as compared to children living in urban areas (6.4%) (Table 6.1.4a).

The cumulative incidence of malnutrition (wasting) as represented by weight for height classifications is presented in Table 6.1.4b. The wasting cumulative incidence was 0.7% and cumulative incidence was 1% in boys as compared to 0.4% in girls. There was also a slight difference in the cumulative incidence of wasting in rural and urban areas (0.5% vs 0.9%). This implies that the urban children are better in their height growth than rural children. Therefore, wasting appears to be lower in rural areas than in the urban areas. The cumulative incidence was also higher for boys within the rural and urban areas (Table 6.1.4b).

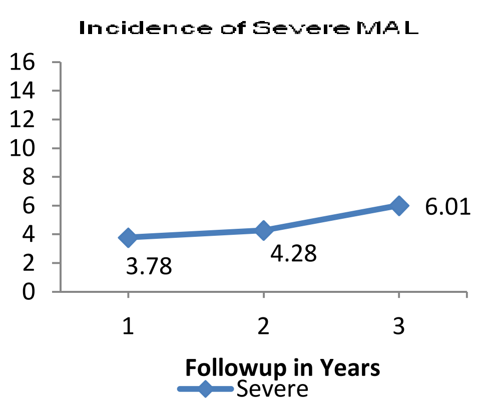
Table 6.1.1c: Incidence of Severe Malnutrition using BMI Categories for all children:

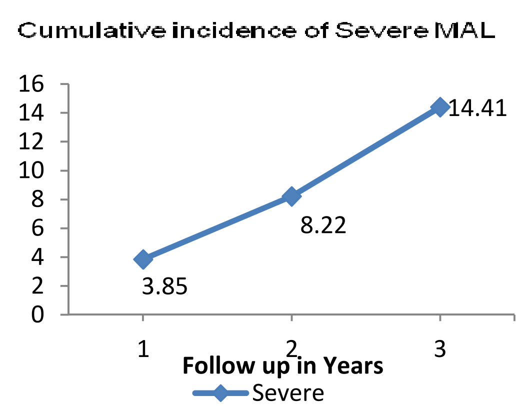
From (Time t)	Incidence of Malnutrition according to BMI classification (t+1)								
Baseline to ist Year (0- 12 months):	**Normal n**	**%**	**Moderate n**	**%**	**Severe n**	**%**	**Total**	**ISM**	**95% CI**
Normal	1110	88.9	113	9.0	26	2.1	1249	3.78	2.97 – 4.78
Moderate	289	58.0	169	33.9	40	8.0	498		
Severe	144	28.3	176	34.6	188	37.0	508		
1st Year to 2nd Year (12.1 – 24 months):									
Normal	1279	86.5	171	11.6	28	1.9	1478	4.28	3.45 – 5.28
Moderate	134	30.5	251	57.2	54	12.3	439		
Severe	17	6.8	82	32.8	151	60.4	250		
2nd Year to 3rd Year (24.1 to 36 months):									
Normal	1202	82.6	239	16.4	15	1.0	1456	6.01	5.04 – 7.15
Moderate	98	19.3	307	60.4	103	20.3	508		
Severe	8	3.5	60	26.2	161	70.3	229		
Incidence Density of Severe Malnutrition per year								4.69	3.82 – 5.74

***ISM – Incidence of severe malnutrition**

Figure 6.1.1: Incidence of Severe Malnutrition of all Children using BMI classification

Overall:





Note: MAL – Malnutrition

Table 6.1.2a: Prevalence of Acute Malnutrition (Underweight) at Baseline using Weight-for-Age by area and sex of the child

	Total	Prevalence of Underweight Malnutrition at Baseline					
		Normal		Moderate		Severe	
		n	%	n	%	n	%
Overall	2494	829	33.2	1090	43.7	575	23.0
Male	1271	360	28.3	570	44.8	341	26.8
Female	1223	469	38.3	520	42.5	234	19.1
Rural	1195	387	32.4	531	44.4	277	23.2
Male	594	165	27.8	271	45.6	158	26.6
Female	601	222	36.9	260	43.3	119	19.8
Urban	1299	442	34.1	559	43.0	298	22.9
Male	677	195	28.9	299	44.1	183	27.0
Female	622	247	39.7	260	41.8	115	18.5

Table 6.1.2b: Cumulative Incidence of Acute Malnutrition (Underweight) during Follow-ups using Weight-for-Age for 3.5 years

	Total	Cumulative Incidence of (underweight) Malnutrition during follow-up					
		Normal		Moderate		Severe	
		n	%	n	%	n	%
Overall	15629	6496	41.6	7447	47.6	1686	10.8
Male	8016	2873	35.8	3903	48.7	1240	15.5
Female	7613	3623	47.6	3544	46.6	446	5.9
Rural	7642	2690	35.2	3936	51.5	1016	13.3
Male	3836	1147	29.9	1985	51.7	704	18.4
Female	3806	1543	40.5	1951	51.3	312	8.2
Urban	7987	3806	47.7	3511	44.0	670	8.4
Male	4180	1726	41.3	1918	45.9	536	12.8
Female	3807	2080	54.6	1593	41.8	134	3.5

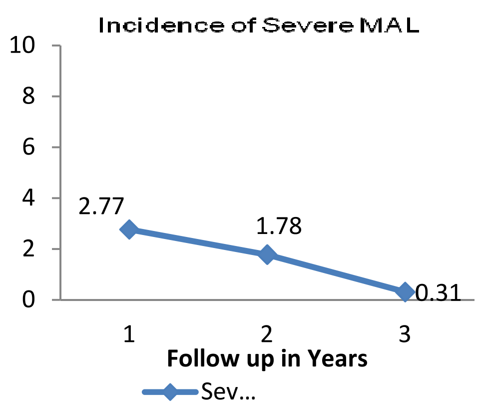
Table 6.1.2c: Incidence of Severe Acute Malnutrition (underweight) using Weight-for-age categories for all children

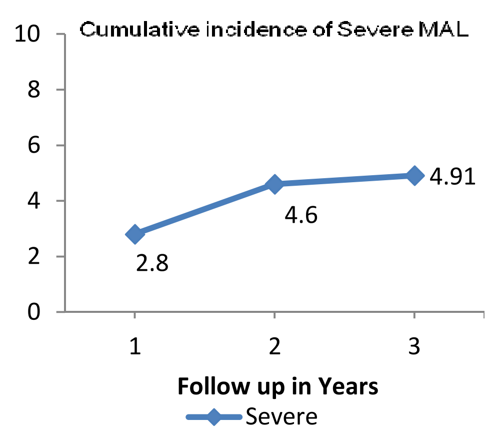
From (Time t)	Incidence of acute malnutrition (underweight) (t+1)								
Baseline to 1st year (0 – 12 months):	**Normal n**	**%**	**Moderate n**	**%**	**Severe n**	**%**	**Total**	**ISM**	**95% CI**
Normal	651	86.8	93	12.4	6	0.8	750	2.77	2.09 – 3.66
Moderate	247	25.1	696	70.7	42	4.3	985		
Severe	25	4.8	223	42.8	273	52.4	521		
1st Year to 2nd Year (12.1 to 24 months):									
Normal	791	88.7	98	11.0	3	0.3	892	1.78	1.26 – 2.50
Moderate	100	10.4	828	86.4	30	3.1	958		
Severe	2	0.6	87	27.4	228	71.9	317		
2nd Year to 3rd Year (24.1 to 36 months):									
Normal	830	92.0	72	8.0	0	0.0	902	0.31	0.12 – 0.69
Moderate	119	11.5	907	87.9	6	0.6	1032		
Severe	1	0.4	117	45.2	141	54.4	259		
Incidence Density of Severe Malnutrition per year								1.62	1.16 – 2.28

***ISM – Incidence of severe malnutrition**

Figure 6.1.2: Incidence of Severe Acute Malnutrition (underweight) of all Children using Weight-for-age Classification:

Overall:





Note: MAL – Malnutrition

Table 6.1.3a: Prevalence of Chronic Malnutrition (stunted) at Baseline using Height-for-Age by area and sex of the child

	Total	Prevalence of Chronic Malnutrition at Baseline					
		Normal		Moderate		Severe	
		n	%	n	%	n	%
Overall	2494	1078	43.2	773	31.0	643	25.8
Male	1271	534	42.0	382	30.1	355	27.9
Female	1223	544	44.5	391	32.0	288	23.5
Rural	1195	421	35.2	377	31.5	397	33.2
Male	594	208	35.0	180	30.3	206	34.7
Female	601	213	35.4	197	32.8	191	31.8
Urban	1299	657	50.6	396	30.5	246	18.9
Male	677	326	48.2	202	29.8	149	22.0
Female	622	331	53.2	194	31.2	97	15.6

Table 6.1.3b: Cumulative Incidence of chronic Malnutrition (stunted) by Area and Sex of the child using Height-for-age classification for 3.5 years

	Total	Cumulative Incidence of Chronic Malnutrition during follow-up					
		Normal		Moderate		Severe	
		n	%	n	%	n	%
Overall	15628	7160	45.8	5252	33.6	3216	20.6
Male	8015	3507	43.8	2577	32.2	1931	24.1
Female	7613	3653	48.0	2756	35.1	1285	16.9
Rural	7642	2973	38.9	2736	35.8	1933	25.3
Male	3836	1429	37.3	1306	34.0	1101	28.7
Female	3806	1544	40.6	1430	37.6	832	21.9
Urban	7986	4187	52.4	2516	31.5	1283	16.1
Male	4179	2078	49.7	1271	32.7	830	19.9
Female	3807	2109	55.4	1245	32.7	453	11.9

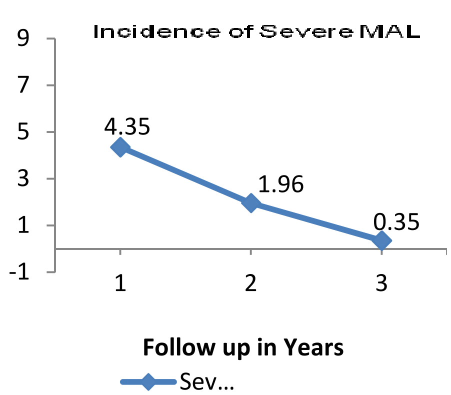
Table 6.1.3c: Incidence of chronic Severe Malnutrition (stunted) over three Years follow up using Height-for-Age classification

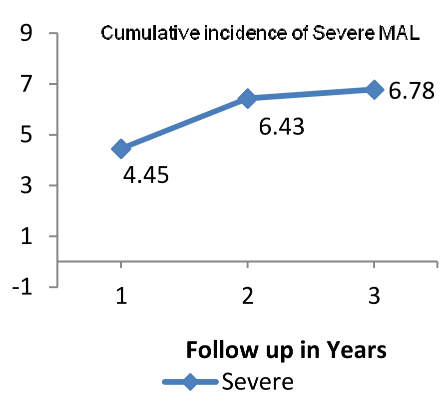
From (Time t)	Incidence of Stunting (t+1)								
Baseline to 1st Year (0-12 months):	**Normal** n	**%**	**Moderate** n	**%**	**Severe** n	**%**	**Total**	**ISM**	**95% CI**
Normal	893	90.7	87	8.8	5	0.5	985		
Moderate	68	9.8	556	80.3	68	9.8	692	4.35	3.47 – 5.44
Severe	1	0.2	95	16.4	482	83.4	578		
1st Year to 2nd Year (12.1 – 24 months):									
Normal	876	94.2	53	5.7	1	0.1	930		
Moderate	87	12.4	582	83.1	31	4.4	700	1.96	1.38 – 2.77
Severe	0	0.0	102	19.0	435	81.0	537		
2nd Year to 3rd Year (24.1 to 36 months):									
Normal	943	97.8	21	2.2	0	0.0	964		
Moderate	150	19.9	596	79.3	6	0.8	752	0.35	0.14 – 0.78
Severe	0	0.0	146	30.6	331	69.4	477		
Incidence Density of severe malnutrition per year								2.22	1.66 – 3.00

***ISM – Incidence of severe malnutrition**

Figure 6.1.3: Incidence of Severe Malnutrition (stunted) of all Children using Height-for-Age Classification:

Overall:





Note: MAL – Malnutrition

Table 6.1.4a: Prevalence of Malnutrition (wasted) at Baseline using Weight-for-Height (wasted) classification:

	Total	Prevalence of Malnutrition (wasted) at Baseline					
		Normal		Moderate		Severe	
		n	%	n	%	n	%
Overall	2494	1868	74.9	520	20.9	106	4.3
Male	1271	937	73.7	269	21.2	65	5.1
Female	1223	931	76.1	251	20.5	41	3.4
Rural	1195	965	80.8	207	17.3	23	1.9
Male	594	478	80.5	102	17.2	14	2.4
Female	601	487	81.0	105	17.5	9	1.5
Urban	1299	903	69.5	313	24.1	83	6.4
Male	677	459	71.4	146	23.5	51	7.5
Female	622	444	71.4	146	23.5	32	5.1

Table 6.1.4b: Cumulative Incidence of Malnutrition (wasted) during follow-ups using Weight-for-Height (wasted) classification for 3.5 years:

	Total	Cumulative Incidence of Malnutrition (wasted) during follow-up					
		Normal		Moderate		Severe	
		n	%	n	%	n	%
Overall	15370	13165	85.7	2092	13.6	113	0.7
Male	8008	6633	82.8	1292	16.1	83	1.0
Female	7362	6532	88.7	800	10.9	30	0.4
Rural	7533	6433	85.4	1059	14.1	41	0.5
Male	3836	3192	83.2	618	16.1	26	0.7
Female	3697	3241	87.7	441	11.9	15	0.4
Urban	7837	6732	85.9	1033	13.2	72	0.9
Male	4172	3441	82.5	674	16.2	57	1.4
Female	3665	3291	89.8	359	9.8	15	0.4

Table 6.1.4c: Incidence of Severe Malnutrition (wasted) using Weight-for-height classification for all children

From (Time t)	Incidence of Malnutrition (Wasted) (t+1)								
Baseline to 1st Year (0 - 12 months):	**Normal n**	**%**	**Moderate n**	**%**	**Severe n**	**%**	**Total**	**ISM**	**95% CI**
Normal	1611	95.5	72	4.3	4	0.2	1687		
Moderate	280	59.2	189	40.0	4	0.8	473		
Severe	51	53.7	37	38.9	7	7.4	95	0.37	0.18 – 0.75
1st Year to 2nd Year (12.1 – 24 months):									
Normal	1774	95.2	85	4.6	4	0.2	1863		
Moderate	108	37.5	172	59.7	8	2.8	288		
Severe	7	43.8	4	25.0	5	31.2	16	0.56	0.31 – 0.99
2nd Year to 3rd Year (24.1 – 36 months):									
Normal	1758	92.2	149	7.8	0	0.0	1907		
Moderate	70	28.0	174	69.6	6	2.4	250		
Severe	4	23.5	5	29.4	8	47.1	17	0.28	0.11 – 0.63
Average Incidence of Severe Malnutrition per year								0.40	0.2 – 0.79

***ISM – Incidence of severe malnutrition**

Figure 6.1.4: Incidence of Severe Malnutrition (Wasted) of all Children using Weight-for-Height Classification:

Overall:

Note: MAL – Malnutrition

6.2 Risk Factor Analysis:

6.2.1. Repeated Measure Analysis using Generalized Estimating Equations (GEE) using BMI classification:

The association of risk factors with malnutrition using BMI classification is shown in Table 6.2.1a and Table 6.2.1b. Table 6.2.1a is the unadjusted analysis showing the association of each of risk factor with malnutrition where as Table 6.2.1b is the adjusted analysis. The Table 6.2.1b was performed using the risk factors that were significant at 25% level in the unadjusted analysis. From the Table 6.2.1a, we observed that male children had higher moderate and severe malnutrition than female children. The odds of having moderate and severe malnutrition was found to be 1.4 (95% CI: 1.2-1.6) times significantly higher for male children as compared to female children. The malnutrition was significantly higher if children were living in a house which had no separate kitchen. The odds of having malnutrition was nearly 1.2 (1.0 – 1.4) times significantly more if there was no separate kitchen as compared to having separate kitchen. There were six risk factors that had p value less than 0.25. These risk factors are sex of the child, defecation, type of floor, presence of a separate kitchen in the house, consanguineous marriage, type of house were significant at 25% level. The risk factor 'type of roof' had one of its category – 'RCC/Pukka' significant at 25% level and hence 'type of roof' was included for multivariable analysis (Table 6.2.1a).

From the multivariable analysis, we observed that children who had defecated in toilets were 6% (3% - 8%) less likely to have severe malnutrition over time as compared to those who had defecated out in the open fields. Male children were also significantly associated with severe malnutrition over time as compared to female children over time (OR: 1.03; 95% CI:1.01-1.05) (Table 6.2.1b).

Table 6.2.1a: Bivariate analysis Generalized Estimating Equations (GEE) for Malnutrition using BMI classification by socio-demographic and household variables:

Variables	BMI classification						Odds Ratio	95% CI		P value
	Normal		Mild/Moderate		Severe					
	N	%	N	%	n	%				
Sex of the child										
Male	5468	58.9	2322	25.0	1497	16.1	1.41	1.24	1.60	<0.001
Female	5934	67.2	2021	22.9	881	10.0				
Area of Residence										
Rural	5569	63.0	2158	24.4	1110	12.6	1.03	0.91	1.17	0.638
Urban	5833	62.8	2185	23.5	1268	13.7				
Birth Order										
1	2073	61.2	874	25.8	438	12.9	1.00			
2	2405	63.4	833	22.0	553	14.6	0.94	0.76	1.15	0.523
≥3	6897	63.4	2614	24.0	1373	12.6	0.92	0.78	1.09	0.327
Mother's Education										
Illiterate/ Literate	6568	62.4	2558	24.3	1393	13.2	1.08	0.87	1.33	0.478
Primary/Middle School	3400	63.2	1223	22.7	753	14.0	1.05	0.84	1.32	0.648
High school/College	1187	64.2	475	25.7	187	10.1	1.00			
Father's Education										
Illiterate/ Literate	3378	62.3	1313	24.2	731	13.5	1.03	0.87	1.23	0.734
Primary/Middle	4688	62.9	1786	24.0	983	13.2	1.03	0.87	1.21	0.727
High school/ College	2849	63.1	1073	23.8	590	13.1	1.00			
Number of Family Members										
<=4	1549	62.2	621	24.9	321	12.9	1.04	0.85	1.28	0.674
5 – 6	5204	62.4	2022	24.2	1114	13.4	1.07	0.93	1.23	0.367
>6	4610	64.0	1666	23.1	928	12.9	1.00			
Fuel for cooking										
Drug/Firewood	9884	63.0	3749	23.9	2063	13.1	1.04	0.86	1.25	0.693
Gas/Kerosene	1479	63.2	560	23.9	300	12.8				
Defecation										
Within premises/latrine	4904	61.8	1870	23.6	1159	14.6	1.12	0.99	1.28	0.077
Open field	6459	63.9	2439	24.1	1204	11.9				
Type of Floor										
Kucha	4389	61.7	1771	24.9	957	13.4	1.09	0.96	1.25	0.185
Pukka	6974	63.9	2538	23.2	1406	12.9				
Presence of a Separate Kitchen										
Yes	8629	64.0	3201	23.7	1655	12.3				
No	2734	60.1	1108	24.4	708	15.6	1.19	1.02	1.38	0.027

Contd…

Variables	Categorized BMI						Odds Ratio	95% CI		P value
	Normal		Moderate		Severe					
	n	%	n	%	n	%				
Consanguineous Marriage										
Yes	4147	61.7	1651	24.6	926	13.8	1.09	0.50	1.24	0.311
No	7214	63.7	2670	23.6	1439	12.7				
Type of House										
Brick and cement	5805	62.8	2214	24.0	1220	13.2				
Brick and/or mud	3091	65.9	1059	22.6	538	11.5	0.91	0.77	1.07	0.248
Others	2467	60.1	1036	25.2	605	14.7	0.80	0.67	0.97	0.023
Type of roof										
Thatched	4282	62.2	1726	25.1	875	12.7	1.21	1.01	1.45	0.043
Tiled	4490	62.4	1671	23.2	1033	14.4	0.97	0.84	1.13	0.720
RCC/Pukka	2317	66.2	803	22.9	382	10.9	1.00			
Follow-up										
0	1391	55.8	542	21.7	561	22.5	1.06	0.98	1.15	0.156
1	1577	69.6	437	19.3	252	11.1	0.53	0.48	0.57	<0.001
2	1549	68.3	464	20.4	256	11.3	0.59	0.55	0.65	<0.001
3	1527	68.1	487	21.7	229	10.2	0.57	0.53	0.62	<0.001
4	1502	66.3	523	23.1	239	10.6	0.64	0.60	0.69	<0.001
5	1374	61.7	597	26.8	255	11.5	0.78	0.71	0.80	<0.001
6	1317	59.6	608	27.5	284	12.9	0.85	0.81	0.90	<0.001
7	1165	54.1	685	31.8	302	14.0	1.00			

Table 6.2.1b: Adjusted GEE analysis for Malnutrition by socio-demographic and household variables

Variables	Odds Ratio	95% CI		P value
Sex of the child				
Male	1.27	1.05	1.40	0.009
Female				
Defecation				
Within premises/latrine				
Open field	1.77	1.30	1.91	<0.001
Presence of a Separate Kitchen				
Yes	0.78	0.62	0.99	0.042
No				
Type of Floor				
Kucha	1.03	0.79	1.35	0.825
Pukka				
Type of House				
Brick and/or cement	1.00			
Brick and/or mud	1.43	1.03	2.00	0.034
Others	1.20	0.92	1.56	0.181
Type of Roof				
Thatched	1.01	0.74	1.37	0.962
Tiled	1.23	0.99	1.53	0.062
RCC /Pukka	1.00			
Follow	1.02	0.96	1.08	0.510
Interaction of sex of child with follow-up	1.03	1.01	1.05	0.013
Interaction of defecation with follow-up	0.94	0.92	0.97	<0.001
Interaction of kucha type of floor with follow-up	0.99	0.96	1.03	0.812
Interaction of no separate kitchen with follow-up	1.01	0.97	1.04	0.702
Interaction of brick and cement type of house with follow-up	1.02	0.97	1.07	0.446
Interaction of brick and mud type of house with follow-up	1.01	0.97	1.05	0.606
Interaction of thatched roof with follow-up	1.01	0.96	1.05	0.738
Interaction of tiled roof house with follow-up	0.99	0.96	1.02	0.637

6.3 Markov Chain

6.3 First Mean Passage Time and Monte Carlo Simulation using BMI classification:

6.3.1 Overall Transition Probability and Mean Passage time for malnutrition according to BMI classification:

The transition probability matrix is presented in Table 6.3.1a. The transition probability matrix presents the probability of transition from the previous state of malnutrition at previous time to the next state of malnutrition at the current time for all children. The transition probability from severe state in the previous time (t) to normal state (t+1) was 0.10. If a child was in the severe state at time t, then the transition to moderate state at t+1 was 0.28. If a child was in moderate state at t, then the probability of transition to normal state at t+1 was 0.31 (Table 6.3.1a).

The mean number of years with 95% CI for different states of malnutrition using BMI classification is presented in the table 6.3.1b. The mean number of years taken for a child who is in the severe state in the previous time to transit to normal state at time t+1 was 2.7 (2.3 – 3.1) years and to transit to moderate state at time t+1 was 2.3 (1.8 – 2.9) years. The average number of years taken to transit from moderate state to normal was 2.0 (1.6 – 2.3) years (Table 6.3.1b).

Table 6.3.1a: Overall Transition Probability Matrices of Malnutrition according to BMI classification

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Overall	Normal	0.87	0.12	0.009	9784
	Moderate	0.31	0.56	0.14	3513
	Severe	0.10	0.28	0.61	1973

6.3.1b: Overall Mean Passage Time (years) and 95% Confidence Interval using BMI classification

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Overall:			
Normal	0.76 (0.57, 0.96)	3.92 (3.21, 4.62)	13.60 (11.15, 16.06)
Moderate	1.97 (1.59, 2.35)	2.43 (1.46, 2.58)	10.52 (8.10, 12.94)
Severe	2.73 (2.31, 3.14)	2.35 (1.81, 2.90)	4.89 (3.02, 6.73)

6.3.2 Transition Probability and Mean Passage time for malnutrition by sex according to BMI classification:

The probability of transition from severe state in the previous time (t) to normal state at time t+1 was 0.10 among male children and 0.12 among female children. The transition probability from severe to moderate state of malnutrition from time t to t+1 was 0.25 among male and 0.33 among female implying that female children transited more from severe to normal or moderate than male children. Similarly, the transition probability from moderate to normal state from t to t+1 was similar across male and female children (0.30 vs 0.31 respectively). (Table 6.3.2a).

The mean passage time for transiting from one state of malnutrition to another by sex of the child showed that, on an average, the number of years taken by male children to transit from severe state from the previous time to normal state at current time was slightly more as compared to female children (Male children: 2.9 (2.5 – 3.4) years; Female children: 2.4 (2.1 – 2.8) years). Also, the average number of years from moderate state at time t to normal state at time t+1 was 2.1 (1.7 – 2.5) years among male children and 1.8 (1.5 – 2.2) years among female children. Hence there is a difference in the mean time of transition for a male and female child with female children taking shorter times to become normal from malnourished (severe or moderate) state as compared to male children (Table 6.3.2b).

The test of hypothesis to test if sex of the child had different transition is presented in table 6.3.2c using the log linear model. The difference in the likelihood (deviance) when sex of the child was included in the model was 14.4 at 4 degrees of freedom suggesting that there was a significant difference in the transition from one state to another state between male children and female children (Table 6.3.2c).

Table 6.3.2a: Transition Probability Matrices of Malnutrition according to BMI classification by sex of the child:

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Males	Normal	0.85	0.14	0.01	4705
	Moderate	0.30	0.55	0.15	1879
	Severe	0.10	0.25	0.65	812
Females	Normal	0.88	0.11	0.006	5079
	Moderate	0.31	0.57	0.12	1634
	Severe	0.12	0.33	0.54	730

Table 6.3.2b: Mean Passage Time (years) and 95% Confidence Interval by Sex of the child using BMI classification

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Males:			
Normal	0.82 (0.60, 1.05)	3.49 (2.86, 4.11)	11.17 (9.19, 13.17)
Moderate	2.09 (1.67, 2.50)	1.91 (1.41, 2.41)	8.54 (6.58, 10.50)
Severe	2.95 (2.50, 3.40)	2.41 (1.90, 2.92)	3.71 (2.26, 5.14)
Females:			
Normal	0.71 (0.55, 0.88)	4.45 (3.63, 5.27)	17.26 (14.07, 20.42)
Moderate	1.84 (1.50, 2.18)	2.16 (1.53, 2.80)	13.59 (10.53, 16.64)
Severe	2.45 (2.09, 2.82)	2.27 (1.67, 2.85)	7.12 (4.56, 9.60)

Table 6.3.2c: Results of Log linear Model by sex of the child:

Model	Model	LR	df	Deviance (G^2)	Difference in df	P value
Model 1	Saturated Model	0.000	-	-	-	-
Model 2	Sex of the child	14.368	4	Model 2 – Model 1 = 14.368	4	0.006

Note: LR – Likelihood ratio

6.3.3 Transition Probability and Mean Passage Time for malnutrition by area of residence according to BMI classification:

The transition matrix showing the probability of transition from one state of malnutrition from time t to another state at time t+1 by area of the residence is shown in Table 6.3.3a. The probability of transition of malnutrition from moderate to normal in rural area was found to be 0.26 where as in the urban it was 0.35. The transition probabilities from severe state to normal state were 0.06 among children from rural and 0.14 for children living in urban areas. The transition probability from severe state to moderate state was 0.28 for children living in rural areas and urban areas (Table 6.3.3a).

The first mean passage time of transition in rural and urban areas is presented in Table 6.3.3b. The average number of years taken to transit from moderate state at the previous time t to normal state at time t+1 was 2.4 (1.9 – 2.9) years in the rural areas and 1.7 (1.4 – 2.0) years in the urban areas. The average number of years to transit from severe to normal in rural and urban areas was 3.4 (2.9 – 3.9) years and 2.3 (2.0 – 2.6) years respectively. The average number of years to transit from severe to moderate state was 2.2 (1.7 – 2.7) years among children in rural areas and 2.4 (1.9 – 3.0) years among children living in urban areas.

The result of log linear analysis showing if area of residence was significant factor associated with the transition of malnutrition from one state to another state is presented in Table 6.3.3c. The deviance value comparing full model with model including area of residence with 4 degrees of freedom showed that there was a significant association (Table 6.3.3c).

Table 6.3.3a: Transition Probability Matrices for malnutrition according to BMI classification by area of residence of the child

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Rural	**Normal**	0.87	0.12	0.009	4774
	Moderate	0.26	0.59	0.14	1729
	Severe	0.06	0.28	0.65	920
Urban	**Normal**	0.86	0.13	0.008	5010
	Moderate	0.35	0.52	0.13	1784
	Severe	0.14	0.28	0.58	1053

Table 6.3.3b: Mean Passage Time (years) and 95% Confidence Interval by Area of Residence using BMI classification

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Rural:			
Normal	0.82 (0.58, 1.06)	3.98 (3.27, 4.69)	12.74 (10.43, 15.05)
Moderate	2.12 (1.93, 2.89)	1.86 (1.32, 2.39)	9.55 (7.32, 11.77)
Severe	3.41 (2.89, 3.93)	2.20 (1.71, 2.69)	4.00 (2.35, 5.66)
Urban:			
Normal	0.73 (0.57, 0.89)	3.86 (3.17, 4.57)	14.44 (11.77, 17.12)
Moderate	1.68 (1.36, 2.00)	2.17 (1.59, 2.76)	11.46 (8.88, 14.05)
Severe	2.30 (1.96, 2.65)	2.45 (1.88, 3.01)	5.77 (3.66, 7.88)

Table 6.3.3c: Results of Log linear Model by the area of residence:

Model	Model	LR	Df	Deviance (G^2)	Difference in df	P value
Model 1	**Saturated Model**	0.000	-	-	-	-
Model 2	**Area of Residence**	39.151	4	Model 2 – Model 1 = 39.151	4	<0.001

LR – likelihood ratio

6.3.4 Transition Probability and Mean Passage Time for malnutrition by presence of a separate kitchen according to BMI classification:

The transition probability matrix by presence of a separate kitchen is shown in Table 6.3.4a. The transition probability from severe state to normal state for children in those houses where there was a separate kitchen was found to be 0.12. However, it was only 0.06 if the children lived in houses having no separate kitchen. The transition probability for children from severe at time t to moderate state of malnutrition at time t+1 if the house did had a separate kitchen was 29% and it was 25% if the house did not have a separate kitchen in the house (Table 6.3.4a). The probability of transition from moderate state to normal state from time t to t+1 was found to be 0.32 for the children living in house where there was a separate kitchen and 0.25 for children living in houses where there was no separate kitchen. The transition probability matrix showed that the probability of transition from severe state to normal or moderate states of malnutrition was higher among children living in house with separate kitchen as compare to the children living in house without separate kitchen.

The mean passage time for transiting from one state of malnutrition to another state from time t to t+1 is presented in Table 6.3.4b. The average number of years taken to transit from moderate state to normal state from time t to t+1 was 1.8 (1.5 – 2.1) years for those children living in houses where there was a separate kitchen and it was 2.7 (2.1 – 3.2) years for children living in houses with no separate kitchen. The transition time to transit from severe state to normal state for children living in houses with separate and no separate kitchen was 2.4 (2.1 – 2.8) years and 3.7 (3.2 – 4.3) years respectively. The transition time taken to move from severe state to moderate state for children living in house with and without separate kitchen was 2.4 (1.8 – 2.9) years and 2.4 (1.9 – 2.9) years respectively. This implied that the time taken to move from severe state of malnutrition to normal state was slower for children living in house without separate kitchen as compared to children living in houses with separate kitchen. The possible reason for the slower transition times from severe state to normal state for children who lived in houses where there was no separate kitchen might be no proper ventilation and as a result were suffering from respiratory infections which led to malnutrition.

The result of the log linear analysis is shown in Table 6.3.4c showing the association of presence of a separate kitchen and transition. This shows that there was a significant association of presence of a separate kitchen and transitions (Table 6.3.4c).

Table 6.3.4a: Transition Probability Matrices for malnutrition according to BMI classification by 'Presence of separate kitchen'

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Yes	Normal	0.87	0.12	0.009	7387
	Moderate	0.32	0.55	0.13	2597
	Severe	0.12	0.29	0.58	1374
No	Normal	0.86	0.13	0.008	2363
	Moderate	0.25	0.58	0.17	889
	Severe	0.06	0.25	0.69	586

Table 6.3.4b: Mean Passage Time (years) and 95% Confidence Interval by 'Presence of a separate kitchen within household' using BMI classification

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Yes:			
Normal	0.74 (0.56, 0.91)	3.96 (3.24, 4.68)	14.70 (12.03, 17.38)
Moderate	1.79 (1.46, 2.13)	2.09 (1.51, 2.66)	11.67 (9.01, 14.32)
Severe	2.45 (2.09, 2.81)	2.36 (1.80, 2.93)	5.78 (3.66, 7.86)
No:			
Normal	0.87 (0.59, 1.14)	3.83 (3.13, 4.52)	11.03 (9.06, 12.99)
Moderate	2.67 (2.11, 3.22)	1.87 (1.35, 2.39)	7.84 (6.01, 9.67)
Severe	3.75 (3.16, 4.34)	2.36 (1.86, 2.86)	3.13 (1.82, 4.41)

Table 6.3.4c: Results of Log linear Model:

Model	Model	LR	Df	Deviance (G^2)	Difference in df	P value
Model 1	**Saturated Model**	0.000	-	-	-	-
Model 2	**Presence of a separate Kitchen**	13.683	4	Model 2 – Model 1 = 13.683	4	0.008

LR – likelihood ratio

6.3.5 Transition Probability and Mean Passage Time for malnutrition by defecation according to BMI classification:

The transition probability matrix showing the probabilities of transition by defecation is shown in Table 6.3.5a. The probability of transition from severe state of malnutrition at a previous time (t) to normal state of malnutrition at the current time (t+1) was found to be 0.13 for those children living in houses where defecation was within the premises of the house or latrine. Children living in houses where defecation was in the open fields had transition probability from severe to normal as 0.08. The transition probability from severe to moderate state of malnutrition for children living in houses where defecation was within the premises of house or latrine and in the open fields was 0.27 and 0.29 respectively. The probability of transition from moderate state of malnutrition to normal for children living in houses where defecation was within the premises of the house or latrine and open fields was 0.33 and 0.29 respectively. (Table 6.3.5a).

The mean passage time for transition from normal to moderate or severe malnutrition or vice versa is presented in Table 6.3.5b. If the defecation was in the within the premises of the hosue or latrine then the number of years taken to transit from moderate to normal was 1.8 years (1.4 – 2.1). The average number of years taken to transit from moderate to normal state was 2.2 years (1.7 – 2.6) if the defecation was in the open fields The average number of years taken to transit from severe state of malnutrition to normal state from time t to t+1 for the children living in houses where defecation was within the premises of the house or latrine and open fields was 2.4 (2.1 – 2.8) years and 3.0 (2.6 – 3.5) years respectively. If the children lived in houses where defecation was within the premises of the house or latrine then the time taken to transit from severe to moderate state of malnutrition was 2.5 (1.8 – 3.1) years and if the children lived in houses where defecation was in the open fields, then the time taken to transit from severe state to moderate state of malnutrition was found to be 2.2 (1.7 – 2.7) years. This implied that children living in houses where defecation was within the premises or latrine and open fields to transit from severe state to moderate state of malnutrition was almost similar (Table 6.3.5b).

The result of log linear analysis is shown in Table 6.3.5c. The result shows that there is a significant change in the transition probabilities from one state of malnutrition to another state of malnutrition across defecation (Table 6.3.5c).

Table 6.3.5a: Transition Probability Matrices for malnutrition according to BMI by 'Defecation':

	Grading (t)	Grading (t+1) Normal	Moderate	Severe	Total no.
Within the premises or Latrine	Normal	0.87	0.12	0.008	4189
	Moderate	0.33	0.53	0.14	1525
	Severe	0.13	0.27	0.59	962
Open Fields	Normal	0.87	0.12	0.009	5561
	Moderate	0.29	0.58	0.14	1961
	Severe	0.08	0.29	0.63	998

Table 6.3.5b: Mean Passage Time (years) and 95% Confidence Interval by 'defecation' using BMI classification

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Within premises of household:			
Normal	0.74 (0.56, 0.91)	3.95 (3.24, 4.66)	14.29 (11.65, 16.92)
Moderate	1.77 (1.44, 2.10)	2.16 (1.57, 2.75)	11.15 (8.59, 13.72)
Severe	2.42 (2.06, 2.79)	2.51 (1.77, 3.08)	5.45 (3.40, 7.52)
Open Fields			
Normal	0.79 (0.56, 1.01)	3.91 (3.20, 4.62)	13.11 (10.72, 15.50)
Moderate	2.16 (1.73, 2.59)	1.92 (1.38, 2.47)	10.06 (7.77, 12.36)
Severe	3.05 (2.58, 3.51)	2.20 (1.70, 2.70)	4.45 (2.70, 6.22)

Table 6.3.5c: Results of Log linear Model:

Model	Model	LR	Df	Deviance (G^2)	Difference in df	P value
Model 1	**Saturated Model**	0.000	-	-	-	-
Model 2	**Defecation**	15.689	4	Model 2 – Model 1 = 15.689	4	0.003

LR – Likelihood ratio

6.3.6 Transition Probability and Mean Passage Time for malnutrition by type of fuel used for cooking according to BMI classification:

The transition probability matrix of transition from one state of malnutrition to another state by type of fuel used for cooking is shown in Table 6.3.6a. The transition probability from transition from severe to normal state if the type of fuel used for cooking was firewood or coal or cow dung was found to be 0.10 where as the transition to normal state from severe state of malnutrition if the type of fuel used for cooking was gas or kerosene was found to be slightly higher as 0.14. The transition probability from moderate state to normal state when type of fuel used for cooking was firewood and gas or kerosene was 0.30 and 0.33 respectively. The transition probability from severe state to moderate state of malnutrition from time t to t+1 for children living in houses where type of fuel used for cooking was firewood, coal or cow dung as compared to children living in houses where type of fuel used for cooking was gas or kerosene was found as 0.28 and 0.29 respectively (Table 6.3.6a).

The mean passage time for moving from one state of malnutrition to another state by type of fuel used for cooking is presented in Table 6.3.6b. The mean number of years taken to move from severe state of malnutrition at time t to normal state at time t+1 was 2.8 (2.4 – 3.2) years for children living in houses where type of fuel used for cooking was firewood or coal or cow dung and 2.3 (2.0 - 2.7) years for children living in houses where type of fuel used for cooking was gas or kerosene. The average number of years to move from severe state to moderate state of malnutrition among children living in houses where type of fuel used for cooking was firewood or coal or cow dung and gas or kerosene was very similar which was 2.3 (1.9 – 3.2) years and

2.6 (1.9 – 3.2) years respectively. The average number of years to move from moderate state to normal state of malnutrition among children living in houses where type of fuel used for cooking was firewood or coal or cow dung and gas or kerosene was 2.0 (1.6 – 2.4) years and 1.7 (1.4 – 2.1) years respectively (Table 6.3.6b).

The result of log linear is shown in Table 6.3.6c. The result showed that transition probabilities across the different state of malnutrition from time t to t+1 by type of fuel used for cooking was not significantly changing (Table 6.3.6c).

Table 6.3.6a: Transition Probability Matrices for Malnutrition according to BMI by 'Type of Fuel Used for Cooking':

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Firewood/Cow dung/coal	Normal	0.87	0.12	0.009	8488
	Moderate	0.30	0.56	0.14	3021
	Severe	0.10	0.28	0.62	1706
Gas/Kerosene	Normal	0.88	0.12	0.005	1262
	Moderate	0.33	0.53	0.13	465
	Severe	0.14	0.29	0.57	254

Table 6.3.6b: Mean Passage Time (years) and 95% Confidence Interval by 'type of fuel used for cooking' using BMI classification:

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Firewood/Cow Dung			
Normal	0.77 (0.57, 0.97)	3.89 (3.19, 4.59)	13.30 (10.90, 15.71)
Moderate	2.00 (1.61, 2.39)	2.00 (1.45, 2.55)	10.28 (7.91, 12.64)
Severe	2.79 (2.37, 3.21)	2.33 (1.80, 2.87)	4.71 (2.90, 6.49)
Gas/Kerosene			
Normal	0.71 (0.55, 0.87)	4.19 (3.42, 4.97)	16.12 (13.15, 19.09)
Moderate	1.74 (1.42, 2.06)	2.23 (1.62, 2.86)	12.51 (9.67, 15.34)
Severe	2.31 (1.96, 2.66)	2.56 (1.95, 3.17)	6.44 (4.10, 8.73)

Table 6.3.6c: Results of Log linear Model:

Model	Model	LR	Df	Deviance (G^2)	Difference in df	P value
Model 1	**Saturated Model**	0.000	-	-	-	-
Model 2	**Type of fuel used for cooking**	3.835	4	Model 2 – Model 1 = 3.835	4	0.429

LR – Likelihood ratio

6.4 Markov Regression:

The risk factor analysis for the Markov model is presented in this section. The Markov regression analysis was performed using the transition probabilities and transition intensity rates. The Markov regression analysis involves Markov property which is nothing but the present state being influenced by only the previous state.

6.4.1 Markov Regression using Transition Probabilities for malnutrition using BMI classification:

Unadjusted Markov regression analysis using the transition probabilities or BMI classification is presented in Table 6.4.1a and the risk factors that were significant at 25% level of significance were included for multivariable analysis.

The adjusted analysis using transition probabilities is presented in Table 6.4.1b.

Table 6.4.1a: Unadjusted Markov regression analysis considering the transition probabilities using BMI classification:

Risk Factors	OR	Robust SE	P value
Sex of the child Male Female	1.20	0.06	0.002
Previous state (M_{it})	12.55	0.05	<0.001
Interaction of sex of the child and previous state	0.96	0.06	0.500
Area of Residence Rural Urban	0.75	0.06	<0.001
Previous state	11.36	0.04	<0.001
Interaction of area and previous state	1.21	0.06	0.002
Father's Education			
Illiterate/Literate	0.91	0.06	0.140
Primary/Middle	0.87	0.08	0.091
High school/ above	1.00		
Previous state	11.94	0.06	<0.001
Interaction father's education and previous state	1.03	0.04	0.394
Mother's Education			
Illiterate/Literate	1.05	0.08	0.588
Primary/Middle	1.04	0.09	0.687
High school/above	1.00		
Previous state	13.20	0.08	<0.001
Interaction of mother's education and previous state	0.96	0.05	0.447

Contd..

Variables	**OR**	**Robust SE**	**P value**
Consanguineous Marriage			
Yes	1.01	0.06	0.889
No			
Previous state	2.19	0.04	<0.001
Interaction of consanguineous marriage and previous state	1.05	0.06	0.455
Birth Order			
1	1.00		
2	0.89	0.07	0.084
≥3	0.87	0.08	0.104
Previous state	11.47	0.06	<0.001
Interaction of birth order and previous state	1.06	0.04	0.153
Number of Family Members			
≤4	1.00		
5 – 6	0.90	0.07	0.151
≥7	0.77	0.09	0.005
Previous state	10.17	0.06	<0.001
Interaction of family members and previous state	1.17	0.04	<0.001
Type of House			
Brick and/or cement	1.00		
Brick and/or mud	0.94	0.06	0.310
Others	0.95	0.08	0.500
Previous state	11.94	0.04	<0.001
Interaction of type of house and previous state	1.05	0.04	0.159

Contd..

Variables	**OR**	**Robust SE**	**P value**
Defecation Practice Within the premises Open fields	1.26	0.06	<0.001
Previous state	12.81	0.04	<0.001
Interaction of defecation and previous state	0.93	0.06	0.257
Type of Fuel Firewood/ Coal/Cow dung Gas/Kerosene	0.84	0.09	0.046
Previous state	11.70	0.08	<0.001
Interaction of type of fuel and previous state of malnutrition	1.06	0.09	0.484
Type of Roof Thatched Tiled RCC/Pukka Others	 1.04 1.09 1.03 1.00	 0.16 0.17 0.18	 0.803 0.589 0.857
Previous state	12.06	0.08	<0.001
Interaction of risk factor and previous state of malnutrition	1.01	0.04	0.718
Presence of a separate Kitchen No Yes	0.82	0.07	0.010
Previous state	11.70	0.04	<0.001
Interaction of presence of a kitchen and previous state of malnutrition	1.24	0.07	0.002
Type of Floor Kucha Pukka	0.95	0.06	0.438
Previous state	12.30	0.04	<0.001
Interaction of type of floor and type of floor and previous state of malnutrition	1.01	0.06	0.850

The unadjusted analysis using Markov regression for malnutrition using BMI classification is shown in Table 6.4.1a. This model was constructed using the previous state of malnutrition at time t-1 as a covariate and current state of malnutrition as the outcome.

A male child on an average has 1.2 times the odds of having severe malnutrition as compared to female child (p value = 0.002) When the interaction of sex of the child and previous state of malnutrition was included there was no effect of interaction (p = 0.5) suggesting that association of sex of the child was similar irrespective of the previous state of malnutrition. The previous state of malnutrition had nearly 12 (p<0.001) times the odds of having severe malnutrition. The interaction of area of residence and previous state of malnutrition was significantly associated with current state of malnutrition suggesting that area of residence was significantly associated with current state of malnutrition irrespective of the previous state of malnutrition (OR=1.21, p value = 0.002). If there was no separate kitchen then there were nearly 1.2 times the odds of having severe malnutrition as compared to having separate kitchen in the household whatever may be the previous state of malnutrition (interaction of previous state of malnutrition and presence of separate kitchen, p value = 0.002). If the child lived in a house where defecation was within the household or latrine then there was on an average, 1.3 times the odds of having severe malnutrition (p<0.001). Number of family members was also a significant factor that was associated with malnutrition at present irrespective of the previous state of malnutrition suggesting that increase in number of members more than the 4 per family then there was an increased risk of severe malnutrition. (OR: 1.17; p <0.001) (Table 6.4.1a).

The adjusted regression analysis using transition probabilities for BMI classification has been presented in Table 6.4.1b.

The previous state of malnutrition was very highly associated to present state of malnutrition satisfying Markovian Property after adjusting for other variables included for multivariable analysis (OR = 9.3; 95% CI: 7.8 – 11.0). Presence of no separate kitchen and its interaction with the previous state of malnutrition was associated with severe malnutrition at the current after adjusting for other risk factors and confounders (OR = 1.3 (95%CI: 1.1 – 1.5), p =0.004). Similarly, number of family members with previous state of malnutrition was significant after adjusting for other risk factors OR = 1.17(95% CI: 1.1 – 1.3, p = 0.002).

Table 6.4.1b: Adjusted Markov Regression analysis with Transition Probabilities using BMI classification:

Variables	OR	95% CI		Robust SE	P value
Presence of a separate Kitchen					
No					
Yes	0.91	0.77	1.08	0.08	0.302
Defecation					
Within the premises	1.26	1.06	1.48	0.08	0.007
Open fields					
Birth Order					
1	1.00				
2	0.91	0.79	1.05	0.07	0.194
≥3	0.94	0.78	1.13	0.09	0.504
Area of Residence					
Rural	0.96	0.79	1.16	1.00	0.660
Urban					
Sex of the child					
Male	1.21	1.07	1.25	0.06	0.003
Female					
Father's Education					
Illiterate/Literate	0.99	0.88	1.11	0.06	0.868
Primary/Middle	1.03	0.90	1.17	0.07	0.641
High school/ above	1.00				
Number of Family Members					
≤4	1.00				
5 – 6	0.93	0.80	1.08	0.06	0.354
≥7	0.81	0.66	0.99	0.08	0.046
Previous State	9.30	7.85	11.01	0.09	<0.001
Interaction of no presence of kitchen and previous state	1.28	1.08	1.51	0.08	0.004
Birth order and previous state of malnutrition interaction	1.01	0.92	1.10	0.05	0.890
Interaction of area and previous state of malnutrition	1.08	0.93	1.25	0.07	0.290
Interaction of number of family members and previous state of malnutrition	1.17	1.06	1.29	0.05	0.002

6.4.2 Transition Intensity Matrix for malnutrition using BMI classification:

The risk factor analysis using transition intensity rate is shown in the Table 6.4.2a. This Table 6.4.2a is an unadjusted analysis. The outcome is analyzed as a binary variable with categories 'Normal' and 'Malnutrition'. The hazard ratios for each risk factor from time t to t+1 were obtained. The adjusted analysis is shown in Table 6.4.2b.

Table 6.4.2a: Unadjusted Markov Regression Analysis using transition intensity matrix using BMI classification:

Variables		t+1			
		Normal		Moderate/Severe Malnutrition	
	t	Hazard Ratio	95% CI	Hazard Ratio	95% CI
Sex of the child:					
Male	Normal Malnutrition	1.00		1.00	
Female	Normal Malnutrition	1.11	(0.98, 1.26)	0.76	(0.67, 0.86)
Area of residence:					
Rural	Normal Malnutrition	1.00		1.00	
Urban	Normal Malnutrition	1.43	(1.26, 1.62)	1.11	(0.98, 1.26)
Presence of a separate kitchen:					
Yes	Normal Malnutrition	1.00		1.00	
No	Normal Malnutrition	0.64	(0.55, 0.75)	0.97	(0.85, 1.11)
Defecation					
Within the premises / toilet	Normal Malnutrition	1.00		1.00	
Open fields	Normals Malnutrition	0.84	(0.74, 0.95)	0.97	(0.85, 1.09)
Type of fuel used for cooking:					
Cowdung / Coal	Normal Malnutrition	1.00		1.00	
Gas / Kerosene	Normal Malnutrition	1.12	(0.94, 1.34)	0.96	(0.79, 1.16)

The hazard ratio of transition from normal to malnutrition for female children was 0.76 which means that females were 24% (14% - 33%) significantly less likely to be transiting from normal to malnutrition state as compared to male children. Children from urban area of residence transited significantly faster from malnutrition to normal state (HR = 1.43; 95% CI: 1.26- 1.62) as compared to children from rural areas of residence. If there was a separate kitchen in the household, those children had a hazard ratio of 1.6 (1.3 – 1.8) times of transiting from malnutrition state to normal state as compare to children who lived in houses where there was no separate kitchen (Table 6.4.2a).

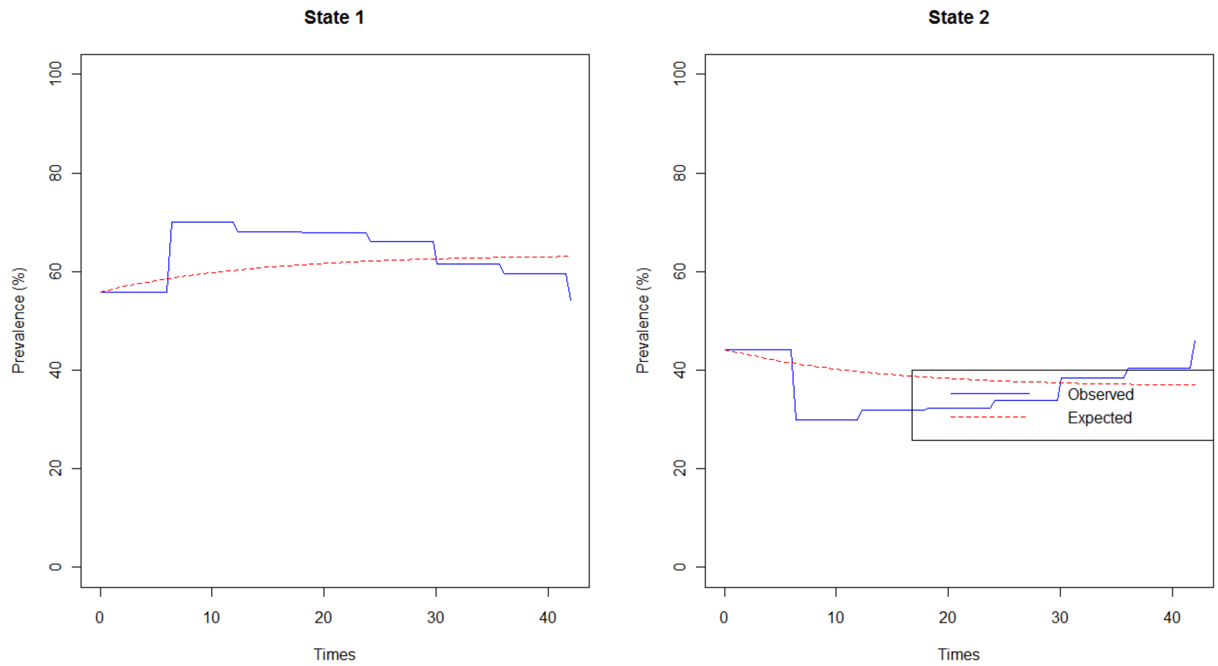
The adjusted analysis as shown in Table 6.4.2b also showed that children living in urban areas had transited faster from malnutrition state to normal state as well as from normal to malnutrition state (HR = 1.8; 95% CI: 1.4-2.2, HR = 1.3; 95% CI: 1.0-1.6) respectively. Female children transiting slower from normal to malnutrition state as compared to male children (HR = 0.76; 95% CI: 0.67 – 0.86). If the defecation was in the open fields, then the rate of transition was 1.6 (95% CI: 1.2 – 1.9) times more likely from malnutrition to normal as compared to defecation within the premises of the household (Table 6.4.2b).

The observed and expected plots are slightly different for states 1 (normal state) and states 2 (malnourished state classified being moderate or severe state). The expected plot has slightly underestimated the prevalence of "normal" children in state 1 where as in state 2, the prevalence of malnutrition children it has been overestimated (Figure 6.4.2).

Table 6.4.2b: Adjusted Markov Regression Analysis using transition intensity matrix using BMI classification:

Variables	t	t+1			
		Normal		Moderate/Severe malnutrition	
		Hazard Ratio	95% CI	Hazard Ratio	95% CI
Sex of the child					
Male	Normal			1.00	
	Malnutrition	1.00			
Female	Normal			0.76	(0.67, 0.86)
	Malnutrition	1.11	(0.98, 1.25)		
Area of residence:					
Rural	Normal			1.00	
	Malnutrition	1.00			
Urban	Normal			1.28	(1.04, 1.58)
	Malnutrition	1.78	(1.44, 2.19)		
Presence of a separate kitchen:					
Yes	Normal			1.00	
	Malnutrition	1.00			
No	Normal			0.97	(0.83, 1.15)
	Malnutrition	0.69	(0.58, 0.82)		
Defecation:					
Within the premises / toilet	Normal			1.00	
	Malnutrition	1.00			
Open fields	Normal			1.18	(0.96, 1.47)
	Malnutrition	1.55	(1.25, 1.92)		
Type of fuel used for cooking:					
Cowdung / Coal	Normal			1.00	
	Malnutrition	1.00			
Gas / Kerosene	Normal			0.92	(0.75, 1.13)
	Malnutrition	0.99	(0.81, 1.20)		

Figure 6.4.2 Observed and expected plot:



6.5 Generalized Estimating Equations for Height-for-Age classification:

The risk factor analysis was also done using Height-for-Age classification which is displayed in tables 6.5.1 and 6.5.2 respectively.

The unadjusted GEE analysis using height-for-age classification is presented in table 6.5.1. This table shows that on an average male children are at higher risk of being severely malnourished (OR = 1.1; 95% CI: 1.0 – 1.2) as compared to female children. Children living in the urban areas were 29% (23% - 33%) less likely to be severely malnourished as compared to children living in rural areas. If the mothers' of children were illiterate or literate, then there was a high risk (OR = 2.0; 95% CI: 1.7 – 2.4) of their children being severely malnourished as compared to children whose mothers had at least high school education. If also the father's of children were illiterate or literate, then there was a high risk (OR = 1.7; 95% CI: 1.5 – 1.9) of their children being severely malnourished as compared to children whose mothers had at least high school education. Children who lived in houses where type of fuel used for cooking was firewood or coal or cow dung then they had a high risk of having severe malnutrition (OR = 1.7; 95% CI: 1.5 – 2.0) as compared to children who lived in houses where type of fuel used for cooking was gas or kerosene. Children living in houses that had no separate kitchen also suffered high risk of severe malnutrition (OR = 1.2; 95% CI: 1.1 – 1.3) as compared to children living in houses that had a separate kitchen.

The table 6.5.2 shows the adjusted GEE model using Height-for-Age classification. On an average male children have significantly higher odds of being malnourished as compared to female children (OR:1.1; 95% CI: 1.0-1.2). Children from the rural area had lower severe malnutrition as compared to urban area (OR = 0.8; 95% CI: 0.7 - 0.9). On an average, children whose mothers were illiterate or literate had 1.4 (1.2 – 1.7) times the odds of having severe malnutrition as compared to mothers who had high school or above education. On an average, children whose mothers had primary or middle education also had higher odds of severe malnutrition as compared to mothers who had high school or above education (OR = 1.4; 95%CI: 1.1 – 1.8). On an average, father's whose children had primary or middle school education was 1.2 (1.0-1.3) times more likely to be malnourished as compared to children whose fathers had high school or above education. Children who lived in family that had more than 6 members was significantly associated with severe malnutrition over time (OR = 1.02; 95% CI: 1.002 – 1.04). (Table 6.5.2).

Table 6.5.1: Bivariate analysis Generalized Estimating Equations (GEE) for Malnutrition as classified using Height-for-Age classification by socio-demographic and household variables:

Risk Factors	Classified using Height-for-Age						Odds Ratio	95% CI		P value
	Normal		Mild/ Moderate		Severe					
	n	%	n	%	n	%				
Sex of the child										
Male	4041	43.5	2959	31.9	2286	24.6	1.11	1.03	1.20	0.006
Female	4197	47.5	3066	34.7	1573	17.8				
Area of Residence										
Rural	3394	38.4	3113	35.2	2330	26.4				
Urban	4844	52.2	2912	31.4	1529	16.5	0.71	0.66	0.77	<0.001
Birth Order										
1	1688	49.9	1112	32.9	585	17.3	1.00			
2	1642	43.3	1320	34.8	828	21.8	1.15	1.01	1.29	0.012
≥3	4883	44.9	3561	32.7	2440	22.4	1.14	1.03	1.27	0.033
Mother's Education										
Illiterate/ Literate	4307	40.9	3536	33.6	2676	25.4	1.98	1.66	2.38	<0.001
Primary/Middle School	2555	47.5	1878	34.9	942	17.5	1.65	1.37	2.00	<0.001
High school/College	1211	65.5	471	25.5	167	9.0	1.00			
Father's Education										
Illiterate/ Literate	1977	36.5	1862	34.3	1583	29.2	1.69	1.51	1.89	<0.001
Primary/Middle	3313	44.4	2605	34.9	1538	20.6	1.40	1.25	1.56	<0.001
High school/ College	2625	58.2	1301	28.8	586	13.0				
Number of Family Members										
<=4	1223	49.1	859	34.5	408	16.4	1.00			
5 – 6	3873	46.4	2763	33.1	1704	20.4	1.08	0.96	1.22	0.185
>6	3142	43.6	2386	33.1	1676	23.3	1.15	1.02	1.30	0.023
Fuel for cooking										
Drug/Firewood	6754	43.0	5431	34.6	3511	22.4	1.74	1.49	2.02	<0.001
Gas/Kerosene	1484	63.5	577	24.7	277	11.8				
Defecation										
Within premises /latrine	4312	54.4	2359	29.7	1261	15.9	0.71	0.65	0.76	<0.001
Open fields	3926	38.9	3649	36.1	2527	25.0				

Contd..

Risk Factors	Classified using Height-for-Age						Odds Ratio	95% CI		P value
	Normal		Moderate		Severe					
	n	%	n	%	n	%				
Type of roof										
Thatched	2756	40.0	2401	34.9	1726	25.1	1.40	1.25	1.58	<0.001
Tiled	3411	47.4	2445	34.0	1337	18.6	1.17	1.04	1.32	0.010
RCC/Pukka/others	2071	52.3	1162	29.3	725	18.3	1.00			
Type of Floor										
Kucha	2756	38.7	2523	35.5	1838	25.8	1.29	1.20	1.39	<0.001
Pukka	5482	50.2	3485	31.9	1950	17.9				
Presence of a Separate Kitchen										
Yes	6547	48.6	4307	31.9	2630	19.5				
No	1691	37.2	1701	37.4	1158	25.5	1.23	1.14	1.33	<0.001
Consanguineous Marriage										
Yes	2939	43.7	2246	33.4	1539	22.9	1.07	0.99	1.16	0.078
No	5268	46.5	3744	33.1	2310	20.4				
Type of House										
Brick and cement	2480	52.9	1439	30.7	768	16.4	1.00	1.08	1.32	<0.001
Brick and/or mud	4175	45.2	3090	33.4	1974	21.4	1.20	1.24	1.54	<0.001
Others	1583	38.5	1479	36.0	1046	25.5	1.38			
Follow-up										
0	1078	43.2	773	31.0	643	25.8				
1	995	43.9	711	31.4	560	24.7				
2	963	42.4	740	32.6	566	24.9				
3	958	42.7	752	33.5	533	23.8	0.96	0.96	0.97	<0.001
4	997	44.0	777	34.4	490	21.6				
5	1066	47.9	751	33.7	409	18.4				
6	1095	49.6	773	35.0	341	15.4				
7	1086	50.5	748	34.8	317	14.7				

Table 6.5.2: Adjusted GEE analysis for Malnutrition by socio-demographic and household variables using Height-for-Age classification:

Risk Factors	Odds Ratio	95% CI		P value
Sex of the child				
Male	1.10	1.01	1.19	0.020
Female				
Defecation				
Within premises/latrine	0.95	0.81	1.11	0.517
Open field				
Area of Residence				
Rural	0.80	0.68	0.92	0.003
Urban				
Presence of a Separate Kitchen				
No	0.98	0.88	1.10	0.781
Yes				
Type of Floor				
Kucha	1.06	0.93	1.21	0.405
Pukka				
Type of House				
Brick and/or cement	1.00			
Brick and/or mud	0.95	0.85	1.07	0.423
Others	0.85	0.71	1.01	0.051
Type of Roof				
Thatched	1.11	0.93	1.33	0.239
Tiled	1.08	0.95	1.23	0.240
RCC/Pukka	1.00			
Type of Fuel used for cooking				
Firewood/Cow Dung	1.19	1.00	1.42	0.055
Gas / Kerosene				
Mother's Education				
Illiterate/Literate	1.45	1.19	1.69	<0.001
Primary/Middle school	1.37	1.11	1.77	<0.001
High school/ College	1.00			
Father's Education				
Illiterate/Literate	1.17	1.03	1.32	0.012
Primary/Middle school	1.30	1.14	1.49	<0.001
High school/ College	1.00			
Birth Order				
1	1.00			
2	1.15	1.02	1.31	0.026
>=3	1.03	0.91	1.16	0.618

Contd…

Risk Factors	**Odds Ratio**	**95% CI**		**P value**
Number of Family Member				
<=4	1.00			
5-6	1.07	0.93	1.23	0.373
>=6	1.02	0.90	1.17	0.745
Interaction of male children with follow-up	1.00	0.99	1.01	0.425
Interaction of defecation within premises with follow-up	0.99	0.97	1.01	0.267
Interaction of kucha type of floor with follow-up	0.99	0.97	1.00	0.097
Interaction of kitchen present with follow-up	1.00	0.99	1.01	0.842
Interaction of brick and mud type of house with follow-up	1.02	1.00	1.04	0.060
Interaction of brick and other type of house with follow-up	1.01	0.99	1.02	0.165
Interaction of thatched roof with follow-up	1.01	0.99	1.03	0.430
Interaction of tiled roof with follow-up	1.00	0.99	1.02	0.553
Interaction of illiterate or literate education of mother with follow-up	1.01	0.98	1.03	0.513
Interaction of primary or middle school educated mother with follow-up	1.00	0.98	1.02	0.921
Interaction of illiterate or literate father with follow-up	1.01	0.99	1.02	0.258
Interaction of primary or middle educated father with follow-up	1.01	0.99	1.03	0.373
Interaction of first child with follow-up	0.99	0.98	1.01	0.287
Interaction of 5-6 members in a family with follow-up	1.02	1.00	1.03	0.050
Interaction of >6 members in a family with follow-up	1.02	1.002	1.04	0.029

6.6 Markov Chain using Height-for-Age classification (stunted):

6.6.1 Transition Probability and Mean Passage Time – overall

The transition probabilities from one state of malnutrition to another state are presented in table 6.6.1a for the overall using Height-for-age classification. The transition probability from severe state at time t to moderate state at time t+1 is 0.14 and from severe to normal is 0.001. The transition probability from moderate to normal is 0.10. (Table 6.6.1a).

The overall first mean passage times are presented in table 6.6.1b which were obtained using height-for-age classification. The transition from severe state of malnutrition to normal state and moderate state takes about 10 (8.4 – 11.3) years and 4 (2.9 – 4.3) years respectively (Table 6.6.1b).

Table 6.6.1a: Overall Transition Probability Matrices of Malnutrition according to Height-for-Age classification

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Overall	Normal	0.95	0.05	0.001	6849
	Moderate	0.10	0.85	0.04	5059
	Severe	0.001	0.14	0.86	3361

Table 6.6.1b: Overall Mean Passage Time (years) and 95% Confidence Interval using Height-for-Age classification

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Overall:			
Normal	0.83 (0.43, 1.24)	9.85 (7.98, 11.70)	41.32 (33.82, 48.87)
Moderate	6.39 (5.02, 7.76)	1.66 (0.81, 2.50)	32.39 (24.99, 48.87)
Severe	9.86 (8.38, 11.34)	3.64 (2.93, 4.35)	5.09 (1.53, 8.56)

6.6.2 Transition Probability and Mean Passage Time using Height-for-Age classification (stunted) by sex of the child:

The transition probabilities for male children and female children using Height-for-age classification are presented in Table 6.6.2a. The probability of transition for male and female children is similar when moving from severe to normal state of malnutrition from time t to t+1 (0.002 and 0.001 respectively). The probability of transition from severe state to moderate for male child was 0.13 and for female child was 0.15 (Table 6.6.2a).

The first mean passage times for male and female children are presented in Table 6.6.2b. The mean passage time when moving from severe state of malnutrition to normal state among male children was 10 (8.5 – 11.5) years and among female children was 9.7 (8.2 – 11.1) years. The average number of years taken to move from severe state of malnutrition to moderate state of malnutrition among male and female children was 3.9 (3.2 – 4.7) and 3.3 (2.6 – 3.9) years respectively. The average number of transitions from severe to normal and severe to moderate state of malnutrition was almost similar among male and female children (Table 6.6.2b).

The hypothesis testing to see if the transitions differed for male and female children was tested using log linear model whose results are presented in Table 6.6.2c. Log linear models suggest there was no difference in the transitions between male and female children (Table 6.6.2c).

6.6.3 Transition Probability and Mean Passage Time using Height-for-Age classification (stunted) by area of residence:

The Table 6.6.3a presents the transition probabilities by area of residence. The probability of transition from severe state to normal state for children living in rural was 0.004 and 0.003 for children in the urban areas. The transition from severe state to moderate state was slightly higher for children living in urban as compared to children in rural areas (0.17 vs 0.12). The transition probability from moderate state of malnutrition to normal state for children in the rural and urban areas was 0.08 and 0.12 respectively (Table 6.6.3a).

The transition time from severe to moderate state of malnutrition for children in rural areas takes about 4.2 (3.4 – 5.0) years where as for those children in the urban areas it takes about 3.0 (2.4 – 3.6) years. However, the transition from severe state to normal state is faster for children living in urban areas than for those children in rural area which is about 8.1 (6.9 – 9.3) years as compared to about 11.9 (10.1 – 13.7) years. The time taken to transit from moderate state to normal state of malnutrition for children living in rural area is 7.7 (6.1 – 9.4) years and it was 5.3 (4.2 – 6.4) years for children living in urban area. This transition time matrix shows that children in the urban areas return to normal from severe state of malnutrition faster than children living in rural areas (Table 6.6.3b).

The result of the log linear model for area of residence is presented in Table 6.6.3c. The log linear model showed that the transition probabilities are different for children across rural and urban area ($p = 0.028$) (Table 6.6.3c).

Table 6.6.2a: Transition Probability Matrices of Malnutrition according to Height-for-age classification by sex of the child:

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Males	Normal	0.94	0.05	0.002	3353
	Moderate	0.11	0.84	0.05	2476
	Severe	0.002	0.13	0.87	1997
Females	Normal	0.95	0.05	0.001	3496
	Moderate	0.09	0.87	0.04	2583
	Severe	0.001	0.15	0.84	1364

Table 6.6.2b: Mean Passage Time (years) and 95% Confidence Interval by Sex of the child using Height-for-age classification

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Males:			
Normal	0.86 (0.42, 1.29)	9.11 (7.38, 10.84)	35.22 (28.78, 41.74)
Moderate	6.28 (4.89, 7.68)	1.69 (0.87, 2.52)	27.28 (21.15, 33.52)
Severe	9.99 (8.46, 11.52)	3.93 (3.18, 4.69)	4.11 (1.19, 6.98)
Females:			
Normal	0.81 (0.40, 1.21)	10.70 (8.68, 12.72)	49.60 (40.41, 58.77)
Moderate	6.52 (5.16, 7.86)	1.64 (0.75, 2.52)	39.43 (30.58, 48.33)
Severe	9.71 (8.24, 11.15)	3.27 (2.65, 3.88)	6.63 (2.15, 11.11)

Table 6.6.2c: Results of Log linear Model:

Model	Model	LR	Df	Deviance (G^2)	Difference in df	P value
Model 1	Saturated Model	0.000	-	-	-	-
Model 2	Sex of the child	1.656	4	1.656 - 0	4	0.799

LR – Likelihood ratio

Table 6.6.3a: Transition Probability Matrices of Malnutrition according to Height-for-age classification by area of residence:

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Rural	**Normal**	0.96	0.04	0.0003	2803
	Moderate	0.08	0.88	0.03	2598
	Severe	0.004	0.12	0.88	2022
Urban	**Normal**	0.94	0.05	0.002	4046
	Moderate	0.12	0.82	0.05	2461
	Severe	0.003	0.17	0.83	133

Table 6.6.3b: Mean Passage Time (years) and 95% Confidence Interval by Area of residence using Height-for-age classification

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Rural:			
Normal	0.83 (0.37, 1.30)	11.52 (9.32, 13.75)	53.47 (43.74, 63.30)
Moderate	7.75 (6.09, 9.40)	1.59 (0.69, 2.48)	42.33 (32.69, 51.95)
Severe	11.88 (10.08, 13.66)	4.21 (3.41, 5.00)	5.59 (1.36, 9.80)
Urban:			
Normal	0.81 (0.44, 1.17)	8.94 (7.24, 10.62)	34.99 (28.63, 41.27)
Moderate	5.32 (4.20, 6.44)	1.76 (0.90, 2.59)	27.10 (20.76, 33.45)
Severe	8.12 (6.88, 9.34)	3.04 (2.43, 3.64)	5.22 (1.91, 8.56)

Table 6.6.3c: Results of Log linear Model:

Model	Model	LR	Df	Deviance (G^2)	Difference in df	P value
Model 1	**Saturated Model**	0.000	-	-	-	-
Model 2	**Area of residence**	10.863	4	10.863 - 0	4	0.028

LR – Likelihood ratio

6.6.4 Transition Probability and Mean Passage Time using Height-for-Age classification (stunted) by presence of a separate kitchen:

The transition probability matrix from one state of malnutrition at time t to another state of malnutrition at time t+1 by the presence of a separate kitchen is presented in Table 6.6.4a. The probability of transition from severe state to normal for children living in house where a separate kitchen was present was 0.002 and 0 if there was no separate kitchen. The probability of transition from moderate to normal state of malnutrition was 0.11 for those children living in house where there was a separate kitchen as compared to 0.083 for those children living in houses when there was no separate kitchen. The probability of transition from severe to moderate for children living in house where a separate kitchen as present and for children living in house where separate kitchen was not present was 0.14 and 0.13 respectively (Table 6.6.4a)**.**

The mean passage time for transition from one state to another by presence of a separate kitchen is presented in Table 6.6.4b. The number of years taken to transit from severe to moderate was about 3.5 (2.8 – 4.2) years for children living in house where there was a separate kitchen and about 3.7 (3.0 – 4.4) years for children living in house where there was no separate kitchen. The number of years to transit from severe to normal when for children living in house that had a separate kitchen was about 9.2 (7.8 – 10.6) years whereas the number of years to transit from severe to normal state of malnutrition for children living in houses that had no separate kitchen was about 11.6 (9.8 – 13.4) years (Table 6.6.4b).

The result of log linear model comparing the presence and absence of a separate kitchen is presented in Table 6.6.4c. The deviance measure for the log linear model was 2.47 at 4 degrees of freedom which suggested that there was no difference in the transition from one state to another in the presence and absence of a separate kitchen **(Table 6.6.4c).**

6.6.5 Transition Probability and Mean Passage Time using Height-for-Age classification (stunted) by defecation:

The transition probability matrix by defecation is presented in Table 6.6.5a. The transition from severe state to normal state of malnutrition was 0.002 for children living in houses when the defecation was within the premises of the house and 0.001 for children in houses when the defecation was in the open fields. The transition probability from severe to moderate state was 0.17 for children in house where defecation was within the premises as compared to 0.11 for those children where defecation was in the open fields. The transition probability from moderate to normal was 0.13 for children where defecation was within the premises and 0.08 for children living in houses where the defecation was in the open fields (Table 6.6.5a).

The mean number of years for transiting from one state to another is presented in Table 6.6.5b. The mean number of years for moving from severe state to normal state of malnutrition was 7.8 (6.6 – 9.0) years for those children in houses where defecation was within the premises of the house and for children where the defecation was in the open fields it took about 12.6 (10.7 – 14.6) years. The number of years to move from moderate to normal state was about 4.9 (3.9 –

6.0) years for children in houses where defecation was within the premises as compared to children in houses where defecation was in the open fields which was 8.2 (6.4 – 9.9) years. (Table 6.6.5b).

The result of the log linear analysis with defecation in the model is presented in Table 6.6.5c. The deviance value was 35.6 at 4 degrees of freedom which was statistically significant ($p<0.001$) suggesting that the transition probabilities differed across the children who had different defecation habits (Table 6.6.5c).

6.6.6 Transition Probability and Mean Passage Time using Height-for-Age classification (stunted) by type of fuel used for cooking:

The transition probability matrix by type of fuel used for cooking is presented in Table 6.6.6a. The transition from severe to normal state of malnutrition was 0.002 and 0 for those children who lived in houses where the type of fuel used for cooking was firewood or cow dung and gas or kerosene respectively. The transition percentage from moderate state to normal state of malnutrition for those children who lived in houses where the type of fuel used for cooking was firewood or cow dung and gas or kerosene was 0.09 and 0.18 respectively (Table 6.6.6a).

The mean passage time of transition from one state of malnutrition to another state by type of fuel used for cooking is presented in Table 6.6.6b. The number of years taken to transit from severe state to normal state malnutrition for those children who lived in houses where the type of fuel used for cooking was firewood or cow dung and gas or kerosene was 10.4 (8.8 – 12.1) years and 7.2 (6.2 – 8.2) years respectively. The transition in terms of years from moderate state to normal for those children who lived in houses where the type of fuel used was firewood or cow dung and gas or kerosene was about 7 (5.5 – 8.5) years and about 3.5 (2.7 – 4.3) years respectively The mean number of years taken to transit from severe state of malnutrition to moderate state of malnutrition among those children who lived in houses where the type of fuel used for cooking was firewood or coal or cow dung and for those who lived in houses where the type of fuel used for cooking was gas or kerosene was 7.0 (5.5 – 8.5) years and 3.5 (2.7 – 4.3) years respectively. This transition time matrix suggested that children living in houses where the type of fuel used for cooking was gas or kerosene transited took lesser time to transit to normal state from severe state of malnutrition as compared to children living in houses that used firewood or coal or cow dung for cooking food (Table 6.6.6b).

The result of the log linear model by type of fuel used for cooking is presented in Table 6.6.6c. The deviance value comparing the saturated and reduced model with type of fuel was 15.3 at 4 degrees of freedom which was statistically significant ($p = 0.004$) (Table 6.6.6c).

Table 6.6.4a: Transition Probability Matrices of Malnutrition according to Height-for-age classification by presence of a separate kitchen:

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Yes	Normal	0.95	0.05	0.001	5454
	Moderate	0.11	0.85	0.04	3617
	Severe	0.002	0.14	0.85	2286
No	Normal	0.94	0.06	0.0007	1395
	Moderate	0.083	0.87	0.04	1428
	Severe	0.00	0.13	0.87	1015

Table 6.6.4b: Mean Passage Time (years) and 95% Confidence Interval by Presence of a separate kitchen using Height-for-age classification:

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Yes:			
Normal	0.80 (0.41, 1.18)	10.13 (8.02, 12.04)	42.70 (34.83, 50.52)
Moderate	5.90 (4.65, 7.14)	1.76 (0.86, 2.67)	33.67 (26.02, 41.25)
Severe	9.21 (7.80, 10.59)	3.54 (2.84, 4.25)	5.48 (1.71, 9.18)
No:			
Normal	0.95 (0.41, 1.48)	8.88 (7.18, 10.57)	37.62 (30.80, 44.39)
Moderate	7.85 (6.19, 9.50)	1.39 (0.72, 2.10)	29.16 (22.45, 35.85)
Severe	11.58 (9.80, 13.38)	3.73 (3.05, 4.41)	4.41 (1.29, 7.54)

Table 6.6.4c: Results of Log linear Model:

Model	Model	LR	Df	Deviance (G^2)	Difference in df	P value
Model 1	Saturated Model	0.000	-	-	-	-
Model 2	Presence of a separate kitchen	2.468	4	2.468 - 0	4	0.650

LR – Likelihood ratio

Table 6.6.5a: Transition Probability Matrices of Malnutrition according to Height-for-age classification by Defecation:

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Within the premises of the house	Normal	0.95	0.05	0.002	3583
	Moderate	0.13	0.82	0.05	1995
	Severe	0.002	0.17	0.83	1097
Open fields	Normal	0.95	0.05	0.0006	3266
	Moderate	0.08	0.88	0.04	3050
	Severe	0.001	0.11	0.89	2153

Table 6.6.5b: Mean Passage Time (years) and 95% Confidence Interval by defecation using Height-for-age classification:

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Within the premises of the house			
Normal	0.76 (0.44, 1.09)	9.54 (7.73, 11.33)	39.49 (32.22, 46.72)
Moderate	4.89 (3.86, 5.92)	1.92 (0.99, 2.84)	31.21 (24.06, 38.24)
Severe	7.77 (6.60, 8.93)	3.03 (2.44, 3.63)	5.83 (2.09, 9.50)
Open fields			
Normal	0.90 (0.38, 1.43)	10.20 (8.25, 12.14)	43.61 (35.64, 51.53)
Moderate	8.16 (6.39, 9.93)	1.52 (0.69, 2.33)	33.89 (26.07, 41.62)
Severe	12.64 (10.68, 14.60)	4.71 (3.80, 5.61)	4.21 (0.91, 7.53)

Table 6.6.5c: Results of Log linear Model:

Model	Model	LR	Df	Deviance (G^2)	Difference in df	P value
Model 1	**Saturated Model**	0.000	-	-	-	-
Model 2	**Defecation**	35.580	4	35.580 - 0	4	<0.001

LR – Likelihood ratio

Table 6.6.6a: Transition Probability Matrices of Malnutrition according to Height-for-age classification by type of fuel used for cooking:

	Grading (t)	Grading (t+1)			
		Normal	Moderate	Severe	Total no.
Firewood or Cow dung	Normal	0.95	0.05	0.002	5615
	Moderate	0.09	0.86	0.04	4553
	Severe	0.002	0.14	0.86	3047
Gas or Kerosene	Normal	0.95	0.05	0	1234
	Moderate	0.18	0.78	0.04	492
	Severe	0	0.13	0.86	320

Table 6.6.6b: Mean Passage Time (years) and 95% Confidence Interval by type of fuel used for cooking using Height-for-age classification

Time t	Time (t+1) Mean Passage Time (95% CI) (in years)		
	Normal	Moderate	Severe
Firewood or Cow dung:			
Normal	0.87 (0.41, 1.33)	9.76 (7.91, 11.59)	38.50 (31.62, 45.41)
Moderate	6.99 (5.52, 8.46)	1.57 (0.76, 2.38)	29.78 (22.91, 36.62)
Severe	10.45 (8.84, 12.07)	3.66 (2.95, 4.36)	4.72 (1.45, 7.92)
Gas or Kerosene:			
Normal	0.67 (0.44, 0.89)	10.28 (8.31, 12,24)	75.37 (61.10, 89.18)
Moderate	3.48 (2.69, 4.26)	2.52 (1.34,3.66)	65.08 (50.69, 79.25)
Severe	7.20 (6.18, 8.23)	3.72 (3.04, 4.39)	9.24 (2.43, 15.99)

Table 6.6.6c: Result of Log linear Model:

Model	Model	LR	Df	Deviance (G^2)	Difference in df	P value
Model 1	Saturated Model	0.000	-	-	-	-
Model 2	Type of fuel used for cooking	15.299	4	15.299 - 0	4	0.004

LR – Likelihood ratio

6.7. Markov Regression Analysis using transition probabilities for Height-for-Age classification:

Markov regression analysis using transition probabilities and transition intensities are presented in section 6.7. The Markov regression analysis for Height-for-Age classification using transition probabilities is presented in Table 6.7.1a and 6.7.1b. The Table 6.7.1a shows the unadjusted analysis and Table 6.7.1b shows the adjusted analysis. The unadjusted analysis presented in Table 6.7.1a has been done for three models. The adjusted analysis was performed with the significant at $p = 0.15$ obtained from the third model.

Table 6.7.1a: Unadjusted Markov regression analysis considering the transition probabilities using Height-for-Age classification:

Risk Factors	OR	95% CI		Robust SE	P value
Sex of the child Male Female	1.08	0.90	1.30	1.00	0.387
Previous state (M_{it})	145.34	127.67	165.46	9.63	<0.001
Interaction of sex of the child and previous state	1.01	0.87	1.16	0.07	0.934
Area of Residence Rural Urban	1.23	1.02	1.47	0.11	0.031
Previous state	172.89	151.23	197.65	11.78	<0.001
Interaction of area and urban area of residence	0.72	0.62	0.83	0.05	<0.001
Father's Education Illiterate/Literate Primary/Middle High school/ above	1.45 1.26 1.00	1.14 1.08	1.85 1.48	0.17 0.10	0.002 0.004
Previous state	144.20	124.90	166.49	10.63	<0.001
Interaction of father'[s education and previous state	0.99	0.91	1.10	0.05	0.992
Mother's Education Illiterate/Literate Primary/Middle High school /above	1.40 1.22 1.00	1.08 0.99	1.83 1.50	0.18 0.12	0.012 0.058
Previous state	139.05	114.92	168.24	13.27	<0.001
Interaction of mother's education and previous state	1.02	0.92	1.14	0.05	0.639

Contd..

Variables	OR	95% CI		Robust SE	P value
Consanguineous Marriage					
Yes	1.03	0.85	1.24	0.09	0.773
No					
Previous state	142.46	126.67	160.21	10.03	<0.001
Interaction of consanguineous and previous state	1.07	0.92	1.24	0.90	0.370
Birth Order					
1	1.00				
2	1.06	0.88	1.28	0.10	0.533
≥3	1.05	0.83	1.32	0.12	0.686
Previous state	145.36	122.78	172.08	12.70	<0.001
Interaction of birth order and previous state	1.00	0.91	1.10	0.05	0.950
Number of Family Members					
≤4	1.00				
5 – 6	1.07	0.88	1.29	0.10	0.496
≥7	1.07	0.82	1.41	0.14	0.611
Previous state	145.30	122.68	172.08	12.68	<0.001
Interaction of family members and previous state	1.00	0.90	1.11	0.05	0.936
Type of House					
Brick and/or cement	1.00				
Brick and/or mud	1.13	0.96	1.34	0.09	0.130
Others	1.20	0.92	1.57	0.16	0.166
Previous state	140.36	121.56	162.07	10.43	<0.001
Interaction of type of house and previous state	1.04	0.94	1.15	0.05	0.477

Contd..

Variables	OR	95% CI		Robust SE	P Value
Defecation Practice Within the premises Open fields	0.97	0.81	1.17	0.90	0.787
Previous state	157.11	138.73	177.93	10.03	<0.001
Interaction of defecation and previous state	0.83	0.71	0.95	0.06	0.010
Type of Fuel Firewood/ Coal/Cow dung Gas/Kerosene	1.38	1.06	1.81	0.20	0.018
Previous state	146.70	118.26	181.97	16.34	<0.001
Interaction of risk factor and previous state	0.99	0.79	1.23	0.11	0.912
Type of Roof Thatched Tiled RCC/Pukka/ Others	1.33 1.25	1.12 0.99	1.58 1.60	0.15 0.12	0.001 0.059
Previous state	157.58	135.12	183.77	12.59	<0.001
Interaction of risk factor and previous state	0.93	0.85	1.03	0.04	0.163
Presence of a separate Kitchen No Yes	1.20	0.97	1.48	0.12	0.099
Previous state	146.71	131.17	164.10	8.59	<0.001
Interaction of risk factor and previous state of malnutrition	0.97	0.82	1.14	0.81	0.686

Contd…

Risk Factors	**OR**	**95% CI**		**Robust SE**	**P value**
Type of Floor Kucha Pukka	1.09	0.90	1.32	0.11	0.360
Previous state	142.49	126.55	160.43	8.79	<0.001
Interaction of risk factor and previous state of malnutrition	1.05	0.90	1.22	0.08	0.518

The unadjusted analysis using height-for-age using the transition probabilities is presented in Table 6.7.1a.

The unadjusted analysis showed that previous severe state of malnutrition had high odds of having severe malnutrition at the current state. In other words, malnutrition at the previous time was the main risk factor that affected the malnutrition at the current time. Apart from the previous time malnutrition there were some other risk factors associated with present state of malnutrition such as area of residence, education of mother and father of the children, defecation, presence of a separate kitchen and type of fuel used for cooking. Children from urban areas were 20% less likely to be severely malnourished irrespective of previous state of malnutrition. Children whose father were illiterate or literate were 1.4 (1.1 – 1.8) times the odds of being severely malnourished as compared to children whose fathers had high school or college education. Children whose father had primary or middle school education also had higher odds of malnutrition as compared to children whose father had education of high school or above (OR = 1.3; 95% CI: 1.1-1.5). Children whose mothers were illiterate of literate had higher odds of severe malnutrition (OR = 1.4; 95% CI: 1.1 – 1.8) as compared to children whose mothers had atleast high school education. Children who lived in houses where defecation was within the household premises or latrines were 17% less likely to be severely malnourished irrespective of previous state of malnutrition as compared to children who defecated in open fields (5% - 20%) (Table 6.7.1a).

The adjusted regression analysis using transition probabilities is presented in Table 6.7.1b using height-for-age classification.

If the children defecated within the house or latrines then those children were 35% less likely to be severely malnourished as compared to children who defecated in the open fields after adjusting for other risk factors (11% - 52%). Children who lived in rural areas had higher odds of severe malnutrition as compared to children in urban areas (OR = 2; 95% CI: 1.4-3.0). Father's education was also associated with severe malnutrition at the current state. Children whose fathers were illiterate or literate were 1.3 times more likely to be malnourished as compared to children whose fathers had at least high school education (1.1 – 1.5) Children who lived in urban areas were 31% less likely to be severely malnourished irrespective of previous state of malnutrition as compared to children living in rural area**.** (Table 6.7.1b).

Table 6.7.1b: Adjusted Markov Regression analysis with Transition Probabilities using Height-for-age classification:

Variables	OR	95% CI		Robust SE	P value
Presence of a separate Kitchen					
No					
Yes	1.04	0.89	1.21	0.597	0.78
Defecation Practice					
Within the premises					
Open fields	0.65	0.48	0.89	0.007	0.99
Mother's Education					
Illiterate/Literate	1.21	0.99	1.47	0.064	0.12
Primary/Middle	1.18	0.97	1.43	0.097	0.11
High school/ above	1.00				
Area of Residence					
Rural					
Urban	1.96	1.43	2.68	<0.001	0.29
Type of House					
Brick and/or cement	1.00				
Brick and/or mud	1.03	0.90	1.19	0.624	0.07
Others	1.11	0.90	1.34	0.335	0.12
Father's Education					
Illiterate/Literate	1.26	1.07	1.49	0.005	0.10
Primary/Middle	1.15	0.99	1.32	0.051	0.83
High school/ above	1.00				
Type of fuel used for cooking					
Firewood/Coal/Cow Dung					
Gas/Kerosene	1.13	0.94	1.36	0.192	0.11
Type of roof					
Thatched	1.11	0.87	1.43	0.394	0.14
Tiled	1.02	0.82	1.26	0.873	0.11
RCC/Pukka/Others	1.00				
Previous State	176.22	148.89	208.57	<0.001	15.41
Interaction of defecation within premises and previous time malnutrition	1.24	0.98	1.58	0.073	0.15
Interaction of type of roof with previous time malnutrition	0.98	0.89	1.08	0.645	0.05
Interaction of urban area of residence with previous time malnutrition	0.59	0.46	0.75	<0.001	0.07

6.8. Markov Regression Analysis using intensity rate for Height-for-Age classification:

The Markov regression analysis using instantaneous rate of transition by height-for-age classification is presented in the section 5.8. The unadjusted analysis using intensity rate is presented in Table 6.8.1a where as adjusted analysis is presented in Table 6.8.1b. The Markov regression using intensity rates was performed considering only two categories which are “normal” and “moderate or severe” malnutrition.

The unadjusted analysis using transition intensity matrix is presented in Table 6.8.1a. The unadjusted analysis showed that children residing in the rural areas and transiting from normal to malnutrition had 1.4 times the hazards of being malnourished as compared to children living in urban areas (1.2 – 1.7). At the same time, children living in the urban areas transited faster to normal from malnutrition as compared to children living in rural areas (HR = 1.8; 95% CI: 1.5-2.1). Children who defecated within the premises of the house or used latrines transited faster from malnutrition to normal as compared to children who defecated in the open fields (HR = 1.8; 95% CI: 1.5 – 2.2). Children who lived in a household where coal or firewood was used for cooking were slower to transit from malnourished state to normal state (HR = 0.4; 95% CI: 0.3 – 0.6). Children whose parents were illiterate or literate and who had primary or middle school education had slower rate of transition from malnutrition to normal. The Markov regression analysis using transition intensity rates also showed that if the household had a thatched or tiled roof then those children had slower rate of transition from malnutrition to normal as compared to children living in households with RCC or pukka type of roof (Table 6.8.1a).

The adjusted analysis using the Markov regression is presented in Table 6.8.1b.

The adjusted analysis showed that children residing in urban areas were more likely (HR = 2.2) to be transited from normal to malnutrition state as compared to children residing in rural areas (1.5 – 3.2). Children who lived in a house without separate kitchen also had higher rates of transition to malnutrition state from normal (HR = 1.5; 95% CI: 1.0-2.5). If children lived in houses where fuel used for cooking was cow dung or firewood, then there was a slower odds of transition from malnutrition state to normal state as compared to children living in houses where fuel used for cooking was kerosene or gas (HR = 0.7; 95% CI: 0.6 – 0.9). Children whose parents had less than high school education showed slower transition rates as compared to children whose parents had at least high school education. Children who lived in thatched type of houses had higher hazard of transiting from normal to malnutrition state as compared to children living in RCC or pukka houses (HR = 1.3; 95% CI: 1.1 – 1.5) (Table 6.8.1b).

Table 6.8.1a: Unadjusted Markov Regression Analysis using transition intensity matrix using Height-for-age classification:

Variables	t	t+1			
		Normal		Moderate/Severe Malnutrition	
		Hazard Ratio	95% CI	Hazard Ratio	95% CI
Sex of the child:					
Male	Normal			1.18	0.96 – 1.45
	Malnutrition	0.98	0.83 – 1.16		
Female	Normal			1.00	
	Malnutrition	1.00			
Area of residence:					
Rural	Normal			1.39	1.21 – 1.73
	Malnutrition	1.80	1.51 – 2.14		
Urban	Normal			1.00	
	Malnutrition	1.00			
Presence of a separate kitchen:					
Yes	Normal			1.00	
	Malnutrition	1.00			
No	Normal			1.07	0.84 – 1.37
	Malnutrition	0.70	0.58 – 0.87		
Defecation					
Within the premises / toilet	Normal			1.13	0.92 – 1.39
	Malnutrition	1.83	1.54 – 2.18		
Open fields	Normals			1.00	
	Malnutrition	1.00			
Type of fuel used for cooking:					
Cowdung / Coal	Normal			0.98	0.75 – 1.29
	Malnutrition	0.45	0.35 – 0.56		
Gas / Kerosene	Normal			1.00	
	Malnutrition	1.00			

Contd…

Variables	t	t+1			
		Normal		Moderate/Severe Malnutrition	
		Hazard Ratio	**95% CI**	**Hazard Ratio**	**95% CI**
Type of Floor:					
Kucha	Normal Malnutrition	0.62	0.51 – 0.74	0.90	0.72 – 1.12
Pukka	Normal Malnutrition	1.00		1.00	
Consanguineous Marriage:					
Yes	Normal Malnutrition	0.89	0.75 – 1.07	1.07	0.86 – 1.32
No	Normal Malnutrition	1.00		1.00	
Mother's Education:					
Illiterate/Literate	Normal Malnutrition	0.39	0.29 – 0.43		
Primary/Middle School	Normal Malnutrition	0.62	0.59 - 0.65	1.00	0.94 – 1.07
Higher Sec or above	Normal Malnutrition	1.00		1.00	
Father's Education:					
Illiterate/Literate	Normal Malnutrition	0.39	0.38 – 0.40	1.00	0.98 – 1.02
Primary/Middle School	Normals Malnutrition	0.62	0.59 – 0.64	1.01	0.94 – 1.06
Higher Sec or above	Normals Malnutrition	1.00		1.00	

Contd…

Variables	t	t+1			
		Normal		**Moderate/Severe Malnutrition**	
		Hazard Ratio	**95% CI**	**Hazard Ratio**	**95% CI**
Number of Family Members:					
<=4	Normal			1.00	
	Malnutrition	1.00			
5-6	Normal			0.99	0.86 – 1.14
	Malnutrition	0.90	0.80 – 1.01		
>6	Normal			0.98	0.77 – 1.25
	Malnutrition	0.82	0.67 – 1.00		
Type of House:					
Brick & cement	Normal			1.00	
	Malnutrition	1.00			
Brick & mud	Normal			0.99	0.93 – 1.05
	Malnutrition	0.71	0.67 – 0.75		
Others	Normal			0.97	0.95 – 1.00
	Malnutrition	0.50	0.49 – 0.52		
Type of Roof:					
Thatched	Normal			1.12	1.08 – 1.16
	Malnutrition	0.55	0.54 – 0.57		
Tiled	Normal			1.06	0.97 – 1.15
	Malnutrition	0.74	0.69 – 0.78		
RCC/Pukka/Others	Normal			1.00	
	Malnutrition	1.00			
Birth Order:					
1	Normal			1.00	
	Malnutrition	1.00			
2	Normals			0.98	0.90 – 1.07
	Malnutrition	0.89	0.83 – 0.97		
>=3	Normal			0.96	0.89 – 1.04
	Malnutrition	0.80	0.75 – 0.85		

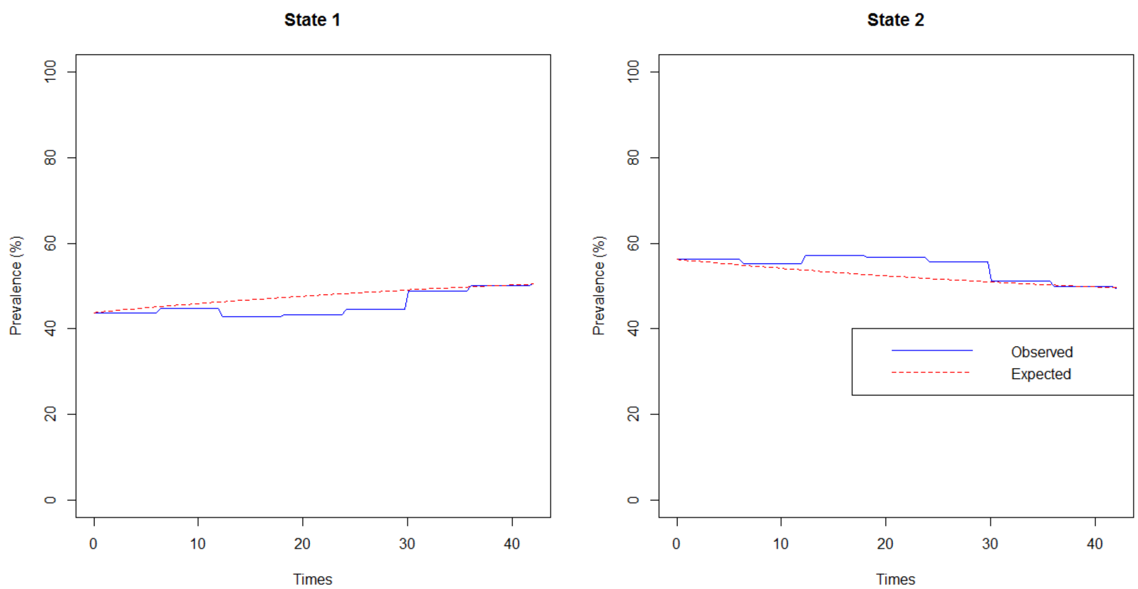
Table 6.8.1b: Adjusted Markov Regression Analysis using transition intensity matrix using Height-for-age classification:

Variables	t	t+1			
		Normal		Moderate/Severe Malnutrition	
		Hazard Ratio	95% CI	Hazard Ratio	95% CI
Area of residence:					
Rural	Normal			2.16	1.46 – 3.19
	Malnutrition	1.25	0.90 – 1.74		
Urban	Normal			1.00	
	Malnutrition	1.00			
Presence of a separate kitchen:					
Yes	Normal			1.00	
	Malnutrition	1.00			
No	Normal			1.47	1.04 – 2.08
	Malnutrition	1.17	0.90 – 1.54		
Defecation:					
Within the premises / toilet	Normal			0.73	0.49 – 1.08
	Malnutrition	1.12	0.80 – 1.58		
Open fields	Normal			1.00	
	Malnutrition	1.00			
Type of fuel used for cooking:					
Cowdung / Coal	Normal			1.08	0.77 – 1.50
	Malnutrition	0.73	0.56 – 0.95		
Gas / Kerosene	Normal			1.00	
	Malnutrition	1.00			
Mother's Education:					
Illiterate/Literate	Normal			0.93	0.85 – 1.02
	Malnutrition	0.67	0.65 – 0.69		
Primary/Middle	Normal			0.96	0.95 – 0.99
	Malnutrition	0.82	0.81 – 0.83		
High school or above	Normal			1.00	
	Malnutrition	1.00			
Father's Education:					
Illiterate/Literate	Normal			1.03	0.57 – 1.19
	Malnutrition	0.62	0.55 – 0.69		
Primary/Middle	Normal			1.01	0.89 – 1.19
	Malnutrition	0.78	0.76 – 0.81		
High school or above	Normal			1.00	
	Malnutrition	1.00			

Contd…

Variables	t	t+1			
		Normal		**Moderate/Severe Malnutrition**	
		Hazard Ratio	**95% CI**	**Hazard Ratio**	**95% CI**
Type of Floor:					
Kucha	Normal			0.71	0.49 – 1.02
	Malnutrition	0.96	0.71 – 1.30		
Pukka	Normal			1.00	
	Malnutrition	1.00			
Type of House:					
Brick and Cement	Normal			1.00	
	Malnutrition	1.00			
Brick and Mud	Normal			1.06	0.99 – 1.07
	Malnutrition	0.95	0.94 – 0.99		
Others	Normal			1.14	1.01 – 1.28
	Malnutrition	0.90	0.83 – 0.98		
Birthorder:					
1	Normal			1.00	
	Malnutrition	1.00			
2	Normal			0.98	0.96 – 1.00
	Malnutrition	0.94	0.91 – 0.96		
>=3	Normal			0.95	0.94 – 0.99
	Malnutrition	0.88	0.86 – 0.89		
Type of Roof:					
Thatched	Normal			1.26	1.08 – 1.47
	Malnutrition	0.90	0.83 – 0.97		
Tiled	Normal			1.12	1.08 – 1.17
	Malnutrition	0.95	0.94 – 0.96		
RCC/Pukka/Others	Normal			1.00	
	Malnutrition	1.00			

Figure 6.8.1 Observed and Expected plots:



The observed and expected percentages in each state are presented in figure . State 1 represents the predicted percentage of "normal" children and state 2 represents the predicted percentage of "malnourished" children by the model. The predicted prevalence of "normal" cases is slightly overestimated by the model where as the predicted number of "malnutrition" cases is also slightly underestimated by the model between 10-30 months.

6.9 Comparison of GEE, Markov Regression with transition probabilities and transition intensity rates for malnutrition using BMI classification:

6.9.1. Comparison of Longitudinal Data Analysis and Markov Regression Models using only two categories (normal and moderate or severe) of BMI classification:

The results comparing GEE or random effects model to the results obtained from Markov regression using transition probabilities and intensity rates using binary outcome is shown in Table 6.9.1. The Markov regression using transition probabilities and intensity rate showed that male children on an average were having higher risk of malnutrition as compared to female children. The odds of having malnutrition in male children was 1.4 (1.2 – 1.5) times more as compared to female child using Markov regression analysis with transition probabilities and the hazard of moving from normal to malnutrition state in male children was 1.2 (1.1 – 1.4) times as compared to female children. The GEE analysis also showed male children were more likely to be malnourished. The Markov regression model showed area of residence was significantly associated with current state of malnutrition irrespective of the previous state of malnutrition (OR = 2.2; 95% CI: 1.6 – 3.0). The Markov regression analysis using transition probabilities and intensity rates showed presence of a separate kitchen was an important risk factor irrespective of the previous state of malnutrition (OR = 1.5; 95% CI: 1.1 – 1.9) for current state of malnutrition. The previous state of malnutrition was very highly associated with current state from the Markov model (OR = 12.6; 95% CI: 8.3 – 19.3). The table 6.9.1 that shows the comparison of the Markov regression and GEE showed that there were many variables that turned out to be significant in the Markov regression model as compared to GEE model findings (Table 6.9.1).

6.9.2 Comparison of Longitudinal Data Analysis and Markov Regression Models using ordinal categories of BMI classification:

The Table 6.9.2 shows the comparison of results obtained using GEE and Markov Regression using transition probabilities. Both the results show male children are on an average have significantly high odds of being severe malnutrition as compared to female children (GEE: OR = 1.3 (1.2 – 1.6); MR: OR = 1.2 (1.1 – 1.4)). The GEE analysis showed presence of a separate kitchen has 25% less likely to have severe malnutrition as compared to no separate kitchen (10% - 90%). The Markov regression analysis showed that the previous state of malnutrition was very highly significantly associated with the current state of malnutrition. In addition to that, urban area of residence, separate kitchen not present and number of family members were also significant factors associated with severe malnutrition irrespective of the previous state of malnutrition**.** (Table 6.9.2).

Table 6.9.1 Adjusted Risk Factor analysis considering BMI classification into two categories which are 'normal' and 'moderate or severe' Malnutrition:

Risk Factors	Generalized Estimating Equations		Markov Regression using Transition Probabilities		Markov Regression using Transition Intensity rate			
					Normal to Moderate or Severe malnutrition		Moderate or Severe to normal	
	OR	95% CI	OR	95% CI	HR	95% CI	HR	95% CI
Sex of the Child								
Male	1.3	1.1 – 1.5	1.4	1.2 – 1.5	1.2	1.1 – 1.4		
Female								
Area of Residence								
Rural							1.8	1.4 – 2.2
Urban					1.3	1.0 – 1.6		
Defecation								
Within the premises								
Open Fields							1.5	1.2 – 2.0
Presence of a separate kitchen								
Yes								
No	1.2	1.0 – 1.5					0.7	0.6 – 0.8
Interaction of rural area with previous state of malnutrition	1.1	1.0 – 1.1	2.2	1.6 – 3.0				
Interaction of malnutrition state at time t-1 (previous state) with no presence of a kitchen			1.5	1.1 – 1.9				
Interaction of defecation with previous state of malnutrition			1.2	1.1 – 1.8				
Previous state of malnutrition (malnutrition state at time t-1)			12.6	8.3 – 19.3				

Table 6.9.2 Adjusted Risk Factor analysis considering BMI classification into ordinal categories (Malnutrition –"Normal", "Moderate" and "Severe"):

Risk Factors	Generalized Estimating Equations		Markov Regression using Transition Probabilities	
	OR	95% CI	OR	95% CI
Sex of the child				
Male	1.27	1.05 – 1.40	1.21	1.07 – 1.25
Female				
Defecation				
Within the premises	1.77	1.30 – 1.91	1.26	1.06 – 1.48
Open fields				
Presence of a separate Kitchen				
Yes				
No	1.28	1.01 – 1.61		
Interaction of defecation with followup	0.94	0.92 – 0.97		
Interaction of sex of the child with followup	1.03	1.01 – 1.05		
Previous state of malnutrition (malnutrition state at time t-1)			9.30	7.85 – 11.01
Interaction of no presence of kitchen and previous state			1.28	1.08 – 1.51
Interaction of family members with previous state of malnutrition			1.17	1.06 – 1.29

6.10. Comparison of GEE, Markov Regression with transition probabilities and transition intensity rates for malnutrition using Height-for-Age classification:

6.10.1 Comparison of Longitudinal Data Analysis and Markov Regression Models using only two categories (normal, moderate or severe) of Height-for-Age classification:

The results comparing the results from GEE and Markov Regression analysis with transition probabilities and intensity rates is shown in tables 6.10.1.

The comparison of the risk factors using GEE and Markov regression using transition probability matrix and transition intensity matrix is presented in Table 6.10.1 with height-for-age classification. The GEE analysis, Markov regression analysis using intensity and probability rates showed that on an average mother's education as an important risk factor for moderate or severe malnutrition. GEE analysis showed that on an average father's education as also a risk factor for malnutrition. The previous state of malnutrition was highly associated with current state from Markov regression (OR = 30.0; 95% CI: 25.0 – 36.0). The mother's education was associated with current state of malnutrition irrespective of previous state from the Markov regression analysis. The hazard of moving from normal to malnourished state (moderate or severe state of malnutrition) for children living in houses where there was no separate kitchen was 1.5 (1.0 – 2.1) times as compared to children living in houses with no separate kitchen. (Table 6.10.1).

6.10.2 Comparison of Longitudinal Data Analysis and Markov Regression Models using ordinal categories of BMI classification:

The results comparing GEE and Markov regression using transition probabilities with Height-for-Age classification classified in an ordinal scale is presented in Table 6.10.2. The common risk factor that was present in GEE and Markov regression analysis was area of residence. Markov regression analysis showed Area of residence and its interaction with previous time malnutrition state, father's education as the other risk factors for severe malnutrition where as GEE analysis showed males being at higher risk and interaction of family members with follow-up time to be associated with severe malnutrition (Table 6.10.2).

Table 6.10.1 Adjusted Risk Factor analysis considering height-for-age classification into two categories which are 'normal' and 'moderate or severe' Malnutrition:

Risk Factors	Generalized Estimating Equations		Markov Regression using Transition Probabilities		Markov Regression using Transition Intensity rate			
					Normal to Moderate or Severe malnutrition		Moderate or Severe to normal	
	OR	95% CI	OR	95% CI	HR	95% CI	HR	95% CI
Mother's Education								
Illiterate/Literate	1.65	1.21 – 2.25	1.09	1.01 – 1.24	0.93	0.85 – 1.02	0.67	0.65 – 0.69
Primary/Middle	1.40	1.00 – 1.96	1.05	1.01– 1.41	0.96	0.95 – 0.99	0.82	0.81 – 0.83
High school/above	1.00		1.00		1.00		1.00	
Father's Education								
Illiterate/Literate	1.37	1.09 – 1.72					0.62	0.55 – 0.69
Primary/Middle	1.59	1.21 – 2.09					0.78	0.76 – 0.81
High school/above	1.00						1.00	
Presence of a separate Kitchen								
Yes								
No					1.47	1.04 – 2.08		
Type of Roof								
Thatched								
Tiled								
RCC/Pukka/Others								
Area of residence								
Rural					2.16	1.46 – 3.19		
Urban								
Type of Fuel used for cooking								
Firewood/Cow dung							0.73	0.56 – 0.95
Gas/Kerosene								
Birth order								
1							1.00	
2							0.94	0.91 – 0.96
>=3							0.88	0.86 – 0.89
Interaction of mother's education with previous time malnutrition state			0.63	0.59 – 0.73				
Interaction of area and previous time malnutrition state	1.05	1.01 – 1.09						
Previous time malnutrition state			30.01	24.99 – 36.04				

Table 6.10.2 Adjusted Risk Factor analysis considering Height-for-age classification into three ordinal categories (Malnutrition – "Normal", "Moderate" and "Severe"):

Risk Factors	Generalized Estimating Equations		Markov Regression using Transition Probabilities	
	OR	**95% CI**	**OR**	**95% CI**
Sex of the child				
Male	1.10	1.01 – 1.19		
Female				
Area of Residence				
Rural			1.96	1.43 – 2.68
Urban	0.80	0.68 – 0.92		
Defecation				
Within the premises			0.65	0.48 – 0.89
Open fields				
Mother's Education				
Illiterate/Literate	1.45	1.19 – 1.69		
Primary/Middle school	1.37	1.11 – 1.77		
High school or above	1.00			
Father's Education				
Illiterate/Literate	1.17	1.03 – 1.32	1.26	1.07 - 1.49
Primary/Middle school	1.30	1.14 – 1.49	1.15	0.99 – 1.32
High school or above	1.00		1.00	
Interaction of greater than 6 members in a family	1.02	1.00 – 1.04		
Interaction of urban area of residence with the previous time malnutrition state			0.59	0.46 – 0.75
Previous state of malnutrition (malnutrition state at time t-1)			176.22	148.89 – 208.57

6.11 Comparison of coverage probability and length of the confidence interval:

The Table 6.11a shows the coverage probability and length of the confidence interval for the two important risk factors as they were significant using Markov Regression with transition probabilities and GEE analysis using BMI classification.

The table 6.11aa shows that coverage probability for presence of a separate kitchen was 95% when Markov Regression analysis with transition probabilities was performed and the average length of the confidence interval was smaller than that compared to GEE analysis. The risk factors "presence of a separate kitchen","defecation" had higher coverage probabilities using Markov regression models as compared to GEE. Also, the length of CIs using Markov regression was smaller as compared to GEE (Table 6.11a).

Similar findings were also obtained using height-for-age classification which is shown in table 6.11b. The risk factors had higher coverage probabilities when Markov regression analysis was performed as compared to GEE analysis (Table 6.11b).

Table 6.11a Comparison of coverage probability and length of the confidence interval using BMI classification:

Risk Factors	Coverage Probability		Length of the Confidence Interval	
	Markov Regression with transition probabilities	**Generalized Estimating Equations**	**Markov Regression with transition probabilities**	**Generalized Estimating Equations**
Presence of a separate Kitchen	0.948	0.912	0.114	0.131
Defecation	0.916	0.910	0.159	0.174

Table 6.11b Comparison of coverage probability and length of the confidence interval using height for age classification:

Risk Factors	Coverage Probability		Length of the Confidence Interval	
	Markov Regression with transition probabilities	**Generalized Estimating Equations**	**Markov Regression with transition probabilities**	**Generalized Estimating Equations**
Area of residence	0.960	0.646	1.620	1.429
Defecation	0.864	0.804	0.529	1.434

DISCUSSION

7. DISCUSSION

The present study used BMI Z scores to classify malnutrition as suggested by practicing pediatricians and child psychiatrists and hence there are not many studies that have used this index to classify malnutrition. One study used BMI Z score to define underweight (15). The study of underweight or stunting (malnutrition) is particularly important as the deficiency of micronutrients in a child's early years may result in a lower attention span, decreased ability to concentrate and poor memory. In other words, deficiency of nutrients in children is known to have severe impact on the cognitive development of children (86). There are several studies that have shown some relation between child malnutrition and cognitive development. Hence it is important to study prevalence of malnutrition (underweight and stunting) among children and risk factors associated with malnutrition as the cognitive development is not a straightforward effect of malnutrition but an interaction of several risk factors with malnutrition (87, 88, 89).

Prevalence:

There are very few studies that reported the prevalence of malnutrition (underweight or stunting or wasting) among children in the age groups 5-7 years. Most of the studies reported are mostly for children under five years of age. There is one study reported from Chile that reported malnutrition among children aged 6 years. This study used underweight defined by BMI Z score. The study was a cross-sectional survey of the children entering grade one from the years 1987 – 2002. The study found the change in the prevalence of stunting and underweight over the years. The study showed that the prevalence of stunting and underweight was above 20% (15). The present study reported the prevalence of severe underweight as 22.5% which was similar to the finding from the previous study. According to the National Family Health Survey (NFHS – 3) carried out in 2005-2006, child malnutrition rates on India reported that 46% under three are underweight (90). The NFHS - 3 reported around 16% to be severely underweight using weight-for-age Z scores (90) where as the present study reported 23% underweight using weight-for-age Z scores and 22.5% using BMI z scores. The prevalence of moderate underweight from the present study using BMI z score was about 22%. When the weight-for-age Z score was used the prevalence of moderate underweight was 43% which was similar to NFHS 3 moderate prevalence (around 40.3%). The prevalence of severe stunting was 25.8% in the present study and this around 20% which was similar to the present study. There was a study that reported the changes in the malnutrition levels in India in the period 1998 – 2005 (NFHS - 2 to NFHS - 3). This study reported that the percentage of stunted children increased from 38.4% to 45.5% and has been a marginal worsening in underweight children where the prevalence increased from 45.9% to 47%. The prevalence of wasted improved from 19.1 to 15.5% (91, 92, 93). This finding suggested that though the present study collected information in the year 1997, the results are still valid as there has not been a drastic reduction in the prevalence of malnutrition in India. The other important finding from the present study was the incidence density of malnutrition (underweight, stunting and/or wasting) was obtained as this study was a prospective study which was assessed after every six months for seven follow-up apart from baseline. Most of the studies were cross-sectional studies or case-control studies and therefore they do not provide the incidence of malnutrition (94, 95, 96).

Risk Factors:

The present study showed that male children were having a higher risk of malnutrition as compared to female children. This finding was dissimilar to the finding from the NHFS 3 data which reported that there was a difference in severe malnutrition (underweight) across male and female children with female children having higher risk of malnutrition as compared to male children (90). There was another study reported from case control study from Bangladesh that did not show any significant difference in malnutrition (underweight) among male and female children. The study reported from Bangladesh involved children less than 2 years of age (4). Moreover, this study from Bangladesh involved severely malnourished children who were enrolled in a hospital and therefore there are chances of prevalence to be slightly overestimated as they are children suffering from diarrhea and thereby severely malnourished.

A cross sectional study reported in Uganda showed area of residence was an important factor on the nutritional health. This difference was present as the study reported that children from urban areas had better immunization rates than children from the rural areas (20). The present study also showed that children from urban area had better health status as compared to rural areas which implied that children from rural areas had higher risk of malnutrition as compared to children living in urban areas. There was a contradictory finding that reported that children from urban areas had high risk of malnutrition as compared to children from rural areas, however, it was also reported that these children from urban areas came from slum areas of Kampala (21).

There were two studies that reported malnutrition as a cause of poor socio economic status of a family (4, 21). This study also brought out similar finding which has been indirectly reflected by the “presence of a separate kitchen in the house” which implied that if there was no separate kitchen in the house where the child lived then he/she has a high chance of malnutrition (underweight and/or stunting) as compared to child living in a house with a separate kitchen. There was another study that reported hygiene practice was influencing the risk of malnutrition among children. The study conducted in Uganda reported that children from dirty and very dirty households are more underweight that children whose households were clean (20). This study also reflected this finding through defecation which implied that if children were from household where defecation was within the household were less likely than those children living in houses where defecation was outside the premises of the house. The present study also showed mother's education was associated with malnutrition of children. In other words, if the mother was illiterate then her child had high risk of being severely stunted as compared to a child whose mother had at least high school education. This type of risk factor was also reported from a study from Kampala and Bangladesh. The study reported that mother's education had a positive association to stunting of their children (94, 4). It appears that the impact of mother's education on nutrition status is through its ability to improve socio-economic status, health facility utilization and greater involvement in child care. However, there was a contradictory finding from study reported again from Kampala which reported that there was no correlation between the education level of mothers and nutrition status of children (21). The present study has also suggested that increase in the number of family size (>6 members) was negatively associated with malnutrition (underweight). There was a study which was similar in a way which reported that increase in the birth order had higher chance of underweight (4). The other risk factor from the present study was "type of fuel used for cooking". The present study showed that families that used firewood or coal or cow dung for cooking as fuel had higher chances of their children being malnourished (underweight and/or stunted) as compared to children in families that used gas or kerosene for cooking as fuel. This was also an indicator of socio-economic status of the family. A similar finding was reported from a study from Uganda which was that families that had alternative sources of fuel in addition to firewood or paraffin had fewer children underweight than those who had only firewood for fuel (20).

Generalized Estimating Equations and Random Effects Model:

The GEE analysis is done when there the outcome is repeatedly collected from the same subject and in the present study malnutrition (underweight or stunted) is repeatedly observed for each child for eight times. Hence there is a need to adjust for the correlations between the outcomes measured over time. The present study used autoregressive correlation structure for adjusting for correlations and these correlations were decided upon looking at the cross-tabulations over the eight time points from baseline. There has been an article that describes that though GEE is robust to wrongly specified correlation structure still the estimates are improper for a wrong choice of correlation structure (28). In the present study autoregressive 1 correlation structure was chosen.

There has been a paper that compared GEE to multilevel modeling in genetic association analysis. The study showed that compared to GEE, MLM had less underestimated odds ratio (50). Another study comparing the risk factors smoking intervention trial also compared results from GEE to random effects model which concluded that test statistic findings are similar which means that risk factors that turned out to be significant were similar in both the models (49). This finding was similar to the present study finding. The risk factors using GEE and Random effects model was similar as there was a high correlation of the responses at the “child” level. The ICC at the household level was very small. The random slope at the “child” level was assessed using difference in the likelihood of the random intercept alone and, random intercept and random slope model at the "child" level which was not significant for the a priori specified risk factors. Hence the final model was mainly the random intercept at the "child" level which means that adjustment is needed only due to the repeated follow-up of each child in the study which is the same adjustment done using GEE analysis. Therefore, results from GEE and random effects model were similar. This is shown in the appendix table 5. The R square using GEE model was 34% which is above our satisfaction but the main purpose was to compare GEE to Markov regression model.

Markov Chain and First Mean Passage Time:

The principle of Markov chain is that the current state is dependent on the previous state. Using the principle, the first mean transition time can be calculated. The first mean passage time provides the transition time taken to move from one state to another state of malnutrition. In other words, this provides us the time taken for a severely underweight or stunted child to get normal. This estimate is useful if we wanted to know on an average how long will severely or moderately malnourished children will get normal and if any factors affect this time so that appropriate intervention can be provided to that particular child. This type of time was not reported so far in any longitudinal studies. The present study reported the transition probability to move from one state to another. The present study also reported the transition probability from one state at the previous time to state of malnutrition at the current time. The transition probability from normal state of malnutrition to severe state of malnutrition was found to be 0.009 (0.9%). A study reported the transition probability in non-insulin diabetes study where the study had three progression states which were asymptomatic, symptomatic and death state. Once a person transited to "death" state he/she cannot transit to a symptomatic or an asymptomatic state. The progression rate from symptomatic state to death state was found to be 2.27% (52). However, the non-insulin diabetes study was done in absorbing state which means once transited to death state there is no back transition to previous states but the present study was for non-absorbing states which means transition probabilities can be obtained from severe state to normal state as well. Another study calculated the "mean survival time" from the transition probabilities from one state to another. There were four states of transition in the study of systemic lupus erythematosus (55). Most of the studies that applied Markov model were for absorbing states (52, 53, 54, 55, 60) and for continuous time Markov models (57). The present study has been applied to discrete time Markov model and also for non-absorbing states. The present study showed that children who were living in houses with a separate kitchen severely underweight took nearly two and a half years to return to normal where as children who were living in houses without separate kitchen took nearly four years to return to normal. Another finding from the present study was that children who were severely underweight in whose house had defecation within the premises of the house took 2.4 (2.1 – 2.8) years to return to normal than those children who were living in houses where defecation was in the open fields (3 (2.6 – 3.5) years). The present study also showed differences across risk factors in the duration of time for stunted children as well. If a child was severely stunted, then it took about 9 (7.8 – 10.6) years to return to normal state if that child was living in a house with a separate kitchen and it took nearly 12 (9.8 – 13.4) years to return to normal from severely stunted state for the child who lived in a house without a separate kitchen. Similarly, if a child lived in a house where type of fuel used for cooking was firewood or coal or cow dung then that child had taken nearly 10 (8.8 – 12.1) years to move from severe stunting to normal when compared to a child who lived in a house where fuel used for cooking was gas or kerosene who took about 7 (6.2 – 8.2) years to return to normal state. These kinds of findings were not available in any studies and also important to understand how certain risk factors were responsible for differences in the duration to return to normal from moderate or severe state of malnutrition.

Markov Regression:

The present study accounts for the risk factors of malnutrition at the current state irrespective of the previous state of malnutrition which is a Markovian property. The study showed that area of residence where the children lived was a significant factor that was associated with severe underweight irrespective of the previous state of malnutrition. Similarly, if there was a no separate kitchen then there was a significant association of severe underweight at the current time irrespective of the previous state of malnutrition at the previous time. A study with covariables explored the effects of factors that influenced the onset, progression and regression of diabetic retinopathy among subjects with insulin-dependent diabetes mellitus (60). The present study explored many variables and the interactions with previous state of malnutrition. The present study also specified the risk of transition from normal to severe malnutrition state at the current time. The present study showed that if there was no separate kitchen in the household where children live are at a high risk of transiting from normal to malnourished state rather than transiting from malnourished state to normal state of malnutrition. This finding indirectly reflected the socio economic status of the family (21). A similar finding was also reported in a study where one of the main risk factor of malnutrition was poor socio economic status of the family (4). There was another study that had risk factor associated with malnutrition to be the father's occupation as rickshaw driver which indicated the low socio economic status of the family. A study observed the risk factors for functionality disability changes with time and the data was a multi-state Markov model. After adjusting for covariates, it was found that ages, sex and size of infarct had no effect on transition rate. This analysis was done by calculating the hazard ratios from one state to another. The present study showed that absence of a separate kitchen within the house was one of the important risk factor that determined the risk of association with current time severe underweight despite any association of previous time's state of underweight with the current time's state of underweight. The number of family members especially when >6 also determines a high risk of severe underweight. This was reflected in a study that reported a high risk of underweight was observed when there was an increase in birth order (4). Children from urban areas had lower risk of severe stunting despite the previous state of stunting where as mother's education had an average effect on severe stunting promoting the fact that if mother's were educated, then there was sign of improving socio-economic status of the family and also better child care. The finding of the present study was similar to the finding from NFHS 3 which reported that there was a high risk factor of malnutrition when mothers' of the children were illiterate. (20, 21, 91, 95).

Comparison of GEE and Markov Regression models:

The Markov regression analysis was performed only using a first order Markov model. The Markov regression analysis was performed only considering the children who had at least one transition in the seven follow-up. Hence the Markov regression analysis considered slightly less number of observations than the GEE or Random effect model analyses. Robust standard errors were obtained for Markov regression analysis with transition probabilities which were compared to the asymptotic standard errors and they were similar, suggesting that first order Markov model was valid. The Markov regression using intensity rate matrix could not converge when three states of malnutrition was considered. The convergence was not achieved as the transition probability was very small especially when the transition was from normal to severe states. Hence the comparison of Markov regression using intensities, probabilities and GEE was performed using “normal state” and “malnutrition state” which is “moderate state or severe state”. The comparison of the 95% confidence intervals from Markov regression and GEE was only done for risk factors that were significant from the adjusted model (61). Another study reported the risk factors for progression of liver fibrosis (62). A study compared Markov model regression, Markov regression with random effects to find the risk factors for transition in Bacterial Vaginosis among women (64) and it was found that Markov regression with random effects had better fit to the data. The present study also had comparison of models however; the models were GEE and Markov regression models. The observed and expected plots of malnutrition were not very different suggesting that Markov model has fitted well to the data. There was no study that compared results of GEE and Markov regression model especially in terms of coverage probabilities and standard errors using simulations. However, our findings support that coverage probability was higher in Markov regression than GEE.

Limitations:

The main limitation of the present study was that data being a secondary data, that is, the period when the data was collected was during 1982 - 1986 and it has been 25 years. Markov regression modeling outcome is more evident if there are time varying covariates where as the present study has most of the covariates not changing over time. In order to prove further Markov regression is consistently better, this concept has to be experimented with various datasets with varying incidence densities and, time intervals for follow-up.

SUMMARY AND CONCLUSION

8. SUMMARY AND CONCLUSION

8.1 Summary:

Background: Malnutrition refers to many diseases each with a specific deficiency in one or more nutrients and each characterized by cellular imbalance between the supply of nutrients and energy on the one hand, and the body's demand for them to ensure growth maintenance. Malnutrition is an important indicator of child health. It is now recognized that 6.6 million out of 12.2 million deaths among children under-five – or 54% of young child mortality in developing countries – is associated with malnutrition. India has the highest percentages of undernourished children in the world. During 1982, seven localities and 22 villages were selected for this study. These localities and villages were selected from Vellore town and KV Kuppam development block sampling frames respectively. All children aged 5-7 years were screened for signs of malnutrition by consultant pediatricians. The children from rural and urban areas of Vellore town were screened at baseline and followed up for every six months for 7 times. Malnutrition was assessed based on these indicators which are BMI Z scores, Height-for-age. The BMI Z scores were classified as "normal" if the BMI Z scores were >-2 standard deviations, "moderate" when Z scores were between -2 and -3 standard deviations and, "severe" if the Z scores were <-3 standard deviations (67). The main hypothesized risk factors for the study were 'defecation practices at household level' (within the household; in the open fields), 'type of fuel used for cooking in the house' (firewood or cow dung or coal; gas or kerosene) and 'presence of a separate kitchen within the household premises' (yes; no). The other confounders that were seen important that have to be adjusted were sex of the child (male; female) and area of residence (rural; urban). Some other covariates that were also included for Generalized Estimating

Equations and Markov Regression using transition probabilities are education of mother and father (illiterate or literate; primary or middle school; high school or above), consanguineous marriage of the parents whose children were included in the study (yes; no), type of roof (thatched; tiled; RCC or pukka), type of house (brick and cement; brick and mud; others) and birth order (1; 2; >=3), number of members in a family (<=4; 5-6; >6), type of floor (kucha; pukka).

Aims and Objectives: The main aim was to find the risk factors for malnutrition.

The objectives of the study are: (i) To estimate the first mean passage time which indicates the average time spent by a child to move from one state to another, to find risk factors of using GEE and Random Effects model, to find risk factors of protein energy malnutrition using transition probabilities, to find the risk factors by calculating the transition intensity matrices and to compare the results obtained from GEE and Markov regression models using transition probabilities and transition intensities.

Prevalence and Incidence: The overall prevalence of severe underweight was 22.5%. The prevalence of severe underweight was higher among children (25%) than female children (19.9%). The prevalence of severe underweight was lower among children living in the rural areas as compared to children living in the urban areas (16.5% vs 28% respectively). The overall incidence for severe underweight was 11.6% and higher incidence rate was observed among male children than female children. The incidence density of severe underweight was around 5% per year. The prevalence of severe stunting was 25.8%. Higher prevalence of severe stunting was found among male children than female children (27.9% vs 23.5% respectively). The incidence was also higher among children in the rural areas (33.2%) as compare to children living in the urban areas (18.9%). The cumulative incidence of stunting was 20.6% and the incidence density for stunting was about 2% per year.

Mean Passage Time and Risk Factors: The overall transition probability from normal state of underweight to moderate was 0.12 and severe state was 0.009. The transition probability of moving from severe underweight to moderate underweight and normal weight was found to be 0.28 and 0.10 respectively. The average number of years taken to transit from severe state of underweight to normal state was about 2.7 (2.3 - 3.1) years. The mean number of years taken to transit from severe underweight to normal across male and female children was almost similar. The MPT from severe underweight to normal in the urban areas was less as compared to MPT in the rural areas. It was also found that children who lived in houses with no separate kitchen and children living in houses that used firewood or cow dung for cooking had had lower transition time from severe underweight to normal as compared to children living in houses that had separate kitchen or used gas or kerosene for cooking. The probability of transition from severe stunting to normal was 0.001. The overall MPT from severe stunting to normal was around nine and a half years (8.4 years – 11.3 years). There was no difference across male children and female children in MPT for stunting. The average number of years taken to move from severe stunting to normal was higher among children in the rural areas as compared to children living in urban areas (11.9 years vs 8.1 years). The mean first passage time in the present study clearly indicates how late or early a person transits from one state of an outcome to another state of that outcome when the child experiences a “risky” factor of the exposure in non-absorbing state models. This is useful in chronic disease epidemiology where a motive is to find out how long would the transition time be, on an average, for a person to transit from one state to another. So that appropriate treatment procedures can be provided. Similar findings of “longer time” were observed when the children lived in houses without separate kitchen, defecated in the open fields and used firewood or cow dung for cooking. The risk factors for severe underweight obtained using GEE were defecation and sex of the child. The risk factors for severe underweight using Markov regression other than the two risk factors mentioned using GEE were family members, presence of a separate kitchen and the state of underweight at the previous time. The risk factors that turned out important for severe stunting using GEE were area of residence, mothers’ education, fathers’ education. The factors important for severe stunting using Markov regression were defecation, area of residence, father’s education, and the state of stunting at the previous time. The transition times cannot be estimated using GEE or Random Effect Models as these models do not account for the fact that the current state of malnutrition is mainly due to the state of malnutrition at the previous time (a Markov Chain principle). Markov regression using transition probabilities involves modeling the outcome state at the current state conditioning on the state of the outcome at the previous time and other covariates. Hence if the previous state of outcome is highly correlated, then it is important to perform Markov regression. Markov regression using intensity rates involves modeling the outcome for a specific transition. This model is very specific to what had been the state of malnutrition in the previous time. If there was a specific hypothesis relating to the specific transition, then the model using the ‘transition rates’ would be better. GEE analysis considers the correlations of the different states of malnutrition overtime and adjusts for that correlation. In most longitudinal data analysis, it is worth considering risk factors that are associated with the movement to current state from previous state. It is essential to test whether current state depends on the state at previous time i.e., if the state of malnutrition at the previous time was significantly associated with current state of malnutrition, then the Markov regression using transition probability is appropriate. In this study, state of malnutrition at the previous time was significantly associated with current state of malnutrition. The standard errors obtained from the Markov regression using transition probabilities were smaller than those compared to GEE analysis and had better coverage probability with shorter length. The risk factor profile for GEE and Markov regression were different with relatively more risk factors using Markov regression as this may be due to higher SEs obtained from GEE analysis when adjusted for the correlation structure. The simulations performed for underweight and stunting showed better coverage probabilities and shorter length of confidence intervals when Markov regression analysis was performed than GEE.

8.2 Conclusion:

In any longitudinal study with discrete non-absorbing outcome, it is essential to estimate the duration of time spent in each state of the outcome. This will help us to study the impact of duration of stay with other risk factors. In longitudinal data if the current state of the outcome depends on the state of the outcome at the previous time, then Markov regression is the best approach to find the risk factors. GEE approach evaluates the overall correlation structure and therefore more likely to have larger standard errors and thereby likely to deal with false positive findings.

IMPACT OF THE STUDY

9. IMPACT OF THE STUDY

1. In longitudinal studies with discrete outcome, it is recommended that First Mean Passage Time need to be calculated to study the impact of this duration on the subsequent outcome. The traditional survival analyses that will provide median time, if calculated will under estimate this duration.
2. In longitudinal studies with discrete outcome, if the current state of outcome depends on the immediate previous state of the outcome, then the Generalized Estimating Equations or Multi Level Modeling method to study the risk factors need not be used. If used this procedure will under estimate the risk factors. In such situations, Markov Regression will provide better estimate of Standard Errors and therefore, better coverage probability for the 95% Confidence Interval (CI) and narrow 95% CI.
3. This study found that ‘presence of a separate kitchen’ was one of the important factors that had an impact on the mean transition time from one state to another state of malnutrition. If the child lived in a house that had no separate kitchen, then there was a long transition time to move from severe state of malnutrition to normal state as compared to a child living in a house that had a separate kitchen.
4. This study also found ‘defecation’ as an important factor that had on the mean transition time. The children who lived in houses where defecation was in the open fields had higher transition time as compared to those children who lived in houses where defecation was within the premises of the house.
5. The children whose mothers’ had low education turned out to be an important factor. The analysis showed that children whose mothers’ were illiterate had a high risk of being

severely malnourished as compared to children whose mothers' had high school or higher level of education.

6. There was also an impact on malnutrition due to the type of fuel used for cooking. If the fuel used for cooking was firewood or coal or cow dung then there was a high risk of children living in such houses to be malnourished as compared to children living in houses where gas or kerosene was used as fuel for cooking.
7. The overall impact of the study findings were that social factors like "mothers' education", and economic factors which are reflected by the "type of fuel used for cooking", "presence of a separate kitchen" were important and hence the recommendation would be to plan interventions in children living in such environments to reduce protein energy malnutrition among children.

REFERENCES

10. REFERENCES

1. Bellamy C. The State of the World's Children 2003. United Nations Children Education Fund (UNOCEF); 2003.

2. Ramachandran Prema. Maternal and Child nutrition: new dimensions of the dual nutrition burden. Indian Journal of Medical Research. 2009;130:575 - 578.

3. Mamoun Nuha, Homedia Susan, Mabyou Mustafa. Prevalence, Types and Risk Factors for Malnutrition in Displaced Sudanese Children. American Journal of Infectious Disease. 2005;1(2):84–6.

4. Nahar Baitun, Ahmed Tahmeed, Brown Kenneth H, Hossain Md. Iqbal. Risk Factos Associated with Severe Underweight among Young Children Reporting to a Diarrhoea Treatment Facility in Bangladesh. Journal of Heath Population Nutrition. 2010;28(5):476 - 483.

5. The South-East Asia Nutrition Research-cum-Action Network. World Health Organization; 2001 p. 1–30.

6. King H. Maurice. Nutrition for Developing Countries. Nairobi: Oxford University Press; 1976.

7. Ghosh Shanti, Bhargava S.K, Madhavan Dhantha, Tasker A.D. Bhargava Vijaya, Nigam S.K. Intra-Uterine Growth of North Indian Babies. Pediatrics. 1971;47(5):826–30.

8. Victoria G Cesar, Adair Linda, Fall Caroline, Hallal C Pedro, Martorel Reynaldo, Richter Linda, Sachdev Harshpal Singh. Maternal and Child undernutrition: consequences for adult health and human capital. The Lancet. 2008;371(9609):340–57.

9. Jeyaseelan L, Lakshman M. Risk Factors for Malnutrition in south Indian children. Journal of Biosocial Science. 1997;29(1):93–100.

10. Beck J. Robert, Pauker G. Stephen. The Markov Process in Medical Prognosis. Medical Desision Making. 1983;3(4):419 - 458.

11. Beck J. Robert. Markov Models of Natural History. Journal of Clinical Epidemiology. 1988;41(7):619–21.

12. Hubbard RA, Zhou XH. A comparison of non-homogeneous Markov regression models with application to Alzheimer's disease progression. Journal of Applied Statistics. 2011;38(10):2313–26.

13. Andersen Per Kragh, Borgan Ornulf, Gill D. Richard, Keiding Liels. Statistical Models Based on Counting Process. II. New York;

14. Onis de M, Monteiro C, Akre J, Clugston G. The worldwide magnitude of protein-energy malnutrition: an overview from the WHO Global Database on CHild Growth. Bulletin of the World Health Organization. 1993;71(6):703–12.

15. Kain J, Uauy R, Lera L, Taibo M, Albala C. Trends in Height and BMI of 6-Year-Old Children during the Nutrition Transition in Chile. Obesity Research. 2005 Dec;13(2):2178–86.

16. Warraich HJ, Javed F, Faraz-ul-Haq M, Khawaja FB, Saleem S. Prevalence of Obesity in School-Going Children of Karachi. PLoS ONE. 2009 Mar 24;4(3):e4816.

17. Seetharaman N, Chacko T, Shankar S, Mathew A. Measuring malnutrition -The role of Z scores and the composite index of anthropometric failure (CIAF). Indian Journal of Community Medicine. 2007;32(1):35.

18. Mian RMA, Ali M, Ferroni PA, Underwood P. The Nutritional Status of School-Age Children in an Urban Squatter Settlement in Pakistan. Pakistan Journal of Nutrition. 2002;1(3):121–3.

19. Nandy Shailen, Irving Michelle, Gordon David, Subramanian S.V, Smith Davey George. Poverty, child undernutrition and morbidity: new evidence from India. Bulletin of the World Health Organization. 2004;83(3):210–5.

20. Kikafunda JK, Walker AF, Collett D, Tumwine JK. Risk Factors for Early Childhood Manutrition in Uganda. Pediatrics. 1998;102(e45):1–8.

21. Owor M, Tumwine J.K, Kikafunda J.K. Socio-economic Risk Factors For Severe Protein Energy Malnutrition Among Children in Mulago Hospital, Kampala. East African Medical Journal. 2000 Sep;77(9):471–5.

22. Ferrari Alexandre Archanjo, Solymos Bernardes, Castillo Rosa Mercedez, Sigulem Dirce Maria. Risk factors for protein-energy malnutrition in pre-school shantytown children in São Paulo, Brazil. Sao Paulo Medical Journal. 1998 Mar;116(2):1654–60.

23. Hanley JA, Negassa A, Edwards MD deB, Forrester JE. Statistical Analysis of Correlated Data using Generalized Estimating Equations: An Orientation. American Journal of Epidemiology. 2003 Feb;157(4):364 - 375.

24. Cannon MJ, Warner L, Taddei JA, Kleinbaum DG. What can go wrong when you assume that correlated data are independent: an illustration from the evaluation of a childhood health intervention in Brazil. Stat Med. 2001 May 15;20(9-10):1461–7.

25. Ballinger GA. Using Generalized Estimating Equations for Longitudinal Data Analysis. Organizational Research Methods. 2004 Apr;7(2):127–50.

26. Carroll RJ. Variances Are Not Always Nuisance Parameters. Biometrics. 2003 Jun 1;59(2):211–20.

27. Sabo RT, Chaganty NR. What can go wrong when ignoring correlation bounds in the use of generalized estimating equations. Stat Med. 2010 Oct 30;29(24):2501–7.

28. Overall JE, Tonidandel S. Robustness of Generalized Estimating Equation (GEE) Tests of Significance against Misspecification of the Error Structure Model. Biometrical Journal. 2004;46(2):203–13.

29. Rao Chaganty N, Joe H. Efficiency of generalized estimating equations for binary responses. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2004 Nov 1;66(4):851–60.

30. Chaganty NR, Joe H. Range of correlation matrices for dependent Bernoulli random variables. Biometrika. 2006 Mar 1;93(1):197–206.

31. Twisk JWR. Applied multilevel analysis: a practical guide. Cambridge University Press; 2006.

32. Pan Wei and Connett E. John. Selecting the Working Correlation Sturcture in Generalized Estimating Equations with Application to the Lung Health Study. Statistica Sinica. 2002;12:475–90.

33. Shults Justine, Sun Wenguang, Tu Xin et al. A comparison of several approaches for choosing between working correlation structures in generalized estimating equations analysis of longitudinal binary data. Statistics in Medicine. 2009 May;28:2338–55.

34. Pan Wei. Akaike's Information Criterion in Generalized Estimating Equations. Biometrics. 2001 Mar;57(1):120–5.

35. Hin L, Wang Y. Working-correlation-structure identification in generalized estimating equations. Statistics in Medicine. 2009 Feb 20;28(4):642–58.

36. Gosho M, Hamada C, Yoshimura I. Criterion for the Selection of a Working Correlation Structure in the Generalized Estimating Equation Approach for Longitudinal Balanced Data. Communications in Statistics - Theory and Methods. 2011;40(21):3839–56.

37. Wang Y, Carey V. Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. Biometrika. 2003;90(1):29–41.

38. Odueyungbo A, Browne D, Akhtar-Danesh N, Thabane L. Comparison of generalized estimating equations and quadratic inference functions using data from the National Longitudinal Survey of Children and Youth (NLSCY) database. BMC Medical Research Methodology. 2008 May 9;8(1):28.

39. Chaganty N.R, Mav D. Estimation methods for analyzing longitudinal data occurring in biomedical research. Chapman and Hall/CRC Press; 2008.

40. Carlin JB, Wolfe R, Brown CH, Gelman A. A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. Biostat. 2001 Dec 1;2(4):397–416.

41. Greenland Sander. When should Epidemiologic Regressions Use Random Coefficients? Biometrics. 2000;56:915–21.

42. Dickinson LM. Multilevel Modeling and Practice-Based Research. The Annals of Family Medicine. 2005 May;3:52–60.

43. Hedeker Donald, Berbaum Michael, Mermelstein Robin. Location-Scale Models for Multilevel Ordinal Data: Between-and Within-Subjects Variance Modeling. Journal of Probability and Statistical Science. 2006 Feb;4(1):1–20.

44. McMahon JM, Pouget ER, Tortu S. A guide for multilevel modeling of dyadic data with binary outcomes using SAS PROC NLMIXED. Comput Stat Data Anal. 2006 Aug;50(12):3663–80.

45. Musca SC, Kamiejski R, Nugier A, Méot A, Er-Rafiy A, Brauer M. Data with Hierarchical Structure: Impact of Intraclass Correlation and Sample Size on Type-I Error. Frontiers in psychology. 2011;2(April):6.

46. Engel B. A simple illustration of the failure of PQL, IRREML and AHPL as approximate ML methods for mixed models for binary data. Biometrical Journal. 1998;40:141–54.

47. McCoach D. Betsy, Adelson L. Jill. Dealing with Dependence (Part 1): Understanding the Effects of CLustered Data. Sage. 2010;54(2):152–5.

48. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: what are the differences? Stat Med. 2009 Jan 30;28(2):221–39.

49. Hu FB, Goldberg J, Hedeker D, Flay BR, Pentz MA. Comparison of Population-Averaged and Subject-Specific Approaches for Analyzing Repeated Binary Outcomes. Am. J. Epidemiol. 1998 Apr 1;147(7):694–703.

50. Zhao Y, Yu H, Zhu Y, Ter-Minassian M, Peng Z, Shen H, et al. Genetic Association Analysis Using Sibship Data: A Multilevel Model Approach. PLoS ONE. 2012 Feb 1;7(2):e31134.

51. Longford NT. Simulation-based diagnostics in random-coefficient models. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2001 Jan 1;164(2):259–73.

52. Kuo HS, Chang HJ, Chou P, Teng L, Chen TH. A Markov chain model to assess the efficacy of screening for non-insulin dependent diabetes mellitus (NIDDM). Internation Journal of Epidemiology. 1999 Apr;28(2):233–40.

53. Leviton A, Schulman J, Kammerman L, Porter D, Slack W, Graham JR. A probability model for headache recurrence. Journal of Chronic Diseases. 1980;33(7):407–12.

54. Silverstein MD, Loftus EV, Sandborn WJ, Tremaine WJ, Feagan BG, Nietert PJ. Clinical course and costs of care for Crohn’s disease: Markov model analysis of a population-based cohort. Gastroenterology. 1999 Jul;117(1):49–57.

55. Marc D. Silverstein, Daniel A.Albert, Noritn M. Hadler. Prognosis in SLE: Comparison of Markov Model to Life Table Analysis. Journal of Clinical Epidemiology. 1988;41(7):623 - 633.

56. Duffy SW, Chen H, Tabar L, Day NE. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. Statistics in Medicine. 1995 Jul 30;14(14):1531–43.

57. Murali R, Jeyaseelan L, Rajaratnam S, John L, Ganesh A. Systemic lupus erythematosus in Indian patients: prognosis, survival and life expectancy. Natl Med J India. 1997 Aug;10(4):159–64.

58. Joly P, Commenges D. A Penalized Likelihood Approach for a Progressive Three-State Model with Censored and Truncated Data: Application to AIDS. Biometrics. 1999;55(3):887–90.

59. McCormick TH, Raftery AE, Madigan D, Burd RS. Dynamic Logistic Regression and Dynamic Model Averaging for Binary Classification. Biometrics [Internet]. 2011 Aug 12 [cited 2012 Mar 9]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/21838812

60. Marshall G, Jones RH. Multi-state models and diabetic retinopathy. Statistics in Medicine. 1995 Sep 30;14(18):1975–83.

61. Pan S-L, Wu H-M, Yen AM-F, Chen TH-H. A Markov regression random-effects model for remission of functional disability in patients following a first stroke: a Bayesian approach. Stat Med. 2007 Dec 20;26(29):5335–53.

62. Boyd A, Lasnier E, Molina JM, Lascoux-Combe C, Bonnard P, Miailhes P, et al. Liver fibrosis changes in HIV-HBV-coinfected patients: clinical, biochemical and histological effect of long-term tenofovir disoproxil fumarate use. Antivir. Ther. (Lond.). 2010;15(7):963–74.

63. Sutradhar R, Barbera L, Seow H, Howell D, Husain A, Dudgeon D. Multistate Analysis of Interval-Censored Longitudinal Data: Application to a Cohort Study on Performance Status Among Patients Diagnosed With Cancer. Am. J. Epidemiol. 2011 Feb 15;173(4):468–75.

64. Sanders KL, Thoma ME, Yu K, Albert PS. An evaluation of the natural history of bacterial vaginosis using transition models. Sex Transm Dis. 2011 Dec;38(12):1131–6.

65. Atashili J, Smith JS, Adimora AA, Eron J, Miller WC, Myers E. Potential impact of antiretroviral therapy and screening on cervical cancer mortality in HIV-positive women in sub-Saharan Africa: a simulation. PLoS ONE. 2011;6(4):e18527.

66. Hardt J, Herke M, Schier K. Suicidal ideation, parent-child relationships, and adverse childhood experiences: a cross-validation study using a Graphical Markov Model. Child Psychiatry Hum Dev. 2011 Apr;42(2):119–33.

67. Mei Zuguo & Grummer-Strawn M Laurence. Standard deviation of anthropometric Z-scores as a data quality assessment tool using the 2006 WHO growth standards: a cross country analysis. Bulletin of the World Health Organization. 2007 Jun;85(6):441–8.

68. Crowder M. On Repeated Measures Analysis with Misspecified Covariance Structure. Journal of the Royal Statistical Society. Series B (Statistical Methodology). 2001 Jan 1;63(1):55–62.

69. Lee J-H, Herzog TA, Meade CD, Webb MS, Brandon TH. The use of GEE for analyzing longitudinal binomial data: A primer using data from a tobacco intervention. Addictive Behaviors. 2007 Jan;32(1):187–93.

70. Cui J. QIC program and model selection in GEE analyses. The stata journal. 2007;7(2):209–220 ST - QIC program and model selection in G.

71. Cui J, Qian G. Selection of Working Correlation Structure and Best Model in GEE Analyses of Longitudinal Data. Communications in Statistics Simulation and Computation. 2007;36(5):987–96.

72. Crowder M. On the use of a working correlation matrix in using generalised linear models for repeated measures. Biometrika. 1995 Jun 1;82(2):407–10.

73. Twisk JWR. Applied longitudinal data analysis for epidemiology: a practical guide. Cambridge University Press; 2003.

74. Carey V, Zeger SL, Diggle P. Modelling multivariate binary data with alternating logistic regressions. Biometrika. 1993 Sep 1;80(3):517–26.

75. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986 Mar 1;73(1):13–22.

76. Diggle P. Analysis of longitudinal data. Oxford University Press; 2002.

77. Preisser JS, Qaqish BF. Robust Regression for Clustered Data with Application to Binary Responses. Biometrics. 1999 Jun 1;55(2):574–9.

78. Greenland S. Principles of multilevel modelling. Int. J. Epidemiol. 2000 Feb 1;29(1):158–67.

79. Catalano P.J. Bivariate modeling of cluster continuous and ordered categorical outcomes. Statistics in Medicine. 1997;16:883–900.

80. Fielding A, Yang M, Goldstein H. Multilevel ordinal models for examination grades. Statistical Modelling. 2003 Jul 1;3(2):127–53.

81. Lesaffre E, Spiessens B. On the Effect of the Number of Quadrature Points in a Logistic Random-Effects Model: An Example. Journal of the Royal Statistical Society. Series C (Applied Statistics). 2001 Jan 1;50(3):325–35.

82. Corporation S. Stata Base Reference Manual: A-J. Stata Corp.; 2005.

83. Kemeny JG, Snell JL. Finite Markov Chains. Springer-Verlag;

84. Hunter JJ. Variances of First Passage Times in a Markov Chain with applications to mixing times. Res. Lett Inf Math. 2006;10:17–48.

85. Kalbfleisch JD, Lawless F Jerald. The Analysis of Panel data under a Markov assumption. Journal of American Statistical Association. 1985 Dec;80(392):863–71.

86. Grantham-McGregor Sally. A Review of Studies of the Effect of Severe Malnutrition on Mental Development. Journal of Nutrition. 1995;125.

87. Berkman DS, Lescano AG, Gilman RH, Lopez SL, Black MM. Effects of stunting, diarrhoeal disease, and parasitic infection during infancy on cognition in late childhood: a follow-up study. Lancet. 2002 Feb 16;359(9306):564–71.

88. Black MM. Micronutrient Deficiencies and Cognitive Functioning. J. Nutr. 2003 Nov 1;133(11):3927S–3931S.

89. Behrman JR. The Impact of Health and Nutrition on Education. World Bank Res Obs. 1996 Feb 1;11(1):23–37.

90. International Institute for Population Sciences. National Family Health Survey - 3. Mumbai; 2000 Oct.

91. International Institute for Population Sciences. National Family Health Survey - 2. Mumbai; 1998 99.

92. Kumar Shiva A.K. Why are Levels of Child Malnutrition Not Improving? Economic and Political Weekly. 2007 Apr 14;

93. Kumar AKS. Why Are Child Malnutrition Levels Not Improving? The Hindu. MADS; 10.

94. Jitta J, Migadde M, Mudusu J. Determinants of malnutrition in underfives in Uganda: An indepth secondary analysis of Uganda demographic and health Survey data. Kampala Uganda: Ministry of Health and Child Health Development Centre Makerer University. Ministry of Health and Child Health Development Centre Makerer University. 1992;24–107.

95. Anderson Kojo Alex, Bignell Whitney, Winful Sophie, Soy iri Ireneous. Risk Factors for Malnutrition among Children 5-years and Younger in the Akuapim-North District in the Eastern Region of Ghana. Current Research Journal of Biological Sciences. 2010;2(3):183–8.

96. Rahyan Md. Israt, Khan M. Sekander Hayat. Factors causing Malnutrition among Five Children in Bangladesh. Pakistan Journal of Nutrition. 2006;5(6):558–62.

APPENDIX

11. APPENDIX

Table 1a: Incidence of Severe Malnutrition using BMI Categories for all children by Age of the Child at recruitment

Age of the child at Baseline = 5 years

From (Time t)	To (t+1)								
Baseline:	**Normal** n	**%**	**Moderate** n	**%**	**Severe** n	**%**	**Total**	**ISM**	**95% CI**
Normal	692	89.1	67	8.6	18	2.3	777	4.15	3.08 – 5.55
Moderate	137	52.7	98	37.7	25	9.6	260		
Severe	65	27.5	80	33.9	91	38.6	236		
1st Year:									
Normal	738	86.7	96	11.3	17	2.0	851	3.79	2.79 – 5.11
Moderate	76	32.8	132	56.9	24	10.3	232		
Severe	10	7.6	41	31.1	81	61.4	132		
2nd Year:									
Normal	682	81.7	141	16.9	12	1.4	835	6.21	4.92 – 7.79
Moderate	49	17.7	171	61.7	57	20.7	277		
Severe	5	4.3	32	27.8	78	67.8	115		

***ISM – Incidence of severe malnutrition**

Tables 6a, 6b and 6c present the incidence of severe malnutrition in each year from baseline by age of the children at inception. The incidence of severe malnutrition was nearly 4% in the first two years and increased to 6% from second to third year when the age of child at inception was 5 years. The incidence of malnutrition was 3%, 4% and 5% respectively in the three years of follo-up from baseline when the age of the child at inception was 6 years (table 6b).

There were very few cases in the age group 7 years at inception. The severe malnutrition incidence was 4% and 6% in the three years of follow-up from baseline.

Tables 7a and 7b represent the severe malnutrition incidence for 3 years from baseline. The incidence of severe malnutrition was lower for girls over the three years (3%, 3% and 4%) while the incidence of severe malnutrition was higher for boys as compared to girls (4%, 5% and 8%) respectively.

Table 1b: Incidence of Severe Malnutrition using BMI Categories for all children by Age of the Child at recruitment

Age of the Child at Baseline = 6 years

From (Time t)	To (t+1)								
Baseline:	**Normal** n	%	**Moderate** n	%	**Severe** n	%	**Total**	**ISM**	**95% CI**
Normal	380	82.2	43	10.0	8	1.9	431	3.15	2.02 – 4.84
Moderate	131	64.2	61	29.9	12	5.9	204		
Severe	72	30.4	87	36.7	78	32.9	237		
1st Year:									
Normal	487	86.5	65	11.5	11	2.0	563	4.81	3.47 – 6.60
Moderate	53	28.5	108	58.1	25	13.4	186		
Severe	7	7.3	36	37.5	53	55.2	96		
2nd Year:									
Normal	473	84.2	87	15.5	2	0.4	562	5.34	3.94 – 7.18
Moderate	45	21.8	122	59.2	39	18.9	206		
Severe	2	2.2	25	27.2	65	70.7	92		

***SI – Severely malnourished**

Table 1c: Incidence of Severe Malnutrition using BMI Categories for all children by Age of the Child at recruitment

Age of the Child at Baseline = 7 years

From (Time t)	To (t+1)								
Baseline:	**Normal** n	%	**Moderate** n	%	**Severe** n	%	**Total**	**ISM**	**95% CI**
Normal	38	92.7	3	7.3	0	0.0	41	4.00	0.9 – 11.58
Moderate	21	61.8	10	29.4	3	8.8	34		
Severe	7	20.0	9	25.7	19	54.3	35		
1st Year:									
Normal	54	84.4	10	15.6	0	0.0	64	5.88	2.22- 13.36
Moderate	5	23.8	11	52.4	5	23.8	21		
Severe	0	0.0	5	22.7	17	77.3	22		
2nd Year:									
Normal	47	79.7	11	18.6	1	1.7	59	5.92	4.67-17.91
Moderate	4	16.0	14	56.0	7	28.0	25		
Severe	1	4.5	3	13.6	18	81.8	22		

***SI – Severely Prevalence**

Table 2a: Incidence of Severe Malnutrition using BMI Categories for all children by Sex of the child

Male:

From (Time t)	To (t+1)								
Baseline:	**Normal** n	%	**Moderate** n	%	**Severe** n	%	**Total**	**ISM**	**95% CI**
Normal	525	87.1	58	9.6	20	3.3	603	4.40	3.21 – 5.99
Moderate	155	59.4	88	33.7	18	6.9	261		
Severe	73	24.7	109	36.9	113	38.3	295		
1st Year:									
Normal	607	84.2	95	13.2	19	2.6	721	5.50	4.22 – 7.13
Moderate	76	31.3	133	54.7	34	14.0	243		
Severe	12	8.1	42	28.2	95	63.8	149		
2nd Year:									
Normal	569	80.6	127	18.0	10	1.4	706	7.87	6.33 – 9.73
Moderate	45	16.5	161	59.0	67	24.5	273		
Severe	6	4.0	32	21.5	111	74.5	149		

Table 2b. Incidence of Severe Malnutrition using BMI Categories for all children by Sex of the child

Female:

From (Time t)	To (t+1)								
Baseline:	**Normal** n	%	**Moderate** n	%	**Severe** n	%	**Total**	**ISM**	**95% CI**
Normal	585	90.6	55	8.5	6	0.9	646	3.17	2.19 – 4.56
Moderate	134	56.5	81	34.2	22	9.3	237		
Severe	71	33.3	67	31.5	75	32.2	213		
1st Year:									
Normal	672	88.8	76	10.0	9	1.2	757	3.04	2.11 – 4.35
Moderate	58	29.6	118	60.2	20	10.2	196		
Severe	5	5.0	40	39.6	56	55.5	101		
2nd Year:									
Normal	633	84.4	112	14.9	5	0.7	750	4.16	3.07 – 5.61
Moderate	53	22.6	146	62.1	36	15.3	235		
Severe	2	2.5	28	35.0	50	62.5	80		

***SI – Severely Prevalence**

Table 3a: Incidence of Severe Malnutrition using BMI Categories for all children Area of Residence

Rural:

From (Time t)	To (t+1)								
Baseline:	**Normal** n	%	**Moderate** n	%	**Severe** n	%	**Total**	**ISM**	**95% CI**
Normal	582	85.7	76	11.2	21	3.1	679	5.56	4.23 – 7.26
Moderate	92	41.6	100	45.2	29	13.1	221		
Severe	21	11.7	65	36.3	93	52.0	179		
1st Year:									
Normal	604	89.2	60	8.9	13	1.9	677	3.38	2.38 – 4.78
Moderate	73	30.5	148	61.9	18	7.5	239		
Severe	6	4.2	50	35.2	86	60.6	142		
2nd Year:									
Normal	606	85.8	94	13.3	6	0.8	706	5.25	4.00 – 6.84
Moderate	45	16.9	176	66.2	45	16.9	266		
Severe	3	2.6	39	33.9	73	63.5	115		

Table 3b: Incidence of Severe Malnutrition using BMI Categories for all children by Area of Residence

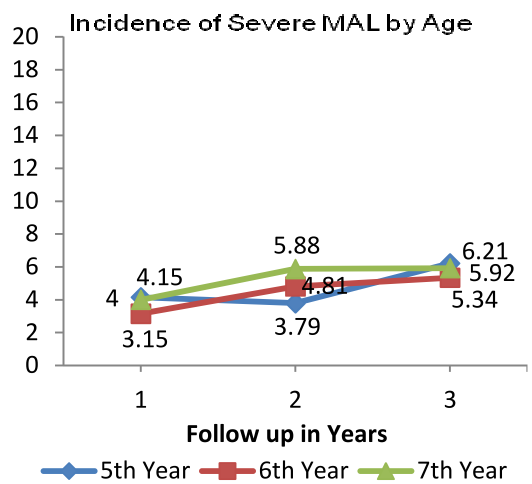
Urban:

From (Time t)	To (t+1)								
Baseline:	**Normal** n	%	**Moderate** n	%	**Severe** n	%	**Total**	**ISM**	**95% CI**
Normal	528	92.6	37	6.5	5	0.9	570	1.89	1.14 – 3.07
Moderate	197	71.1	69	24.9	11	4.0	277		
Severe	123	34.4	111	33.7	95	28.9	329		
1st Year:									
Normal	675	84.3	111	13.9	15	1.9	801	5.09	3.89 – 6.65
Moderate	61	30.5	103	51.5	36	18.0	200		
Severe	11	10.2	32	29.6	65	60.2	108		
2nd Year:									
Normal	596	79.5	145	19.3	9	1.2	750	6.75	5.34 – 8.50
Moderate	53	21.9	131	54.1	58	24.0	242		
Severe	5	4.4	21	18.4	88	77.2	114		

***SI – Severely Prevalence**

The tables 3a and 3b are the incidence of severe malnutrition among children in rural and urban areas. The incidence of severe malnutrition in urban was low in the first year from baseline (2%) but increased in second year (5%) and third year (7%). The incidence of severe malnutrition was similar over the three years (5% in the first year from baseline, 4% in the second year and 5% in the third year).

Figure 1: Incidence of Severe Malnutrition by age of the child at inception using BMI classification:



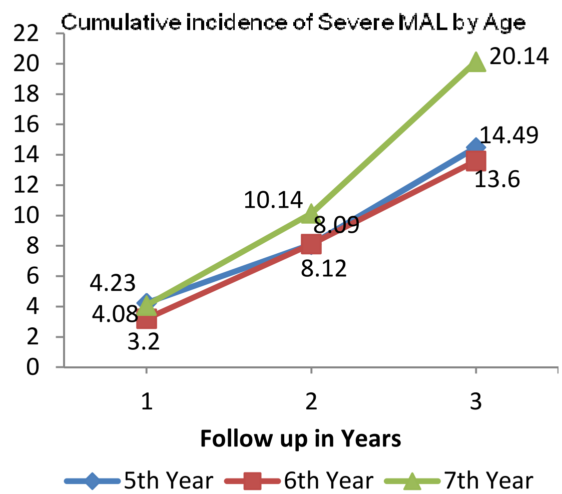
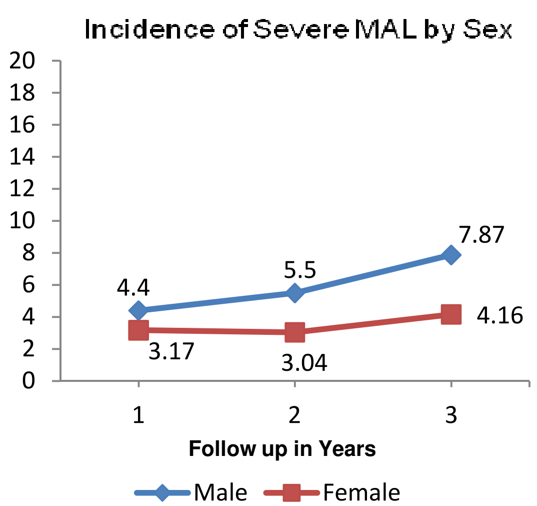
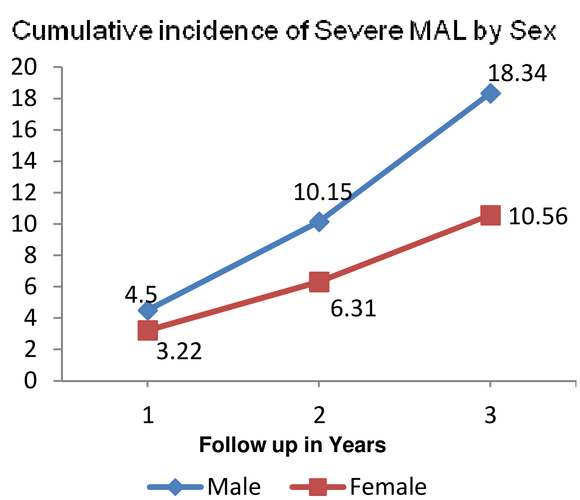


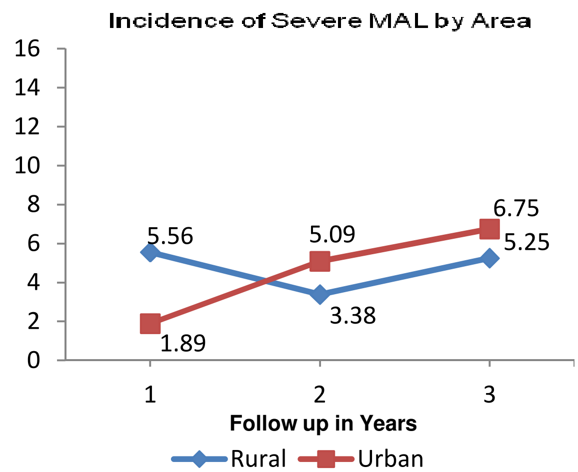
Figure 2: Incidence of Severe Malnutrition by Sex of the Child using BMI classification:

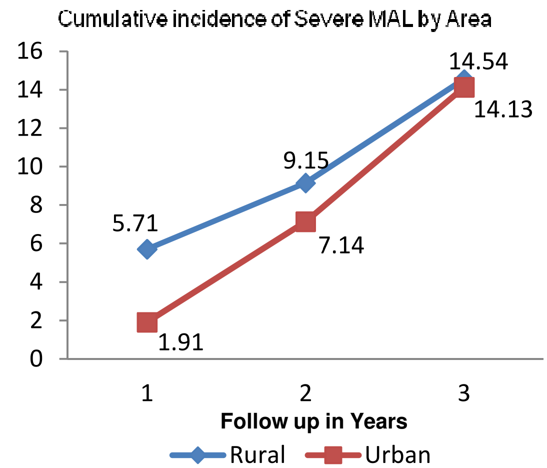




Note: MAL – Malnutrition

Figure 3: Incidence of Severe Malnutrition by Area of Residence using BMI classification:





Note: MAL – Malnutrition

Table 4a: Incidence of severe malnutrition using Height-for-Age classification by Age of the Child at baseline

At 5th Year:

From (Time t)	**To (t+1)**								
Baseline:	**Normal** n	**%**	**Moderate** n	**%**	**Severe** n	**%**	**Total**	**ISM**	**95% CI**
Normal	470	88.7	56	10.6	4	0.8	530		
Moderate	37	9.4	312	79.2	45	11.4	394	5.30	4.02 – 6.95
Severe	1	0.3	49	14.0	299	85.7	349		
1st Year:									
Normal	462	94.7	25	5.1	1	0.2	488		
Moderate	53	13.5	322	81.7	19	4.8	394	2.27	1.45 – 3.50
Severe	0	0.0	72	21.6	261	78.4	333		
2nd Year:									
Normal	504	98.4	8	1.6	0	0.0	512		
Moderate	103	24.4	315	74.6	4	0.9	422	0.43	0.12 – 1.14
Severe	0	0.0	103	35.2	190	64.8	293		
Incidence density of severe malnutrition per year									

At 6th Year:

From (Time t)	To (t+1)								
Baseline:	**Normal** n	%	**Moderate** n	%	**Severe** n	%	**Total**	**ISM**	**95% CI**
Normal	376	93.1	27	6.7	1	0.2	404	3.13	2.03 – 4.76
Moderate	26	9.7	222	82.8	20	7.5	268		
Severe	0	0.0	43	21.5	157	78.5	200		
1st Year:									
Normal	367	93.9	24	6.1	0	0.0	391	1.64	0.88 – 2.95
Moderate	30	10.8	238	85.3	11	3.9	279		
Severe	0	0.0	25	14.3	150	85.7	175		
2nd Year:									
Normal	390	97.3	11	2.7	0	0.0	401	0.29	0.01 – 1.11
Moderate	45	15.0	253	84.3	2	0.7	300		
Severe	0	0.0	38	23.9	121	76.1	159		
Incidence density of severe malnutrition per year									

At 7th Year:

From (Time t)	To (t+1)								
Baseline:	**Normal** n	%	**Moderate** n	%	**Severe** n	%	**Total**	**SI**	**95% CI**
Normal	47	92.2	4	7.8	0	0.0	51	3.7	0.083 – 10.77
Moderate	5	16.7	22	73.3	3	10.0	30		
Severe	0	0.0	3	10.3	26	89.7	29		
1st Year:									
Normal	47	92.2	4	7.8	0	0.0	51	1.28	0.01 – 7.59
Moderate	4	14.8	22	81.5	1	3.7	27		
Severe	0	0.0	5	17.2	24	82.8	29		
2nd Year:									
Normal	49	96.1	2	3.9	0	0.0	51	0.00	0.00 – 5.43
Moderate	2	6.7	28	93.3	0	0.0	30		
Severe	0	0.0	5	20.0	20	80.0	25		
Incidence density of severe malnutrition per year									

Table 4b: Incidence of severe malnutrition using Height-for-Age classification by Sex of the Child

Male

From (Time t)	To (t+1)								
Baseline:	**Normal**		**Moderate**		**Severe**		**Total**	**SI**	**95% CI**
	n	**%**	**n**	**%**	**n**	**%**			
Normal	446	90.5	44	8.9	3	0.6	493	5.73	4.34 – 7.53
Moderate	30	8.7	269	78.2	45	13.1	344		
Severe	1	0.3	41	12.7	280	87.0	322		
1st Year:									
Normal	425	92.8	32	7.0	1	0.2	458	2.27	1.41 – 3.58
Moderate	37	11.0	282	83.9	17	5.1	336		
Severe	0	0.0	46	14.4	273	85.6	319		
2nd Year:									
Normal	456	98.9	5	1.1	0	0.0	461	0.36	0.07 – 1.11
Moderate	85	23.0	281	76.2	3	0.8	369		
Severe	0	0.0	91	30.5	207	69.5	298		
Incidence density of severe malnutrition per year									

Female:

From (Time t)	To (t+1)								
Baseline:	**Normal**		**Moderate**		**Severe**		**Total**	**SI**	**95% CI**
	n	**%**	**n**	**%**	**n**	**%**			
Normal	447	90.9	43	8.7	2	0.4	492	2.98	2.00 – 4.38
Moderate	38	10.9	287	82.5	23	6.6	348		
Severe	0	0.0	54	21.1	202	78.9	256		
1st Year:									
Normal	451	95.6	21	4.4	0	0.0	472	1.67	0.97 – 2.82
Moderate	50	13.7	300	82.4	14	3.8	364		
Severe	0	0.0	56	25.7	162	74.3	218		
2nd Year:									
Normal	487	96.8	16	3.2	0	0.0	503	0.35	0.07 – 1.06
Moderate	65	17.0	315	82.2	3	0.8	383		
Severe	0	0.0	55	30.7	124	69.3	179		
Incidence density of severe malnutrition per year									

Table 4c: Incidence of severe malnutrition using Height-for-Age classification by Area of the Residence

Rural:

From (Time t)	**To (t+1)**								
Baseline:	**Normal**		**Moderate**		**Severe**		**Total**	**SI**	**95% CI**
	n	**%**	**n**	**%**	**n**	**%**			
Normal	367	95.3	18	4.7	0	0.0	385	2.89	1.88 – 4.41
Moderate	45	13.2	275	80.6	21	6.2	341		
Severe	1	0.3	61	17.3	291	82.4	353		
1st Year:									
Normal	379	93.3	26	6.4	1	0.2	406	2.14	1.29 – 3.47
Moderate	22	6.4	306	89.2	15	4.4	343		
Severe	0	0.0	49	15.9	260	84.1	309		
2nd Year:									
Normal	397	97.5	10	2.5	0	0.0	407	0.25	0.01 – 0.97
Moderate	55	14.0	336	85.5	2	0.5	393		
Severe	0	0.0	71	24.7	216	75.3	287		
Incidence density of severe malnutrition per year									

Urban:

From (Time t)	**To (t+1)**								
Baseline:	**Normal**		**Moderate**		**Severe**		**Total**	**SI**	**95% CI**
	n	**%**	**n**	**%**	**n**	**%**			
Normal	526	87.7	69	11.5	5	0.8	600	5.47	4.18 – 7.11
Moderate	23	6.6	281	80.1	47	13.4	351		
Severe	0	0.0	34	15.1	191	84.9	225		
1st Year:									
Normal	497	94.8	27	5.2	0	0.0	524	1.82	1.10 – 2.95
Moderate	65	18.2	276	77.3	16	4.5	357		
Severe	0	0.0	53	23.2	175	76.8	228		
2nd Year:									
Normal	546	98.0	11	2.0	0	0.0	557	0.44	0.13 – 1.16
Moderate	95	26.5	260	72.4	4	1.1	359		
Severe	0	0.0	75	39.5	115	60.5	190		
Incidence density of severe malnutrition per year									

Table 5: Bivariate (unadjusted) analysis for malnutrition (ordinal outcome) by socio-demographic and household variables with random intercept and random slope at child level and random intercept at household level

Variables	'Naïve' Regression OR (SE)	Random Intercept at Child level OR (SE)	Random Intercept and Random slope at Child level OR (SE)	Random Intercept and, Random slope at Child level and Random Intercept at Household level OR (95% CI)
Sex of the child				
Female	0.68 (0.02)	0.41 (0.07)	0.47 (0.06)	0.47 (0.37, 0.59)
Male				
ICC	-	0.76	-	0.0011
Area of Residence				
Rural				
Urban	1.02 (0.03)	1.25 (0.21)	1.16 (0.14)	1.20 (0.94,1.52)
ICC	-	0.763	-	0.0001
Birth Order				
1	1.00	1.00		
2	0.95 (0.04)	0.85 (0.23)	0.88 (0.17)	0.87 (0.60, 1.26)
≥3	0.92 (0.04)	0.82 (0.18)	0.83 (0.13)	0.83 (0.61,1.13)
ICC	-	0.763	-	0.02
Mother's Education				
Illiterate/ Literate	1.11 (0.06)	1.19 (0.34)	1.21 (0.44)	1.21 (0.78,1.80)
Primary/Middle School	1.10 (0.06)	1.14 (0.35)	1.18 (0.25)	1.18 (0.82,1.79)
High school/ College	1.00	1.00		
ICC	-	0.765	-	0.0008
Father's Education				
Illiterate/ Literate	1.04 (0.04)	0.99 (0.21)	1.02 (0.15)	1.01(0.75, 1.35)
Primary/Middle School	1.01 (0.04)	0.99 (0.23)	1.01 (0.16)	0.97 (0.71,1.33)
High school/ College	1.00	1.00	1.00	1.00
ICC	-	0.765	-	0.0034
Number of Family Members				
<=4	1.00	1.00	1.00	1.00
5 – 6	0.99 (0.04)	0.85 (0.23)	1.13 (0.30)	1.10 (0.77, 1.58)
>6	0.94 (0.04)	0.82 (0.18)	0.95 (0.25)	0.96 (0.66,1.38)
ICC	-	0.763	-	0.00006

Contd….

Variables	'Naïve' Regression OR (SE)	Random Intercept at Child level OR (SE)	Random Intercept and Random slope at Child level OR (SE)	Random Intercept and, Random slope at Child level Random Intercept at Household level OR (95% CI)
Fuel for cooking Drug/Firewood Gas/Kerosene	1.01 (0.04)	1.05 (0.26)	1.07 (0.18)	1.07 (0.77,1.50)
ICC	-	0.763	-	0.0008
Defecation Within premises/latrine Open field	1.12 (0.03)	1.47 (0.25)	1.35 (0.16)	1.36 (1.08,1.73)
ICC	-	0.763	-	0.0009
Type of roof Thatched Tiled RCC/Pukka Others	 0.88 (0.08) 0.89 (0.09) 0.74 (0.07) 1.00	 0.68 (0.35) 0.75 (0.39) 0.40 (0.22) 1.00	 0.76 (0.26) 0.77 (0.29) 0.45 (0.18) 1.00	 0.53 (0.23,1.20) 0.50 (0.22,1.13) 0.23 (0.10,0.52) 1.00
ICC	-	0.762	-	0.02
Type of Floor Kucha Pukka	1.09 (0.03)	1.18 (0.21)	1.16 (0.14)	1.15 (0.91,1.46)
ICC	-	0.763	-	0.032
Presence of a Separate Kitchen Yes No	 1.00 1.20 (0.04)	 1.00 1.40 (0.28)	 1.00 1.35 (0.19)	 1.00 1.38 (1.05,1.83)
ICC	-	0.762	-	0.0006
Consanguineous Marriage Yes No	1.09 (0.03)	1.18 (0.21)	1.16 (0.14)	1.17 (0.92, 1.50)
ICC	-	0.763	-	0.001
Type of House Brick and/or mud Brick and cement Others	 1.15 (0.04) 1.00 1.29 (0.05)	 1.37 (0.28) 1.00 1.65 (0.41)	 1.30 (0.19) 1.00 1.52 (0.26)	 1.30 (0.98,1.72) 1.00 1.52 (1.08,2.14)
ICC	-	0.76	-	0.002

Note: ICC – Intraclass Correlation Coefficient